Comparing verbal and visual rhetoric
Andersson, Fred

Published in:
Cognitive Semiotics

E-pub ahead of print: 29/05/2024

Document Version
Proof

Document License
CC BY

Please cite the original version:
Fred Andersson*

Comparing verbal and visual rhetoric – and the danger of “empirical shorthands”

https://doi.org/10.1515/cogsem-2024-2007
Published online ▪▪▪

Abstract: This paper will evaluate the quantitative content analysis of visual rhetoric which has been put forward by George Rossolatos in the context of audio-visual and marketing research. Rossolatos bases his taxonomy of visual rhetorical figures on Groupe µ’s adaptation of classical rhetoric in \textit{Rhétorique générale} (1970). With the aim of creating a systematic model for marketing analysis, he and his collaborators have coded and classified an extensive number of audiovisual commercials from the top international strata of economically valuable brands. For comparison, I have during a period of four years given undergraduates the task of collecting and categorizing a minimum of ten audio-visual commercials, each of which should exemplify one of the figures referred to by Rossolatos. The outcomes of the task indicate that a random explorative study with a minimum of operationalization and training of coders is enough for obtaining results akin to those reported by Rossolatos. This may be regarded as a positive outcome from a pedagogical point of view, but it may also raise the suspicion that Rossolatos’s criteria are merely intuitive. If the incidence of visual rhetorical figures is to be studied as an independent content variable of audio-visual enunciations, the values/figures should be described with greater caution.

Keywords: visual rhetoric; quantitative content analysis; cognition; Groupe µ; George Rossolatos

1 Preliminary remarks on the relationship between visual semiotics and content analysis

A first draft of this paper was originally presented at the 12th conference of the IAVS–AISV (International Association of Visual Semiotics), “Visual semiotics goes cognitive”, at Lund University, Sweden, 22–24 August 2019. Organized by Göran Sonesson and his collaborators at the former Centre for Cognitive Semiotics in Lund, and

*Corresponding author: Fred Andersson, Faculty of Arts, Psychology and Theology, Åbo Akademi University, Turku, Finland, E-mail: franders@abo.fi
dedicated to the two branches of semiotic studies that Sonesson significantly influenced and shaped during his long career of interdisciplinary scholarship – cognitive semiotics and visual semiotics – the conference was marked by his personal presence and dedication.

My own choice of topic was motivated by my application, in pedagogical contexts, of Sonesson’s program for a visual *rhetoric of the lifeworld* – a revision of some of the analytical models proposed by the Groupe µ, of which one of the remaining members – Jean-Marie Klinkenberg of the University of Liège – chaired the session in which I presented my paper. The discussion afterwards focused mainly on the role of empirical studies and reliable quantitative measurements in semiotics. In his work, Göran Sonesson repeatedly returned to the question whether visual semiotics could not only benefit from the findings of experimental psychology and psycholinguistics, but also design its own semiotic experiments, tailored to any semiotically relevant research questions at hand. Some notable examples of controlled experimental design in the context of visual semiotics are found in the work of Hartmut Espe (e.g. Espe 1985) and Martin Krampen (e.g. Krampen et al. 2007), who were both affiliated with the *Universität der Künste* in Berlin. Focusing on the identification of semantic differentials in the spirit of C. E. Osgood, or testing optimal conditions for the recognizability of pictographs and other simplified visual representations, Espe and Krampen contributed to the methodology that in media studies and the social sciences is known as VCA (Visual Content Analysis).

**1.1 VCA according to Gillian Rose**

In a widely used book on methods for visual analysis and interpretation, Gillian Rose characterizes quantitative VCA as “counting what you think you see” (Rose 2007: 59). The methodology she describes is, however, a quite simplified variety of content analysis: in a selected corpus of pictures, for example the 598 photographs from *National Geographic* in an earlier study by Lutz and Collins, the researcher identifies an open range of relevant “categories” of content and then counts the incidence of those in the material (Figure 1). The categories must be “exhaustive” (covering all relevant aspects of interest for the study), “exclusive” (the categories should not overlap) and “enlightening” (the categories should fulfil criteria of validity and coherence, Rose 2007: 65). Rose provides a list of 22 categories that were defined in Lutz and Collins’ study (Rose 2007: 66), but the list represents a mix of different data types and the internal order between those is not accounted for; we find in the list categories of nominal data (world location, type of activity, ethnic group, etc.),
binominal or “either-or” data (dressed or not-dressed, male or not-male, Westerner or non-Westerner, etc.), and ratio data (estimated age of persons, number of people in depicted groups, etc.).

