

Institutions as Multimodal Accomplishments Towards the Analysis of Visual Registers

Jancsary, Dennis; Meyer, Renate; Höllerer, Markus A.; Boxenbaum, Eva

Document Version

Accepted author manuscript

Published in:

Multimodality, Meaning, and Institutions

DOI:

[10.1108/S0733-558X2017000054A004](https://doi.org/10.1108/S0733-558X2017000054A004)

Publication date:

2018

License

Unspecified

Citation for published version (APA):

Jancsary, D., Meyer, R., Höllerer, M. A., & Boxenbaum, E. (2018). Institutions as Multimodal Accomplishments: Towards the Analysis of Visual Registers. In M. A. Höllerer, T. Daudigeos, & D. Jancsary (Eds.), *Multimodality, Meaning, and Institutions* (pp. 87-117). Emerald Group Publishing. <https://doi.org/10.1108/S0733-558X2017000054A004>

[Link to publication in CBS Research Portal](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us (research.lib@cbs.dk) providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 18. Jun. 2025

INSTITUTIONS AS MULTIMODAL ACCOMPLISHMENTS: TOWARDS THE ANALYSIS OF VISUAL REGISTERS

Dennis Jancsary, WU Vienna

dennis.jancsary@wu.ac.at

Renate E. Meyer, WU Vienna & Copenhagen Business School

renate.meyer@wu.ac.at

Markus A. Höllerer, WU Vienna & UNSW Sydney Business School

markus.hoellerer@wu.ac.at

Eva Boxenbaum, PSL Research University – MINES ParisTech

& Copenhagen Business School

eva.boxenbaum@mines-paristech.fr

Abstract. In this article, we develop and advance an understanding of institutions as multimodal accomplishments. We draw on social semiotics and the linguistic concept of metafunctions to establish the visual as a specific mode of meaning construction. In addition, we make semiotic modes conducive to institutional inquiry by introducing the notion of distinct ‘modal registers’ – specialized configurations of linguistic signs within a particular mode that are adapted and applied in the reproduction of institutions or institutional domains. At the core of our article, we operationalize metafunctions to develop methodology for the analysis of visual registers. We illustrate our approach with data from Corporate Social Responsibility (CSR) reporting in Austria.

Keywords. Institutions; meaning; metafunctions; multimodality; social semiotics; visual analysis; visual registers; corporate social responsibility; Austria

Introduction

In this article, we aim at developing a systematic approach to capturing the social meaning structures instantiated in visual and multimodal aspects of the social and organizational world. Building on insights from social semiotics, we systematically lay out how institutions can be understood – and studied – as multimodal accomplishments. From previous institutional research we know that institutions are complex collective achievements that are inherently multidimensional, that is, they are reproduced and act back on their producers on multiple dimensions of the lifeworld. Essentially, institutions have been conceptualized as both symbolic and material (e.g., Friedland & Alford, 1991): they permeate discourse (e.g., Phillips, Lawrence, & Hardy, 2004), practices (e.g., Mohr & Lee, 2000), rituals (e.g., Sillince & Barker, 2012), and material objects (e.g., Jones, Boxenbaum, & Anthony, 2013). Recent research indicates that institutions also have an affective and embodied dimension in addition to a cognitive dimension (e.g., Lok, Creed, DeJordy, & Voronov, 2017; Toubiana & Zietsma, 2016). Spanning all these dimensions, institutions are inherently meaningful for the actors who create, reproduce, or transform them, and are constructed based on a multiplicity of different sign systems. Hence, institutions are inevitably multimodal accomplishments in which multiple sign systems are co-present and interact in the construction of meaning. To fully understand institutional dynamics and their organizational consequences, research is challenged to develop conceptual and methodological insights that are able to capture such multimodal character of institutions.

Yet, there is no common body of knowledge in organizational research as to what constitutes a distinct semiotic mode, how individual modes are connected to institutions, how they interact – and how they can best be studied. We build on the emerging insight that different modes constitute distinct forms of creating and communicating meaning (e.g., Bell & Davison, 2013; Meyer, Höllerer, Jancsary, & van Leeuwen, 2013). Verbal language is, by far, the best

researched semiotic mode in institutional analysis. In addition, the visual mode has recently been gaining much attention (e.g., Bell, Warren, & Schroeder, 2014), and we suggest that further extending and systematizing our understanding of the ‘visual construction of meaning’ (Höllerer, Jancsary, Meyer, & Vettori, 2013) is a useful starting point for a genuinely multimodal agenda for institutional inquiry. In more detail, our aim in this article is to develop conceptual and methodological insights that enable future research to (a) adequately capture and understand the specific character of visual communication; (b) systematically use visual data also in larger-scale analyses of meaning structures (e.g., on the field level); (c) usefully contrast and integrate visual and verbal instantiations of institutions; and (d) help to better understand the multimodal character of institutions. To advance this agenda, we initiate a more systematic discussion of how the visual works as a sign system, and how it can be utilized for institutional analyses and, more broadly, for organization research.

Conceptually, we build on social semiotics (e.g., Kress & van Leeuwen, 2006) and systemic functional linguistics (e.g., Halliday & Hasan, 1989) and define a mode as a “socially shaped and culturally given semiotic resource of meaning making” (Kress, 2010: 79). Different modes may accomplish similar communicative results. We tackle these similarities between modes through the prism of ‘metafunctions’ – essential performative effects that any mode must achieve in order to be considered *as* a mode. However, each mode achieves these metafunctions in its very own way. For instance, friendliness or anger is expressed differently in visual communication than it is in verbal communication. We further suggest that the realization of metafunctions in particular institutional domains is achieved in the form of ‘modal registers’, defined as collective adaptations of the meaning-making resources of a semiotic mode according to the specific social/institutional context of use (e.g., Matthiessen, 2015). Registers provide insights into the content, style, and structure of institutionalized meanings. Researchers in the institutional theory tradition have pointed to the existence of verbal registers (e.g., Jones & Livne-Tarandach, 2008), material registers (e.g., Jones, Maoret, Massa, & Svejenova, 2012),

and emotional registers (e.g., Toubiana & Zietsma, 2016). We offer a way to systematically analyze ‘visual registers’ that encompasses the particular visual and aesthetic elements of an institution or institutional domain.

Such conceptual understanding of different modes and their manifestation in texts enables us to specify metafunctions more systematically and in detail for the visual mode. It is also the basis for the development of a methodological approach that highlights an inventory of genuinely visual elements of meaning construction. Our ultimate aims are to enable analyses of visual ‘imageries-of-motive’ (Höllerer *et al.*, 2013) with the same level of granularity as verbal vocabularies and to facilitate an appreciation of how visual imageries reproduce, maintain, and challenge institutions through their content, style, and composition. This, we argue, is merely a first step toward capturing not only the *multidimensional*, but also the *multimodal* character of institutions.

Our article proceeds as follows: first, we briefly outline the social semiotic basis for our conceptual and methodological development. We introduce the concept of a visual ‘mode’ and elaborate on semiotic metafunctions. We then turn to modal registers and present pioneering research on visual registers. In the second part of our article, we explain how metafunctions operate in the visual mode and propose a set of methodological steps to capture the visual registers of institutional domains. We illustrate selected aspects of our methodology with data from the Austrian Corporate Social Responsibility (CSR) discourse. To conclude our article, we outline its main contributions and point out some promising directions for future research.

Conceptual foundations: Metafunctions and modal registers

To make visual and multimodal text relevant and systematically accessible for institutional inquiry, it is necessary to clarify how it affects institutional dynamics. An institution, according

to Berger and Luckmann (1967), is characterized by a reciprocal typification of actor and act. During institutionalization, individual actors and their actions become objectivated as social roles and practices, and at the same time, the subjective meanings undergirding behaviors become social meanings and, eventually, social knowledge. To express these institutionalized meanings, specific sign systems drawing on semiotic modes are developed, become typified, and sediment. We define the typified linguistic instantiations of an institution in a particular mode as a modal register of that institution.

It follows that a modal register lies at the intersection of institution and mode. While a mode denotes a sign system, or ‘language’, available in the context of *culture*, a modal register is an instantiation and variation of that mode within the context of an *institutional domain*, which is itself a sub-system of the broader culture (Matthiessen, 2015). Figure 1 summarizes the relationships between these key terms. Contexts influence language, while lower levels of instantiation are nested in – and may constitute variations of – higher levels.

----- *Insert Figure 1 about here* -----

Below, we first establish the visual mode of communication and its specific way of addressing basic functions of meaning construction, before applying this insight to the study of institutions and, more specifically, their visual registers.

Delineating the visual mode through semiotic metafunctions

Building on Kress and van Leeuwen’s (2006) visual social semiotics, we understand the visual as a distinct semiotic mode, i.e., a “socially shaped and culturally given semiotic resource of meaning making” (Kress, 2010: 79). Accordingly, a mode as a specific sign system is collectively shared by members of a culture (see, for instance, Kress & van Leeuwen’s [2006] discussion of the visual mode in Western culture). Existing literature distinguishes, for instance,

image, writing, gesture, scent, and music as different modes. Kress (2010) proposes that every mode offers unique resources for meaning construction, but that these different resources are employed for accomplishing the same basic semiotic functions. Such shared ‘metafunctions’ (see also Halliday & Hasan, 1989) are construed as a way of ‘unpacking’ and comparing semiotic modes. They are vital performances that every mode must fulfill in order to work as a full system of communication. We introduce the three metafunctions here briefly and elaborate on them in more depth in the methodology section.

