Semiotic Review 12: Dialogues between Continental Semiotic and Linguistic Anthropology | article published September 2025 |
DOI: https://doi.org/10.71743/bwf80g96

Enunciative Praxis and Enregisterment in the Domain of Generative Artificial Intelligence

Maria Giulia Dondero
mariagiulia.dondero@uliege.be

Abstract: This article focuses on Generative Artificial Intelligence (GenAI) in the area of text-to-image translation. I address the question of stereotyped visual styles using the theory of *enunciative praxis* formulated in continental semiotics (and, more specifically, in so-called Paris School Semiotics). I relate this theory with the concept of *enregisterment* in linguistic anthropology in order to describe the way in which visual forms undergo a process of sedimentation and typification and then are transposed to feed into other forms. The first section of the article presents the way in which a painting (as a stratified form of other forms), as well as a corpus of interrelated paintings, may be studied through the methodology of enunciative praxis and enregisterment; the second section develops the study of databases and algorithms as tools for a renewed approach to semiotic theory of enunciation and visual language.

Keywords: generative artificial intelligence; text-to-image models; multimodal translation; text; enunciation; enregisterment; style

Introduction

This paper focuses on the relation between image generation and databases, considering databases as a crucial instrument (in association with algorithmic models) for understanding visual language in our current times. In particular, I focus on image styles and their processes of generation, deformation, and circulation. Helpful in understanding Generative Artificial Intelligence (GenAI), and especially the area of text-to-image

translation, I address and discuss the concept of *enunciative praxis.* This discussion draws upon the traditional theorization in continental semiotics and notably in the so-called Paris School semiotics. I then relate the notion of enunciative praxis with discussions of the concept of *enregisterment* in linguistic anthropology in order to describe the way in which visual forms undergo a process of typification and are then transposed to feed into other forms. This overarching process is examined and exemplified by an analysis of traditional painting and its relationship to generative AI models.

The first section of my paper will thus present enunciative praxis and enregisterment as a methodology through which an image such a painting (as a stratified form of other forms), as well as a corpus of interrelated paintings, may be studied as part of a genealogy; the second section develops the case of databases and algorithms as tools for a renewed approach to semiotic theory of enunciation and the current functioning of visual language.

If the first section serves as an exemplification of enunciative praxis, conceived as a methodological theorization capable of revealing how a *singular* visual artifact—such as a Renaissance painting—makes visible various typifications and particularizations of forms on the concentrated surface of the image, then the second section addresses the production of visual images as a human practice intersecting with machinic logic. This logic is partially concealed to GenAI users, and the analyst must reconstruct it by studying the outputs generated by the machine. This second part takes as its corpus not a singular image that compresses different styles and motifs from various historical periods, but rather the generative process itself—specifically, the production of *multiple visual proposals* (or, more technically, predictive forms) understood as outputs of a particular mode of extraction and compression of historical painterly styles. This part aims to highlight the process of typification and renewal of visual forms, exemplified by operations mediated through annotated image databases, algorithmic architectures, statistical analysis and the multimodal text-to-image translation processes specific to each generative model. To this end, I focus particularly on the GenAI Midjourney because, in my view, it is the best tool for producing and studying the typification and deformation of historical artistic styles (especially when compared to the GenAIs DALL•E and Stable Diffusion). In order to describe the process by which *singular stereotypical* forms are generated from annotated databases—that is, from more or less stabilized correspondences between words and visual patterns—enunciative praxis and enregisterment offer the most appropriate theoretical and methodological tools for studying this process of sedimentation and renewal as they are both concerned with processes of generalization (typification) and characterization (singularization) of visual forms. In particular, linguistic anthropology allows us to specify the notions composing the enunciative praxis, and notably what is called in Paris School semiotics *virtualization* and

actualization, as it focuses on the organization of these processes into particular divisions and sectors of discursivity according to social domains, professions, styles, et cetera.

Enunciative Praxis and Enregisterment

According to a very general definition, the theory of enunciation allows us to conceive of a mediation between Saussurean *langue*—a system of virtualities—and *parole*—the actualizations of *langue* in discourse. Enunciation, thus, broadens the spectrum of the virtualities of language. Since the work of Émile Benveniste (1971), the theory of enunciation has greatly evolved and has enabled to take into account the fact that *langue* is not a fixed system of grammatical rules but a schematization of historically attested texts, built on the practices of speakers. In Jacques Fontanille's work (e.g., 2006 [1998]), this relation between *langue* and *parole* gives way to the theory of enunciative praxis, the latter being conceived of as a dynamic between the *sedimentation* of existing structures of signification and the *creativity* inherent to any ongoing semiotic process. In this sense, enunciative praxis is not the sum of all discourses performed but the locus of a *discursive schematization* that makes it possible to account for the "thickness" of our linguistic performances, caught between projection (protention) and memory (retention).

This dynamics of protention and retention in discourse practices has the merit of valuing the complexity of our semiotic operations caught between more or less quick or slow cultural processes of innovation and stabilization, and sometimes institutionalization—or blurring/refusal—of textual regularities; in doing so, enunciative praxis multiplies the steps of this process from two to four, thereby strengthening its descriptive power for analytic description. This four-step process has been visualized by Fontanille in the following figure (Figure 1).

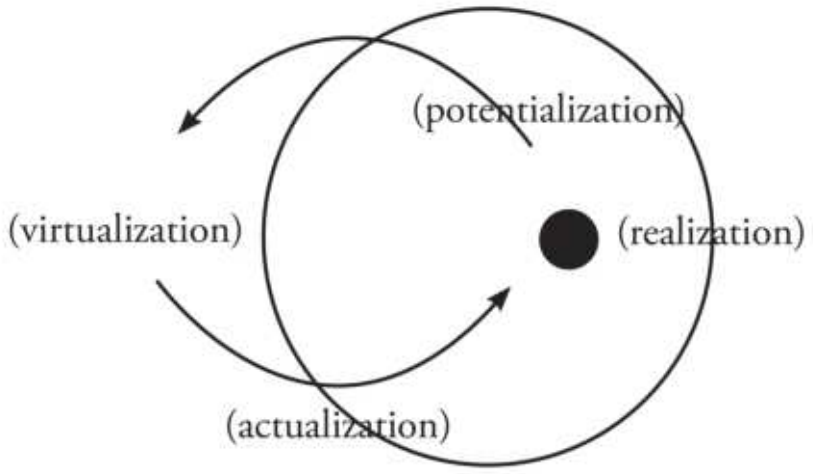

Figure 1. Schema of the modes of existence in enunciative praxis; from Fontanille 2006[1998]:199

As shown in Figure 1, the *process* of virtualization replaces the virtuality pole representing Saussurean *langue*. In fact, the pure virtualities of Saussurean *langue* are beyond the scope of the schema of enunciative praxis theory because they are abstracted from the dynamics of discursive praxis.