If the identification or coding of these data categories in the pictures are to be performed by external coders, and not simply by the researchers themselves, it can be expected that different categories will require very different levels of discrimination and abstraction. In addition, the potential number of discriminatory alternatives will be virtually unlimited for some categories, for example in the case of Lutz and Collins’ category “surroundings of people photographed”. What kind of training will coders receive beforehand for ambiguities to be resolved and for reliability and validity to be secured? And besides: How long or short could an open list of categories be if it is to be “exhaustive” and “exclusive” as well as coherent – all categories being related to a common research question in such a manner that the study validly measures what it sets out to measure? Rose doesn’t provide any straightforward answers to such questions, but she writes that “[c]odes must be defined as fully as possible and a pilot study should ensure that two different coders using the same codes produce the same results from the same set of images. If they do not, the codes must be refined so that they do” (Rose 2007: 68).

Figure 1: Average number per article of photographs showing “Westerners in non-Western settings”, according to Lutz and Collins’ study of *National Geographic*. Data for each year between 1950 and 1985. Graph reproduced from Rose (2007: 69).
The procedures described by Rose do not, however, amount to a structured quantitative content analysis, and are not in accordance with the research paradigm for such analyses that has been developed in press studies, media studies and other social sciences since the mid twentieth century. Originally, quantitative content analysis was exclusively oriented towards printed text or verbal transcripts – “counting what you think you read/hear”. As in any common questionnaire, the codes of structured content analysis are not defined as an open list of “categories”, but as clearly defined variables comprising exhaustive sets of mutually exclusive choices or values. A standard coding exercise in press studies would be to provide coders with short extracts from newspaper articles belonging to different journalistic genres or newsbeats; in coding, those extracts would be identified as representative of different values of the variable “text genre”. The reliability and replicability of the results of such procedures are calculated as either “inter-coder reliability” (agreement between different coders at the same time of coding), “intra-coder reliability” (agreement between different times of coding by the same coder), or both.

1.2 Structured VCA and socio-semiotics

The application of this methodology in the field of visual media studies, and the development of quantitative visual content analysis or VCA, has been fraught with epistemological and methodological difficulties, because pictures do not unambiguously conform to predefined standards of text segmentation. A solution was proposed by Bell (2001), who in addition to such socio-cultural context variables as Gender, Occupational Role and Domestic/Public Setting added certain dimensions of pictorial/spatial “grammar”, defined by Erving Goffman in his studies of publicity pictures and by Gunther Kress and Theo van Leeuwen in their framework for socio-semiotic pictorial analysis.

Thus, the variable of proxemics – or perceived “social distance” – between the standard spectator and actors in the picture, comprises the values “(a) Intimate. (b) Close personal. (c) Far personal. (d) Close social. (e) Far social. (f) Public” (Bell 2001: 29). The variable of perceived social behaviour or attitude, especially in publicity pictures, comprises “(a) Offer/ideal: the model depicted offers herself/himself as an idealized exemplar of a class or attribute, looking away from the viewer […] (b) Demand/affiliation (equality): model looks at the viewer directly, smiling. (c) Demand/submission: model looks down at the viewer, not smiling. (d) Demand/seduction: model looks up at the viewer, head canted, smiling or ‘pouting’ and (e) None of the above” (Bell 2001: 31). The variable of modality, or perceived level of
realism/truth value, comprises a coding choice between standard or neutral modality, factual modality, and fantasy modality (Bell 2001: 20–21).

1.3 Automatic visual information retrieval

Training protocols for coders in VCA, and taxonomies developed for labelling and retrieving files in image databases, have contributed to the development of advanced algorithms for image search and image analysis in the field of machine learning and AI. A unified taxonomy of variables and values for Automatic Image Content Extraction (AICE) in photographic images has recently been proposed by Männistö et al. (2022). The taxonomy covers six main groups of variables, including technical variables and variables of low-level visual features, with the following subgroups: (1) Technical variables, (2) Compositional variables, (3) Modality, (4) Content participants: 4.1. General, 4.2. Persons/actors (4.2.1. Characteristics of single persons, 4.2.2. Characteristics of multiple persons), 4.3. Objects/goals for action (main motif vs side motif, object categories), 4.4. Settings/events, (5) Interaction: 5.1. Spatiality, 5.2. Distances, 5.3. Activity, 5.4. Interaction, (6) Visual similarity with other images and image clusters (Männistö et al. 2022: 8–10).

