Multimodal Coherence Revisited: Notes on the Move From Theory to Data in Annotating Print Advertisements

Hartmut Stöckl* and Jana Pflaeging

Department of English and American Studies, English and Applied Linguistics, University of Salzburg, Salzburg, Austria

The present paper proposes a preliminary annotation framework for the study of multimodal coherence in print ads. The framework is firmly based on a review of current theorizing on multimodal cohesion and coherence. Following guidelines for good empirical practice, the annotation scheme is tested on 50 ad samples, seeking to illustrate genre-specific qualities of multimodal coherence and to identify difficulties in applying conceptual frameworks to data. The paper employs six dimensions of multimodal coherence: layout, cohesive ties, info-linking, relational propositions, mode-centricity and multimodal rhetorical operations. In conclusion, we outline some of the more general challenges that multimodal annotation raises when it makes close contact with data.

Keywords: multimodal coherence, corpus, annotation, print advertising, empirical multimodality research

INTRODUCTION – BACKGROUND AND AIMS

Despite an increasing intensity of work in multimodality research (see e.g., Klug and Stöckl, 2016; Bateman et al., 2017; Wildfeuer et al., 2019; Pflaeging et al., 2021), multimodal coherence must still count as an under-researched area of study. The present article seeks to improve our grip on multimodal coherence, both with a view to its conceptual modeling and its empirical analysis. The point of departure in this methodological sketch is a view of multimodality as "textual combinations of different modes and their integration in terms of structure, discourse semantics, and rhetorical function" (Stöckl, 2019a, p. 50). This definition interprets the very nature of multimodality as residing in intermodal cohesion/coherence. It also highlights the three facets of text-connectedness, namely:

(1) **Structure:** structure-based mode linking in the form of cohesive ties,
(2) **Discourse semantics:** sense continuity generated from meanings expressed in the different modes,
(3) **Rhetorical function:** a genre-specific, task-based functionality resulting both from structural and semantic texture.

In relation to the corpus of print advertisements (see Stöckl, 2021) underlying the method-building here, this triad may be concretized: The ads construct cohesive ties between words/expressions in ad copy, headlines or slogans and visual elements in images (sometimes also involving typographic resources). Based on these ties, but even in their absence, recipients will relate verbal and visual propositions in order to construct plausible and contextually relevant multimodal meaning(s)

in the form of an overall discourse hypothesis. All this semiotic work between advertiser and consumer is performed with a knowledge of the genre conventions of advertising in mind, which stipulate that any multimodal coherence must be instrumental in constructing an argument. Such commercial or pro-social arguments contain claims about and descriptions of the superior quality of the commodity/brand/service in question and offer reasons or evidence to support these claims (see Ripley, 2008).

The main aim of this article is to interrogate relevant theory in order to suggest and critically reflect a possible framework for the corpus-based analysis of *multimodal coherence in print ads*. We deliberately shift the focus away from multimodal cohesion with its emphasis on formal mode-connectedness (see Sanchez-Stockhammer and Schubert, 2022) to a consideration of coherence, which we regard as the more comprehensive concept with greater explanatory reach and power. In contrast to case studies, which allow for broad discursive interpretations, we acknowledge the need in corpus-based studies to work with clear analytical criteria and to suggest concrete ways of annotating for various multimodal properties. A central tenet in our methodological suggestions is the general recognition of the genre-sensitivity of multimodal coherence. This means that the modeling must be able to explain the special ways in which print advertisements construct multimodal coherence and how these meaning-making processes are constrained and guided by the properties of the genre.

Our explorations of multimodal coherence in the advertising genre are based on a corpus of print advertisements sampled from recent volumes of *Lürzer's International Archive (2019/2020)*, an internationally acclaimed monthly that records current trends in avant-garde advertising. It holds graphically outstanding, often award-winning campaigns that represent a particular type of advertising, which can be called visually refined, rhetorically articulate and verbally minimalist. *Lürzer's International Archive (2019/2020)* receives submissions from all corners of the globe and covers a total of 32 product types. It thus reflects very many different ad cultures and all the thematic variety there is in the advertising business.

The corpus on which our research is based was originally compiled for a study on multimodal argumentation (see Stöckl, 2021). This research interest resulted in a focused selection of ads that use so-called doctored images, i.e., computer-generated printed images (Stöckl, 2021, p. 194), which are particularly interesting for their visual rhetorical qualities. Initially, the corpus comprised 232 print ads from the six 2019 volumes of *Lürzer's International Archive (2019/2020)* and has since been expanded to include samples from 2020 as well, currently counting 377 ads. For the purposes of the present paper, a sub-set of 50 ads was established and explored with a view to aspects of multimodal coherence.

Section A Review of Theory below reviews some of the main previous work in the study of multimodal coherence (and cohesion), gauging promising directions and preparing guidelines for our own proposal. Section Making Contact with Data outlines some core principles of sound empirical work in multimodality and introduces several strategies to

support the move from theoretical constructs to a fine-grained description of a given data set. Using a sufficiently wide and varied selection of sample ads (see **Appendix**), Section Toward Annotation exemplifies the individual criteria in the annotation for multimodal coherence and explains their inter-relatedness. Section Summary and Conclusions is a critical reflection of the framework proposed, which identifies some general challenges in working the empirical cycle from theory to data.

A REVIEW OF THEORY

Since its inception by Halliday and Hasan (1976), studies in text-connectedness have variably and often interchangeably used the notions of *cohesion* and *coherence*. Coherence may be regarded as a text-external property that "is ascribed by participants" (Bublitz, 1989, p. 39), or "as an internal property of discourse, which characterizes it as constructed out of component parts" (Taboada, 2019, p. 205). Moreover, discourse coherence has pragmatically been understood as a felicity condition in the comprehension of texts (Taboada, 2019, p. 205) or, in a logical-semantic view, as "the absence of non-sequiturs" (Taboada and Habel, 2013, p. 66). There is a consensus to refer to the means of text-internal connectedness as cohesion and to any sense-continuity constructed with or without the help of formal signals or cues as coherence.

With this terminological demarcation in place, the interrelation between cohesion and coherence becomes relevant. Carrell (1982, p. 484) strongly cautions against "taking certain aspects of linguistic form as the cause, and not the effect, of coherence." On the other hand, as Tanskanen (2006) shows, it is also evident that cohesive devices support the coherent interpretation of a text in various ways. Taboada (2019, p. 212) suggests the interesting concept of a "space of coherence relations" to reconcile cohesion and coherence. She proposes entity relations between items in the structure (cohesion) and propositional relations between concepts (coherence), which jointly make up the *texture* of a text. The space she talks about is "the space between structural organization at the genre level and the syntactic level" (Taboada, 2019, p. 213).

In standard views (see e.g., Bateman et al., 2017, p. 128–135; Stöckl, 2019a, p. 50–61), multimodality is defined as the integration of semiotic modes in a process of formal (by configuring elements in space and time), semantic (by creating cohesive ties and discourse relations between elements) and rhetorical (by responding to a task-driven goal) mode-linking. The result of such inter-semiotic integrational processes is a multimodal structure, or "texture" (see Liu and O'Halloran, 2009), which naturally follows the demands of coherence. In this sense, the concepts of intra-modal cohesion/coherence readily extend to multimodal communicative artifacts and events. Put differently, the very notion of *multimodality* necessitates or inherently involves the notions of *coherence* and *cohesion*.

Not surprisingly then, Bateman's seminal theory review of multimodal coherence (see also Bateman, 2014a, p. 117–136, 159–221; Bateman, 2014b) collects and orders classic approaches to mono-/intra-modal coherence in order to show

their applicability to text-image relations. It essentially comprises cohesion/cohesive ties, clause-relations (grammar), logico-semantics, relational propositions (as discussed in Rhetorical Structure Theory, RST) and rhetorical figures. Interestingly, few of these approaches are flagged as being explicitly about coherence; this is especially true of the ones based on rhetorical figures. Below, we briefly review selected work on text-image relations representative of the various approaches, highlighting strengths and weaknesses.

Royce (1998) may be regarded as the starting point of attempts to adopt the linguistic notion of *cohesion* as developed by Halliday and Hasan (1976) to multimodal text. Based on the general complementarity of verbal and visual meanings, Royce argues it is inter-semiotic sense relations like repetition, hyponymy or collocation that build multimodal cohesion on the ideational meta-functional level. The study also suggests that the modes interact on the interpersonal (e.g., forms of address) and the compositional (e.g., salience and info-value) levels and underlines that studying the formal structure of both modes is the prerequisite for modeling their complementarity. Tseng (2013) works in the same functionalist paradigm and employs notions of *cohesive chains* to film structure, also extending the framework to action patterns. Sanchez-Stockhammer and Schubert (2022) is work that continues the paradigm of multimodal cohesion, studying the specific nature of cohesive ties relative to the requirements of different genres.

Kong (2006) can be seen as a typical study of multimodal coherence as essentially built through discourse relations. It proposes a taxonomy of "word-image connections" (Kong, 2006, p. 209) that combines multiple layers: logico-semantic relations, status relations, spatial arrangement of relations, meta-functions, as well as evaluative and metaphorical orientation of the relations. Bateman (2014a, p. 188–190) labels this approach "grammatical" as it explicitly draws on Hallidayan functional grammar, yet it also employs notions of RST and shows an awareness of layout. Similar work that is based on logico-semantics has been promoted by Unsworth (2007) and has been shown to lend itself to studying the comprehension of multimodal texts (e.g., Daly and Unsworth, 2011).

Liu and O'Halloran (2009) are one of the few to establish a close and explicit link between inter-modal cohesion and multimodal coherence. Their proposal of intersemiotic texture as "the crucial property of coherent multimodal texts" (Liu and O'Halloran, 2009, p. 367) contains a comprehensive framework for tracking inter-semiotic cohesive devices that is linked with a generalized account of logico-semantics. Liu and O'Halloran argue that for coherence to emerge from inter-semiotic cohesion, recipients need to understand texture as cues for co-contextualizing or recontextualizing semiotic choices in different modes. Multimodal sense-making is understood here as "interaction and negotiation across different modes" (Liu and O'Halloran, 2009, p. 385), which is quite an adequate, even if rather general idea of the generation of coherence.

While not explicitly placed in the field of coherence and distancing itself from classic approaches to cohesion (cf. Bateman, 2008, p. 145), Bateman's GeM-model (Bateman, 2008) has all the makings of a multi-level analytical framework for the

study of multimodal documents. Its six layers (in the revised version, cf. Bateman, 2009, p. 60) help describe how modes are linked to produce coherent interpretations of the structures, semantics and functions of multimodal texts. Layout and navigation attend to the spatial configuration of elements and the use readers make of these; linguistic and content structure look at wording/grammar and propositional content; rhetorical and genre structure direct attention to the ways in which propositions relate and texts follow generic stages. The model, which has been designed with an eye to annotating corpus data, places much emphasis on the role of layout and relational propositions (RST) in generating coherent discourse interpretations. It also links any specifics of multimodal meaning-making to genre, which provides the overriding functional and structural constraints.

By contrast, Taboada and Habel (2013) is a single-level approach to multimodal coherence, which aims to model the phenomenon on the basis of rhetorical relations alone, i.e., relational propositions between verbal utterances and pictorial content. There is no doubt that the application of RST to text-image relations is generally reasonable and productive. However, the study also reveals limitations to do with the inherently difficult delimitation of the various relations and the vagueness and inexplicitness of pictorial propositions. The approach is strong in internal consistency but it could be sharpened by an attention to the verbal and visual elements entering into a relation, i.e., cohesive ties.