Whereas the process of *virtualization* covers every discourse that has been concretely sedimented and schematized and that is thus solicitable in discursive events, *actualization* coincides with the process going from sedimented forms to action through the recalling of acquired competences and skills to be used by the enunciator. Realization is the action of putting something into discourse through particular discursive forms (*mise en discours*) and, finally, *potentialization* is the reverse process which follows realization.[1] The latter is a process of putting significations attached to certain forms on standby so that they may subsequently be *virtualized* and stocked/sedimented—or not; they may disappear, that is, remain a single-use phrase or a single-use strategy in a work of art (or in another kind of semiotic artifact).

In a nutshell, I would say that the whole path of enunciative praxis covers cultural processes based on a sequence that we could describe as the transformation of semiotic habits (virtualized) operated by social groups, which they appropriate (actualization), acting in turn as producers of partly renewed forms of speech (realization). In the potentialization process, these realized appropriations/propositions undergo a sort of selection (englobing forms of exclusion): the selected forms may become *enregistered*, to use the term from linguistic anthropology, as stabilized schemas that can subsequently be used and transformed once more.

Generally speaking, this theorization can be compared with other approaches in language sciences, such as interactional linguistics and linguistic anthropology.

As for the first approach, we can state that Fontanille's theory highlights the fact that each discourse, including images and other nonlinguistic texts, has a *discursive depth*, based on what the linguistic anthropologist Charles Goodwin (1994, 2018) calls the "substrate." In Goodwin's work, the substrate is a *reservoir* that is *partly removed from the field of practice in course of realization, but that can be partly re-solicited by a movement of appropriation* and actualization. As I discuss in a publication devoted to Goodwin's work (Dondero 2024b), this movement of actualization should be understood as an act of selection in relation to everything that has previously been virtualized and schematized and that is available and re-utilizable. There, I suggested the idea that Goodwin can be seen as an heir of Saussure and Benveniste in his description of a dynamics characterized by this tension between some norms available and the particular actions social groups perform in relation to these norms (alignment, reference, rejection, etc.). But Goodwin may

also be conceived of as a precursor of the four modes of existence formulated by Fontanille, for at least two reasons. The first is that Goodwin participates in the construction of the *specification* of two of the poles of enunciative praxis (the substrate and the moment of interaction). The substrate is a reservoir of available forms (comparable to the virtualization mode of existence) that selected groups of people, for instance professional groups, can use in their (expert) performances. More specifically, Goodwin takes into account the process of step-by-step acquisition and of sharing a set of competences (ways of seeing, what he calls "professional vision")[2] within professional groups such as geologists, lawyers, or law-enforcement officers—that I would liken to the process of actualization. In this sense, Goodwin specifies virtualized discursive forms through the study of the specific expertise and habits of different social groups, making it possible to distinguish varied registers of speaking and seeing within general social discourse. The second reason is that, contrary to Saussure and Benveniste, Goodwin works on this dynamics of substrate and performance not only by studying natural language but by attending to englobing visual forms such as scientific drawings and diagrams, multimodal textualizations, and also gestures, prosody and rhythms in interactions, scientific compartments of knowledge, and so on. In this sense, manners of tracing lines, bodily movements, and ways of looking, appreciating, and discerning (in sciences, for instance) become sorts of "micro-languages" inheriting virtualized social domain-linked usages that are specified into local (scientific, professional, etc.) proto-grammaticalizations and sets of variations.

In addition, and reflecting the theme of this issue, there is also an axis of dialogue between enunciative praxis and the concept of discursive enregisterment and typification in linguistic anthropology. This dialogue can be useful to detail some processes outlined in the enunciative praxis theory. In Asif Agha's (2003) historical analysis of "Received Pronunciation" in the United Kingdom, as well as in his other works on "enregistered voices" (2005), *enregisterment* is defined as the process through which forms of speech "come to be socially recognized (or enregistered) as indexical of speaker attributes by a population of language users" (2005:38). Enregisterment is possible through a process he calls, after Goffman, *role alignment*. Role alignment characterizes how *models* of speaking (in particular, typified personae), in being invoked in discourse, are also (dis)aligned to by participants in such events of discourse. In this sense, the process of enregisterment not only describes the process of realization (wherein sedimented forms—models of speech—are enacted in particular contextualized events of discourse), but also partly coincides the final step of potentialization, where such singular forms are already circulating and made available across events of discourse.[3] Enregisterment and potentialization are both processes of "grammaticalization" (i.e., conventionalization) and of a gradual augmentation of social recognizability and stabilization of discursive forms that coincide with a progressive institutionalization of semiotic habits or discursive forms.[4] As already

stated, potentialization also includes the negative process of refusal or dispersion and disappearance: some discursive propositions are only performed by a very small number of people, are never repeated, or even consciously refused or studiously avoided. But even in cases of oblivion or refusal (role dis-alignment), we encounter a process of cancellation that is essential to the entire enunciative praxis.

In a certain sense, if enregisterment includes the final step of a potentialization process which tends to the virtualization/sedimentation of the selected discursive forms, in the potentialization mode, we can also detect another (sub)process of enunciative praxis that has been formulated in linguistic anthropology: *entextualization* (Bauman and Briggs 1992; Silverstein and Urban 1996). Entextualization denotes the process by which a text emerges, which is also to say, how such an emergent form is detached from its field of happening and made available for fixation and circulation (it's de- and re-contextualization) across other moments of semiosis.[5] In this sense, entextualization represents the inchoative moment of potentialization and enregisterment its final moment. Seen in this way, the concepts of enregisterment and entextualization can help continental semiotics detail the process which leads from different moments of potentialization to virtualization, the latter being focused on its accomplished grammaticalization and future availability.

The differences between the two approaches are numerous, of course, as enunciative praxis and enregisterment are heirs to different theoretical objectives, corpora, and styles of thought. For example, Agha's discussion lays much emphasis on the micro-level of social interactions ("we cannot understand macro-level changes in registers without attending to micro-level processes of register use in interaction" [Agha 2005:38]). Interaction coincides with the mode of existence of realization in Fontanille's schema. By contrast, for the Paris School semiotics, realization is traditionally conceived of as a "text-artifact" (in the sense used in linguistic anthropology), with a dense and stabilized texture of forms inscribed on a material substrate (a canvas, a piece of paper, and so on) rather than as a time-bound social interaction.[6] Secondly, enregisterment concerns not only forms of typification from a diachronic perspective but also social processes of differentiation in synchrony (a kind of multiple and multi-faceted acts of distinction in the sense of Bourdieu 1984), whereas in the Paris School semiotics, pragmatics is not a matter of social interaction but instead is included in what, after Benveniste, is called *semantics* (see Fabbri 1998). In fact, in Greimasean semiotics, social interactions are mostly studied post facto, through so-called "uttered enunciations" as manifested in text-artifacts (see Fontanille 1989; Dondero 2020), for example, by attending to communication devices ("traces" of enunciation) contained in texts (verbal, visual, audiovisual) such as forms of address (pronouns and anaphors in natural language; frontality and side view in pictures) and point-of-view and spatial and temporal indexical signs (here/there,