Some of the greatest challenges for automatic visual information retrieval are associated with the variables listed as (5) by the authors: spatial context and spatial relationships within the scene, distances between actors in the scene and between actors and camera, kinaesthetic action schemes (such as gestures, postures), and indications of interpersonal or human-object interaction. However, the application of simplified models for extraction of perceptual invariants, for example skeleton models for pose analysis, now generates data that will be possible to match against existing image archives and corpuses by means of, as the authors write, “automatically clustering […] poses according to the AICE variable ‘6.2 Appearance-based grouping’” (Männistö et al. 2022: 18).

1.4 Automatic image analysis in cultural analytics and semiotics

Within his widely recognized framework for Cultural Analytics, Lev Manovich has utilized the ImageJ package of processing tools in scientific imaging (imagej.net) for extracting low-level visual features – such as average brightness levels and average number of contoured segments – in image files that represent painted artworks. From these data, visualizations of the stylistic development of individual oeuvres of visual art can be produced. In the field of visual semiotics, Manovich's
methodology has been applied in a study by Reyes-Garcia and Sonesson (2019) that aimed at demonstrating a certain level of plastic consistency and order in Mark Rothko’s non-figurative art.

The work of Manovich and his group has also been frequently referenced in publications by Maria-Giulia Dondero and other participants in a series of round-tables and seminars, conducted at the Paris Cité University and the University of Liège, on the implications of computer-aided image analysis for visual semiotics. In a recent article, Dondero (2022) discusses the technology of generating deep-fake videos by means of mapping individual physiognomic and kinaesthetic data unto data derived from another individual. Dondero identifies this technology with a shift in the cultural function or statut of portraiture, and traces the technical and ideological origins of face-recognition algorithms back to Galton’s invention of the composite portrait in the nineteenth century.

### 1.5 Metaphorical and rhetorical content in images

These recent applications and theoretical reflections in semiotics are, however, limited to accounts of the material or enunciative level of images, and the processes of low-level feature extraction that can be related to it. Only to a limited extent has Dondero and other members of the Paris/Liège group considered content analysis in the stricter sense of the analysis of semantic and iconographic content. Concerning, on the other hand, such taxonomies as that of Männistö et al., they lack an account of content dimensions pertaining to the sphere of metaphorical, poetic, or in a wider sense rhetorical expression. One likely reason for this is that automatic image retrieval or content extraction is still struggling with the more basic problems of spatial context and bodily kinaesthetics, but too limited contacts with art historians and other scholars in the humanistic field can also be a factor.

Because of the abundance of rhetorical figures defined in classical rhetoric and poetics, a full content analysis of all rhetorical figures that may appear in verbal speech and printed text would comprise a variable of between 400 and 500 values. This taxonomy, representing some 2 500 years of teaching and scholarship in the ancient art of rhetoric, can be broken down into more manageable variables, determined by whether the figures depend on word morphology, syntactic construction, semantic units (or semes) or discursive patterns of reasoning (see Section 2.1 below). Nevertheless, coding tasks are typically limited to only a selection of the most common rhetorical figures.

Compared to the long history of verbal rhetoric, the development of theories of visual rhetoric that transcend the established four categories of iconographic analysis – visual symbols, attributes, personifications and allegories – is a very
recent phenomenon. There is no consensus regarding the question if, and in that case how, visual rhetoric operates in a manner that can be compared to the structural characteristics of verbal figures. As we shall see, many epistemological problems are associated with the schematic application of verbal rhetorical categories in visual analysis, and before those problems are sorted out the retrieval process should preferably not be left to automatic algorithms. The increasing reliance on statistical measures may also result in an overuse of what I have here chosen to term “the empirical shorthand”. I borrow this notion from a pioneer of Visual Sociology, Douglas Harper, who writes the following apropos the statistical outcomes of coding content variables in corpuses of documentary photographs:

Numerical data are, in fact, shorthand ways of summarizing material that is often far from hard-edged. We categorize reactions to an idea or experience as equally spaced attitudes ("strongly agree", "agree", "no opinion", "disagree", "strongly disagree") when in fact they are reactions that are not equally spaced or even similar in intensity. [...] Sociologists are so used to empirical shorthand that it is seldom questioned. (Harper 2012: 8–9)