Forceville (2020) makes a strong plea for viewing coherence as contextually relevant interpretations of multimodal textures guided by a host of knowledges (i.e., genre, discourse topic, common logic etc.) and resulting in a discourse hypothesis that requires plausibility testing. To Forceville, coherent interpretations need to overcome the pragmatic indirectness and semantic under-specification typical of much multimodal meaning-making, which requires recovering propositional form (explicatures) and drawing inferences (implicatures) from the sign combinations presented. While offering no simple annotation categories, the approach sensitizes us to the dynamics of coherence generation.

Tseronis (2021) is a recent example of work aimed at capturing what may be called multimodal rhetorical figures, i.e., "a minimal unit of meaning consisting of the combination of a certain form with specific content, which can be realized in a variety of modes or combinations thereof" (Tseronis, 2021, p. 377–378). Rather than purely ornamental, such figures have been shown to directly facilitate the argument in promotional genres (see Kjeldsen, 2012). The more general idea here is that just as relational propositions build coherent discourse structures, so multimodal rhetorical figures, too, establish often crucial text structures that are fundamental for a coherent message interpretation (see Rossolatos, 2013 for rhetorical approaches to advertising).

Finally, Ewerth et al. (2021) present computational approaches to interpreting multimodal data from the advertising and the news genre. Their study shows that prominent types of multimodal relations may be learnt by software, which can then be used to recognize them and quantify their occurrence. The five general classes of image-text relations are defined by differences in cross-modal mutual information, semantic

correlation and status relations. The category of *cross-modal mutual information* relies on the identification of verbal and visual elements that provide the information in a cohesive tie. Semantic correlations range from coherent over uncorrelated to contradictory. With computer-vision becoming more easily available and more powerful, this approach also points ways to the automated annotation of multimodal corpus data.

The history of multimodal coherence research generally reveals two main camps: approaches opting for a single-layer framework (e.g., just cohesive ties or just relational propositions) and others that employ a multiple-layer framework, seeking to find an effective combination of analytical criteria.

The review of major approaches to the coherence of text-image relations allows a number of general conclusions to be drawn for theory development and empirical research alike. These will be the guiding principles of our own proposal for studying multimodal coherence in print ads (see Section Toward Annotation).

1. It is advantageous to explicitly link the study of cohesive ties with the ways in which discourse structures are established. There is a need in studies of multimodal coherence to model which elements of text and image are picked up by higher order semantic structure. Typologies of the verbal and image elements that typically function in cohesive ties are a worthwhile spin-off of such studies.
2. Rather than work with one analytical dimension only, it seems advisable to combine a number of them. This does better justice to the very notion of *multimodal coherence* and prevents us from missing out on central observations. Multiple-layer frameworks may come at the cost of greater redundancy, but this can be phased out later in the annotation. Logico-semantic and relational propositions seem vital ingredients in any framework just as are cohesive ties.
3. While not seemingly semantic in nature, layout may prescribe or suggest likely ways of mode integration. It is plausible that a given genre like the printed ad employs a number of layout patterns, which may be correlated with types of status, for instance, or help with determining relational propositions or rhetorical figures in RST.
4. Naturally, annotations will be based on the sets of categories the various approaches make available. Before anything can be annotated, we must settle on a coherent interpretation of the multimodal ad in question. Such discourse hypotheses may be facilitated by a relevance-theoretical reconstruction of propositions and their relations. Also, concrete texts may point up the necessity to expand a set of categories or to revise it.
5. The single most important factor guiding the deployment of coherent multimodal structures and their interpretation is genre. Therefore, annotations ought to capture the genre-typical features on the various levels of coherence description.

MAKING CONTACT WITH DATA

The study of a print-ad corpus with a view to multimodal coherence naturally involves the application of theoretical concepts to concrete texts. In light of current developments in multimodality research, especially the empirical turn (Pflaeging et al., 2021, p. 16), several preliminary remarks seem in order before we turn to aspects of applying theoretical constructs in empirical analysis.

Formed around the practical realization that communication naturally involves *several* semiotic modes, the development of theory and methods in multimodality has always been tied to observations made about authentic data. It can thus be considered empirical (see Pflaeging et al., 2021, p. 9). Despite the long tradition in data-based research, the underlying methodological assumptions and strategies pursued in empirical multimodality research are often only vaguely articulated (Bateman, 2016, p. 37), especially in those areas situated toward the qualitative pole of the empirical continuum. In much previous work on the conceptualization and analysis of multimodal coherence, for instance, a systematic, reliable and data-sensitive application of abstract theoretical constructs has often been taken for granted, but such applications are rare and hardly consistent. This situation has also been subject to substantial criticism of existing proposals for the analysis of text-image relations (see e.g., Forceville, 1999; Bateman, 2014a).

A more productive engagement with multimodal data can be achieved if the logic of the *empirical research cycle* is considered and implemented with due care. To this end, Pflaeging et al. (2021, p. 19–23) propose several quality criteria for good empirical practice. They include established criteria such as objectivity, reliability, and validity (see e.g., Krippendorff, 2004, Ch. 12, 13; also Moosbrugger and Kelava, 2014). While these criteria apply to the empirical research process as a whole, it is useful to subdivide the cycle into two distinct phases. The first half of the cycle accounts for a move *from theory to data*, whereas the second half ensures that analytical results *feed back into theory-building* (see Pflaeging et al., 2021, p. 19). In the present paper, a focus is set on the first half of the empirical cycle, which moves aspects of data annotation into the foreground.

The application of a theoretical framework to data requires what is commonly called *operationalization*. This means that theoretical concepts need to be made "accessible", or "measurable", in order to base hypotheses on a more stable footing. Achieving a viable operationalization of analytical concepts for empirical multimodality research is far from trivial. When dealing with multimodal coherence, especially, we usually seek to get a firm grip on communicative phenomena that are not measurable as such. Instead, just what nuances of meaning are conveyed in a multimodal text is often a matter of a recipient's very own discourse interpretation. Discourse interpretations, in turn, are complex abductive processes that are cued by lexis, visual image elements and grammar, but they also rely on knowledge of discourse topic as well as genre and world knowledge. Therefore, a reliable, recoverable description of such phenomena poses a number of challenges to multimodality scholars who are aiming for an empirically solid description of their data.

Given the comparably high degree of complexity and subjectivity in interpreting discourse, it might not come as a surprise that in previous work on multimodal coherence, several

dimensions of the empirical cycle have remained underexposed. While efforts have been made to establish small-scale taxonomies for empirical analysis, discussions often lack a more explicit articulation of methodological premises, decisions and detailed criteria for application. Also, sufficiently elaborated taxonomies that allow analysts to move gradually from abstract theoretical concepts to more specific, data-sensitive categories are scarce. Such aspects need attending to more in future empirical work (see Bateman, 2016, 2019).

Moving forward, it proves beneficial to draw on existing proposals from disciplines and research fields that have also faced the challenge of operationalizing concepts that seem hard to measure. In empirical social science (see Häder, 2010) as well as in empirically-oriented communication studies (see Krippendorff, 2004; Drisko and Maschi, 2015; also Springer et al., 2015), for instance, scholars have formulated requirements and steps of an empirically sound operationalization of theoretical concepts. Our approach has also been loosely inspired by *Legitimation Code Theory* (LCT, Maton and Chen, 2016), which is meant to frame the more specific methodological steps adopted from work in the social and communication sciences. According to LCT, academic disciplines construct frameworks of *theoretical knowledge* (in ref. to Bernstein, 2000, p. 208, qtd. in Maton and Chen, 2016, p. 29), expressed through comparably abstract categories and their interrelations (Bateman, 2019, p. 300). The application of such abstract frameworks, then, runs the risk of imposing theoretical distinctions onto data in a "cookie-cutter" style (see Maton and Chen, 2016, p. 29; also Bateman, 2019, p. 303). This effect can be avoided, however, by pursuing several strategies, generally referred to as *external languages of description* (Maton and Chen, 2016, p. 29–31).

In the early stages of an empirical research project, these external languages of description serve to suggest particular foci for analysis (Maton and Chen, 2016, p. 30), that is, phenomena worthy of investigation. This step already requires extensive knowledge of existing theoretical proposals, which helps shape an informed gaze. A thorough understanding of a research area's core interests and leading objectives also secures an openness to aspects of the data that are yet to be sufficiently described. In the present study, a survey of previous work on multimodal coherence drew our attention to relational propositions, multimodal cohesive ties and their dependency relations as core themes in the research literature. Taking into account existing frameworks and prominent patterns in our data, we set a particular focus on discourse relations between text and image. As layout has also been shown to play a part in generating multimodal coherence (see e.g., Bateman, 2008; Hiippala, 2013), it complements our framework as a minor analytical focus. Finally, our analysis also covers multimodal rhetorical figures. Not a standard focus in a study of multimodal coherence, they are part and parcel of advertising discourse (see e.g., Durand, 1987; Phillips and McQuarrie, 2004 for visual rhetoric) and were thus included in response to the descriptive demands of our data set. Even though we hope that our analytical agenda already promises a detailed account of multimodal coherence, further descriptive dimensions can always be added, such as argument structure, for instance.

The development of external languages of description also involves establishing elaborate conceptual taxonomies that fill the gap between higher-level theoretical constructs and more specific conceptual variants attestable in a given data set. During analysis, concepts are "divided into or reconceptualized as categories which, through engagement with data, are recursively divided into subcategories" (Maton and Chen, 2016, p. 30). The resulting category systems typically span several levels of abstraction. Intermediate levels of specificity may already suffice to describe particular types of corpus data, whereas lower-level concepts need to be introduced to account for very particular corpora. In the present article, our discussion mainly aims at intermediate-level categories and their operationalization. We believe that even at this level, concepts ought to be neatly defined and their interrelations specified with due care. Also, any other guidelines and criteria that support the annotation process need to be made explicit and benefit from an illustration with appropriate examples from the corpus data (see also Maton and Chen, 2016, p. 31). Only in this fashion do annotations become "reliably recoverable" (Bateman, 2019, p. 305).

Based on previous work on how to best operationalize theoretical concepts into analytical criteria for empirical analysis (Häder, 2010, p. 51–52, 56; Springer et al., 2015, p. 55, 88–89; also Maton and Chen, 2016), we suggest four points that can guide the move from theory to data further:

1. **Detailed Taxonomies:** When applied in annotation, taxonomies ought to be sufficiently differentiated, offering both adequate horizontal scope and vertical depth of terms.
2. **Distinct Concepts:** Individual concepts must be defined clearly, making sure they do not overlap and offer useful instructions to the coders.
3. **Indicators:** It is desirable to have at one's disposal indicators for recognizing or "measuring" the data so as to be able to make and justify the necessary distinctions between analytical categories.
4. **Exemplification:** Terms and criteria may best be understood and demarcated in the analysis, if they can convincingly be demonstrated on examples drawn from the corpus.
5. **Documentation:** Finally, it is crucial to document the annotation process by specifying definitions and analytical criteria in a codebook, and by outlining relevant steps in the form of instructions.