once/now, etc.). Moreover, the concept of enregisterment in linguistic anthropology was first worked out through attention to natural language discourse (though see, e.g., Agha 2011 on commodity registers; Chumley 2016 on painting; Murphy 2016 on furniture design; Nakassis 2023b on cinema) while Paris School semiotics begins not necessarily from natural language but from non-linguistic texts, especially the domain of painting in art world.[7] To simplify, we can say that while linguistic anthropology focuses more on how the (inter)discursive differentiation of social voices (which align with and refract social attributes such as gender, class, caste, and profession) relates to social differentiation itself, continental semiotics tends to focus on describing the alignment of visual forms within the framework of image genealogies in the domain of the arts, in order to study more or less collective styles or movements. Here, the ultimate objective is to analyze the relationship between individual images and the typification and singularization of *forms*, as well as to detect the influences and perdurances of one style in another.[8] In this sense, the two disciplines are respectively interested in understanding the relation between unique, event-specific semiotic forms (voices, motifs) and a sort of "contagion" between them that becomes recognizable, and reorganized, as belonging to certain *social groups* (for linguistic anthropologists) or to a certain *group of artworks* (for continental semiotics).

The Discursive Depth and Thickness of the Image

Enunciative praxis, entextualization, and enregisterment are thus crucial instruments to study image genealogy and the typification and renewal of styles. In this section, I show how a single image can be seen as the condensation of the "modes of existence" (virtualization, actualization, realization, potentialization) that characterize enunciative praxis. I then address Generative Artificial Intelligence (GenAI), particularly text-to-image translation, which—contrary to the single image that compactifies different styles and periods of time—raises the question of the entire process of enunciative praxis, from sedimentation to realization, and back, from a diachronic perspective.

In Paris School semiotics, a painting can be seen as a realization, that is, a "concentration" of visual forces and different styles, each singularizing a different period of time. In fact, an image, understood as *parole* in Saussurean terms, or as an utterance or text in the Greimasean sense,[9] can encompass within itself traces of other forms, present in various degrees of intensity and extension, and which, in their hybridization, contribute to changing both themselves and the other forms present on the canvas.

The four modes of existence constitute the discursive "thickness," the intensity and extension of presencing, of some form. Attending to this thickness enables us to describe each text, in this case a painting, in relation to cultural practices that are, on the one hand, undergoing transformations in the process of its (the painting's) entextualization and recontextualization and, on the other hand, synchronically *within* the painting, that is, in

the dialogic relations of forms ("voices" we might say, with Agha) within the text to each other.[10] Each image can thus be conceived of as realizing, to different degrees of presence, various styles and enunciative influences that come from multiple periods of time. These "stylistic periods" may be considered as enunciative agents that can be described through the analytic method of tensive semiotics formulated by Fontanille and Zilberberg in *Tension et signification* (1998): within an image, the specificity of a certain historically situated style is built in a graded relation to other historical styles and can be understood as a complex texture of forces involved in (by-degrees) agreement with, or in opposition, to another force. This does not mean, however, that a picture—as a space of concentration and presentification of various influences from other artists and past styles—prevents the emergence of unique stylistic characteristics defining its singularity. More generally, each artistic style is constituted by a particular way of selecting and organizing spatial compositions from the works of others, by a specific manner of intersecting with moments drawn from the evolution of another stylistic trend (whether at its inception, peak, or decline), or by a deliberate rejection of either broad aspects or specific details of a previous or contemporaneous style. It is thus possible to state that each artwork exemplifies a unique encounter of voices, and that these voices, brought together and activated within a painting, must be considered as singular appropriations and modifications of a particular moment or characteristic of other styles temporal development.

In this respect, the French art historian Henri Focillon's seminal ideas regarding images are of particular interest, especially the *stratification* of multiple periods of time within images and the consequent *stratification of forms* originating from different times within the same period of history. In his seminal book *The Life of Forms in Art* (1992[1934]), Focillon explains that two or more different styles can coexist in works of art: certain elements can be seen as an anticipation on their current times, while inversely, other elements can be seen as conservative.

To understand a painting, thus, it is useful to study the *degrees of the intensity of the presence* of each of these different styles and motifs and also the way in which these enter or inhabit the visual discourse of an image. Every work of art is a kind of polyphony of voices made of more or less strong influences of various pictorial styles, the latter being intended as cultural stocks and as references. Like Focillon, the art historian Aby Warburg (2012) also considers it possible for the parts of an image to *not be synchronous*, that is to say, that every image is informed by different styles: those that precede it and those that are in preparation.[11] In fact, one part of an artwork can, for one audience, be totally experimental and not fully recognizable or acceptable while, for another (or at a later moment), this same part or treatment of matter in a painting can be potentialized and accepted by others, such as a community of artists, and thereby virtualized as a style to

imitate.[12] The complexity of paintings has been shown by Warburg through the famous example of Ghirlandaio Cappella Tornabuoni's *Birth of the Baptist* (1486-1490; Figure 2).

Figure 2. Ghirlandaio, *Birth of the Baptist*, 1486-90, Cappella Tornabuoni, Santa Maria Novella, Florence

This painting exemplifies what Warburg calls the conflicting styles within the Renaissance style, marked by the coexistence of both Apollonian mathematical logic and Dionysian thinking.[13] In fact, in this religious scene of the Catholic Florentine Renaissance, the pious women assisting the birth of the Baptist are depicted in the same style: monumental. Their bodies are static, their dresses are heavy, and the gestures depicted as slow. But there is a figure that is disrupting this scene and its rhythm: the servant at the extreme right of the painting, who is depicted making a strange movement (see Didi-Huberman 2001). This movement is considered very erotic for a religious painting and contrasts with the bodily attitudes of the other women present in the image.

This painting, in short, includes two different styles that were available in the virtualized universe of visual forms at the time of its making: a Renaissance Christian style (the religious women are still, they are covered by long garments and their dynamicity is weak) and the Greco-Roman style of pagan religions. As such, the servant appears as a dynamic apparition[14] semi-covered with garments that seem to move with the wind.[15] This figure

of the servant is an heir to the figure of the nymph coming from the profane universe of Greek and Roman art.[16] Compare the classical *basso-relievo* in Figure 3 with the detail of the servant in Ghirlandaio's painting.