The *ideational metafunction* addresses how a mode represents the ‘world out there’ (relationship between the ‘text’ and the world) by depicting particular objects and/or actions as well as settings or circumstances and by establishing conjunctions between these elements. The ideational metafunction thereby constitutes the subject matter or *content of representation*. Whereas verbal representations are ‘symbols’ and denote the world conventionally (e.g., the word ‘dog’), visual representations can, through their iconic (visual similarity; e.g., a photo of a dog) and indexical (visual reference, e.g., footprints of a dog) forms, more closely approximate actual sensory experience (Peirce, 1991). A representation can create *conjunctions between elements* within a textual composition in a narrative form, i.e., through “unfolding actions and events, processes of change, transitory spatial arrangements” (Kress & van Leeuwen, 2006: 59), or in a conceptual form by means of “representing participants in terms of their more generalized and more or less stable and timeless essence” (Kress & van Leeuwen, 2006: 79), for instance, in taxonomies. In general, all semiotic modes can use the narrative form and/or the conceptual form to represent the world.

The *interpersonal metafunction* captures how a mode addresses audiences (relationship between text and viewer) and establishes attitudes towards what is represented. While the verbal mode uses specific grammatical constructions (e.g., pronouns, imperative voice, question marks) to address readers directly, the visual mode can ‘draw’ the viewer into the image and

assign her a specific position and perspective within the composition (e.g., Kress & van Leeuwen, 2006) through ‘gaze’ or angle. In this way, visual text expresses the *embodied positions* of viewers. In addition, given their ability to ‘encode’ information in diverse visual designs (e.g., color, brightness, saturation, shapes, etc.), visual texts also convey distinct *coding orientations*. Coding orientations are part of the interpersonal metafunction, as they influence whether particular audiences are likely to understand and accept the claims inherent in the message. While an artistically literate viewer may be able to see the ‘true’ form of a person in an abstract painting, border control officers examining passport photographs have very different criteria for determining ‘truth’.

The *textual metafunction*, finally, represents how a mode relates elements internally as well as externally, combining them into coherent texts. This metafunction makes a text “operationally relevant” and “makes the difference between a message and a mere entry in the grammar or the dictionary” (Halliday, 1976: 24). Such function includes, for instance, the construction of textual boundaries (e.g., ‘paragraphs’ in written texts, ‘pauses’ in spoken texts, or ‘frames’ in visual texts) and the salience of specific elements (e.g., center-margin, color, brightness, and perspectival design in visual text).

Each semiotic mode achieves these metafunctions, but does so differently. Whereas verbal language has received a lot of attention, the workings of other modes have been explored to a significantly lesser degree. Hence, first of all, a more systematic characterization of how other modes realize metafunctions is vital for understanding how meaning is constructed, maintained, or transformed. Second, and with regard to institutional analysis, understanding the modal registers, i.e., the typified linguistic instantiations, that characterize specific institutions or institutional domains is an important step in capturing institutions as multimodal accomplishments.

The visual registers of institutions and institutional domains

Institutions are hard to study empirically, since they need to be inferred from their manifestations in the social world. Berger and Kellner (1984: 33) make this point by emphasizing that institutions, as such, are never directly accessible empirically. Similarly, Friedland (2009: 51) suggests that institutional objects “are only known through their conjoint conceptual and practical specificity”. Institutional theory has long established that institutions and institutional domains are characterized by distinct resources for meaning construction. For instance, it has been shown that institutionalized accounts (Scott & Lyman 1968; Creed, Scully, & Austin, 2002; Meyer, 2014) or vocabularies-of-motive (Mills, 1940) accompany – and vary across – institutions. Different modes provide unique resources for meaning construction, and their use is socially organized and regulated (Kress, 2010). In differentiated societies, accordingly, institutional domains develop and prescribe specific modal registers – distinct repertoires of mode-specific signs used to realize metafunctions. Or, in other words, a register is “a language functioning in an institutional domain” (Matthiessen, 2015: 5). Institutions and institutional domains, accordingly, both provide and restrict the specific registers which, in turn, reproduce them.

Previous research has identified a variety of linguistic resources to illuminate institutions, which bear strong similarity to registers. Institutionalists have analyzed resources for meaning construction within verbal registers, most notably vocabularies (e.g., Loewenstein, Ocasio, & Jones, 2012; Meyer, Egger-Peitler, Höllerer, & Hammerschmid, 2014; Ocasio, Loewenstein, & Nigam, 2015) and styles (e.g., Friedland, Mohr, Roose, & Gardinali, 2014). More recently, institutional scholars have ventured beyond verbal registers. For instance, Jones *et al.* (2012) use artifact codes, i.e., the material features that are selected and combined to enact categories. Toubiana and Zietsma (2016) propose that institutional logics also have emotional registers, defined as prescriptions about the appropriate use and expression of emotions. Akin to

vocabularies-of-motive, Höllerer *et al.* (2013) elaborate on distinct imageries-of-organizing, which reproduce institutions visually. Similarly, Jones, Meyer, Jancsary, and Höllerer (2017: 632) point to visual registers, arguing that “[i]n order to fully grasp the visual potential, future research needs to address how different institutional or field-level logics are encoded and evoked through variations in style, perspective, color, or, more general, through different aesthetic codes”. The analytical query into the meaning of institutions can thus privilege specific registers and potentially involve multiple registers, each ‘fed’ by a specific semiotic mode.

While research on verbal registers has greatly advanced in recent years, access to visual registers is just beginning to develop. Despite promising initial insights, research has thus far yielded no systematic way of linking visuals to institutions. That is, analytical approaches that link visuals to meanings and social situations usually focus on the content of visual communication, ignoring its ‘grammar’. One consequence is that success in the endeavor to use multiple registers simultaneously to study meaning construction remains elusive. This shortcoming obstructs the development of systematic methods for multimodal inquiry, i.e., capturing multiple registers drawing from different modes to shed light on the phenomenon under study.

Some important or even pioneering conceptual work on the visual register exists. Based on a substantial body of literature on visuals and emotions, Lefsrud, Graves, and Phillips (2013; for an empirical application, see also Lefsrud, Graves, and Phillips, this volume) develop propositions on how the interplay between written text and visuals operates in legitimacy work. In a similar vein, Meyer, Jancsary, Höllerer, & Boxenbaum (forthcoming) develop theory regarding the effect of visual communication on the institutionalization of novel ideas. They identify a set of visual as well as verbal affordances and formulate propositions about how, and under what conditions, each mode positively affects a novel idea’s transition through

consecutive stages of institutionalization in ways that outperform the other mode. Another line of inquiry examines the nature of institutions empirically through the lens of visual registers. Based on concepts from structural linguistics, Höllerer *et al.* (2013) study the meaning of Corporate Social Responsibility from an analysis of visuals in corporate reporting documents. Through an in-depth coding of the content, style, and accompanying verbal text of more than 1,500 visuals, they reconstruct discursive topoi and the institutional logics that pervade them. Further advancing this research stream, Delmestri, Oberg, and Drori (2015) analyze the content and design elements of over 800 university logos to shed light on national differences in their meaning and use. Moreover, Drori, Delmestri, and Oberg (2016) show how the shifting prevalence of narratives associated with logos provides insight into processes of institutional change in the university sector. Finally, Höllerer, Jancsary and Grafström (2017) examine the interplay between visual and verbal elements in a large number of newspaper articles on the Global Financial Crisis (GFC). They suggest that the visual elements in multimodal compositions enhance the perceived validity of theorization and representation of the GFC and link the crisis to generalized global myths and socially shared categories.

These studies illustrate the potential of taking visual registers into account when studying institutional meaning and its effects. We suggest that visual registers offer some unique features that enable institutionalists to study these topics in ways that are complementary to the insights generated through other registers. On a methodological level, we suggest that different registers are complementary to one another in the sense that they accomplish the same metafunctions in different ways. More precisely, in the case of the visual register, we propose that meaning is expressed through distinct combinations of (a) *content of representations* and *conjunction between elements* (i.e., components of the ideational metafunction); (b) *embodied positions* of viewers and distinct *coding orientations* (i.e., components of the interpersonal metafunction); and (c) the *composition of boundaries and salience* (i.e., components of the textual

metafunction). Accordingly, we develop a methodology that aims at capturing the meaning of an institution by extracting relevant information from its visual register.