TOWARD ANNOTATION

This section aims at drafting a codebook for exploratory annotations, taking into account previous work on multimodal coherence and our engagement with 50 print ads[1]. As advocated in the previous section, this requires us to be specific about the descriptive scope of analytical categories and to provide strategies for their reliable application. It also entails complementing existing concepts with further analytical categories (at various

---

[1]The 50 advertisements break down into product categories as follows: *accessories* (1), *audio & video* (1), *automotive* (6), *children* (1), *cosmetics* (1), *alcoholic drinks* (3), *soft drinks* (3), *food* (8), *house & garden* (5), *pharmaceuticals* (4), *publishers* (1), *retailers* (1), *services* (5), as well as *social & environment* (10).

FIGURE 1 | Taxonomic overview of concepts for an analysis of layout in print ads.
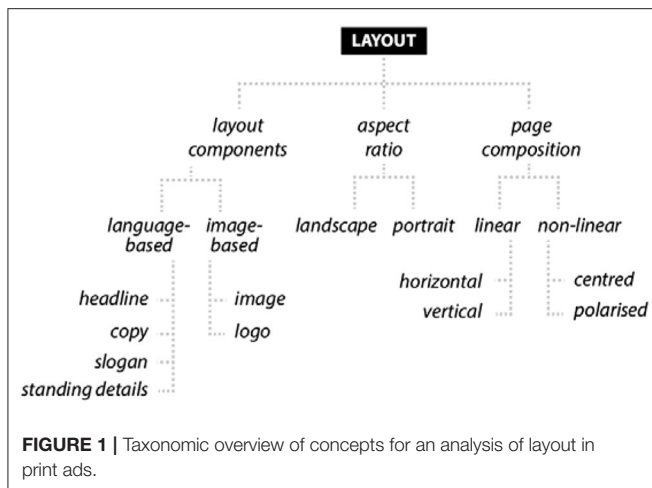
levels) to characterize the various properties of multimodal coherence-building in print advertisements. We do not claim to offer a comprehensive account that is fully sensitive to patterns in our data set. What we offer is a critical appreciation of available taxonomies, which involves a discussion of how to best adjust and apply the dimensions and criteria of analysis to the print-advertising genre. Also, by focusing on several analytical levels, we seek to highlight the scope of variation and the ways in which the various levels of annotation are interrelated. Finally, our work identifies problem areas and points out necessary adjustments in explanatory theory on coherence.

Layout

In previous accounts of print artifacts, a description of layout has only been a minor focus of analysis. Much can be gained, however, if existing approaches are integrated into a broader framework and elaborated further with a view to reliable application (see **Figure 1**). First and foremost, the page space of a print ad is an object of visual perception. Based on Gestalt laws and further principles described in perceptual psychology (Arnheim, 1976; also Bateman, 2008, p. 57–74), recipients recognize and establish distinctions between visually salient masses and volumes, which can be termed *layout components*. As a first step in annotating for layout, coders would thus pay particular attention to instances of shape-defining color contrast between areas of white space and actual compositional elements.

As a second step, layout components are further specified by grouping them in accordance with the semiotic modes they rely on (see **Figure 1**). Compositional elements that are realized through image-based modalities, e.g., the photographic image or vector graphics, and those that rely on the semiotic mode of written language (or text) constitute two major types of components in print advertising. Within text-based elements, we may find a meaningful use of typographical contrast, which yields additional types of elements at yet lower levels of abstraction (see Bateman, 2008, p. 115, 117–121; also Pflaeging, 2021, p. 85–86). Typically, these visual distinctions are functionally motivated and

FIGURE 2 | Example 1: Rao's Homemade, BSSP, Sausalito, CA (USA), Lürzer's Archive 2/19: 36.

correspond to standard generic stages of the print advertising genre, i.e., headline/claim, copy, slogan, and standing details. In an example from our corpus, an advertisement for pasta sauce (see **Figure 2**, rao's homemade_food_usa_la_2-19_36), compositional elements are either image- or text-based. Taking into account the workings of layout, typography and color as well, we can establish five perceptually distinct types of components: a headline/claim (*What homemade sorcery is this?*), a slogan (*Rao's Homemade. Delicious speaks for itself*), standing details (*@dardenunited*), copy (*Simple ingredients. Rich flavor*) as well as a composite image, which combines vector- and photographic elements. It is also interesting that the ad in question configures the semiotic resources of typography and layout according to the conventional genre pattern of playing cards. Such a graphic evocation of genres has also been called *genre embedding* or *mixing* and generally contributes to a coherent interpretation (cf. Stöckl, 2014, p. 291–292). In the example, sorcery is cued and specified by the graphic design; it relates to fortune, perhaps even the supernatural, associated with some card games like tarot.

As a third step in layout analysis, we can ask how much page space has been attributed to layout elements, which is a strong indicator of the perceptual prominence and rhetorical importance of a particular mode. Estimating just how much space a particular layout component covers (relative to the available page space) involves measuring its dimensions manually or with the help of page recognition software. In our sample ad, image-based components take up the largest part of the page space, which increases its visual salience and suggests its centrality to the argument.

Finally, we can describe in what ways elements are distributed across the page. The placement of layout components in close spatial proximity to one another has been shown to suggest some form of rhetorical unity (Bateman, 2008, 2011; Hiippala, 2013; Pflaeging, 2021, p. 82–90). In the example, the clustering of the typographically distinct elements *Rao's Homemade* and *Delicious speaks for itself* suggest more immediate rhetorical relations between them. Looking at page composition more generally, certain types of *aspect ratio* seem to co-occur with particular ways of arranging elements: Landscape formats lend themselves to horizontal compositions. Portrait formats, in turn, seem to favor vertical arrangements. While gaze movement is influenced by a broad variety of perceptual cues (such as an element's color, shape, size; see Holsanova et al., 2006, p. 84) patterns of spatial arrangement are thought to play their part in guiding perception. Thus, text and image components may appear as arranged in a "linear" sequence with either text or image in primary position, also allowing for an alternation of text and image (see burger king_food_italy_la_3-20_59). Even though we should be careful not to simplistically interpret such relative positions in multimodal layout as prioritizing image or text in multimodal meaning-making (see Section Mode-Centricity), they may say something about perceptual salience and entry points to message construal. Two further patterns become evident in the data, which could be labeled *centered* and *polarized*. Centered layouts position either image or text in central position, keeping the peripheral areas empty so as to create maximum focus and impact. Polarized layouts position generic stages in distinctly segregated parts (corners or margins) while keeping the image central (see rao's homemade_food_usa_la_2-19_36, see **Figure 2**). Rather than suggest a logically-semantically relevant sequence like horizontal and vertical pattern, both central and polarized layouts strive for maximum salience and optimal chunking of information.

Advertising layout is anything but straightforward despite the distinct efforts obvious in the patterns discussed above to create either a modal sequence or a modal focus. Two reasons for the complexity of layout in ads are first, the integration of writing into visual image elements (VIEs), such as on packaging (see keraone_cosmetics_ colombia_la_6-19_39, shiner bock_drinks_alc_usa_la_1-19_35, see **Figure 7**) and second, the typo-pictorial play that transfers lines of writing into objects (or vice versa). In one ad, for instance, the claim *COPD doesn't happen overnight* takes the visual shape of cigarette smoke emanating from a heavy smoker (see stiolto_pharma_usa_la_5-20_82).

FIGURE 3 | Taxonomic overview of concepts for an analysis of *cohesion* in print ads.
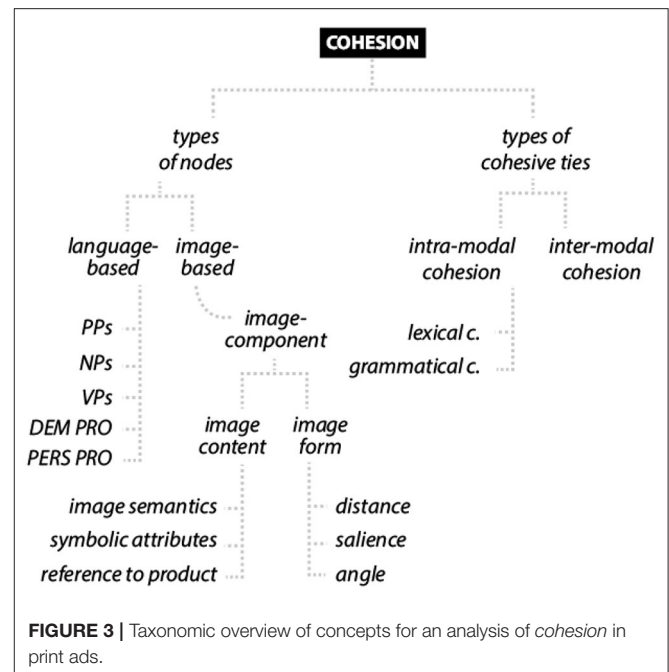
Cohesive Ties

Rather than apply ready-made types of cohesion (see Halliday and Hasan, 1976) or inter-semiotic sense relations (Royce, 1998), we suggest that it is paramount to primarily inspect the mode-specific resources that enter into a cohesive tie. A multimodal cohesive tie is any pair of structural items, e.g., one verbal and the other visual, which establish a semantic dependency relation in discourse interpretation. Instances when a visual image element points to or links with a language item or vice versa would be a clear indicator of such a tie. We believe an inventory of lexical, grammatical and pictorial items that regularly enter into multimodal cohesive ties is useful because it allows us to understand the genre-specific functions of the elements in the ties. In order to compile such a collection of typical cohesive tie constituents, it is necessary to devise a formal and/or semantic classification of the linguistic structures and the types of images or image elements that constitute the ties. For language, a grammatical classification of linguistic material setting up ties seems suitable, as the different structures enable particular semantics. For instance, a prepositional phrase conveniently refers to the location of VIEs relative to one another, e.g., *behind the wall* in freeland_social_thailand_la_5-19_86 refers to the position of a dead deer with the antlers' trophy at the front side of the wall. By contrast, nominal phrases are suited to identifying individual VIEs, whereas idioms enable allusions to visual content or a pun on the image, as in bill-e's_food_usa_la_1-19_51, where *feed your inner pig* cohesively relates to a man whose bared belly button features a pig's tail. Verb phrases may be used to ascribe actions to people represented in an image, or they fill in the details of a depicted event or process (see **Figure 3**).