Figure 3. Comparison between a nymph figure and the servant in *Birth of the Baptist*

The two styles from both the Christian Renaissance and Greek Paganism are present in the painting, but at what intensity, respectively? Noticeably, the Christian sacred scene occupies most of the painting's surface, in the terms of tensive semiotics, the spatial *extension* of the Christian scene. Following Fontanille (2006[1998]), we may define (qualitative or affective) *intensity* as the force of assumption of an enunciation, and (measurable, spatial) *extension* as the capacity of unfolding of the enunciation's figurative declension. Thus, we can affirm that the extension of the Renaissance religious component is high and its intensity low: the rhythm of the disposition of these women is regular and calm. Conversely, the intensity of the pagan style is high and its topological extension low (the servant occupies a small area of the painting and is in a peripheral position). Moreover, this high intensity of the apparition of the servant is a function of the way in which she emerges in the painting as a sort of "*survenir*," that is, as a kind of revolution within the rules and the perceptive effect of duration of the other, more dominant or conventional, part of the painting. In fact, the perceptual visual rhythm of the Renaissance dominant is broken by the movement of the (Pagan) servant.

We can state, thus, that what was available in virtualized form has been performed, or realized—that is, entextualized as a sort of tropic combination of distinct and historically distant models of painting and meaning-making—through different intensities and extensions *within* the image. And what about the future of this stratification of forms and styles and its potentialization/virtualization? This kind of mixed realization has been extensively rejected in Italy (it did not go through a process of enregisterment or virtualization), though we know that in other places of the world so-called religious syncretism of this sort became the norm. In Latin America, for instance, visual forms characterizing different religious communities have been completely accepted and virtualized and used again, so that religious and stylistic syncretism has had a long life. On this topic, the Italian semiotician Massimo Leone (2011) has studied the case of Pedro de Gante and Diego Valadés in Mesoamerica in the 16th century. He describes the "portability of the early-modern Catholic message" (2011:60) in Central and South Americas through the use of the concept of "entextualization," stating that:

> In order to dispel the ghost of a failed evangelization of the world, early-modern Catholic missionaries brought about strategies of religious persuasion based on what current linguistic anthropology calls "entextualization," the idea that "chunks of discourse come to be extractable from particular contexts and thereby made portable. … These chunks of discourse, or 'texts' can thereby circulate and be recontextualized, inserted into new contexts (Keane 2007, p. 14). However, the entextualization of Catholicism put forward by Pedro de Gante was not neutral at all: it implied a semiotic ideology according to which Mesoamerican pictograms could be "culturalized," detached from their traditional religious content and used as independent expressive forms, able to convey Christianity's religious content. (ibid.)

As mentioned beforehand, if entextualization can be seen as the inchoative moment of the potentialization mode, here we can see contextualization as a moment which coincides, in the enunciative praxis process, with the mode of actualization, when a discursive form that has been virtualized is ready to be reemployed, that is, re-contextualized. It is the contact with the new era that allows a text (such as *Birth of the Baptist*) to change or even totally reverse the established meaning of the "original" visual form by re-contextualizing it in some novel, tropic way.

Databases and Algorithmic Models in the Study of Images

Enunciative praxis and enregisterment are also useful analytic concepts in the frame of Generative Artificial Intelligence (GenAI) in order to study processes of automatic translation from text to text (as with large language models) and from text to image (as with so-called diffusion models that translate text to image).

As to the first, Schneider (2022) addresses the enregisterment process at play with automatic translation tools and GenAI. Schneider points out that, in the context of large language models, only texts from a certain section of society and only in certain languages are used and thus, their influence increases, while the presence of minority languages and expressions becomes even smaller: "This does not necessarily happen because of a desire of actors in the digital language industry to enforce social and sociolinguistic hierarchies but *is coproduced by the affordances and materialities of digital language technology, as machine learning tools require a predefined data set to be trained and amplify what is dominant*" (Schneider 2022:381, my emphasis).[17] In what follows, I devote my attention to a similar process in the context of AI that generates visual forms (and is also based on selective data sets) and I consider the history of paintings as the basis for generating new images. In order to test the stereotypical character and the creativity of these models, I started with very famous painters and artists (e.g., such as Vincent Van Gogh and Paul Klee) to see how each one is more or less plastic than the other; the conclusions to be drawn are perhaps unsurprising: while it is very difficult to extract Klee's style, it is very easy to extract Van Gogh's (art gadgetization in museum bookshops and in fashion have already shown this functioning). This result can be understood in relation to the trend highlighted by Schneider (2022), who compares English to other less widely used and less flexible languages. In a sense, Van Gogh can be compared to the use of the English language in archives and databases for at least two reasons: first, it has already been extensively employed for commercial purposes and gadgetization; the forms of his painting have been widely disseminated and made recognizable to everyone. Second, his forms are easily reproducible because his brushstrokes are highly distinctive, undulating, and more predictable than Klee's pictorial style. Klee, in contrast, employed a wide range of techniques and a complex interplay between semi-figurative representation, the plastic grid of composition, and the materiality of the substrate. These three aspects of his artistic production are less easily reproducible, and no commercialization or typification has been made from Klee's forms.

Databases and algorithms today allow us to have a hold on virtualization: before the availability of big data, virtualization was a very abstract concept and consequently we could not test nor operationalize it. It is now possible to rethink and readdress some of the questions having long interested linguistics and semiotics, thanks to databases that may be considered as the concretization and totalization *in the same (virtual!) place* of all canonized artistic images (and their formalized characteristics) produced in the Western world. (Here, we do well to recall Schneider's warnings of the limited, and political, nature of such data sets.) All these images have been translated into lists of numbers, meaning that they are all now readily *comparable and manipulable*. I put forth the hypothesis that these canonized images, that is, sedimented and available forms (virtualizations), are

currently schematized in databases that become manipulable because they are represented in what, in image processing, is called the “latent space.” Latent space is:

> the abstract, multidimensional space in which deep-learning algorithms turn digital objects (e.g., the vast quantities of images and texts that have been uploaded to the internet) into latent representations so that they can be processed and *used to generate* new digital objects (e.g., new images and new texts). Latent representations are made of vectors; that is, long lists of numbers that define the coordinates of the digital objects encoded and embedded in latent space and their relations of distance and proximity within it, just as the three coordinates x, y, and z define the position of a physical object in three-dimensional space and its relations to other physical objects. (Somaini 2023:77, my emphasis)

Confronted with a database, we face not the substance of the visual, but a substance made of numbers, namely a substance that presents itself as a transcription of numbers manipulable by algorithms. In other words, within the collection of digital images, all the images contained in the database are identified by numerical strings.