Methodological development: Extracting meaning structure from visual data

In this section, we aim at developing a methodological approach that is capable of capturing genuinely visual forms of meaning construction. Any methodology targeting meaning must capture how modes realize specific metafunctions in distinct ways. For instance, the verbal mode relies on sequences of words to construct sentences and larger texts, which differs from the spatial representation of elements in the visual mode. Accordingly, a common form of hermeneutical analysis is to follow a sequence of verbal text and extensively interpret the role of each word according to its position within the sequence (i.e., ‘sequential’ analysis, see, e.g., Lueger, Sandner, Meyer, & Hammerschmid, 2005). The spatial and holistic configurations of visual texts, in contrast, require a different set of analytical concepts such as, for instance, vanishing lines, distance, or perspectival information (e.g., Bohnsack, 2008). For space reasons, we focus here on how to code specific realizations of the ideational and interpersonal metafunctions as realized in visual registers. In doing so, we take considerable inspiration from Kress and van Leeuwen (2006). For suggestions on how to combine the analysis of visual data with that of verbal data in a multimodal inquiry, we wish to point to existing literature (e.g., Jancsary, Höllerer, & Meyer, 2016; Machin & Mayr, 2012).

Capturing the ideational metafunction in visual registers

The ideational metafunction focuses on the content of representation as well as the conjunctions between elements of this content (Halliday & Hasan, 1989). The *content of representation* encompasses participants and their attributes (people and objects), processes (actions and

events), and the broader setting(s). The written word as signifier relies on conventional reference to the signified. In contrast, sketches, drawings, and photographs display varying degrees of iconicity (e.g., Rowley-Jolivet, 2004), meaning that they can range from icons to symbols (e.g., Peirce, 1991; see also Zhao, this volume). Their perceived verisimilitude bestows upon iconic visual representation (such as, for example, photos) a particularly fact-like quality (e.g., Graves, Flesher, & Jordan, 1996) and allows for the depiction of characteristics of material artifacts (e.g., texture, physical and spatial properties, etc.). Such iconic representation also entangles the conceptual and illustrative components of vocabularies (e.g., Loewenstein *et al.*, 2012): Iconic visuals inherently communicate social categories as well as examples of these categories at the same time (e.g., Meyer *et al.*, forthcoming).

Conjunctions between elements can take the form of narrative (dynamic relationships between elements) and conceptual ones, with the latter encompassing both analytical (classifications, hierarchies, or attributions) and symbolic (reference to more broadly available ideas) relations (e.g., Kress & van Leeuwen, 2006). The spatial ordering of the visual mode enables the representation of conjunctions between elements that go beyond the sequential structure of verbal text and shows hierarchies and depth more immediately and intuitively, and in a much more condensed manner. Such spatial relating of categories facilitates the communication of collective schemata and, ultimately, the ‘architecture’ of institutions and their logics (e.g., Ocasio *et al.*, 2015).

Coding the ideational metafunction can rely on existing techniques of visual content analysis (e.g., Rose, 2007) and capture [participants]¹, [processes], and [settings] with different granularities. Additionally, [attributes] of each of these elements can also be coded, such as, for instance, gender, age, and ethnicity for people, as well as material, texture, and mobility for physical objects. Similarly, relationships can be classified as [narrative] (e.g., actions and

¹ Terms in square brackets denote codes or coding categories.

events), [analytical] (e.g., comparisons, classifications, hierarchies, or attributions), or [symbolic] (e.g., metaphors, metonymies, or other tropes).

----- *Insert Table 1 about here* -----

Capturing the interpersonal metafunction in visual registers

The interpersonal metafunction denotes the relation between the producer of a sign and the receiver, or reproducer, of that sign (Kress & van Leeuwen, 2006), along with implied attitudes and judgments (Halliday & Hasam, 1989). By allowing the viewer to scrutinize the content of a visual from a specific perspective, distance, and angle, visuals imply *embodied positions*. The interpersonal function, consequently, is a way of assigning the viewer (but also its counterpart in the visual, and potentially the image producer) a specific position in the social fabric. According to Kress and van Leeuwen (2006), the interpersonal metafunction encompasses three central aspects.

First, embodied positions reveal whether the viewer is conceptualized as a passive observer or an active participant in what is going on in the visual. The analytical question is whether some form of ‘contact’ is established between the viewer and people or objects. Direct eye contact and interactive gestures (e.g., pointing at the viewer) ‘draw’ the viewer into the image. If there is no indication of direct contact, the viewer is positioned as a passive observer. We suggest coding contact as strong [strong contact] whenever direct contact is established, as weak [weak contact] when there is no direct contact, and as absent [no contact] whenever there are no people present and there is no obvious interaction with physical objects.

Second, position is characterized by the social distance between viewer and content. Distance implies specific forms of (imaginary) social relations (e.g., intimate, friendly, one of being simple acquaintances, or complete strangers). It also connotes availability and

accessibility of both people and physical objects. Following social semiotics, distances imply an [intimate] gaze when we see a person (usually only a headshot) or object at very close range, with the implication that we could almost touch them. An [interpersonal] gaze means that we see most parts of a person or object (or small groups thereof) and implies that they could be approached with some effort. Finally, an [impersonal] gaze shows one or more people or objects from afar, with the implication that they are out of reach both literally and metaphorically.

Third, embodied positions provide information about implied subjectivities through perspective and angle. Kress and van Leeuwen (2006) suggest that a specific ‘point of view’ enforces a certain subjectivity in the sense of ‘being subjected’ to something. Seeing something from a high vertical angle (‘bird’s eye view’) suggests [viewer power] over what is seen (‘looking down on somebody’). In contrast, a low vertical angle (‘frog’s eye view’) assigns [representation power] to the content (‘looking up to someone’). Eye-level perspective suggests [equality]. Horizontal angle, on the other hand, expresses degree and quality of involvement (Kress & van Leeuwen, 2006). A frontal angle is realized when the plane of the participants aligns with that of the viewer and/or the central vanishing points of the image are within the depicted frame. Such a perspective implies [involvement] in the sense of ‘belonging to the same world’. In contrast, an oblique angle means that the frontal plane of the participants and that of the viewer diverge from each other, implying [detachment] or ‘otherness’.

Embodied positions are highly relevant for institutional analysis. They imply social roles and identities, which are central ‘building blocks’ of institutions and provide information about understandings of actorhood in a distinct institutional domain. They are also representative of a distinct ‘gaze’, for example, of professions (e.g., Styhre, 2010), that enforces specific ways of seeing. Contact with the viewer communicates and further qualifies (e.g., are people smiling or frowning, pleading or commanding, etc.) typified relationships. Distance and perspective provide additional details about subject-object-practice relations (Franzosi, 2010; Friedland,

2009) that reproduce institutions (e.g., which human and non-human objects are close and accessible, while others are out of reach). Vertical perspective reveals institutionalized power relations and ‘orders of worth’ (Boltanski & Thévenot, 2006) by implying core values, exemplars of virtue, and powerful social roles. Horizontal angles, in contrast, show boundary work through mechanisms of inclusion and exclusion (i.e., the definition of particular roles as either ‘self’ or ‘other’).

In addition to suggesting embodied positions, visual registers also typify certain *coding orientations* – “relatively reliable guides to the truth or factuality of messages” that are valid within specific social groups and relate to their central values, beliefs, and social needs (Kress & van Leeuwen, 2006: 154). Visual text expresses such shared conventions through ‘modality markers’ such as color (e.g., saturation, differentiation, and modulation), contextualization (absence or presence of background), representation (abstraction and detail), depth, illumination, and brightness. A [naturalistic] coding orientation refers to an everyday sense of realism. It demands that visuals are as close to actual perception as possible and is most typically realized in snapshot photography. A [sensory] coding orientation deviates from the naturalistic one by enhancing aesthetic properties to trigger enhanced sensory experiences and stimulate affective reactions. Examples include highly stylized photomontages such as are commonly found in advertisements, posters, or glossy magazines. [Abstract] coding orientation reduces the individual to the generic, for instance in sophisticated flowcharts and process models, traffic and warning signs, as well as certain types of (modern) art. A [technological] coding orientation, finally, is mostly concerned with counting, weighing, and measuring, or providing useful ‘blueprints’ of social reality and is therefore employed, for instance, in diagrams and/or networks. Coding orientations are evocative of distinct institutional domains (e.g., science, art, religion, law, etc.). As such, they are highly relevant for further exploring the aesthetic (e.g., Jones *et al.*, 2017) and emotional (e.g., Toubiana & Zietsma, 2016) aspects of institutions.

----- *Insert Table 2 about here* -----

Towards analyses of social meaning structures across metafunctions

We have shown how to code metafunctions in visual texts to study the visual registers of institutions, and we have illustrated some areas of applicability. These elements promise exciting new avenues for institutional inquiry. Much work, however, remains to be done. Since messages unfold their meaning(s) through the *interplay* of metafunctions (Halliday, 1976), and often multimodally, we need to capture this interplay within elements of registers as well as between registers. Recent work in communicative and discursive institutionalism suggests that the differentiated structure of institutions can be methodologically captured through structural analyses of verbal texts and discourses (e.g., Jones & Livne-Tarandach, 2008; Meyer & Höllerer, 2010). We add the analysis of visual texts to this agenda.