While a large bulk of cohesive ties in ads seem to be constituted by lexical items, mainly different types of nouns or
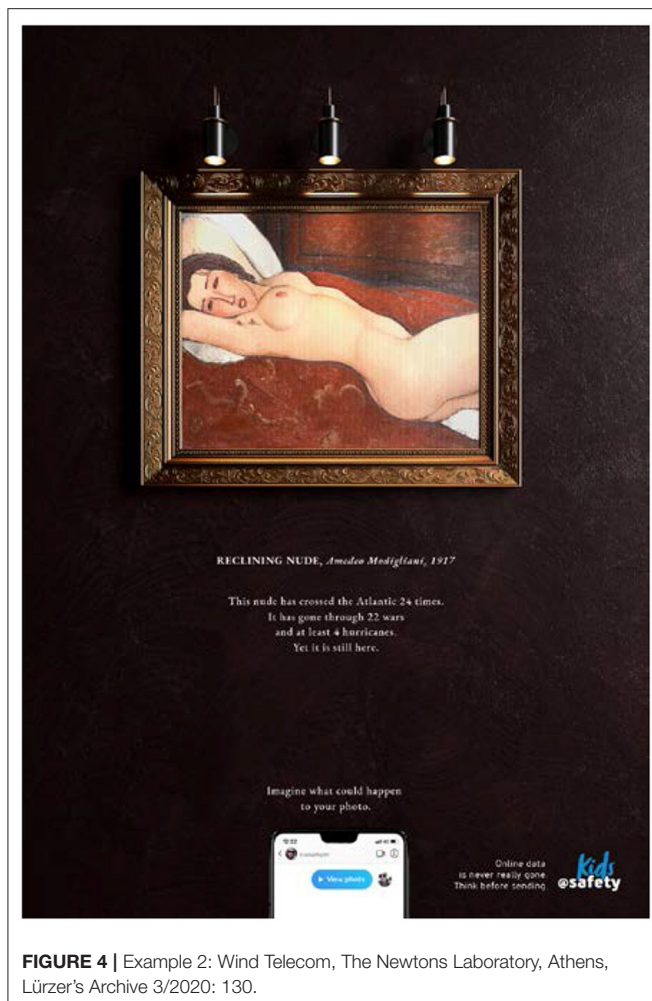
FIGURE 4 | Example 2: Wind Telecom, The Newtons Laboratory, Athens, Lürzer's Archive 3/2020: 130.

names (testimonials, product, brand) and adjectives describing and evaluating qualities of VIEs, demonstrative pronouns are salient cases of grammatical cohesion. "This"/("these"), for instance is generally capable of flexibly pointing to selected VIEs or an entire scene, as in a social advert (the covenant house_social_usa_la_5-20_118), where *this* relates to an instance of abuse on a street and to 3rd Avenue, and *this city* to New York. But 'this' can also refer to a particular image shown in its entirety, as in *this nude* relating to a well-known 1917 Modigliani painting shown in an ad (wind telecommunication_ social_greece_la_3-20_130, see **Figure 4**) to argue that photos stored on mobile phones are in peril from the same dangers that have threatened the classic painting (*Atlantic crossings, wars, hurricanes*). Personal pronouns (he, she, they) are similar grammatical means which effectively relate to people shown in images. A complicating factor in annotating for the linguistic means of setting up cohesive ties is gauging and acknowledging the impact of intra-modal cohesive ties (in the headlines, claims, copies, slogans), which may shape the exact nature of an inter-modal cohesive tie.

A classification of VIEs that become part of an inter-modal cohesive tie is primarily semantic and must respond to a number

of relevant issues. First, it may very generally pinpoint different types of image semantics (see **Figure 3**), such as scenario, action, object, person or fusion to arrive at an idea of the kind of visual semantic complexity that language items cohesively relate to. Fusion, i.e., a morphing of two image elements into one gestalt, seems to offer a particularly rich or complex semantic that text can pick up on. For instance, in a McDonald's ad (mcdonald's_food_germany_la_6-19_45), the image shows an object whose shape is both a burger and the interior of an opera house, allowing the metaphoric claim *A symphony for your senses*.

Second, it is relevant to pay attention to the symbolic attributes (Kress and van Leeuwen, 1996, p. 108–114, see **Figure 3**) of the people and objects depicted as these are vital in making positive evaluations of product and brand. Text might relate to these by way of abstract nouns (e.g., *revolution - Lenin*) or adjectives (e.g., *biodegradable - forest*). An optician's ad, for instance, depicts a man wearing glasses with armor, lance, saddle and windmill, claiming: *Don't worry Sancho, they're just windmills. When you see well, the story changes* (see vimax_accessories_spain_la_3-20_12). The inter-semiotic references here would be impossible without the symbolic attributes to activate the Sancho Panza character from Don Quixote.

A third point concerns visual references to the product (see **Figure 3**), which can be made in the form of decontextualized pack shots, depictions of product use or mere metonymic/metaphorical allusions to the product/brand. Names and class nouns appear to be the main linguistic means to establish inter-semiotic ties with these VIEs. A fusion of coffee cup and drinks can, for instance, provide a direct reference to the product, allowing the simple claim *Proper coffee. Now in a can* (see costa_drinks_soft_uk_1-20_33). Quite a covert visual reference to the product is provided in an ad for batteries, where the product is conspicuously absent and all we see is two sumo wrestlers, whose symbolic attributes facilitate the no-frills claim *powerful batteries* (see nanfu_house_china_la_4-19_70).

A typology of images and VIEs may be further complicated by the fact that it is not only image content but also image design (see **Figure 3**) that can function as a meaningful entity which text may refer to. Such elements of composition, as for instance distance, salience and angle can be referred to by language items, as in a car ad (see ford ranger_automobile_chile_la_1-19_13) where the claim *get out* relates to an extremely low angle shot from down between tower blocks up into the open sky. An exploration of multimodal cohesive ties along the lines sketched here forms the very basis of studying relational propositions (see Section Relational Propositions) as the ties become indicators of elements in rhetorical structure. After all, multimodal message construal must crucially rely on verbal and visual form. Meaning-making will, however, also be affected by rhetorical operations linking text and image (see Section Multimodal Rhetoric). These are often not materialized but must be recovered from world- and discourse knowledge.

Information-Linking

In order to describe how components of a print ad form a coherent whole, we need to move beyond cohesive ties and focus on the level of discourse relations. In social semiotics especially

FIGURE 5 | Taxonomic overview of concepts for an analysis of *information-linking* in print ads.
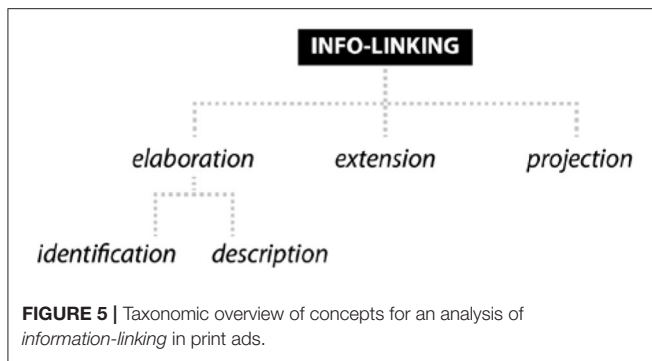
(see e.g., Martinec and Salway, 2005, p. 350; van Leeuwen, 2005, p. 219–230; Unsworth, 2007), it is a widely accepted view that such relations are best understood as instances of information-linking. With regard to text-image relations in particular, several kinds of information-linking have been proposed. Some of them are still drawn directly from frameworks for a description of grammatical relations between clauses (see Halliday and Matthiessen, 2004, p. 376–383; Martinec and Salway, 2005; Unsworth, 2007). Others show a clear orientation toward the discourse-level (van Leeuwen, 2005), which is the stance we are taking here as well.

Following van Leeuwen (2005, p. 230), information-linking can be of two main types (see **Figure 5**). On the one hand, text and image can be linked through an elaboration-relation, that is, cases where "words pick out one of the possible meanings in the image" (van Leeuwen, 2005, p. 229, in ref. to Barthes, 1977 and Halliday, see e.g., Halliday and Matthiessen, 2004). On the other hand, both modes may be related through extension, defined as cases in which "two items - one verbal, one visual - provide different, but semantically related information" (van Leeuwen, 2005, p. 229). In print advertising, both of these relations occur regularly. The coffee ad mentioned earlier (see costa_drinks_soft_uk_1-20_33) is a very simple case of elaboration: *Proper coffee. Now in a can* identifies the two image elements morphed in the visual fusion, "cup of coffee" and "can." Still the image is not redundant, as water droplets on the can convey the idea that the drink is cold/iced coffee. Even in such plain text-image relations, there may be elements of extension, as ads must strive to instrumentalize the images in realizing genre-typical communicative functions or rhetorical tasks. For instance, the telecommunications ad which uses the Modigliani painting (see wind telecommunication_social_greece_la_3-20_130, above) does elaborate it by giving the title (*reclining nude*) and calling it *this nude*, but essentially the copy provides extension, i.e., meta-communicative info on what the painting has survived in history (*crossed the Atlantic 24 times, gone through 22 wars and at least 4 hurricanes*). Most importantly, the copy extends the image to serve in an argument, namely by pointing – in an analogy between art and private photos stored on mobile devices – to the dangers of sending images (of naked people) on social media (*Imagine what could happen to your photo*) and claiming that *online data is never really gone.*
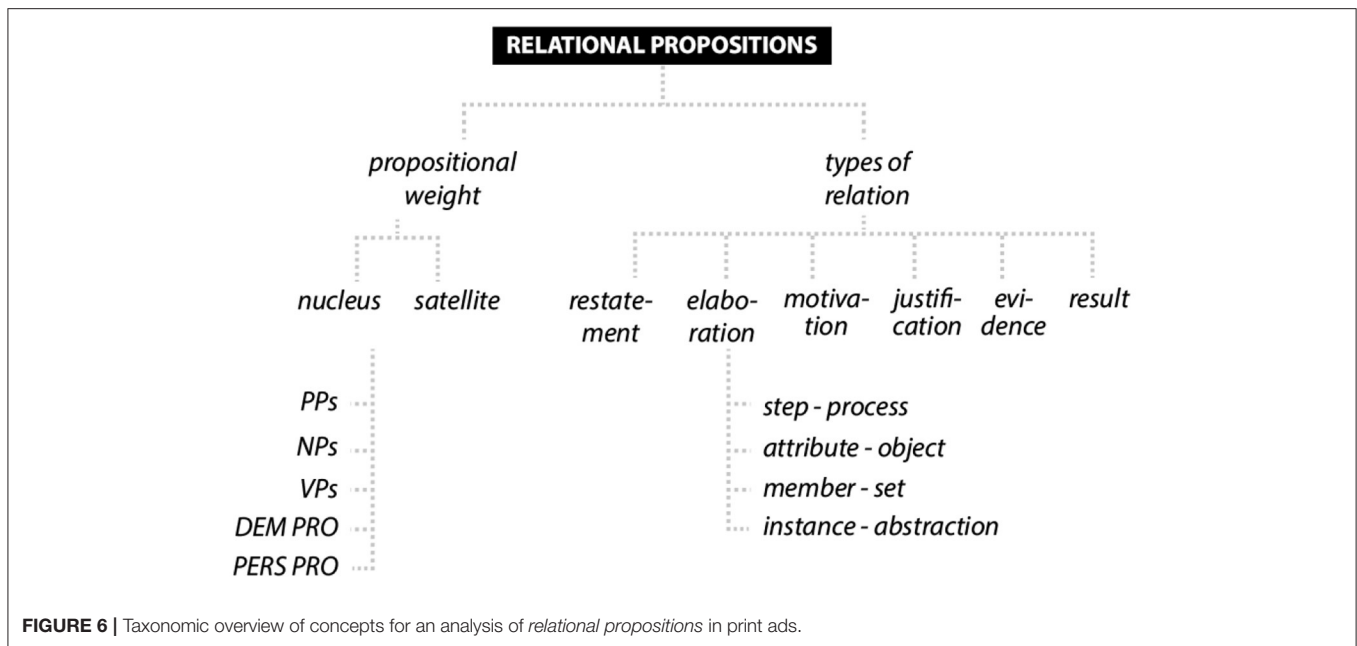
Given the broadness of the concepts, it may not be surprising that elaboration and extension are attestable in a large number

of examples. However, instead of staying at this comparably high level of abstraction, further sub-categories are necessary to account for the variation in discourse patterns we are likely to encounter in specific data sets. To get a firmer grasp of the various phenomena typically subsumed under the headings of *elaboration* and *extension* respectively, van Leeuwen makes a first suggestion. Elaboration, for instance, can occur in the form of explanation, defined as "text makes image more specific," or may be realized as specification, which describes cases where an "image makes text more specific" (van Leeuwen, 2005, p. 230). Extension, in turn, may play out as similarity, contrast or complement. In some cases, however, definitions could benefit from further delineation and specification. Instances where the text content is regarded *similar to* the image content (van Leeuwen, 2005, p. 230) are likely to overlap with *elaboration-*categories. The category *complement*, where the "text/image content adds *further* information" (van Leeuwen, 2005, p. 230), seems rather broad, potentially overlapping with *contrast*.