Now, what happens once all the images have been transposed into strings of numbers within a database? In the case of a GenAI like Midjourney, this database becomes a sort of closed system[18] (a *langue*) that makes the images (utterances/realizations) commensurable with one another. The term *system* here coincides with a collection of texts that are co-present and made commensurable through chains of numbers and manipulable through algorithmic operations on these chains of numbers. In fact, the commensurability of images contained in databases makes them manipulable by algorithms. As such, the quantification and measurement of formal and plastic characteristics of each image (colors, forms, topology, etc.) against the characteristics of all the other images in the dataset allows for the generation of a system of structural differences and similarities within the collection (the “controlled distance” mentioned by Somaini 2023).[19]

As I show below, thanks to databases and algorithms we can now better answer the old question that has obsessed linguists and semiologists such as Émile Benveniste, Roland Barthes, and others (e.g., Christian Metz, René Lindekens): does a visual *langue* exist?[20] In other words, does a global *reservoir* of visual marks exist which we can consider as an ensemble of distinguishable and differentiated marks that are able to construct a strict identification on the plane of expression and, relatedly, when assembled, a meaning on the plane of content (as with so-called natural language)?[21]

Image generation can help us to answer this recurrent question because the major plastic characteristics of all well-known styles have been annotated in databases (in effect, a reservoir has been constructed in the database) and made commutable by the algorithmic models that operate on them. Image generation today thus offers us the possibility to confront how algorithmic models respond to our linguistic requests (we can perform elicitation experiments on the "grammar" of the model) and, further, allow us to investigate how stylistic traits can be solicited through linguistic terms.[22] In this sense, the functioning of these models shows (or reaffirms) the inadequacy of linguistic segmentation of the world to describe an image while also showing how such models do, on their own plastic dimensions, have a sort of "*langue*."

To see this, let us delve into the problem of image generation and the rules of its functioning.

Generating Images from Databases

In this section, I describe the operations realized by generative AI models. Image generators such as Midjourney or DALL•E can produce images from descriptions via natural language prompts, and the opposite is also possible: to obtain a description of a given image by a GenAI in verbal language. All these operations start with a more fundamental kind of translation: that of images and texts into machine-friendly lists of numbers known as "embeddings" (Figure 4).

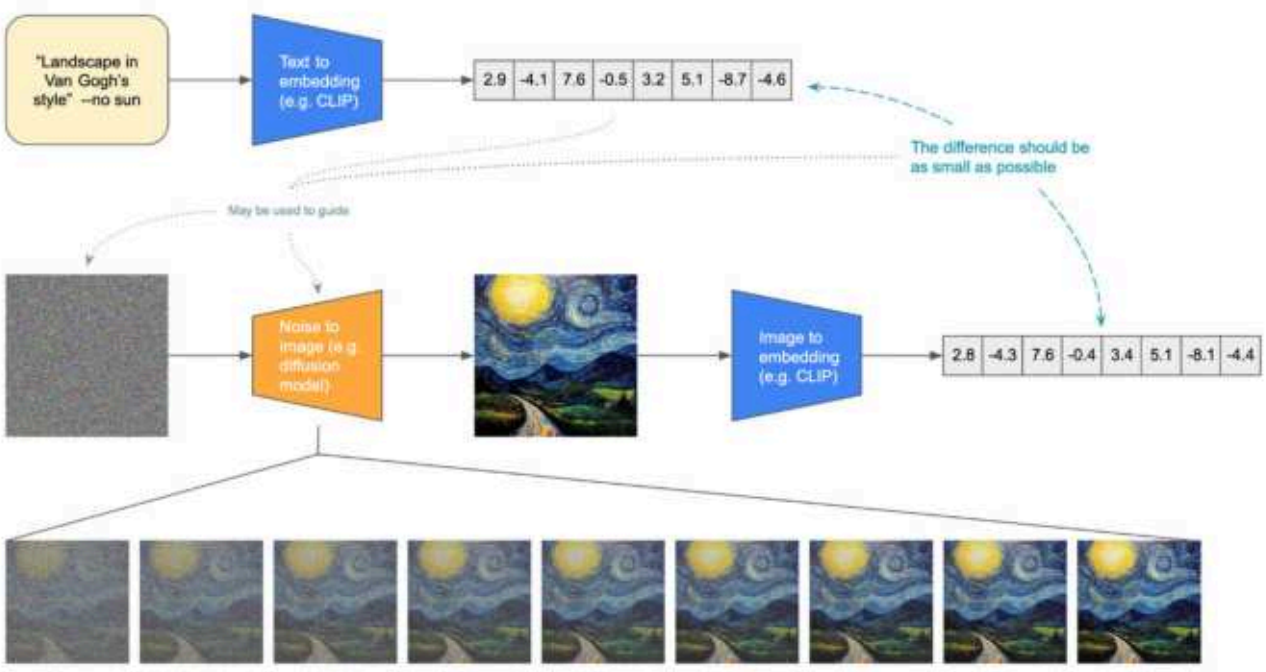

Figure 4. Association between the embeddings of verbal texts and images requires them to be very close to be effectively associated (from A. Deliège)

Current text-to-image generative models use a language processing component that understands the correspondences between natural language and visual forms and transforms prompts into embeddings that are then used by a diffusion-based component to guide the image generation process.[23] These models, which enable the translation between verbal and visual signs, are determined by the organization of the database contents by human labor of analysis, coding, and tagging, by the rules learnt by the system on the training set and by their own architecture (Impett and Offert 2025). These translations manifest through the similarity of the distribution of the verbal and visual embeddings resulting from the projection of the database contents into the latent space by the algorithmic system.

What is important for us here is that the new images are generated through operations performed on *all* the images that are stored and annotated (according to style, author, and genre) within available databases such as Wikiart, Artsy, Google Art and Culture, and the like. Given this, we can test the AI model for the stereotypes of famous painters that the database has learned, and test combinations of styles that reveal, at least partly, how algorithms work on the translation between verbal and visual languages. It is without a doubt only through the forms already known and settled in our cultural perception that it is possible to understand the work of the machine, not only the degree of its much-questioned "creativity,"[24] but also the way in which it transforms the forms we know by making them into averages that transform into stylistic stereotypes.[25]

But there are two things that this process of transformation of styles into averages[26] of multiple singular images does not prevent: the first is that various averages can produce new forms, as demonstrated by many competitions won by AI-generated images and by some very original image generations by Manovich and Arielli (2021–2024). Such processes allow us to estimate the part of the aleatoric in what thus appears as the creativity agency that accompanies all image generation. The second is that, although the machine can extract and mimic various styles, the style of the machine always remains visible. We'll return to the question of style in what follows.

For now, let us briefly explain the functioning of Midjourney. When a written instruction (prompt) is given to the Midjourney platform, four visual translations are by default obtained through the production of four original images. Each of these translations can be understood as different *optimizations* of the prompt given, differentiated according to luminosity, colors, the positioning of the objects, and so on. These optimizations of the given prompt can be considered realized exemplifications of the mode of actualization, that is, a series of variations of combinations of colors and lines that are intended as options or alternative models that embody possible realizations of associations between embeddings. The user/experimenter can thus choose the version that suits him or her

best and decide to continue searching for the image believed to be attainable by giving further instructions: he or she can modify the prompt serving as input or the image through which the quest will continue. This specific kind of image production can thus be described in terms of decision operations and transformation requests realized through verbal and visual instructions supported by the system of embeddings that correlate them.[27]

If the four optimizations represent preliminary options or models for visualizing and articulating the prompt, the subsequent steps of decision-making take on a different character. This process is guided by the user's objectives, of course. In most cases, the ideal image sought—one that aligns most closely with the user's imagination—depends on objectives that can be described as aesthetic. This is the mode of use that is imagined (or at least, declared) by the producers of GenAI as the typical or normative way to engage such platforms (Midjourney's tagline, e.g., is "Create visual masterpieces with Midjourney, an engine for your imagination"). In other domains of experimentation, such as the scientific one explored in this research, there is no need to select or designate a single image as the most suitable; all experimentations and optimizations are significant.