We suggest that capturing specific patterns in the way metafunctions are realized reveals visual registers and facilitates the task of ‘measuring meaning structures’ (e.g., Mohr, 1998) based on visual data. Codes capturing the visual registers of institutions can inform a variety of subsequent analyses. In particular, taking metafunctions as a starting point is an excellent basis for methodological procedures aiming at the identification of broader meaning structures. In our explorative empirical illustration below, we illustrate briefly how codes related to the ideational and interpersonal metafunctions can inform semantic network analytical techniques (e.g., Carley, 1993; Mohr & Duquenne, 1997). Of course, other methods conducive to the identification of meaning structures may be equally suitable.

Another asset of the suggested methodology is its rootedness in social semiotic theory, which acknowledges the distinct differences between modes. While the focus of this article is on capturing visual registers, the conceptual insight that each mode needs to address all three

metafunctions makes our methodology a starting point for genuinely multimodal analysis focusing on differences and interactions between multiple registers drawing from different modes.

Explorative empirical illustration

We now illustrate our considerations by exploring the visual register of reporting on Corporate Social Responsibility (CSR) in Austria. Since the turn of the millennium, there has been increasing public debate on the role and responsibilities of business within society, urging corporations to engage proactively in this discourse (e.g., Höllerer, 2013). New management ideas and forms of corporate communication have emerged that address such a need. CSR broadly denotes social and societal challenges related to conducting business (Hiss, 2009) and stresses values of, for instance, integrity, fairness, and transparency (e.g., Matten & Moon, 2008). However, its actual meaning, content, and scope are far from established universally, and local translations abound (e.g., Höllerer *et al.*, 2013).

In Austria, the Anglo-American terminology of CSR has only relatively recently been adopted (e.g., Höllerer, 2013), despite – or even because of – the existence of a longstanding social consensus about the social responsibility of business. Only in the late 1990s and early 2000s did corporations begin to use the terminology of CSR. Stand-alone annual CSR reports were not issued by any publicly traded corporation in Austria prior to 2001. Over the last decade, reporting on CSR has become increasingly institutionalized, with ever more publicly listed corporations issuing either stand-alone or integrated reports. Still, research shows that the ‘chameleon-like’ character (Meyer & Höllerer, 2016) of CSR – incorporating both similarities to a corporatist model of coordinated market economies and a more liberal model of capitalism – places it in an ambiguous position within the institutional domain of the market. Similarly,

Höllerer *et al.* (2013) have shown that the discourse of CSR in corporate reporting intertwines elements from a variety of institutional domains.

For this explorative illustration, we used a sub-sample of the CSR reports analyzed in a previous study (Höllerer *et al.*, 2013), consisting of 1,023 visuals drawn from 11 corporations and 25 CSR reports published between 2001 and 2008. The sub-sample encompasses both the first and the last report published by each corporation within this timeframe, as well as reports published at regular intervals in between. A final disclaimer is in order: in every empirical study, the methodological approach must be designed and adapted according to the specific research objectives and guiding research questions. Our aim in this article is not to achieve substantial understanding of the institutional embeddedness of CSR, but to illustrate how social semiotics can be employed to reconstruct visual registers. A respective research question would ask which meaning of CSR is contained within the visual register of Austrian CSR reporting. In the following we selectively illustrate what we perceive as the most interesting aspects of such a visual register: the *content of representation* of CSR (one component of the ideational metafunction), and the *embodied positions* (one component of the interpersonal metafunction) enforced on the viewer. We also explore their relationship to the verbal text in the reports to contextualize the usage of the visual register.

Analyzing the ideational metafunction of the Austrian CSR discourse

In a first step, we conducted a relatively simple content analysis of our sample of visual data to reconstruct different aspects of the ideational metafunction. Following the approach outlined in the previous section, we coded participants, processes, settings, and their attributes in the visuals within the reports. For more focus, we coded only the most salient elements in each image. Table 3 provides an overview of the most frequently represented content and its prevalence in the visuals.

----- *Insert Table 3 about here* -----

From this analysis, we gain four immediate insights into the meaning of CSR. First, men are clearly more dominant than women. Adult men are portrayed individually almost twice as often as women. Additionally, all-male groups are more common than mixed groups, and almost ten times more common than all-female groups. Finally, the percentage of women depicted as individual adolescents or children (11%) is more than five times higher than that of men (2%). The ethnicity of people is highly homogenous, with only 4% of images showing non-Caucasian people. Second, industry is visually more dominant than nature, which is surprising given the strong environmental aspect of CSR. Third, there is very little action. Communication strongly surpasses more physical activities. And fourth, the visuals are highly de-contextualized, with almost half of the coded visuals providing no or unclear cues about settings.

In a further step, we coded the most prominent verbal representations in order to contextualize the usage of the visual register. Since there are significant amounts of written text in CSR reports, we restricted our coding to the headlines under which visuals were placed. This approach enabled us to identify the main topics of communication. We found the most prevalent verbal topics to be ecological issues (18%), social issues (15%), preface and CEO testimonials (10%), company presentation (10%), human resource issues (8%), strategy (7%), products and production (6%), sustainability (5%), and economic issues (5%). These topics correspond with the idea of a ‘triple bottom line’, framed by a favorable presentation of the corporation and its strategies.

Analyzing the interpersonal metafunction of the Austrian CSR discourse

We then explored embodied positions as a central aspect of the interpersonal metafunction. Accordingly, we coded for contact, distance, and power relations. Table 4 provides frequencies

and shows that, in general, the visual register spans the whole spectrum of potential relationships. However, while visuals are rather evenly distributed across the dimensions of contact and distance, the power dimension is weakly differentiated, with most visuals suggesting equal power between viewer and represented content. Since the Austrian CSR discourse generally positions the viewer at eye-level, we characterize CSR as egalitarian in the visual register.

----- *Insert Table 4 about here* -----

Elements of embodied positions should not be interpreted in isolation. In fact, what establishes specific relationships between viewers and visuals, also known as a ‘gaze’, is the combination of different dimensions, including contact and distance. We further refined the analysis accordingly.

In our data set, greater distance implies reduced contact. An intimate gaze often establishes eye contact to depicted persons. In contrast, eye contact is uncommon in interpersonal and impersonal gazes. Visuals without people are commonly depicted using an impersonal gaze; however, the viewer is also offered an intimate gaze in a substantial number of cases. Inanimate objects are instead presented as things to be ‘inspected’ (a ‘scientific gaze’) – either holistically from a distance, or closely with a focus on details. People are less frequently subjected to an intimate gaze unless strong contact with the viewer is established. The relevance of power also increases with distance and is most pronounced in visuals without people. Table 5 depicts these combinations of contact and distance in a 3x3 matrix. Viewer power [VP] and representation power [RP] are added to those cells in the matrix in which they are most prevalent. The [VP+] in bold font in the “No Contact/Impersonal” cell denotes the highest percentage of visuals with power implications. In our data, viewer power (i.e., a downward gaze) exists towards objects (no contact) and in relation to people who are not looking back (although such a ‘voyeuristic’ or ‘medical’ gaze is less common), but almost never towards people with whom eye contact is

established. Representation power almost always stems from objects as well, meaning that the viewer is not subjected to the power of people within the visuals (which would create intimidating perspectives).

----- *Insert Table 5 about here* -----

Intersecting metafunctions

This methodology also allows for an examination of how the ideational metafunction (in this case, content representations) maps onto the interpersonal metafunction (in this case, embodied positions). Such a mapping provides further insights, for instance, about which embodied positions apply to which participants and in relation to what kinds of topics. Table 6 provides information on the kinds of visuals that are most prevalent in each of the nine segments of the matrix presented in Table 5. We also include information from chapter headlines in reports to shed light on the topics covered.

----- *Insert Table 6 about here* -----

It is striking that people who engage with the viewer directly ('strong contact' row) have a very specific profile. They appear in unclear or non-descript settings and hardly ever work physically, but engage in communication or leisurely activities. Such direct engagement provides the strongest identification cues, which suggests that the meaning of CSR within the Austrian discourse, as expressed in the visual register, is primarily a 'white collar' phenomenon, where high-status actors address the viewer on an equal footing (the dominance of 'equality' of power). Interestingly, the same is largely true of visuals that do not engage the viewer directly ('weak contact' row). From intimate and interpersonal gazes, we observe primarily white-collar employees working and discussing topics. This impression changes slightly, however, in relation to an impersonal gaze. Here, the 'other' appears not only as people with different ethnic

and cultural backgrounds or of different age, but also as different settings, such as rural environments (see examples in Table 6). Interestingly, such people are presented with weak contact, from a greater distance and from an oblique horizontal angle, which suggests detachment: “The oblique angle says: ‘what you see here is not part of our world; it is their world, something we are not involved with’” (Kress & van Leeuwen, 2006: 136). In this context, it creates boundaries between the individuals who perform and embody CSR and those who are seen as the beneficiaries of these CSR activities.