For researchers who analyse documentary photographic images of social and ethnographic settings, images taken by themselves or by others, Harper recommends that before the image corpus is reduced to an empirical shorthand of statistic measures one should carefully examine each image and consider it as research data in and by itself – not merely as an exponent of predefined values. As an example of how VCA abstracts numerical data from the full visual record I here reproduce a graph from Gillian Rose’s book – Figure 1 above. The graph shows that the number of Westerners depicted in non-Western settings decreased on the pages of *National Geographic* after the middle of the 1960s, reaching a zero level in 1975. Rose quotes Lutz and Collins’ interpretation of this pattern as an inverse measure of the degree of conflict in the non-Western world post 1965 – when UN staff and military personnel dominated the photographic news reports from conflict zones in Asia, Africa and South America, this imagery was edited out to maintain an illusion of harmony (Rose 2007: 69). As readers, we are supposed to accept the graph as a “proof” of a conclusion that most likely was arrived at *ad hoc*. But the graph only shows how the coders have interpreted two content categories – “Westerner” and “non-Western setting”. It tells neither how the coding was performed, nor how the chosen categories relate to other categories and variables that may possibly contradict Lutz and Collins’ conclusion.

What follows here reflects observations collected in a pedagogical context when helping undergraduate students to understand basic principles of verbal and visual rhetoric. In my capacity as lecturer of Visual Studies at Åbo Akademi University in Finland, I have given a course in visual rhetoric every year since 2016. The participant’s backgrounds have been mixed, but a majority have been marketing
students from the faculty of Social sciences and Economics. One of the mandatory tasks, repeated each year, is to practise classification of rhetorical figures in audio-visual commercials.

The classification is done according to George Rossolatos’ suggestion for how an analysis of audio-visual messages could be based on the model put forward by Groupe µ in *Rhétorique Générale/A General Rhetoric* (Groupe µ 1981: 45). For preparation, the participants have read Rossolatos’ paper “An anatomy of the multimodal rhetorical landscape of the world’s most valuable brands” in *International Journal of Marketing Semiotics* (Rossolatos 2013). They have then been given the task of collecting and analyzing a minimum of ten audio-visual commercials, each of which should exemplify one of the figures defined by Rossolatos. With a participation rate of 15–20 participants/year, the exercise had after four years (2019) generated a collection of 500 different examples.

2 A verbal analytical framework: Rossolatos and the *Rhétorique Générale* of Groupe µ

Rossolatos’ terminology and the general theoretical framework of his analyses is explicitly based on *Rhétorique Générale*. A work familiar to scholars and students of rhetoric, but probably less so to the general academic public, *Rhétorique Générale* is a reformulation of the systems and distinctions of classical verbal rhetoric in terms of Modern linguistics and information science. In my teaching, I usually explain the basic structural principle of *Rhétorique Générale* with the aid of diagrams and tables such as that in Figure 2.

2.1 The four categories of change and three new figures

As demonstrated in Figure 2, *Rhétorique Générale* systematizes the different kinds of rhetorical figures in accordance with what is sometimes called “the four categories of change” in classical rhetorical theory – that is, adjunction (or addition), suppression (or subtraction), substitution, and permutation. Along the horizontal axis, figures are classified into metaplasms (figures of morphology), metataxes (figures of syntax), metasememes (tropes and other figures based on *nuclear semes*), and metalogisms (figures of reasoning). All figures in Figure 2 are not placed according to the original classification of Groupe µ, as the table partly reflects George Rossolatos’ individual approach in his books and reports published between 2013 and 2018 (Rossolatos 2013: 75–89, 2014: 337–339).
Moreover, Rossolatos has added three figures defined by himself. He intends these figures and their definitions to cover cases of visual and cinematic rhetoric that do not easily fit into the classical conceptual framework. The figures in question are termed *accolorance* (or repetition of colours), *reshaption* (or repetition of shapes), and the more complicated *pareikonopoeia*.\(^1\) This latter term could be roughly translated as “the making of parallel images”. The commercials classified by Rossolatos as containing the figure of *pareikonopoeia* are characterized by a strong parallelism between scenes with a similar setting and similar spatial characteristics, but different actors or events for each scene.