In an endeavor to provide categories that are more reliably recoverable in analysis, Caple (2013, p. 131) subdivides *elaboration* into identifying/classifying participants, places and points in time and gives a clear orientation as to what units within a text and image support such interpretations (see **Figure 5**). In the photojournalistic image-caption ensembles she analyses, the identification of depicted places can, for instance, be achieved through locative prepositional phrases. In her study of image-caption combinations in science journalism, Pflaeging (2021, p. 120–122, 208–211) complements identification with the concept of *description* to account for instances where the text serves to highlight visible or inherent attributes of image elements, thus expanding the taxonomy horizontally. She also suggests further subtypes of extension, which are naturally harder to operationalize as interpretation is not supported by cross-modal cohesion.

When applied to the case of print advertising, it would be paramount to qualify the kinds of genre-typical extensions the copy produces in linking up to the image. This can likely best be accomplished by treating text as constructing a more or less explicit argument structure, which images are linked into. For example, the pharma ad for an inhaler to treat COPD (see stiolto_pharma_usa_la_5-20_82, above) utilizes the image of an exaggeratedly chain-smoking woman as the problem or cause for its solution (claim): *But in 5 min she can start to breathe better*, which is then substantiated by the following textual proof: *STIOLTO RESPIMAT was shown to improve patients' lung function – no matter how long they have smoked*. In similar ways, images can be extended verbally to function as premises, claim, evidence/proof etc (see Ripley, 2008).

Theoretical accounts of information-linking may also include the relation of *projection*, which presents a markedly different case. Here, the text appears to be a more integral part of the image as it belongs to (or emanates from) the people, animals or fictitious animate beings shown. In an environmental protection ad (see mercado libre_social_colombia_la_2-20_93), a dolphin is represented as saying *Buy straws that never reach the ocean*. As the likely victim of marine pollution by plastic, the speaking dolphin increases the plausibility and urgency of the

FIGURE 6 | Taxonomic overview of concepts for an analysis of *relational propositions* in print ads.

argument. Similar projections lend credibility and authenticity to personal narratives (*My love for gigantic structures was born when the hard helmets I played with dwarfed me*; see avianco cargo_services_colombia_la_6-19_82) or testimonials (*I use metro, the subway store . . .*; see bazar metro_retailer_chile_la_4-19_88) by showing an appropriate character making the relevant statements. It seems obvious then that projection pursues central rhetorical tasks inherent in advertising.

Info-linking may be complicated by the fact that we may wonder about the directionality of elaboration and extension. Had advertising perhaps better be regarded as a genre where images elaborate and extend the semantically more reliable ad copy? Or is it wise to categorically assume that it can only be language that elaborates and extends perceptually dominant images? This question relates to the annotation category of *image-centricity* (see Section Mode-Centricity).

Relational Propositions

Rhetorical Structure Theory (RST, Mann and Thompson, 1986; see Mann and Taboada, 2005–2021) is widely accepted as a suitable framework for the analysis of coherence, which models how various propositions in a stretch of discourse connect semantically. It has been adopted successfully to multimodal coherence, accepting the essentially propositional nature of images (see e.g., Bateman, 2008, p. 151–163; Forceville, 2020, p. 254–255). In relational propositions, two "spans of text" are related to one another so that one has a specific function relative to the other. This means there is a general subordination relation between a nucleus, a more central/independent proposition, and satellite, detailing the nucleus and depending on it (see **Figure 6**).

While RST provides a reliable taxonomy of relations with sufficiently clear definitions for application, it also confronts the analyst with a number of problems. First, whereas relations

can be signaled in different ways, there are also un-signaled relations (Taboada and Mann, 2006, p. 438–442). Second, as "no single taxonomy seems suitable" (Mann and Thompson, 1988, p. 256), the annotators may wonder whether the genre-specificity of the data calls for extension and modification. Taboada and Mann (2006, p. 438) advocate caution with this, but the specific multimodal nature of the advertising genre might make specific sub-types of relations necessary (e.g., types of *evidence*). Finally, a multimodal stretch of discourse may be interpreted as instantiating two relations, for instance justification and evidence. Such multi-level relations (Taboada and Mann, 2006, p. 443) complicate the analysis.

An easy but inadequate way to characterize relational propositions set up by stretches of ad-text and image would be to indiscriminately qualify these as a restatement-relation. Restatement is a relation where a situation expressed in one mode is restated by the other. Such an analysis does not reflect the simple truth that hardly any ad images are redundant (see costa_drinks_soft_uk_1-20_33, above). Another rather fuzzy relation is elaboration, which stipulates that additional info is added to a nucleus and thus specifies it. Elaboration is quite similar to the identical term used in info-linking, where it is also a superordinate concept requiring sub-types. Mann and Taboada (2005–2021) propose six types of elaboration that determine the nature of the specification, for instance, a step in a process, an attribute of an object, a member of a set, or an instance of something abstract (see **Figure 6**). In a Lego toy ad (see lego_children_thailand_la_5-19_29), for example, the abstract claim that the Lego blocks are *for every size of imagination* is elaborated by a visual instance, showing a toy skyscraper in miniature contrasted by a very large one. A beer ad (see shiner bock_drinks_alc_usa_la_1-19_35, see **Figure 7**) claims the brand sells *20 different beers*, a proposition which is elaborated
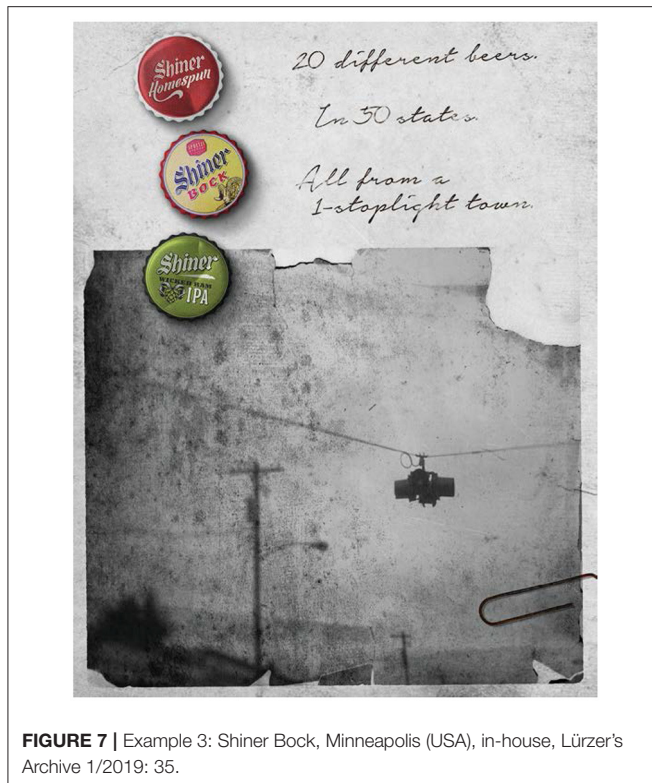
FIGURE 7 | Example 3: Shiner Bock, Minneapolis (USA), in-house, Lürzer's Archive 1/2019: 35.

FIGURE 8 | Taxonomic overview of concepts for an analysis of *mode-centricity* in print ads.

a set of genre-typical multimodal relational propositions. Their description is complicated by the difficulty of fixing visual propositions in ad-images generally and by a lack of signaling devices in a rather covert/indirect but also often minimalist genre.

Mode-Centricity

Despite its intuitive plausibility, the decision on whether the relative status of text and image is equal or unequal in a given ad seems empirically difficult. Following Martinec and Salway (2005, p. 343), two different constellations appear genre-typical: first, unequal status, "when one of them [i.e., modes, HS/JP] modifies the other" and second, equal/complementary status, "when an image and a text are joined equally and modify one another" (see **Figure 8**). It is the first type that raises questions of centricity because here the image may be subordinated to the text (verbiage-centricity, or classic illustration) or the text may be subordinated to the image (image-centricity, or classic anchorage) (see Stöckl, 2020, p. 191–195). We can say for certain from our corpus exploration that it is hardly ever the case that "a whole image is related to a whole text", as Martinec and Salway (2005) stipulate for equal status, which is why there is a strong argument for unequal status generally. This is borne out by the ways in which selected language expressions pick out certain image elements in multimodal cohesive ties (see Section Cohesive Ties). One could take a way out of this system of classification and side with Ewerth et al. (2021, p. 119), who claim that, disregarding concrete details, an interdependent relation between text and image (see **Figure 8**) is the most likely to encounter in ads, where the mode-combination "conveys a new meaning or interpretation which neither of the modalities could have achieved on its own." Their definition attests to a more clearly delineated understanding of just what this type of relative status denotes. With a view to the corpus, however, not all ads seem to fall into this category; thus, it seems necessary to refine existing theoretical proposals further in response to patterns detectable in the data.

Our corpus data strongly suggest that there are in fact several factors contributing to the perceived centrality of one mode or the other (see Pflaeging, 2021, p. 211 for a similar argument). We argued above that image-first layouts and large image formats

by showing three different bottle tops as members of the set of beers.

Alongside motivation, where a proposition is added to increase a recipient's desire to comply with whatever is requested (e.g., *Find danger before it finds you*; see volkswagen_ automobile_argentina_la_4-19_21), and justification, where the appropriateness or acceptability of a speech act is justified by a suitable proposition (e.g., *What homemade sorcery is this???*; see rao's homemade_food_usa_la_2-19_36, see **Figure 2**), evidence seems another ad-typical relational proposition (see **Figure 6**). Here, a recipient's belief in a claim is supported by information that can substantiate that claim as evidence. In an ad for TV-sets (see sharp_audio_video_mexico_ la_6-19_15), for instance, the statement *Your tablet is not a TV* is related to an image that shows a screened box fight giving the impression that all participants are being squeezed into too little space. The image thus serves as direct visual evidence of the claim.

Finally, result appears to be another useful relational proposition in the logic of advertising, where one state/event is described or shown as resulting from another causally related one. The result relation is present, for instance, in a social ad (see foodgroot_social_czech_la_3-19_105) where a process is mentioned in the factual statement. *The average person eats 87 kilos of plastic a year,* and the image shows the final result of this process, a decayed corpse with plastic spilling from the body.