The prevalence of intimate gazes involving both people and inanimate objects reinforces our previous explanation when we look more closely at the content of the visuals. Depicting people up close and establishing eye contact communicate concern and accountability (see also Höllerer *et al.*, 2013). In contrast, zooming into production processes and related objects insinuates an invitation to inspect, which implies transparency. Both gazes serve as sources of legitimacy for the organization issuing the report, since accountability and transparency are two major pillars of public disclosure. Whereas an intimate gaze with eye contact invites the reader to be a ‘partner’, an intimate gaze on objects casts the reader in the role of ‘examiner’. Both assigned roles are related to CSR, the specific genre (CSR reports) and the context (Austria). For instance, viewers are not cast as beneficiaries of CSR, but as evaluators of the company’s CSR performance; subjectivity as depicted in strong contact and direct angle is, to an overwhelming extent, Caucasian and male. The verbal text indicates a similar pattern, but the visual register communicates this meaning in a more strongly embodied way.

Finally, natural imagery is most often invoked without the addition of human actors, and ecological topics are addressed primarily with visuals devoid of people. Although several visuals merge beautiful nature with industrial buildings (see also Höllerer *et al.*, 2013), such nature is still portrayed as mostly ‘untouched’. Interactions between industry and nature happen at an impersonal distance. Nature, in general, becomes more visible as the distance in the visual

increases. This distance implies that business and industry are portrayed as being closer to the reader (the ‘self’), while nature is akin to the ‘other’. In addition, nature is consistently portrayed as utopian. The visual representation of nature emphasizes characteristics such as ‘whole’ and ‘beautiful’, whose meaning pertains to preservation, rather than to restoration. The visual register also avoids any admittance of past wrongdoings in relation to nature. Notably, there are no images of natural habitats impacted by industrial activity: the natural environment is represented as still being intact.

Explorative insights into the structure of the visual register

We conclude our methodological illustration with a brief presentation of how this methodology can be used to examine the wider meaning structure of the Austrian CSR discourse. We have defined a register as a configuration of linguistic signs tied to a distinct institutional domain. The way in which elements of meaning are structured and configured provides insights into shared and institutionalized schemata and templates (e.g., Ocasio *et al.*, 2015). To capture the configuration of meanings, we plotted the coded elements as a network by counting their co-occurrences within individual images (e.g., how often does the depiction of specific participants co-occur with the suggestion of a distinct perspective). We then used the Newman algorithm, which identifies community structures in sparsely weighted networks by optimizing the modularity of partitions (e.g., Newman, 2006), highlighting distinct clusters within the network that are more strongly interconnected internally than they are externally (see also Höllerer *et al.*, 2013). Figure 1 shows the resulting network, which represents the configuration of meaning of CSR in Austria. Circles denote the ideational aspects of the visual data (participants, processes, and settings). Triangles stand for the topics as reconstructed from the verbal chapter headlines. The hexagonal shapes highlight the distinct interpersonal aspects (combinations of

contact, distance, and angle) as represented visually. The coloring indicates the three distinct clusters identified by the algorithm.

----- Insert Figure 2 about here -----

The network shows three clusters, each of which suggests a distinct gaze in relation to specific topics and objects related to CSR. Each cluster represents a particular aspect of the meaning of CSR. Below, we provide a short and preliminary interpretation of each cluster.

Ecological (dark grey cluster): *The scrutinizing gaze towards ecological impact.* In relation to ecological and production issues, a scrutinizing gaze dominates the general company presentation and the more general parts of the reports, such as CEO testimonials and introductions to sustainability. The visual representation is characterized by the absence of humans and the use of angles that suggest power differences. The ideational aspect of images focuses on industrial and natural objects, as well as on their integration. This perspective conveys transparency, from up close (detail) as well as from afar (overview). It positions the viewer as examiner and evaluator of the corporation's ecological impact.

Economic (black cluster): *The partner-like gaze on management.* In relation to economic aspects of CSR, the viewer is positioned on an equal footing with management. This gaze is used for economic and strategic issues and characterized by strong intimacy. This gaze invites the viewer to identify with management actors and, by proxy, with the corporation itself. Since the viewer is invited to identify with management, interests are supposedly aligned, and the scrutinizing gaze on ecological impact is thus perceived from the management's perspective.

Social (light grey cluster): *The benevolent gaze towards the human 'other'.* A third gaze applies to the CSR topic of social and HR issues. It is characterized by humans depicted at greater distances (interpersonal and impersonal), but displays equality through eye-level depictions. The greater distance stresses the 'otherness' from the viewer's perspective. The

verbal headings identify this otherness as belonging to the social dimension; the visual register specifies the human ‘other’ as being diverse in terms of gender, age, and cultural background, and specifically includes non-adults. They are the beneficiaries of corporate CSR activities and not part of the world of the viewer.

Collectively, these three gazes represent the meaning of CSR in Austrian CSR reports as expressed in the visual register. This meaning is extracted from visuals through an analysis of the ideational metafunction (content of representation) and the interpersonal metafunction (embodied position) of the visual mode of communication. They correspond roughly to the ‘triple bottom line’, suggesting that each form of sustainability (ecological, economic, social) is characterized by a specific gaze: industrial activities are integrated with nature to deproblematize them (ecological concern), readers are invited to join and scrutinize management (economic concern), and specific social groups are designated as beneficiaries of CSR (social concern). We complemented this visual analysis with some meanings expressed in the verbal register (e.g., captions), but did not conduct a full-fledged multimodal analysis. Future research could systematically compare similarities and differences between the meanings of CSR as expressed in the visual and the verbal registers. We anticipate that the visual register may contain meaning that is far more embodied and implicit and include aspects of meaning that are difficult to verbalize (e.g., ethnicity or power relations). Hence the need for a methodology that can analyze each register separately, while also enabling comparison between them. We argue that such a multimodal methodology can rely on meta-functions not only to capture, but also to compare and combine the meanings conveyed through different modes of communication.

Conclusions and implications

This article conceptualized institutions as multimodal accomplishments that are instantiated not only by verbal registers (e.g., distinct vocabularies), material registers (e.g., artifact codes), and emotional registers (e.g., different affective expressions), but also by distinct visual registers. For this purpose, we developed a conceptual perspective linking institutions to semiotic modes and specific registers and outlining the social semiotic notion of meta-functions that transcend different modes of communication. We articulated the semiotic metafunctions for the visual mode and applied these insights to develop key features of visual registers in particular institutional domains. Building on these conceptual ideas, we developed further our methodology for analyzing the meanings contained in the visual register, which we illustrated in a mostly visual analysis of meanings in Austrian CSR reports. We propose that this methodology can be applied to the analysis of larger quantities of visual and multimodal data. This application enables a better understanding of visual registers and opens inquiry into how they are linked to other registers, which collectively instantiate institutions.

Our article has two central limitations. First, also due to space restrictions, we focused only on the ideational and the interpersonal metafunctions. Future research should extend our methodology by also developing procedures for capturing the textual metafunction. We specifically call for further methodological development grounded in empirical analysis that may find additional and more nuanced ways in which modes other than the verbal mode address all three metafunctions. This orientation would add additional depth and breadth to our methodological approach. Second, our empirical illustration is, of necessity, rather rudimentary. Given our focus in this article on developing novel methodology, our aim was not to provide any comprehensive insights into the institutional aspects of CSR. Future research is challenged to investigate in more depth how exactly different registers instantiate distinct institutional domains, and how they may differ according to the focal issue, but also according to the communicative media under investigation.

As a first contribution to literature, we address recent calls to further explore the visual construction and reproduction of institutions (e.g., Meyer *et al.*, 2013) by developing the notion of modal registers as contextually bound configurations of linguistic elements that express broader institutional meanings. The study of visual registers can shed light on institutional dynamics, about which we know little, because our theoretical concepts and methodologies are maladapted to studying them. For instance, the implicit power relationship between depicted individuals and the imagined audience can easily reveal social positioning (i.e., center-periphery relations), thus avoiding the potential controversy that may arise in using the verbal register to convey sensitive power relations. Moreover, verbal text is a poor data source for studying what is taken for granted. Once something is fully institutionalized, it is no longer articulated verbally, which makes it almost impossible to detect in written or spoken language. Verbal methodologies cannot readily differentiate between absences that result from taken-for-grantedness versus perceived irrelevance. In visual registers, however, taken-for-granted ideas become visible. If, for instance, health care staff is consistently portrayed as female in visuals, while there is hardly any trace of gender in the associated verbal text, then we may infer that health care staff is implicitly expected to be female. We can thus eliminate the alternative hypothesis that gender is irrelevant for health care delivery. Hence, to study taken-for-grantedness, we require access to multiple registers and need to compare the meaning residing in each register. Along similar lines, how meaning is created at the early stages of institutionalization is poorly accessible through verbal analysis. When new prototypes are being developed and new practices theorized, visual representation may play an important role in generating and consolidating new collective meaning (e.g., Meyer *et al.*, forthcoming). The ability of visuals to engage viewers in creative sensemaking is crucial at this stage of institutionalization, and its contribution to collective meaning formation may far surpass the capacity of verbal text (e.g., Cartel, Colombero, & Boxenbaum, this volume; Boxenbaum, Daudigeos, Pillet, & Colombero, this volume; Höllerer *et al.*, 2017).