Rossolatos’ more technical and lengthy definition of *pareikonopoeia* is “Employment of similar images with different senses; similarity is conferred by the employment of different actors in different filmic sequences/syntagms with similar postures, usually enhanced by the employment of the same production techniques (e.g. all syntagms featuring close-ups or medium shots or alternating close-ups/medium shots), who repeat the same underlying theme under different manifest narratives” (Rossolatos 2013: 79). It is quite clear from this quote that Rossolatos considers the *pareikonopoeia* figure to be intimately connected to what he refers to as “filmic syntagms”, or the syntax of the sequence of scenes in a

\(^1\) In the Peircean framework for analysis of information design proposed by Nicole Amare and Alan D. Manning, both *accolorance* and *reshaption* would probably be defined as “decoratives”, or purely plastic qualities as iconic signs. Cfr. Amare and Manning (2013: 27–30).

---

*Figure 2:* The general rhetoric. Classification of some rhetorical figures according to the main system of Groupe μ in *Rhétorique Générale/A General Rhetoric* (1970/1981), with the addition of three multimodal figures defined by George Rossolatos (underscored). Table by the author (2019).
commercial. Therefore, I have allocated the figure to the column of metataxes in Figure 2, albeit with a question mark.

2.2 Problems of segmentation and semantic hierarchy

Metataxes are the operations occurring above the morphological level in the rhetoric of \textit{figurae verborum}, or figures of grammatical composition. However, I am not copying Rossolatos here. To my knowledge, Rossolatos has in his research never addressed the question if, and in that case how, it would be possible to categorize examples of visual rhetoric as either metaplasm, metataxes, metasememes and metalogisms. I have only found a small terminological remark, in which he defines the operation of substitution as one in which “An expressive unit completely substitutes another in a filmic sequence/syntagm, either morphologically or syntactically or at a semantic level” (Rossolatos 2013: 83).

Oddly, for a reader acquainted with \textit{Rhétorique Générale}, Rossolatos here seems to exclude the possibility of partial substitutions. There are a number of verbal figures that would be categorized as partial substitutions if we follow the taxonomy of Groupe µ, but Rossolatos never clarifies whether he takes the substitutive logic to be the same for both verbal and visual “expressive units.” Moreover, Rossolatos’ definition of a substitution as a figure in which a unit “completely substitutes another” runs contradictory to his decision to define the figure of synecdoche as a substitution, rather than as an adjunction or a suppression. Synecdoche is hardly a substitution in the sense of a unit that “completely substitutes another.” On the contrary, it functions as the representation of a phenomenon by means of a selection of a sub- or superordinate level or part of it (Groupe µ 1981: 102–106), as in the expression “chair” to signify a professorship; this expression still retains the material notion of the chair or \textit{cathedra} in which Medieval professors lectured. This of course concerns semantics, but Rossolatos remains largely silent regarding the criteria for separating semantic operations from syntactic and plastic or morphological ones.

I honestly wonder if this silence is some sort of a strategy, on the part of Rossolatos, to avoid some of the most obvious epistemological difficulties associated with transferring the categories of verbal analytics to the field of visual and spatial analytics. It is hard to see that the very endeavour of finding visual analogies to such figures as anagram, anaphora or antithesis should be less problematic than any other analogies with similar \textit{pan-linguistic} objectives. We should bear in mind, on the other hand, that Rossolatos and other scholars of a similar vein never intend to study semiotic systems that are exclusively verbal or exclusively visual. Their main research object is the multimodal combination of sound, music, words and moving
images, as clearly stated in the title of the article mentioned above: “A Multimodal Rhetorical Landscape of the World's Most Valuable Brands”. The format and aesthetics of the short audiovisual commercial for television and other screen media is certainly a most pertinent frame for studying this multimodality and its psychological and social impact. Still, in this research the linguistic model is often applied as a kind of blueprint against which other modalities are treated, and it is often unclear whether the signifying units attributed to visuals would be in any way salient without verbal anchoring.

Rossolatos’ reference to some Budweiser commercials as examples of what he asserts is “visual anaphora” is a clear example of this tendency (Rossolatos 2013: 110; see also Figure 3). Here, the “visual anaphora” is supposed to consist in the repetition in each filmic segment of either the Budweiser bottle or the Budweiser logo. Why a

Figure 3: Budweiser anaphora. “Budweiser anaphora”, Figure 36 in Rossolatos (2013: 110).
logo should be defined as a “visual” and not a verbal element is not discussed in the article. Neither does Rossolatos discuss or problematize the implied segmentation which would make it possible to differentiate a visual anaphora from, for example, an alliteration or a rhyme, the latter two being metaplasms and not metataxes according to *Rhétorique Générale* (cf. Figure 2).