With a view to Taboada and Habel's (2013) study of image-caption relations in science writing and journalism, our hypothesis is that advertising, too, is characterized by
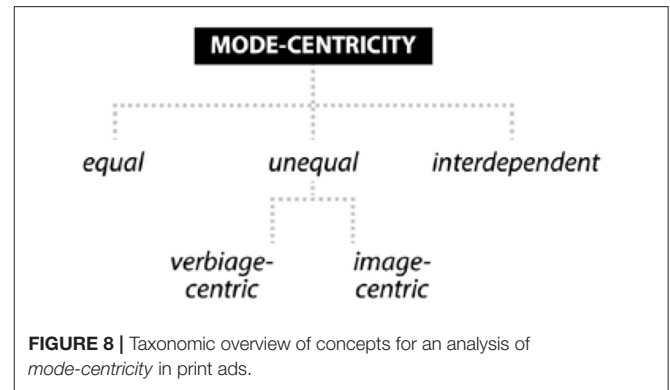
generally could be interpreted as a kind of perceptual image-centricity (see Stöckl, 2019b, p. 20 for a distinction between *image-centrality* and *image-dominance*; also Pflaeging, 2019, p. 116–117). However, this may belie the actual workings of the cohesive ties and the nature of the info-linking in concrete cases, which exhibit quite some variety across the corpus. It appears that long copy potentially makes available more cohesive ties to the image, a constellation that would favor image-centricity. This is the case in the telecommunications ad discussed above (see wind telecommunication_ social_greece_la_3-20_130, see **Figure 4**), where the image obtains centricity as the painting's title and the consistent demonstrative/personal pronoun references (*this nude, it*) presuppose the image they modify. Projection as a type of info-linking would also appear to facilitate image-centricity, as we must see the character speaking first in order to understand the projected speech (see avianco cargo_services_colombia_la_6-19_82). In cases where the origin of the projected speech is not shown, verbiage-centricity seems the more plausible case (see cruz_roja_social_mexico_la_3-19_92).

Image-centricity should also be considered for cases where text remains cryptic without a close inspection and adequate interpretation of the image. This is attested, for instance, in a Burger King ad that runs the following long copy: *On some nights, you can hear the sizzling screams of a flat-top fryer. 522 State Route 31, McHenry, IL 60050. If you dare to get near this scary place, you will get a free flame-grilled Whopper ® on the BK ® App* (see burger king_food_usa_la_5-20_64). The image shows a deserted truck stop against ominous storm clouds. Long-copy may, however, equally clearly instantiate verbiage-centricity, where the image is an illustration of a selected concept from the text. This can be seen in a Smart ad (see smart_automobile_germany_la_3-19_13), where the text is a lengthy quasi-dialogic conversation about cats' needs, arguing that with a small car such as a Smart you can make *the quickest turnarounds*, in case you've forgotten to feed your cat.

In sum, mode-centricity appears to be a tricky category to annotate for. Treating it relative to the levels of layout, cohesive ties and info-linking may be one way of achieving a higher degree of operationalization (see Sections Layout, Cohesive Ties, and Information-Linking).
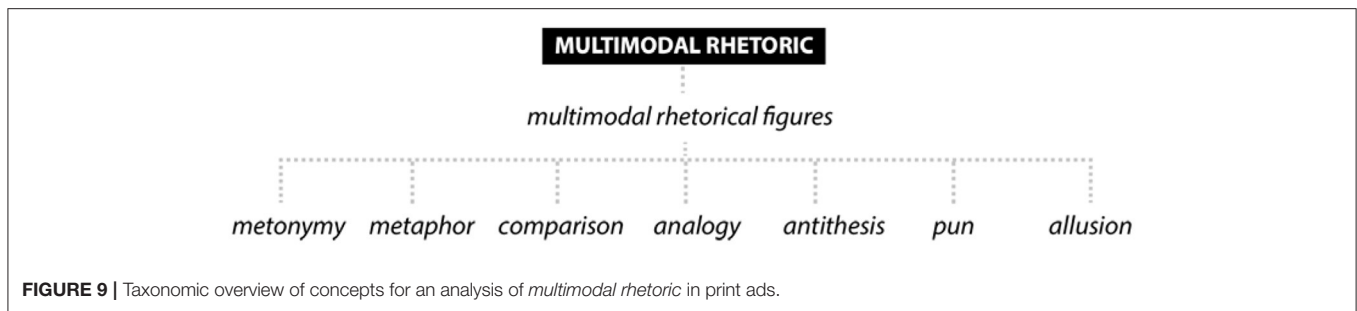
Multimodal Rhetoric

The level of the rhetorical figure has been exploited intensively in studies of advertising (see e.g., Durand, 1987; Phillips and McQuarrie, 2004). The focus has been either on verbal or on visual rhetoric, but a multimodal approach (see Tseronis and Forceville, 2017) must define a rhetorical figure as constructed from a linking or interaction of text and image elements. Forceville (1996, p. 148–161) shows the existence of verbo-visual metaphors in advertising, whose constitution is partly through text, partly through the image. Other figures, for instance antithesis and allusion (Tseronis, 2021), may adopt that same multimodal principle, where a figure as a "minimal unit of meaning" (Tseronis, 2021, p. 377) and its corresponding mental operation is cued as the result of combining two modes. Generally, we follow the principle underlying cognitive theories of metaphor and metonymy (Kövecses, 2002) and regard patterns

of thought to materialize or be instantiated in multimodal rhetorical figures.

Annotating for multimodal rhetorical figures (operations) is fraught with difficulty because verbal (and visual) figures cannot easily be transposed onto text-image relations. Also, existing taxonomies of verbal figures are extensive and complex (cf. Fahnestock, 2011). They provide useful checklists at best, but it may be most convenient to start with standard tropes and focus on what Phillips and McQuarrie (2004) call the meaning operations inherent in them. There is clear evidence in our corpus of multimodal rhetorical phenomena that engage text and image in complementary fashion (see also Stöckl, 2021). These may conveniently be described as meaning operations (or semantic patterns) across the modes, such as comparison/analogy, metonymy, metaphor, punning, and allusion to name the most salient (see **Figure 9**).

Metonymy is a standard but important rhetorical operation, which replaces one sign for another, often involving the product advertised itself. In a beer ad (see shiner bock_drinks_alc_usa_la_1-19_35, see **Figure 7**), *beers* is visually substituted for by three bottle tops in different colors. In a cosmetics ad (see keraone_cosmetics_colombia _la_6-19_39), the product, a *hair straightening treatment,* is metonymically represented by a comb running through a woman's hair. Metaphor, a different kind of replacement operation, attracts more attention as it usually involves a substitution of abstract by concrete concepts, which come from different, non-contiguous domains but correlate in experience. The basis for the substitution is a comparison, as in a Bayer pharma ad (see bayer_pharma_usa_la_5-20_78), where the risk of a heart attack is visualized by an explosive device on the verge of catching fire. In a bin liner ad that guarantees leak-proofness (see glad_house_usa_la_5-19_48), Jon Snow (from *Game of Thrones*) is shown with his lips sealed, thus comparing not leaking to not telling a secret (*my character dies this season*). Analogy is based on a comparison of form or content. It is at work in an ad for a medical lab (see emilio ribas medical lab_pharma_brazil_la_4-19_74) that shows *Monet's landscape in Montecario* and *a microscopic view of human kidney* resembling the painting in texture and color. This is used in the claim that *There is no piece of art as perfect as the human body*. Antithesis, where opposites imply a comparison, features in a vodka ad (see absolut_drinks_alc_ mexico_la_1-20_30) that multimodally contrasts a rainbow flag burning with people reclaiming it, evaluating the two parts of the antithesis as *obsolete* (i.e., against LGBTQ) and *absolut(e)* (i.e., for it). Puns in the corpus appear to be primarily triggered by a linguistic expression, but the double meaning is brought to the fore by the image. So, for example, in the copy of an ad for pasta sauce (see rao's homemade_food_usa_la_2-19_36, see **Figure 2**) *what homemade sorcery is this?* "sorcery" is clearly intended as a homophonic pun on sauce, which the image depicts. Finally, allusion is quite common and manages well to utilize special types of knowledge certain target groups of ads are expected to have. The optician's ad discussed above (see Section Cohesive Ties; see vimax_accessories_spain_la_3-20_12) alludes to Don Quixote multimodally by naming *Sancho* and *the story* and by

FIGURE 9 | Taxonomic overview of concepts for an analysis of *multimodal rhetoric* in print ads.

showing the windmill and the typical attire of the character (lance and armor).

It is obvious, not only with allusion, that understanding multimodal rhetorical figures requires encyclopedic knowledge, a clear sense of how the rhetorical operation is relevant to the overall message and the ability to draw inferences. This is the point where annotation may get quite difficult as we are leaving the domain of semiotic manifestations and are entering the realm of the implicit, indirect and pragmatic. Forceville (2020) accounts well for the richness of such interpretative processes, but annotation must insist on clearly definable operations, something that can best be done by exploring lots of data and defining each multimodal figure precisely (see also Stöckl, 2021, p. 195–199). Rhetorical figures are claimed to play a crucial role in facilitating multimodal argumentation, as Kjeldsen (2012, p. 251) says: "rhetorical figures guide the viewer's construction of the arguments in the ad."

SUMMARY AND CONCLUSIONS

The aim of the present paper was to sketch a preliminary annotation framework for the study of multimodal coherence in print ads. This was to develop from both a critical inspection of relevant theory and close analytical contact with data. Coherence is generally seen as an inherent property of multimodal communication, which must have a structural basis in more or less explicit inter-modal cohesive ties, but which is also constructed from relating propositions that text and image provide in mode-specific ways. Ultimately, the specific kind of coherence we expect to see in advertisements is governed by the rhetorical regime of the genre and its task-based generic stages. We set great store by outlining principles of good empirical practice and strove to adhere to them by cautiously moving from analytical concepts that relevant theory provides to the data in our corpus.

The annotation framework we propose contains all relevant levels for the description of multimodal coherence: layout, cohesive ties, info-linking, relational propositions, mode-centricity (i.e., relative status), and multimodal rhetoric. With regard to all of these dimensions, we sought to make our analytical rationale and coding process transparent by defining most lower-level concepts *explicitly* and by illustrating our annotation work with numerous examples. The investigation of

a comparably large corpus of 50 ads also allowed us to critically move through all levels and most options for annotation with the aim of checking the workability of the scheme and identifying potential problems. We will use the present section to point to some of the general challenges in empirical work of this kind, informed by our own observations and (pragmatic) decisions during analysis.

Granularity and Density of Description. The notion of *granularity* refers to the level of detail with which a text's *texture* is captured: a fine-grained analysis relates smaller textual components, whereas a coarser description relates larger stretches of discourse to one another. Taboada (2019, p. 218) remarks in general that "the issue of granularity in segmentation is open." The tentativeness of empirical work, however, allows for later reduction, once we have found out which textual elements carry multimodal coherence in more foundational ways. Moreover, we are faced with the decision of just how many analytical dimensions are necessary to do justice to existing theoretical proposals as well as patterns observable in the data. For an analysis of multimodal coherence generally, this means we can opt for high density of description, attending to many instances of mode-connectivity, or for a low density of description, singling out a priori links between text and image we judge to be critical and central. We believe high density to be the better approach, seeking to obtain an optimally detailed idea of multimodal coherence.