Our second contribution to research on meaning and institutions is our elaboration of a methodology to access the visual registers of institutions. It acknowledges the multimodal character of institutions and enables the systematic study of the visual register in which institutionalized meaning is expressed. We discuss how the visual mode may operate differently than other modes – and/or that it interacts with other modes in ways we do not yet fully comprehend. For instance, visuals create and reproduce distinct gazes that draw viewers into particular meaning spheres (e.g., Styhre, 2010). Although written text can convey similar information, visuals engage the embodied aspects more directly and make them accessible to research. Visuals may also support the turn to materiality in institutional theory (e.g., Jones *et al.*, 2013; Oliveira, Islam, & Toraldo, this volume) to the extent that they mediate between verbal text and material artifacts. Relative to verbal text, visuals retain a degree of iconicity and abstract less from material reality, which makes visual representation a potentially useful ‘interface’ between the cognitive and the material bases of institutions. Future research should definitely extend our line of inquiry to the material register.

Third, our methodology carries potential for conducting organizational research that spans several modes of communication, but further work is needed to elaborate on how to use our methodology to tap into other, and multiple, registers. While research in organization and management studies has started to recognize the value of visuality and multimodality (e.g., Bell & Davison, 2013; Meyer *et al.*, 2013), we have yet to develop any significant insights into the specific interaction between modes and the modal registers that instantiate institutions. A social semiotic perspective presents a particularly interesting avenue for multimodal research and an excellent opportunity for initiating a broader discussion. In this article, we have focused on the visual mode and visual registers. The concept of metafunctions, however, has the potential to reach much further. Metafunctions permit comparison across different modes of communication. In order to apply our methodology to the analysis of other modes, or to multiple modes at once, further efforts are required to explicate how each mode addresses the same

metafunctions and how meaning is constituted and conveyed in particular registers. Our article provides a starting point for such a characterization, but more work is required to develop a full-fledged methodology for multimodal inquiry. Accordingly, we encourage scholarly efforts that extend the agenda to less studied modes such as sound, scent, and gesture. We also call for future research that systematically compares modes across metafunctions and articulates the interaction of different registers in denoting specific institutions. Further systematization of multimodal inquiry in and around organizations may not only facilitate dialogue across methodological divides, but also help organization research to become inherently multimodal. Finally, our methodology also adds a visual layer to structural approaches to institutions that have previously focused almost exclusively on verbal text. Our analytical indications for studying the visual register align particularly well with the idea of vocabulary structure (e.g., Loewenstein *et al.*, 2012; Ocasio *et al.*, 2015) as a way of analyzing meanings and institutional domains. Visual analysis adds word-to-image relations, since visuals can provide both social categories and examples. In fact, we would claim that the social semiotic notion of metafunctions could more broadly inform future developments in the vocabulary perspective, thereby enabling even more in-depth studies of how the combination of modes provides new insights into the distribution of meanings across social contexts. Relatedly, our methodology for analyzing the visual register also paves the way for comparing the verbal and visual aspects of different institutional logics (e.g., Thornton *et al.*, 2012) as they manifest in discourse and practice. We also call for further methodological development that can enable systematic multimodal inquiry. Our work being preliminary, we suggest future research aims at systematically studying meaning across registers. Future work may extend our methodology to analyze other registers, or it may extend the scope by plotting broader structural and relational patterns into a multimodal grid and apply semantic network analytical techniques or multidimensional scaling methods to capture wider meaning structures. The prospects for multimodal inquiry are, in our view, promising and exciting. We hope to have taken one

important methodological step toward realizing its significant potential for the analysis of institutions and meaning, as well as for organizational studies more broadly.

Acknowledgements

We gratefully acknowledge financial support from the Danish Research Council: DFF-1327-00030. Dennis Jancsary also wishes to thank SCANCOR and the Weatherhead Center for International Affairs at Harvard University for providing an excellent and supportive working environment. We wish to express our gratitude to the participants of the 2nd International Conference: Visuality, Materiality, Multimodality in Copenhagen for their inspiring comments. In particular, we are indebted to our two reviewers for their in-depth engagement with our ideas and their valuable suggestions.

References

- Bell, E., & Davison, J. (2013). Visual management studies: Empirical and theoretical approaches. *International Journal of Management Reviews*, 15(2), 167-184.
- Bell, E., Warren, S., & Schroeder, J. E. (Eds.). (2014). *The Routledge companion to visual organization*. Abingdon, UK, New York, NY: Routledge.
- Berger, P. L., & Kellner, H. (1984). *Für eine neue Soziologie. Ein Essay über Methode und Profession*. Frankfurt/Main: Fischer Wissenschaft.
- Berger, P. L., & Luckmann, T. (1967). *The social construction of reality. A treatise in the sociology of knowledge*. New York, NY: Anchor Books
- Bohnsack, R. (2008). The interpretation of pictures and the documentary method [64 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 9(3).
- Boltanski, L., & Thévenot, L. (2006). *On justification: Economies of worth*. Princeton and Oxford: Princeton University Press.
- Boxenbaum, E., Daudigeos, T., Pillet, J. C., & Colombero, S. (2018). Multimodal construction of a rational myth: Industrialization of the French building sector from 1945 to 1970. *Research in the Sociology of Organizations*, 54(B), 3-36.
- Carley, K. (1993). Coding choices for textual analysis: A comparison of content analysis and map analysis. *Sociological Methodology*, 23, 75-126.
- Cartel, M., Colombero, S., & Boxenbaum, E. (2018). Towards a multimodal model of theorization processes. *Research in the Sociology of Organizations*, 54(A), 153-182.
- Creed, W. E. D., Scully, M. A., & Austin, J. R. (2002). Clothes make the person? The tailoring of legitimating accounts and the social construction of identity. *Organization Science*, 13(5), 475-496.
- Delmestri, G., Oberg, A., & Drori, G. S. (2015). The unbearable lightness of university branding. *International Studies of Management & Organization*, 45(2), 121-136.
- Drori, G. S., Delmestri, G., & Oberg, A. (2016). The iconography of universities as institutional narratives. *Higher Education*, 71, 163-180.
- Franzosi, R. P. (2010). Sociology, narrative, and the quality versus quantity debate (Goethe versus Newton): Can computer-assisted story grammars help us understand the rise of Italian fascism (1919–1922)? *Theory and Society*, 39(6), 593-629.
- Friedland, R. (2009). Institution, practice, and ontology: Toward a religious sociology. *Research in the Sociology of Organizations*, 27, 45-83.
- Friedland, R., & Alford, R. R. (1991). Bringing society back in: Symbols, practices, and institutional contradictions. In W. W. Powell & P. J. DiMaggio (Eds.), *The new institutionalism in organizational analysis* (pp. 232-263). Chicago, IL: University of Chicago Press.
- Friedland, R., Mohr, J. W., Roose, H., & Gardinali, P. (2014). The institutional logics of love: measuring intimate life. *Theory and Society*, 43(3-4), 333-370.