### 3 A visual analytical framework: Groupe µ (1992)

A radically different approach to the analysis of visual rhetoric was proposed by Groupe µ in their 1992 *Traité du Signe Visuel*. Figure 4, a table translated from the *Traité*, has an obvious affinity with the model shown in Figure 2, but instead of a linguistic taxonomy the horizontal axis represents four levels of transformation that are operative in visual semiosis and in simple image manipulations. These are the geometrical, the analytical, the optical and the kinetic transformations. The term “analytical” here refers to manipulations of global characteristics of a picture – easily performed in PhotoShop or any similar editing program – such as colour channels, spatial frequency and the extraction of contours. The optical transformations likewise concern manipulation of luminance contrast, depth of field and foveation (Groupe µ 1992: 156–177).

If visual rhetoric is theorized from the point of view of a specifically visual and spatial conceptual framework, we will be able to identify figures of a rather different character than the classical ones referred to by Rossolatos. Regarding adjunctive operations, for example, the repetition of a Budweiser bottle tells nothing

![Figure 4: Visual transformations. Table VI in Groupe µ (1992: 157), adapted and translated by the author.](image-url)
about how geometrical adjunctions, such as unexpected sizes of objects (homothety), or analytical suppressions, such as a sudden cut to a monochromatic film sequence (subtractive filtering), can be utilized for rhetorical visual effects.

Regarding the theme of the conference in which this paper was originally presented – Visual Semiotics Goes Cognitive – I think this different conception of the four categories of change in rhetoric is very instructive if we want to explain the importance of a cognitive approach. The four categories of change – adjunction, suppression, substitution and permutation – are indeed categories, operative in a taxonomic categorization. As clarified by Groupe µ in their Traité and later in Principia Semiotica, these categories belong to what is often referred to as Cognitive Types, or CT, not necessarily confined to the black box that Umberto Eco regarded the mind to be (Groupe µ 2015: 51–54). One might say that they do not have any sense in themselves, but that they contribute to the extraction of sense from sensory input. They are not specific to any sensory channel or modality, such as auditory speech or visual images. On the contrary, they exemplify categorization as desensorialization. This is lucidly stated by Groupe µ in the following quote from Principia Semiotica; I have here attempted to translate the quote myself, in advance of the authorized English translation of the book that has been promised by the group:

Because the categorization occurs through a densensorialization, and is a product of thought, albeit of a more elementary kind, it tends to eliminate the sensory residues. This is the reason why it is necessarily a-modal (Groupe µ 2015: 226).²

For a more nuanced understanding of the multimodal communication that occupies Rossolatos and other marketing researchers, it would be necessary to address the idea that multimodality is possible precisely because of factors that are a-modal. In this manner, it would be possible to avoid the Scylla and Charybdis of either the simplistic application of verbal models in visual communication research, or the stubborn denial of any parallels between linguistic and visual modes altogether. When working with students and coders in quantitative identification tasks, people like Rossolatos and myself will then be able to retain more interesting observations than those implied in, for example, the repetition of a Budweiser bottle in a commercial in which the voiceover or slogan constantly repeats the words “Budweiser” and “Bud”.

² Also see Stjernfelt (2007: 66–75).
4 Comparison between Rossolatos’ method and structured VCA

I also want to briefly consider the statistical aspect of experimental design and evaluation in the type of quantitative content analyses pursued by Rossolatos. As practised in sociological contexts and in communication studies, quantitative content analysis involves four essential factors: the validity of dependent and independent variables in experimental design, the selection of examples (or stimuli) and coders (or experimental subjects), the clarity of instructions for coders, and the measurement of reliability. The outcome and its reliability is dependent on the validity of variables and the clarity with which the codes have been defined beforehand by the researcher.