This difference implicates different modes of existence that the praxis of enunciation follows. In the context of text-to-image models, for example, an aesthetic end aims to narrow down multiple choices and select a kind of "unique piece," whereas a scientific end (such as our own in this article) values the diversity and potential of multifaceted results/outputs. In a sense, we could say that the aesthetic aim is to achieve the mode of *realization*, while in the scientific domain, the mode of *actualization* is paramount.[28] The scientific approach to AI-generated images enhances our ability to perceive and understand multiplicity and diversity—and ultimately, possibly, the organization of the database. In other words, in the aesthetic domain, the tokens generated by the model serve as instruments to achieve a kind of type or ideal exemplification of this specific "co-enunciation" between embeddings. In contrast, in the scientific domain, all optimizations related to a particular association of embeddings may be regarded as equivalent functional compositions. All these compositions aim to explore the database in the search for stylistic more or less recognizable and stereotyped patterns.[29] In other words, the aestheticization process of these visual optimizations of prompts leads to the election of a single and unique text as intended in continental semiotics (a text-artifact in linguistic anthropology, that is, a picture), while scientific experiments with such GenAI lead to a "text" as a result of the entextualization-contextualization process as intended in linguistic anthropology. As my experimentations with Midjourney are devoted to a scientific objective regarding visual semiotics and, more collaterally, art history, the term *text* according to its conception in linguistic anthropology becomes useful.

The entire procedure leading to image generation might be described through the lens of a co-enunciation process involving the intertwined agencies of the prompt, the training datasets, the database, and the algorithms—a process previously discussed in D'Armenio, Deliège, and Dondero 2024. I would further specify the following agencies at play:

> 1. The initial prompt submitted by the user to the model;
>
> 2. In the case of DALL•E, the text-generating algorithm of ChatGPT. When the user inputs a prompt to DALL•E, the prompt is itself sent to ChatGPT and revised to make it more compatible with DALL•E. We can access this level of translation with the paid version of DALL•E when using the API, that is, the programming mode; having done so, when we write a prompt, ChatGPT can reply as follows: "The exact prompt I sent to DALL•E to generate this image is "...", where what is sent ("...") is different and more detailed than the one the user originally proposed;[30]
>
> 3. The multiple visual results of the prompt, that can evolve through further manipulations and mediations via another prompt addressed to the first output obtained or via other commands;
>
> 4. The annotated database on which the algorithms act through abstraction operations to ensure that an average is obtained: this is the average—or more precisely, the distribution—made by the algorithms of all the image annotations in the database that are selected by the prompt;
>
> 5. The action of the algorithms themselves, which must ensure alignment between the embeddings of the prompt and the embeddings of the image annotations;
>
> 6. The random dimension of the *initialization* of the generative process linked to the diffusion model functioning, which we can attempt to detect by repeating the same prompt hundreds of times. Only through this repetition do we have a chance to measure the agentivity of randomness through the multiple visualizations obtained.[31] More precisely, the randomness in the image generation process using diffusion models relates to the fact that a trained diffusion model has learned, from a noisy image and a piece of text, how to reconstruct an image corresponding to the original (non-noisy) image in the dataset and its associated text. Ideally, then, a well-trained model should be able, at inference time (i.e., when the model is being used), to create an image

> corresponding to any given text starting from “noise” (*without associating it with any specific image from the dataset*, as that association only occurs during training). So, at the start of inference, a noise image is generated and paired with the prompt (more precisely, the embedding of the revised prompt), which is then progressively transformed into a (denoised) image by the diffusion model. This step introduces randomness, because a different initialization (i.e., a different noise seed) will result in a different final image. Depending on the revised prompt and the initialization, the model will draw on different associations between text and visual patterns—all of which are plausible in principle—but within a distribution of possibilities that enables the generation of many different images.[32]

General Realizations

To test the functioning of the machine on visual forms in order to analyze the processes of typification and singularization that are at the basis of enunciative praxis and enregisterment, I decided to test artistic images, notably well-known pictorial styles that are considered to be unique and singular to an author or a movement. Today, many different styles (as many as can be coded in and detected by such large databases) have become partly reproducible thanks to these models that are able to extract styles from all the visual texts that are stocked and annotated in databases.

Let me begin with a quotation from the linguistic and sociocultural anthropologist, Eitan Wilf, who wrote the following lines in 2013, ten years before the democratization of Midjourney, DALL•E, and Stable Diffusion. In his view, these generative models *do not only produce texts, they extract styles.*[33] In his view, even if these generative systems,

> much like a CD or an MP3 file, are technologies of reproduction, they do not reproduce specific texts, or [Peircean] Seconds, but styles, or [Peircean] Thirds. *Their object of reproduction is the principle of generativity* that is responsible for producing the specific texts that are the object of reproduction of the kind of media technologies that have traditionally stood at the center of linguistic and semiotic anthropological research. These systems’ reproduction of style consists both in their ability to *abstract a style* from a corpus of Seconds and to generate new and different Seconds or texts in this style, indefinitely so. (2013:186, my emphasis)

Following Wilf, the images automatically produced by these models may, in the terms of continental semiotics, be conceived of as *sui generis* realizations, that is, “general

realizations." But in which sense? And how? That is, by what process of actualization is this competence to realize generals possible?

Here, we can conceive the case of automatically generated images following some suggestions concerning the entextualization process discussed above. Recall that entextualization produces a "construable/iterable *type*" (Nakassis 2025a, this volume) of pattern that linguistic anthropologists call "text," a form of emergence of "textures that 'dynamically figurate' their own pragmatics (their own enunciation)" (ibid.) through a process of cohering together. As already stated, this process of cohering is the first step of the potentialization process in the enunciative praxis perspective, something that detaches from the here-and-now event to construct various sorts of coherence. But this cohering is also a presentification of generals, the ostension of possible diverse combinations situated between actualization and realization. It's not possible to identify these general realizations with realizations in the sense of definitive embodiments of colors and lines on an organized substrate: on the contrary, they are distributions of qualia that can be "continued" according to different purposes (artistic, scientific, and so on), as explained before.