- Graves, O. F., Flesher, D. L., & Jordan, R. E. (1996). Pictures and the bottom line: The television epistemology of U.S. annual reports. *Accounting, Organizations and Society*, 21(1), 57-88.
- Halliday, M. A. (1976). *System and function in language: Selected papers* (edited by G. Kress). London: Oxford University Press.
- Halliday, M. A., & Hasan, R. (1989). *Language, context, and text: Aspects of language in a social-semiotic perspective*. Oxford, NY: Oxford University Press.
- Hiss, S. (2009). From implicit to explicit corporate social responsibility: Institutional change as a fight for myths. *Business Ethics Quarterly*, 19(3), 433-451.
- Höllerer, M. A. (2013). From taken-for-granted to explicit commitment: The rise of CSR in a corporatist country. *Journal of Management Studies*, 50(4), 573-606.
- Höllerer, M. A., Jancsary, D., & Grafström M. (2017). 'A picture is worth a thousand words': Multimodal sensemaking of the global financial crisis. *Working paper, WU Vienna University of Economics and Business*.
- Höllerer, M. A., Jancsary, D., Meyer, R. E., & Vettori, O. (2013). Imageries of corporate social responsibility: visual recontextualization and field-level meaning. *Research in the Sociology of Organizations*, 39(B), 139-174.
- Jancsary, D., Höllerer, M. A., & Meyer, R. E. (2016). Critical analysis of visual and multimodal texts. In R. Wodak & M. Meyer (Eds.), *Methods of Critical Discourse Studies* (3rd ed., pp. 180-204). Los Angeles, CA: Sage.
- Jones, C., Boxenbaum, E., & Anthony, C. (2013). The immateriality of material practices in institutional logics. *Research in the Sociology of Organizations*, 39(A), 51-75.
- Jones, C., & Livne-Tarandach, R. (2008). Designing a frame: Rhetorical strategies of architects. *Journal of Organizational Behavior*, 29(8), 1075-1099.
- Jones, C., Maoret, M., Massa, F. G., & Svejnova, S. (2012). Rebels with a cause: Formation, contestation, and expansion of the de novo category "Modern Architecture," 1870–1975. *Organization Science*, 23(6), 1523-1545.
- Jones, C., Meyer, R. E., Jancsary, D., & Höllerer, M. A. (2017). The material and visual basis of institutions. In R. Greenwood, C. Oliver, R. E. Meyer & T. Lawrence (Eds.), *The Sage handbook of organizational institutionalism* (2nd ed., pp. 621-646). Los Angeles et al.: Sage.
- Kress, G. (2010). *Multimodality. A social semiotic approach to contemporary communication*. Abingdon, UK, New York, NY: Routledge.
- Kress, G., & van Leeuwen, T. (2006). *Reading images. The grammar of visual design* (2nd ed.). London, UK, New York, NY: Routledge.
- Lefsrud, L., Graves, H., & Phillips, N. (2013). Dirty oil, ethical oil: Categorical illegitimacy and the struggle over the Alberta Oil Sands. *Academy of Management Proceedings*, 2013(1).
- Lefsrud, L. M., Graves, H., & Phillips, N. (2018). Dirty oil or ethical oil? Visual rhetoric in legitimization struggles. *Research in the Sociology of Organizations*, 54(B), 101-142.
- Loewenstein, J., Ocasio, W., & Jones, C. (2012). Vocabularies and vocabulary structure: A new approach linking categories, practices, and institutions. *The Academy of Management Annals*, 6(1), 41-86.
- Lok, J., Creed, W. E. D., DeJordy, R., & Voronov, M. (2017). Living institutions: Bringing emotions into organizational institutionalism. In R. Greenwood, C. Oliver, R. E. Meyer & T. Lawrence (Eds.), *The Sage handbook of organizational institutionalism* (2nd ed., pp. 691-620). Los Angeles et al.: Sage.
- Lueger, M., Sandner, K., Meyer, R. E., & Hammerschmid, G. (2005). Contextualizing influence activities: An objective hermeneutical approach. *Organization Studies*, 26(8), 1145-1168.
- Machin, D., & Mayr, A. (2012). *How to do critical discourse analysis: A multimodal introduction*. Los Angeles et al.: Sage.
- Matten, D., & Moon, J. (2008). "Implicit" and "explicit" CSR: A conceptual framework for a comparative understanding of corporate social responsibility. *Academy of Management Review*, 33(2), 404-424.
- Matthiessen, C. (2015). Register in the round: Registerial cartography. *Functional Linguistics*, 2(9), 1-48.
- Meyer, R. E. (2014). 'Re-localization' as micro-mobilization of consent and legitimacy. In G. S. Drori, M. A. Höllerer & P. Walgenbach (Eds.), *Global themes and local variations in organization and management: Perspectives on glocalization* (pp. 79-89). New York, NY, Abingdon, UK: Routledge.
- Meyer, R. E., Egger-Peitler, I., Höllerer, M. A., & Hammerschmid, G. (2014). Of bureaucrats and passionate public managers: Institutional logics, executive identities, and public service motivation. *Public Administration*, 92(4), 861-885.
- Meyer, R. E., & Höllerer, M. A. (2010). Meaning structures in a contested issue field: A topographic map of shareholder value in Austria. *Academy of Management Journal*, 53(6), 1241-1262.
- Meyer, R. E., & Höllerer, M. A. (2016). Laying a smoke screen: Ambiguity and neutralization as strategic responses to intra-institutional complexity. *Strategic Organization*, 14(4), 373-406.
- Meyer, R. E., Höllerer, M. A., Jancsary, D., & van Leeuwen, T. (2013). The visual dimension in organizing, organization, and organization research. *The Academy of Management Annals*, 7(1), 487-553.

- Meyer, R. E., Jancsary, D., Höllerer, M. A., & Boxenbaum, E. (forthcoming). The role of verbal and visual text in the process of institutionalization. *Academy of Management Review*.
- Mills, C. W. (1940). Situated actions and vocabularies of motive. *American Sociological Review*, 5(6), 904-913.
- Mohr, J. W. (1998). Measuring meaning structures. *Annual Review of Sociology*, 24(1), 345-370.
- Mohr, J. W., & Duquenne, V. (1997). The duality of culture and practice: Poverty relief in New York City, 1888-1917. *Theory and Society*, 26(2), 305-356.
- Mohr, J. W., & Lee, H. K. (2000). From affirmative action to outreach: Discourse shifts at the University of California. *Poetics*, 28(1), 47-71.
- Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23), 8577-8582.
- Ocasio, W., Loewenstein, J., & Nigam, A. (2015). How streams of communication reproduce and change institutional logics: The role of categories. *Academy of Management Review*, 40(1), 28-48.
- Oliveira, F. M., Islam, G., & Toraldo, M. L. (2018). Multimodal imaginaries and the “Big Worm”: Materialities, artefacts and analogies in São Paulo’s urban renovation. *Research in the Sociology of Organizations*, 54(A), 27-62.
- Peirce, C. S. (1991). *Peirce on Signs: Writings on Semiotic*. UNC Press.
- Phillips, N., Lawrence, T. B., & Hardy, C. (2004). Discourse and Institutions. *The Academy of Management Review*, 29(4), 635-652.
- Rose, G. (2007). *Visual methodologies: An introduction to the interpretation of visual materials* (2 ed.). London, UK, Thousand Oaks, CA, New Delhi, India: Sage.
- Rowley-Jolivet, E. (2004). Different visions, different visuals: A social semiotic analysis of field-specific visual composition in scientific conference presentations. *Visual Communication*, 3(2), 145-175.
- Scott, M. B., & Lyman, S. M. (1968). Accounts. *American Sociological Review*, 33(1), 46-62.
- Sillince, J. A. A., & Barker, J. R. (2012). A tropological theory of institutionalization. *Organization Studies*, 33(1), 7-38.
- Styhre, A. (2010). Knowledge work and practices of seeing: Epistemologies of the eye, gaze, and professional vision. *Culture and Organization*, 16(4), 361-376.
- Thornton, P. H., Ocasio, W., & Lounsbury, M. (2012). *The institutional logics perspective: A new approach to culture, structure and process*. Oxford, UK, New York, NY: Oxford University Press.
- Toubiana, M., & Zietsma, C. (2016). The message is on the wall? Emotions, social media and the dynamics of institutional complexity. *Academy of Management Journal*, doi: 10.5465/amj.2014.0208.
- Zhao, W. W. (2018). Protest in style: Exploring multimodal concision in rhetorical artifacts. *Research in the Sociology of Organizations*, 54(A), 119-149.

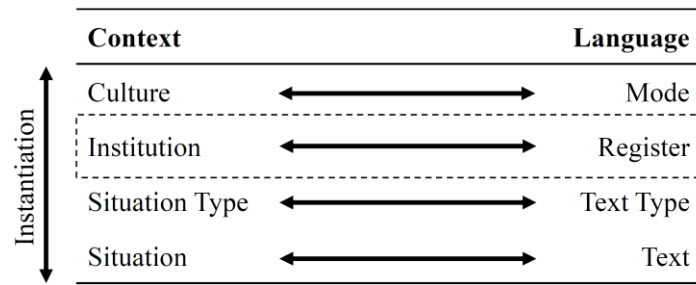


Figure 1: Relationship between mode, register, and institution (adapted from Matthiessen, 2015)

Circles: Ideational aspects of the visual data (participants, processes, and settings)

Triangles: Topics as reconstructed from the verbal chapter headlines

Squares: Interpersonal aspects of the visual data (contact, distance, and angle)