Directionality in Binary Structure. Quite a number of analytical categories are based on binary structures, i.e., connectivity-relations established between two units of discourse structure in different modes. This is true of cohesive ties, relational propositions, status/centricity and rhetorical figures. All these raise questions about the directionality underlying the processes of inter-modal coherence generation. It is our impression from the variety and complexity of the data in the corpus that fixing nucleus and dependent satellite, to borrow these terms from RST, is inherently difficult. A useful general assumption might be to accept relations to reverse and to not place too much emphasis on centricity or directionality, but rather assume overall complementarity. The loss of detail this entails is manageable as we can still fully capture the essential quality of the binary relations.

Single or Multiple Coding. The available annotations from the codebook handle easiest when we decide to allocate one label to every ad in each analytical category. This may work well with

some categories but not with others. For instance, Taboada and Habel (2013, p. 81–84) envisage multiple relational propositions to hold between language and image units and find this produces a more adequate description of coherence. Similarly, we saw in our data that multiple cohesive ties obtain between text and image in one ad. The same may be possible for rhetorical figures that are generally known to combine. While single coding maintains a manageable complexity overall, multiple coding quickly increases it, but at the same time it produces results that do better justice to the material at hand. A trade-off between these factors must be negotiated so as to make the appropriate choices in annotation.

Discourse Hypotheses. In annotation work such as required and illustrated here for multimodal coherence, codings are necessarily based on subjectively derived discourse hypotheses. Analysts must follow their interpretations, which in turn are based not just on linguistic and visual forms, but also crucially on encyclopedic knowledge, discourse/generic knowledge, inferential procedures and an awareness of message-relevance in the given context (see Forceville, 2020). That such hypothesizing is important was highlighted in particular by instances of allusion, punning and metaphor, but it also underlies any interpretation of relational propositions. In the absence of signaling devices in images and with the propositional and syntactic indeterminacy of images (see Messaris, 1997, p. ix–xiii), the scope for alternative interpretations and multiple discourse hypotheses increases further. In order to arrive at inter-subjectively plausible classifications, categories not only need to be distinct (affording a higher degree of reliability), but the decision-making with regard to tricky cases needs to be made explicit too (achieving a more *objective* research design). This seems especially so when corpora grow larger, individual texts fade out of sight and most of the interpretative work naturally becomes invisible. One way of mitigating the influence of subjective judgment in coding is to involve several trained annotators and test in how far the assigned labels overlap (i.e. check for inter-coder reliability, see Krippendorff, 2004, p. 215).

Frequencies and Correlations Between Categories. Given the appropriate quantification of the annotations and their statistical processing, one can hope to see patterns of communicative phenomena at various analytical levels, e.g., typical linguistic structures facilitating cohesive ties or types of info-linking, typical relational propositions or multimodal rhetorical figures etc. For all levels investigated, one could make statements about ad-specific ways of supporting or constructing multimodal coherence. Based on our corpus explorations, we also assume that there are more or less systematic relations between the levels, such as, for instance, between inter-semiotic cohesive ties and relational propositions or between layout and info-linking. It must be a vital part of the further statistical processing of the data to tease out such correlations between analytical categories and glean from them genre-typical ways of designing multimodal coherence.

Adaptation of Existing Frameworks. One gain of completing the empirical cycle would be to adapt original frameworks to the requirements of the data. Such modifications could take three forms: (1) We could decide to drop the less gainful categories or the ones that produce too much redundancy, something we would not envisage at the current stage. (2) Further analytical dimensions may be added horizontally, expanding the scope of the framework. A possible addition to our current analytical agenda could be image type, or generic stages, which could be traced in terms of what mode predominantly realizes which stage. Also, one may want to annotate for what parts of a classic argument structure are realized through which mode. (3) The framework might also be expanded vertically, increasing depth and detail of the annotation in one analytical category. For instance, we were not fully satisfied yet with the scope of annotations available for a description of the semantics of the VIEs that are part of a cross-modal cohesive tie. Also, we recognize the potential to look more closely at the range of multimodal rhetorical figures.

Despite our efforts to achieve a higher degree of objectivity, reliability and validity in this exploratory study, the nature of our empirical work remains tentative. Investigating a larger corpus, using a refined codebook plus an annotation diary and involving multiple annotators will be means to improve the situation, but even the best of intentions will not eradicate the preliminary value of any empirical study. Thus, tentativeness must be heeded above all, both in relation to the framework, which must be open to change, and in relation to our own analytical results, which must be open to correction and refinement.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The data set has been compiled from two recent volumes of *Lürzer's Archive*. The authors claim a fair use of the materials. The ads may be accessed at *Lürzer's International Archive*, see https://www.luerzersarchive.com/.

AUTHOR CONTRIBUTIONS

HS and JP contributed to conception and design of the research. HS sampled the corpus data, drafted the sections Introduction, A review of theory, Toward annotation, and Summary and conclusions. JP wrote the section Making contact with data, contributed several paragraphs to the section Toward annotation, and designed the diagrams. All authors contributed to manuscript revision, read, and approved the submitted version.

ACKNOWLEDGMENTS

REFERENCES

Arnheim, R. (1976). *Art and Visual Perception: A Psychology of the Creative Eye*. University of California Press.

Barthes, R. (1977). "The rhetoric of the image," in *Image, Music, Text*, ed. S. Heath (New York: Hill + Wang), 2–51.

Bateman, J. A. (2008). *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents*. Basingstoke: Palgrave Macmillan.

Bateman, J. A. (2009). "Discourse across Semiotic Modes," in *Discourse, of Course: An Overview of Research on Discourse Studies*, ed. J. Renkema (Amsterdam: Benjamins), 55–66.

Bateman, J. A. (2011). "The decomposability of semiotic modes," in *Multimodal Studies*, eds K. O'Halloran, and B. A. Smith (Hoboken: Taylor and Francis), 17–38.

Bateman, J. A. (2014a). *Text and Image: A Critical Introduction to the Visual/Verbal Divide*. London/New York, NY: Routledge.

Bateman, J. A. (2014b). "Multimodal coherence research and its applications," in *The Pragmatics of Discourse Coherence*, eds H. Gruber, and G. Redeker (Amsterdam: Benjamins), 145–177.

Bateman, J. A. (2016). "Methodological and theoretical issues for the empirical investigation of multimodality," in *Handbuch Sprache im Multimodalen Kontext*, eds N. M. Klug, and H. Stöckl (Berlin: De Gruyter), 36–74.

Bateman, J. A. (2019). "Afterword," in *Multimodality: Disciplinary Thoughts and the Challenge of Diversity*, eds J. Wildfeuer, J. Pflaeging, J. A. Bateman, O. Seizov, and C. I. Tseng (Berlin: De Gruyter), 297–321.

Bateman, J. A., Wildfeuer, J., and Hiippala, T. (2017). *Multimodality: Foundations, Research and Analysis - A Problem-oriented Introduction*. Berlin: De Gruyter.

Bernstein, B. (2000). *Pedagogy, Symbolic Control and Identity: Theory, Research, Critique*. Oxford: Rowman and Littlefield.

Bublitz, W. (1989). Topical coherence in spoken discourse. *Studia Anglica Posnaniensia*. 22, 31–51.

Caple, H. (2013). *Photojournalism: A Social Semiotic Approach*. Basingstoke: Palgrave Macmillan.

Carrell, P. L. (1982). Cohesion is not coherence. *TESOL Q*. 16, 479–488. doi: 10.2307/3586466

Daly, A., and Unsworth, L. (2011). Analysis and comprehension of multimodal texts. *Aust. J. Lang. Lit*. 34, 61–80.

Drisko, J., and Maschi, T. (2015). *Content Analysis*. Oxford: Oxford University Press.

Durand, J. (1987). "Rhetorical figures in the advertising image," in *Marketing and Semiotics: New Directions in the Study of Signs for Sale*, ed. J. Umiker-Sebeok (Berlin: De Gruyter), 295–318.

Ewerth, R., Otto, C., and Müller-Budack, E. (2021). "Computational approaches for the interpretation of image-text relations," in *Empirical Multimodality Research: Methods, Evaluations, Implications*, eds J. Pflaeging, J. Wildfeuer, and J. A. Bateman (Berlin: De Gruyter), 109–138.

Fahnestock, J. (2011). *Rhetorical Style: The Uses of Language in Persuasion*. Oxford: Oxford University Press.

Forceville, C. (1996): *Pictorial Metaphor in Advertising*. London/New York, NY: Routledge.

Forceville, C. J. (1999). Educating the eye? kress and van leeuwen's reading images: the grammar of visual design. *Lang. Lit*. 2, 163–178. doi: 10.1177/096394709900800204

Forceville, C. J. (2020). *Visual and Multimodal Communication: Applying the Relevance Principle*. Oxford: Oxford University Press.

Häder, M. (2010). *Empirische Sozialforschung: Eine Einführung*. Wiesbaden: Springer.

Halliday, M. A. K., and Hasan, R. (1976). *Cohesion in English*. London: Pearson.

Halliday, M. A. K., and Matthiessen, C. M. I. M. (2004). *An Introduction to Functional Grammar*. London: Hodder Arnold.

Hiippala, T. (2013). The interface between rhetoric and layout in multimodal artifacts. *Lit. Linguistics Comput*. 28, 461–471. doi: 10.1093/llc/fqs064

Holsanova, J., Holmqvist, K., and Rahm, H. (2006). Entry points and reading paths on newspaper spreads: comparing a semiotic analysis with eye-tracking measurements. *Vis. Commun*. 5, 65–93. doi: 10.1177/1470357206061005

Kjeldsen, J. E. (2012). "Pictorial argumentation in advertising: visual tropes and figures as a way of creating visual argumentation," in *Topical Themes in Argumentation Theory*, eds F. H. van Eemeren, and B. Garssen (Dordrecht: Springer), 239–255.

Klug, N. M., and Stöckl, H. (2016). *Handbuch Sprache im Multimodalen Kontext*. Berlin: De Gruyter.

Kong, K. (2006). A taxonomy of discourse relations between words and image. *Inf. Des. J*. 14, 207–230. doi: 10.1075/idj.14.3.04kon

Kövecses, Z. (2002). *Metaphor: A Practical Introduction*. Oxford: Oxford University Press.

Kress, G., and van Leeuwen, T. (1996). *Reading Images: The Grammar of Visual Design*. London/New York, NY: Routledge.

Krippendorff, K. (2004). *Content Analysis: An Introduction to its Methodology*. London/Thousand Oaks: Sage.

Liu, Y., and O'Halloran, K. L. (2009). Intersemiotic texture: analyzing cohesive devices between language and images. *Soc Semiot*. 9, 367–388 doi: 10.1080/10350330903361059

Lürzer's International Archive (2019/2020). Vienna, Austria.

Mann, W. C., and Taboada, M. (2005–2021). *Intro to RST - Rhetorical Structure Theory*. Available online at: http://www.sfu.ca/rst/01intro/intro.html (accessed June 22, 2022).

Mann, W. C., and Thompson, S. A. (1986). Relational propositions in discourse. *Discourse Processes*. 9, 57–90. doi: 10.1080/01638538609544632

Mann, W. C., and Thompson, S. A. (1988). Rhetorical structure theory: toward a functional theory of text organization. *Text*. 8, 243–281. doi: 10.1515/text.1.1988.8.3.243

Martinec, R., and Salway, A. (2005). A system for image-text relations in new (and old) media. *Visual Commun*. 4, 337–371. doi: 10.1177/147035720505928

Maton, K., and Chen, R. T.-H. (2016). "LCT in qualitative research: creating a translation device for studying constructivist pedagogy," in *Knowledge-Building: Educational Studies in Legitimation Code Theory*, eds K. Maton, S. Hood, and S. Shay (London/New York, NY: Routledge), 27–48.