Every visual output of a generative model may in fact be considered as a *coherent prediction* and, for this reason, a general realization, something visualizing possible combinations of indexed characteristics. In this sense, the work of Nakassis (2023a) on different kinds of entextualization can be useful. He distinguishes three dimensions of coherence—denotational (coherences of symbolic grounds: information structure, narrative), interactional (coherences of indexical grounds: pragmatic action); and aesthetic (coherences of iconic grounds: imageistic patterns of qualia)—and thus three kinds of text: denotational texts, interactional texts, and image-texts. Regarding aesthetic coherence, Nakassis's proposal is to avoid essentializing the visual substance of expression of text-artifacts (and other substances such as sound or audiovisual) finally aiming at: "Dissolving the language versus image opposition to clarify what traverses both, this is what I have called an image-text" (Nakassis 2023a:75). In this sense, in his use, the term "image" does not only include visual forms. In fact, the image is "distributed "in" and across sign activity, in whatever medium or modality" (2023a:75). Nakassis' view on image-texts (which includes visual pictures)[34] offers crucial advances in linguistic anthropology in order to study the intertwining between social and political facts and their relationship of "representation and presence" with cinema production and reception (Nakassis 2023b). However, in my view, this perspective must be adjusted in relation to the social status of visual images and the values that shape their circulation across various social domains (science, religion, art, advertising, law, politics). For instance, in the context of works of art, especially in the case of "autographic" arts such as painting, the notion of image-text is less applicable insofar as the notion of a "text" in linguistic anthropology describes not a

token-instantiation (a text-artifact) nor the evenemential process of its emergence (entextualization) but a semiotic type, a virtualization that is iterated (recontextualized/actualized) through its realized tokens (what Nakassis, following W. J. T. Mitchell, calls, in distinction to image-texts, “pictures”). This is because autographic arts involve non-reproducible artifacts that matter not only for the specific and unique composition of markings on the substrate (defined by Goodman as dense in syntax and in semantics) but also for their very singular history of instantiation.[35]

In continental semiotics, the term *text* has a different technical meaning, denoting a dense whole of meaning that is fully realized by some token object. An image, thus, such a painting circulating through an artistic status, is a text in the sense of sacralized object of contemplation and (auto)-reflection. Using a similar notion, the French mathematician Thom (1983) describes a painting as a closed space inhabited by forces in cooperation and in conflict. On this view, a painting is a textural concentration of forces that *cannot be dissolved in other pictures nor in natural language*. This for at least two reasons which are related to the autographic mode: the first reason concerns the image as a text in the continental semiotic sense: the painting which circulates as a work of art condenses a unique, non-repeatable (theoretical) proposal about topology, chromatism, and outlines.[36] The second reason is that in the artistic domain in the Western tradition, paintings are valued for their specific and non-reproducible history of instantiation encompassing a certain technique, certain materials, and the quality of the instruments used (the image as a material object).[37]

If paintings as works of art circulating in the art world are text-objects in the continental semiotic sense, because they are unique given their specific kind of density of marks and a specific embodied theory of space (or color or something else), and because of their instantiation history (image as object[38] valuable for its specific substrate, historically-indexed materials and colors, specific application and specific gesture, and rhythms of inscription), on the contrary the AI generated-images in scientific domain are like image-texts in the linguistic anthropological perspective, that is, iterable types of patterns whose substance is hybrid and in constant translation.[39] In fact, images conceived of as outputs from generative models are the cohering of patterns detached from a general context, a context represented by the database as operated on through algorithms.

A Question of Style

Let us turn to the question of styles and their actualization, that is, the capacity or competence to produce *texts*, in the sense of linguistic anthropology: detachable patterns from an encompassing context; in our case, new visual optimizations of the prompt from its database. When you propose a prompt to Midjourney, the prompt operates a selection

within image annotations contained in the entire database; this selection identifies a region of the database, a region in which the prompt, that is then transformed to an embedding (a list of numbers), encompasses the relevant and corresponding image annotations. In this way, the images that are made relevant by the words present in the prompt constitute a semantic region, which associates the image annotations to the different words making up the prompt.

Let us consider some examples of this functioning. Asking Midjourney to generate images of Van Gogh via the following prompt, “A landscape in Van Gogh’s style,” we obtain four default images (Figure 5).

Figure 5. Prompt to Midjourney by the author and colleagues in 2023: “A landscape in Van Gogh’s style”

If we repeat the prompt “A landscape in Van Gogh style”, interestingly, the results are always similar but different; and it is this difference, or *variability* across each set of images that is produced that is important for our discussion (Figure 6). While a user may be interested, ultimately, in one or the other of the GenAI’s output for aesthetic purpose, what is produced by such algorithms are not isolated, unique and singular images per se, but *samplings of regions of the dataset*.

Figure 6. Prompt to Midjourney by the author and colleagues in 2024: "A landscape in Van Gogh style"

In other words, what is produced by the generative models are not images per se but *extractions of features typical of a region of the dataset*—this is why I propose here to call this ensemble of visual marks resulting from a region of a database a "regional style." A region of a dataset encompasses examples of patterns produced by the work of algorithms *exploring* (Meyer 2023) certain domains of the dataset associated with annotations referenced to the style mentioned in the prompt, and operationalized by embeddings. The style of the images obtained from the prompt, thus, are partly determined by the semantic *region* of the database that is solicited by the prompt. To study regional style semiotically, it is necessary to describe the plastic features characterizing the multiple outputs obtained. In this sense, the plastic analysis described by Greimas (1989)—a structuralist analysis of topological, chromatic, and eidetic oppositions and repetitions within an image—must be further developed. It is essential to focus on both the repetition and variation of *isolated* formal features (e.g. the color yellow and its modulations) as well as recognizable objects that appear across various images (for instance, in Van Gogh-inspired landscapes, the recurring motif of a road beginning in the foreground and continuing into the middle ground). Additionally, attention must be paid to *compositional* recurrences and variation across the outputs, such as the use of asymmetrical topological and chromatic divisions of the pictorial surface, curvilinear

outlines, and parallel lines that emphasize the stratification of the landscape levels characteristic of these Van Gogh inspired pictures.

Confronted with a reservoir of diverse plastic categories—such as figure/background, surface/depth, distant viewing/close viewing, linear/curvilinear outlines, and so forth— it becomes possible to identify prevailing compositional trends and the recurrence of specific details or plastic characteristics (e.g., circular celestial bodies populating the sky or chromatic nuances of the sky) associated to the same prompts or slightly different prompts. The regional style encompasses this range of variations of recurring features and objects represented which can be quantitatively assessed through statistical analysis.

Describing a style is, of course, a challenging task, as it traditionally concerns the variation of a microsyntax of marks that remains recognizable across different represented subjects, such as a landscape or a face. However, to what extent does this actorial syntax change when drawing a landscape or a human face? In the case of artificial styles, this question can be tested by prompting, for example, “Van Gogh’s style” but specifying a subject other than a landscape — such as a portrait (Figure 7).