4.1 A non-transparent segmentation and coding process

From Rossolatos’ descriptions of his methodology it is evident, however, that his studies do not belong to this branch of textual or visual quantitative content analysis. He relies heavily on corporate evaluations, such as the BrandZ annual report of the WPP group: “[…] the con-figuration modes of the rhetorical landscape of the most valuable brands, as reported in BrandZ 2012” (Rossolatos 2013: 75). As for methodology, he defines the following three methodological steps in a study published in *Semiotica*:

The first step in this exploratory endeavor consists of segmenting the selected corpus of ad films in individual segments with the employment of the content analytic software atlas.ti […]. The segmentation procedure of the 62 ad films from 13 sub-brands of the 3 key brand players that make up the selected corpus (the UK cereals market) resulted in 321 filmic segments (further details about the featured brands and ad films may be provided upon request). The second step consists in coding the resulting segments with one or more rhetorical figures and producing relevant descriptive statistics with the aid of atlas.ti, as will be shown in the ensuing section. The final step consists in applying multivariate mapping techniques and factor analysis to the content analytic output with regard to examining the overall covariation patterns in the data-pool and interpreting the output. (Rossolatos 2014: 339)

Neither in the quoted article, nor in some previous texts that he refers to, Rossolatos mentions any criteria for his selection, except of “a corpus of 62 ad filmic texts from 13 sub-brands of the 3 major brands in the UK cereals market and 321 ad filmic segments that resulted from the segmentation procedure.” It then seems that his selection procedure is highly heteronomous (dependent on previous corporate evaluations), and probably partly dictated by some unstated stochastic principle.
Apparently, the only coder involved in “coding the resulting segments with one or more rhetorical figures” is the researcher himself; if there is some stage of automatization and machine learning involved, it is not mentioned in the articles. An independent variable is brand type – “13 sub-brands of the 3 major brands in the UK cereals market” – but if only one coder is involved, the only dependent variable would be the “within-subject” one of Rossolatos’ own personal choice between a set of values (figures) in coding. No tests of either intra-coder-reliability or inter-coder-reliability are reported. From the point of view of a more strictly delimited experimental methodology, these studies must then be judged as merely “explorative” – a term which Rossolatos himself indeed uses.

4.2 An empirical shorthand

My use of students as coders at Åbo Akademi between 2016 and 2019 is explorative in a similar manner. The only operative criterion for being selected as a coder is being a student in my course, and the students have simply been instructed to look for random examples at YouTube. The instructions for the task are contained in Rossolatos’ article and in my preparatory lecture. The codes identified by the coders are identical to Rossolatos’ definitions of verbo-visual rhetorical figures in multimodal communication. Part of the outcome is statistics of the reported incidence of each rhetorical figure, as shown in Figure 5, taken from Rossolatos’ second study from 2014.

As seen in Figure 5, the Atlas.ti program for content analysis has been used by Rossolatos for a multivariate mapping of the semantic identity of different brands, and each rhetorical figure has been given an atlas.ti code (“SUB_META”, etc.). The percentages indicative of “above category average incidence” of figures in each specific brand have then been used as input data for a correspondence analysis. The correspondence analysis results in a measure of the “degree of semic similarity” between brands and the identification of semantic clusters (Rossolatos 2014: 348–354).

Rossolatos concludes that “The above analyses are particularly pertinent for projecting brand image structures at the very encoding stage of ad filmic texts, by attending to how a brand’s semic structure alongside rhetorical relata emerge in specific ad filmic segments” (Rossolatos 2014: 354). This might be so. Still, there are clearly some unresolved epistemological problems at “the very encoding stage of ad filmic texts.” Rossolatos tends to put these problems within parentheses and to resort to simple statistics as a shorthand or shortcut, with the aid of which he purports to advance rather far-reaching scientific conclusions. From an
epistemological point of view, however, these conclusions are drawn at the basis of definitions and coding procedures that may in the end be largely subjective.

### 4.3 Comparison between Rossolatos’ studies and the outcomes of the ÅAU student codings

If we look at the simpler statistical level one cannot deny that there are at least some indications of reliability in the procedure, especially concerning the incidence of the four basic operations, or categories of change, in different corpuses. See Figure 6. This table shows the results from the mentioned studies by Rossolatos as compared with the coding of my Åbo Akademi University students. When we look at the incidence of each single figure, however, the results are a bit more problematic. There is, as a matter of fact, a greater degree of coherence between the spontaneous Åbo Akademi student coding and Rossolatos’ (2014) study than between Rossolatos’ (2013, 2014) studies taken separately. See Figure 7.