Messaris, P. (1997). *Visual Persuasion: The Role of Images in Advertising*. Thousand Oaks: Sage.

Moosbrugger, H., and Kelava, A. (2014). "Qualitätsanforderungen an einen psychologischen Test (Testgütekriterien)," in *Testtheorie und Fragebogenkonstruktion*, eds H. Moosbrugger, and A. Kelava, (Heidelberg: Springer), 8–26.

Pflaeging, J. (2019). "On the emergence of image-centric popular science stories in 'national geographic'," in *Shifts toward Image-Centricity in Contemporary Multimodal Practices*, eds H. Stöckl, H. Caple, and J. Pflaeging (New York, NY: Routledge), 97–122.

Pflaeging, J. (2021). *Genre Development in Science Journalism: A Multimodal Discourse Analysis of the National Geographic 'feature article' – 1915, 1965, 2015*. [PhD thesis] Halle (Saale); Salzburg: University of Halle-Wittenberg and University of Salzburg.

Pflaeging, J., Bateman, J. A., and Wildfeuer, J. (2021). "Empirical multimodality research: the state of play," in *Empirical Multimodality Research: Methods, Evaluations, Implications*, eds J. Pflaeging, J. Wildfeuer, and J. A. Bateman (Berlin: De Gruyter), 3–32.

Phillips, B. J., and McQuarrie, E. F. (2004). Beyond visual metaphor: a new typology of visual rhetoric in advertising. *Mark Theory*. 4, 113–136. doi: 10.1177/1470593104044089

Ripley, M. L. (2008). Argumentation theorists argue that an ad is an argument. *Argumentation*. 22, 507–519. doi: 10.1007/s10503-008-9102-2

Rossolatos, G. (2013). Rhetorical transformations in multimodal advertising texts: from general to local degree zero. *Hermes* 50. 97–118. doi: 10.7146/hjlcb.v26i50.97821

Royce, T. D. (1998). Synergy on the page: exploring intersemiotic complementarity in page-based multimodal text. *JASFL Occas. Pap*. 1, 25–48.

Sanchez-Stockhammer, C., and Schubert, C. (2022). Cohesion in multimodal discourse. *Special Issue of Discourse Context Media*.

Springer, N., Koschel, F., Fahr, A., and Pürer, H. (2015). *Empirische Methoden der Kommunikationswissenschaft*. Stuttgart: UTB.

Stöckl, H. (2014). "Typography", in *Interactions, Images and Texts: A Reader in Multimodality*, eds S. Norris and C. D. Maier (Berlin: De Gruyter), 283–295.

Stöckl, H. (2019a). "Linguistic multimodality – multimodal linguistics: a state-of-the-art sketch," in *Multimodality: Disciplinary Thoughts and the Challenge of

Diversity, eds J. Wildfeuer, J. Pflaeging, J. A. Bateman, O. Seizov, and C.-I. Tseng (Berlin: De Gruyter), 41–68.

Stöckl, H. (2019b). "Image-centricity - when visuals take center stage: analyses and interpretations of a current (news) media practice," in *Shifts toward Image-Centricity in Contemporary Multimodal Practices,* eds H. Stöckl, H. Caple, and J. Pflaeging, J (New York, NY: Routledge), 19–41.

Stöckl, H. (2020). "Multimodality and mediality in an image-centric semiosphere: a rationale," in *Visualizing Digital Discourse: Interactional, Institutional and Ideological Perspectives*, eds C. Thurlow, C. Dürscheid, and F. Diémoz (Berlin/Boston: de Gruyter), 189–202.

Stöckl, H. (2021). "Pixel surgery and the doctored image: the rhetorical potential of visual compositing in print advertising," in *Empirical Multimodality Research: Methods, Evaluations, Implications*, eds J. Pflaeging, J. Wildfeuer, and J. A. Bateman (Berlin: De Gruyter), 189–211.

Taboada, M. (2019). The space of coherence relations and their signaling in discourse. *Language Context Text*. 1, 205–233. doi: 10.1075/langct.00009.tab

Taboada, M., and Habel, C. (2013). rhetorical relations in multimodal documents. *Discourse Stud.* 15, 65–89. doi: 10.1177/1461445612466468

Taboada, M., and Mann, W. C. (2006). Rhetorical structure theory: looking back and moving ahead. *Discourse Stud.* 8, 423–459. doi: 10.1177/1461445606061881

Tanskanen, S. (2006). *Collaborating Towards Coherence: Lexical Cohesion in English Discourse*. Amsterdam: Benjamins.

Tseng, C.-I. (2013). *Cohesion in Film: Tracking Film Elements*. Basingstoke: Palgrave Macmillan.

Tseronis, A. (2021). From visual rhetoric to multimodal argumentation: exploring the rhetorical and argumentative relevance of multimodal figures on the covers of the economist. *Visual Commun.* 20, 374–396. doi: 10.1177/14703572211005498

Tseronis, A., and Forceville, C. (2017). *Multimodal Argumentation and Rhetoric in Mass Media Genres*. Amsterdam: Benjamins.

Unsworth, L. (2007). "Image/text Relations and Intersemiosis: Towards Multimodal Text Description for Multiliteracies Education," in *Proceedings of the 33rd ISFC*, eds L. Barbara and T. B. Sardinha (São Paulo: PUCSP), 1165–1205.

van Leeuwen, T. (2005). *Introducing Social Semiotics*. London/New York, NY: Routledge.

Wildfeuer, J., Pflaeging, J., Bateman, J. A., Seizov, O., and Tseng, C.-I. (2019). *Multimodality: Disciplinary Thoughts and the Challenge of Diversity*. Berlin: De Gruyter.

APPENDIX

Table A 1 | Table 3 gives an overview of the 50 advertisements that constitute the corpus on which the exploratory annotations presented in this paper are based.

N°	Identifier	Agency	Archive
01*	*absolut_drinks_alc_mexico_la_1-20_30*	Anónimo, Mexico City	1.2005
02	*aegean_services_greece_la_1-19_86*	Sohosquare, Athens	1.1904
03	*afine_aesthetic_clinic_services_brazil_la_2-19_79*	Capim.AG, Cascavel	2.1917
04*	*avianco cargo_services_colombia_la_6-19_82*	DDB Colombia, Bogotá	6.1953
05*	*bayer_pharma_usa_la_5-20_78*	Harrison & Star, New York	5.2013
06*	*bazar metro_retailer_chile_la_4-19_88*	Inbrax, Santiago de Chile	4.1930
07*	*bill-e's_food_usa_la_1-19_51*	Lewis Communications, Mobile, Alabama	1.1904
08*	*burger king_food_italy_la_3-20_59*	Wunderman Thompson, Milan	3.2022
09*	*burger king_food_usa_la_5-20_64*	MullenLowe, Boston	5.2050
10	*casa & video_retailers_brazil_la_1-20_74*	Propeg, Rio de Janeiro	1.2005
11	*cinemark_publ-media_ecuador_la_1-20_58*	Publicis, Quito	1.2004
12	*colin's hope_social_usa_la_5-19_90*	FCB Health, New York	5.19103
13*	*costa_drinks_soft_uk_1-20_33*	Bartle Bogle Hegarty, London	1.2002
14*	*cruz_roja_social_mexico_la_3-19_92*	Geometry, Mexico City	3.1952
15	*dairyland_drinks_soft_canada_la_1-19_38*	Cossette, Vancouver	1.1901
16*	*emilio ribas medical lab_pharma_brazil_la_4-19_74*	Delantero, Fortaleza	4.1908
17*	*ford ranger_automotive_chile_la_1-19_13*	JWT, Santiago de Chile	1.1902
18*	*foodgroot_social_czech_la_3-19_105*	McCann, Prague	3.1965
19*	*freeland_social_thailand_la_5-19_86*	Ogilvy & Mather, Bangkok	5.19100
20*	*glad_house_usa_la_5-19_48*	FCB, Chicago	5.1918
21*	*keraone_cosmetics_colombia_la_6-19_39*	DrMarketing, Bogotá	6.1912
22	*kfc_food_hong kong_la_5-19_39*	Ogilvy & Mather, Hong Kong	5.1960
23*	*lego_children_thailand_la_5-19_29*	Ogilvy & Mather, Bangkok	5.1905
24	*lemonaid_drinks_soft_austria_la_4-19_32*	Heimat, Vienna	4.1910
25	*makita_house_ukraine_la_5-19_47*	Tough Slate Design, Kiev	5.1917
26*	*mcdonald's_food_germany_la_6-19_45*	Leo's Think Tank, Munich	6.1970
27*	*mercado libre_social_colombia_la_2-20_93*	Sancho BBDO, Bogotá	2.2021
28*	*nanfu_house_china_la_4-19_70*	The Nine, Shanghai	4.1916
29	*peta_social_germany_la_2-19_95*	fischerAppelt, Berlin	2.1931
30	*pilsener_drinks-alc_ecuador_la_5-20_31*	Mullenlowe Delta, Guiayaquil	5.2026
31*	*rao's homemade_food_usa_la_2-19_36*	BSSP, Sausalito	2.1917
32	*sana sana_food_costa rica_la_1-19_57*	Gitanos, San José	1.1910
33	*sea_shepherd_social_germany_la_1-19_102*	Ogilvy, Frankfurt/Main	1.1907
34*	*sharp_audio_video_mexico_la_6-19_15*	Archer Troy, Mexico City	6.1904
35*	*shiner bock_drinks_alc_usa_la_1-19_35*	Shiner Bock, Minneapolis	1.1901
36	*simple green_house_thailand_la_5-19_51*	Ogilvy & Mather, Bangkok	5.1921
37*	*smart_automotive_germany_la_3-19_13*	BBDO, Berlin	3.1913
38*	*stiolto_pharma_usa_la_5-20_82*	FCB Health, New York	5.2017
39	*teyko mitsubishi_automotive_brazil_la_4-19_27*	443ag, Curitiba	4.1929
40*	*the covenant house_social_usa_la_5-20_118*	GSW, New York	5.2079
41	*viastein_house_hungary_la_6-19_50-51*	White Rabbit, Budapest	6.1922
42*	*vimax_accessories_spain_la_3-20_12*	Pajarraco, Barcelona	3.2005
43	*vodafone_services_turkey_la_1-19_85*	Y & R Team Red, Istanbul	1.1903
44	*volkswagen_automotive_argentina_ la_2-19_19*	Geometry, Buenos Aires	2.1911
45*	*volkswagen_automotive_argentina_la_4-19_21*	DDB, Buenos Aires	4.1925
46	*volvo_automotive_poland_la_5-19_22*	Gray, Warsaw	5.1934
47	*walgreen's_pharma_usa_la_6-19_65*	Area 23, New York	6.1912
48	*welti-furrer_services_switzerland_la_1-19_91*	Ruf Lanz, Zurich	1.1909
49*	*wind telecommunication_social_greece_la_3-20_130*	The Newtons Laboratory, Athens	3.2065
50	*wwf_social_bolivia_la_5-20_124*	Humano, Santa Cruz	5.2085

The text identifiers follow the pattern brand_product category_country_source_page. All 28 ads that are referred to in the present paper are indicated with an asterisk (*).