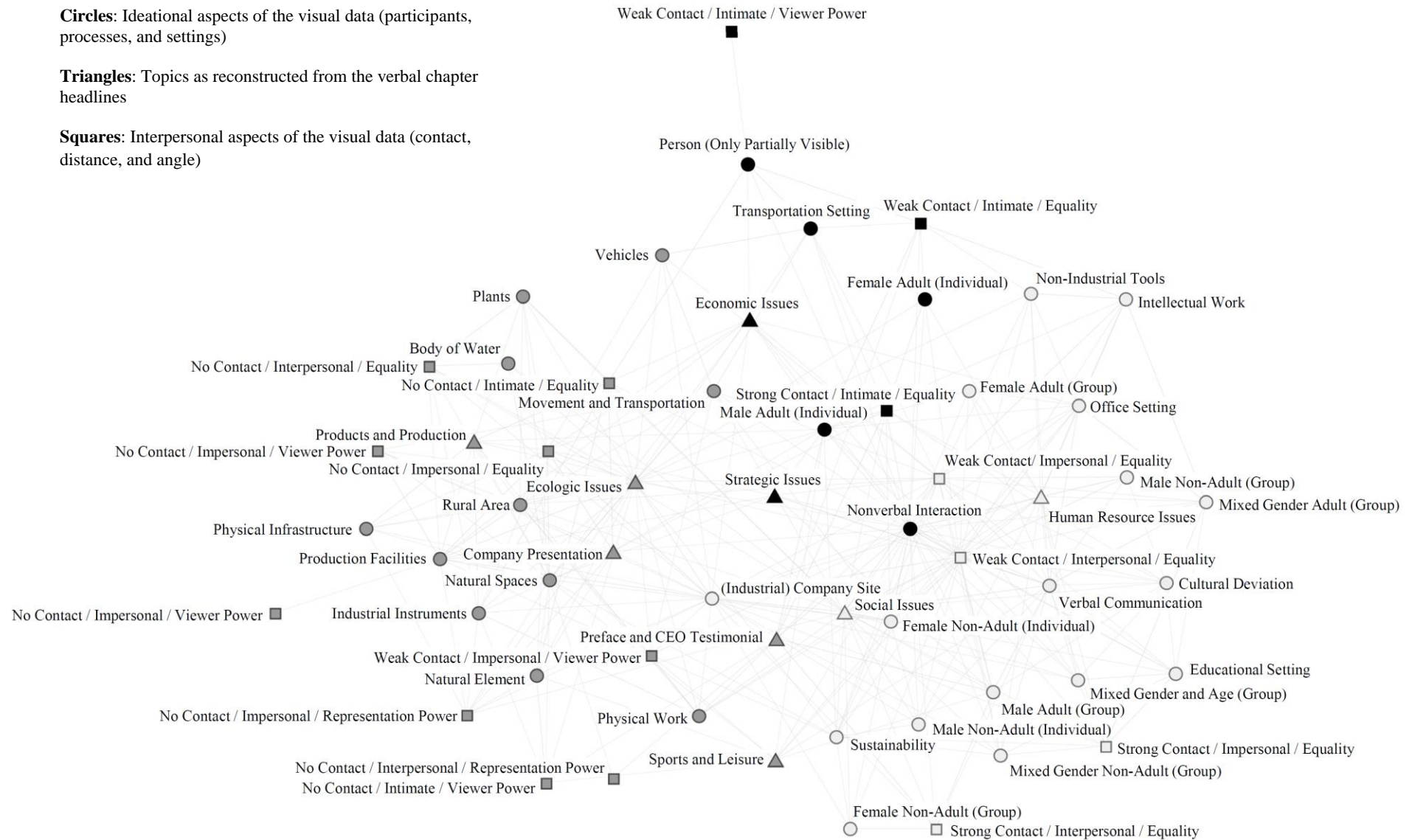


Figure 2: Central 'gazes' suggested in the visual register

Aspects of the visual register	Coding	Relevance for institutional inquiry (exemplary list)
<i>Content of representation</i>	<i>Content:</i> Participants, attributes, processes, settings	Institutions and materiality Category manifestations Central truth claims
<i>Conjunction between elements</i>	<i>Relations:</i> Narrative, conceptual (analytical or symbolic)	Internal institutional ‘architecture’ Category conventions Power relations

Table 1: Capturing the ideational metafunction (summary)

Aspects of the visual register	Coding	Relevance for institutional inquiry (exemplary list)
<i>Embodied position</i>	<i>Contact</i> : Strong contact, weak contact, no contact <i>Social distance</i> : Intimate, interpersonal, impersonal <i>Vertical angle</i> : Viewer power, representation power, equality <i>Horizontal angle</i> : Involvement, detachment	Construction of actorhood Power relations Relational aspects of logics Institutional types of ‘gaze’ Boundaries of institutions and institutional domains
<i>Coding orientations</i>	Naturalistic, sensory, abstract, technological	Aesthetic and emotional aspects of institutions

Table 2: Capturing the interpersonal metafunction (summary)

People	Participants		Processes	Settings
	Nature	Material artifacts		
Male adults (35%)	Plants (3%)	Production facility (12%)	Nonverbal interaction (28%)	Non-descript (47%)
Female adults (14%)	Natural elements (3%)	Industrial instruments (5%)	Verbal communication (16%)	(Industrial) company sites (16%)
Individual body parts (6%)	Bodies of water (2%)	Non-industrial instruments (3%)	Intellectual work (6%)	Natural spaces (12%)
Female non-adults (6%)		Vehicles (3%)	Physical work (5%)	Rural areas (8%)
Male non-adults (6%)		Physical infrastructure (3%)	Movement and transportation (4%)	Office settings (8%)
			Sports and leisure (4%)	Transportation and mobility (3%)
				Educational setting (2%)

Table 3: Content representations (in percent of total images; n=1,023)

Aspects of embodied positions	Codes		
<i>Contact</i>	Strong contact	Weak contact	No contact
	269 (26%)	374 (37%)	388 (38%)
<i>Distance</i>	Intimate	Interpersonal	Impersonal
	360 (35%)	302 (30%)	379 (37%)
<i>Power</i>	Equality	Viewer power	Representation power
	944 (92%)	29 (3%)	57 (6%)




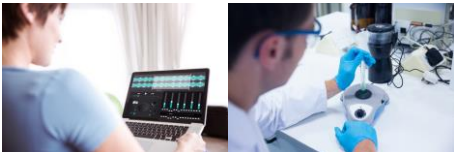


Table 4: Interactional aspects of the visual register (in percent of total images; n=1,023)²

² Please note that some visuals were composed of several images, which means that any visual could potentially be coded with multiple codes per category. Consequently, the numbers add up to more than 100%.

	Intimate	Interpersonal	Impersonal
<i>Strong contact</i>	157	66	52
<i>Weak contact</i>	74 [VP]	176	134 [VP]
<i>No contact</i>	131 [VP]	62 [VP] [RP]	195 [VP+] [RP]

VP = Viewer power; RP = Representation power

Table 5: The ‘gaze’ as combination of contact, distance, and power (n=1,023)

	Intimate	Interpersonal	Impersonal
<i>Strong contact</i>	<p>These visuals show people³ (~94%) engaged in non-verbal interaction (~54%) in non-descript settings (~91%). These are the classic ‘portraits’, either of board members or experts. They are most commonly found in the preface of the report (~22%), the company presentation (~11%) or on social (~12%) and HR issues (~11%).</p> 	<p>These visuals still show primarily people (~76%), but also mobile (~18%) and immobile (~11%) artifacts. People are engaged in non-verbal (~52%) and verbal (~21%) communication in non-descript (~51%), industrial (~15%) and office (~12%) settings. Such visuals either present the whole board of directors or show various people presenting artifacts to the viewer. They primarily appear in the preface (~23%) and on social issues (~20%).</p> 	<p>Again, people (~60%) are the most dominant participants, but also mobile artifacts (~29%). They engage in non-verbal (~40%) and verbal (~38%) interaction, as well as in sports and leisure activities (~13%). Settings are non-descript (~40%), or industrial (~17%) and education (~15%) environments. In such visuals, groups of people are introduced, or they present trophies at celebrations. Their place in the report is in sections on social (~27%), HR (~13%) and ecological (~12%) issues.</p> 
<i>Weak contact</i>	<p>While people (~55%) are the dominant participants, body parts (~27%) and mobile artifacts (~15%) are also shown. People are engaged in non-verbal (~40%) and verbal (~12%) interaction as well as intellectual work (~19%) in non-descript (~54%), and workplace (~12%) settings, or in vehicles (~14%). Such imagery is relevant in the preface (~16%) and sections on social issues (~11%).</p> 	<p>The dominance of people (~84%) is very strong here. They engage in non-verbal (~44%) and verbal (~43%) interaction and intellectual work (~11%) in non-descript (~38%), office (~22%) and industrial (~17%) settings. Such imagery is related to social (~16%), HR (~15%) and ecological (~11%) issues.</p> 	<p>The participants are composed of people (~53%), mobile (~29%), and immobile (~12%) artifacts. Again, actions are verbal (~31%) and non-verbal (~22%) interactions and intellectual work (~15,7%), performed in non-descript (~23%), industrial (~16%), office (~13%), and rural (~13%) settings. Topics are ecological (~22%) and social (~22%) issues.</p> 

³ People and objects are counted only if they are the most salient content of the image – otherwise all images providing any kind of contact would feature people to a degree of 100%.

No contact

When no people are featured, intimate distance is often applied to mobile artifacts (~41%) and natural objects (~12%). Such visuals are usually rather static, but for some natural events (~5%). Settings are non-descript (~80%). Such imagery is found in sections on ecological (~18%) and social (~11%) issues, as well as on production technologies (~11%).

At personal distance, mobile artifacts (~31%), natural objects (~29%), and immobile artifacts (~29%) dominate. Natural events feature in 5% of visuals. Settings are non-descript (~42%), natural (~32%), and industrial (~15%). Related topics are ecological issues (~21%) and production technologies (~11%).

Here, immobile artifacts (~66%) are the dominant participant, followed by natural objects (~14%) and mobile artifacts (~12%). Dynamics in the visuals are constituted primarily by moving vehicles (7%). Settings are industrial (~37%), natural (~28%), and rural (21%). Such imagery features in chapters on ecological (~36%) and social (~11%) issues, as well as in company presentation (~16%).



For copyright reasons, it was not possible to include the original pictures from our data set; still we wish to illustrate the categories of our coding scheme with ‘typical’ images as closely as possible. We therefore use stock photography; all images are used under license from Shutterstock.com.

Table 6: Merging interpersonal and ideational metafunctions