Particularly striking is the extremely high incidence of the figure of anaphora in Rossolatos’ (2013) study. Perhaps this outlier is indicative of a difficulty on the part of the coder (i.e. Rossolatos himself?) to separate anaphora from other kinds of adjunctive figures. We may also ask whether the lack of coherence between the 2013...
and 2014 results indicate not only a reliability problem but also a more fundamental problem related to validity. What type of training and instruction has been fed into the coding process? Have the replicability of the instructions been tested in such a manner that they would, in principle, be possible to feed into a machine learning process? Do the instructions involve routines for separating verbal and visual dimensions, or, differently put, to identify in a systematic manner certain visual features that could be regarded as analogous to certain structural characteristics of verbal rhetorical figures?

**Figure 6:** Quantitative comparison. Comparison between incidences of figures of adjunction, suppression, substitution and permutation in Rossolatos (2013, 2014) and Åbo Akademi University student codings (2016–19). Table by the author (2019).

<table>
<thead>
<tr>
<th></th>
<th>ÅA 2016-19</th>
<th>Rossolatos 2013</th>
<th>Rossolatos 2014</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ADJ (+)</strong></td>
<td>52.46 %</td>
<td>63.00 %</td>
<td>57.20 %</td>
</tr>
<tr>
<td><strong>SUPP (-)</strong></td>
<td>9.96 %</td>
<td>5.00 %</td>
<td>3.30 %</td>
</tr>
<tr>
<td><strong>SUB (+)</strong></td>
<td>31.24 %</td>
<td>20.00 %</td>
<td>34.10 %</td>
</tr>
<tr>
<td><strong>PER (-)</strong></td>
<td>6.39 %</td>
<td>10.00 %</td>
<td>6.50 %</td>
</tr>
</tbody>
</table>

**Figure 7:** Variance between the three instances of coding. Comparison between incidences of selected figures in Rossolatos (2013, 2014) and in the Åbo Akademi University student codings (2016–19). The comparison provides an estimation of intra-coder reliability (coding of Rossolatos) and inter-coder reliability (Rossolatos compared with students) with high degrees of variance marked in bold. Table by the author (2019).
At least in the articles referred to here, Rossolatos provides no information whatsoever about these factors. From my personal experience with students as coders, there is reason to assume, however, that even when coders are explicitly instructed to look for visual equivalents, they tend to focus on the more easily recognizable verbal patterns. Exceptions would be the specifically visual or plastic figures of accolorance and reshaption. In these cases, the results also show a comparatively low degree of variance: 1.47 for accolorance and 0.31 for reshaption, to be compared with 2.38 for pareikonopoeia and 6.46 for anaphora.

5 Conclusion

In this paper I have examined the quantitative approach to rhetorical analysis which has been put forward by George Rossolatos in the context of audio-visuality and marketing research. The examination has been related to both linguistic and cognitive aspects of the possibility of a visual rhetoric. Rossolatos bases his taxonomy on Groupe µ's adaptation of Classical rhetoric in Rhétorique générale. With the aim of creating a systematic model for marketing analysis, he and his collaborators have coded and classified an extensive number of audiovisual commercials from the top international strata of economically valuable brands. Implicit in this approach is the assumption that the figures defined in the Classical tradition, which originally concern only verbal discourse, are expressed both verbally and visually in the commercials. The approach generates a numerical mapping of the frequency and distribution of identified figures in the corpus. However, Rossolatos discusses neither the reliability of the results, nor the specifics of the coding procedure, nor the impact that verbal and aural enunciations may have upon the choice to code a particular visual enunciation as belonging to a specific rhetorical class.

In this sense, the generated data could be regarded as an empirical shorthand that bypasses further control of reliability and validity. The comparison between the statistical outcomes of different studies by Rossolatos, and the results of the Åbo Akademi University student task between 2016 and 2019, revealed reliability problems that are indicative of insufficient distinctions between coded values. The criteria for classifying certain elements of visual editing and storytelling as analogous to certain specific verbal figures should be subject to more careful phenomenological considerations that should preferably be communicated to coders as part of their training.
References


Bionote

Fred Andersson
Faculty of Arts, Psychology and Theology, Åbo Akademi University, Turku, Finland
franders@abo.fi

Fred Andersson is a docent in art history focusing on visual communication. Although he is a historian of art and literature by training and teaches most periods of European cultural history, most of Andersson’s research has been related to more “contemporary” topics. He has studied visual art, visual communication, and visual art criticism after 1950, especially from a semiotic point of view. Since 1998, he has been affiliated with Nordic and international networks for interdisciplinary research in the humanities. He participated in the Lund Circle of Semiotic in the early 2000s.