Figure 7. Midjourney. Prompt by author and colleagues in 2025: “A portrait in Van Gogh’s style”

When looking at these outputs, it seems that we can identify the same microsyntax of Van Gogh's style in both the portrait and the landscape tests. Some portraits are even embedded within typical Van Gogh's landscapes, reproducing the luminous syntax characteristic of his depictions of the sun and other celestial bodies. More generally, the kind of brushstroke remains consistent across both landscapes and portraits, as if the style of the *Starry Night* (1889) had been transferred into faces and especially into Van Gogh's self-portrait series. This suggests that the model has inferred the landscape depicted in Van Gogh's *Starry Night* style[41] in a very precise way during its training phase; but, perhaps, too precisely or too broadly. If we compare, for instance, *Orchard with Cypresses* (1888)[41] and *Portrait of Madame Ginoux (1890)*[42] both by Van Gogh, we can observe important differences in the brushwork. The first painting features very fragmented strokes, with different colors intersecting, while the second is characterized by large areas of uniform color with broader yellow strokes that do not intersect with those of other colors. The model does not appear programmed to replicate these differences either in the chromatic composition of the strokes or in the forms of the strokes. In this sense, we might say that the averaged algorithmic imprint "Van Gogh" characterizes landscapes and portraits in the same manner.

Figure 8. Prompts to Midjourney by author and colleagues in 2024: Left – "A landscape in Picasso's style"; center – "A landscape in Paul Klee's style"; right – "A landscape in Gauguin's style"

Each regional style is mobile and detachable.[43] This mobility can be explained by the fact that these general realizations (technically, model predictions) are always varying in the range of a limited region, that is, a zone of more or less probable combinations of embeddings. Of course, every database encompasses multiple regions corresponding to different author functions (such as painters or named artistic movements) and semantic zones evoked by the prompt (for instance, the semantic zone of "landscape"). It is only by repeating the same thematic prompt (e.g., landscapes) while varying the names of painters and artistic movements (in the style of Picasso, Klee, Gauguin, etc.) that we can

discern what remains stable and recognizable, despite the variations introduced by the stereotyped style of each painter (Figure 8).

What remains invariant across all these stereotyped generations of particular regional styles can be identified in my view as the *model's style*, making it recognizable in *all* the generations produced from different prompts: I call this the "global style" of the generative model. This global style can be assimilated to the social voice in the tradition of linguistic anthropology (Agha 2005) and visually, it corresponds to the average, not of each painter's annotated characteristics, but of the overall characteristics concerning visuality on which the machine has been fed and trained. If regional styles show the declensions of local zones of the database, the global style encompasses the most popular styles of our time, a mix made by a photographic, graphic, and "realistic" styles filtered and inspired by animation, video games, East Asian comics, and so on. The regional style, on the contrary, can be described as enregistered *models of individual voices* (Nakassis 2025b), as simulations of recognizable authors. Here, we say *models* of individual voices as the regional style belonging to Van Gogh's stylistic variations model the specificity of what Van Gogh embodies in art history (i.e., these variations are not an index to a biographic person per se), which is to say, the characteristics that make his work recognizable in contrast with all other painters of his time and other times.

Regarding what I call the "global style," Lev Manovich (2025) notes that: "AI neural nets used for image generation frequently have a default 'house' style. This is the actual term used by Midjourney developers. If one does not specify a style explicitly, the AI will generate it using this 'default' aesthetic." While this may be the case, note that even if a singular style is specified in the prompt, a certain general, global style, "vision of the world," remains recognizable. As for the tests shown earlier, the flashy colors are of course not inherited from 19th- or 20th-century paintings but from other types of pictures distributed in the general database. The style of each model is, in fact, identifiable *when we subtract* all the characteristics of the various regional styles we have invoked in our prompts: what remains after the subtraction of all the specificities characterizing each painter's style mentioned in the prompts corresponds to the global style. Indeed, in Gen-AI universe, style is not defined solely by *differences*, as in the art history tradition but, more importantly, by *subtraction*.

Midjourney tests are useful, thus, for recognizing not only the plasticity of stereotypical forms of various famous painters and enregistered styles but also to reflect on the idea of composition that the machine develops when the prompts contain two different styles, or an image of a painter uploaded on the platform combined with a prompt on another style/author. Manovich and Arielli (2021–2024) mixed styles of different painters, for instance Bosch and Malevich, or Brueghel and Kandinsky. In such cases, the machine

uses abstract artists such as Malevich and Kandinsky as landscape artists, offering the overall topology of the image in their style while hosting small figurines in styles typical of painters such as Bosch and Brueghel. The latter are clearly, in the machine's perspective, considered as painters of multitudes of small (sometimes monstrous) characters and not as painters who possess an encompassing view to organize the surface space as a whole (even if, traditionally, Bosch and Brueghel are considered landscape artists themselves!). We realize that abstract painting is seen by the AI model as an empty geometrical space meant to structure representations that incorporate visual elements characteristic of more figurative painters such as Bosch and Brueghel.

During these series of tests, I also explored various mixings of different painting styles, for instance using Leonardo's *Mona Lisa* associated to the prompt: "Mona Lisa in manneristic Pontormo style" (Figure 9).

Figure 9. Prompt to Midjourney by the author and colleagues in 2024: "Mona Lisa in manneristic Pontormo style"

The result allows us to understand that the database region concerning Pontormo is made of collective scenes (or paintings annotated as "collective scenes") and not portraits; in fact, the Mona Lisa in Pontormo's style is surrounded by various characters, even more numerous when the image is zoomed out as in the images on the right!

Of course, it is possible to continue testing such combinations and adjust the model's system of weights available to users. For instance, we can prompt the model to generate an output with a predominant influence from Leonardo or, conversely, from Pontormo's Mannerist style, using weighted percentages (e.g., 30% Mona Lisa, 70% Pontormo).

Additionally, we can use the “temperature” parameter, which allows for a more “creative” output of Midjourney. This parameter instructs the model, in situations requiring random selection, not to always choose the most probable options, but to also consider other plausible alternatives that may be less obvious.

Final Remarks

In these final lines, I’d like to return to the aforementioned concept of enunciative praxis that makes it possible to understand the relation between singular artistic images, automatically generated image series, and image databases.

The theory of enunciative praxis is traditionally useful for understanding the cultural process of production of new forms and their later stabilization, sedimentation, or disappearance. AI databases have allowed this theory to be tested, for the first time, on a “concrete” corpus that coincides with the database of near all recognized and virtualized, sedimented (enregistered) annotated images and styles (what was called “big data”) in the Western art tradition.[44] As seen throughout this article, the database is a specific kind of virtualization, which makes the relations between discourses testable and manipulable. On the one hand, this virtualization is, for the intents and purposes of our analysis, stable because every image keeps its entirety and its configuration guaranteed by the numbers which identify it; at the same time, it also becomes manipulable, translatable, and remixable even if it is not exactly made of units you can move elsewhere as in a puzzle. The manipulation made possible by algorithms operating in databases create an effect of distributing recognizable characteristics even if such characteristics are not all identifiable by viewers of such images as clear-cut units of meaningfulness.[45] In this sense, experiments with generative models confirm the difficulty already highlighted by Benveniste, Barthes, Lindekens, and other theorists of visual language to be able to isolate elements in visual compositions. Further, while outside of databases it is possible to extract stabilizing and stabilized regularities, through these models and their openness to being tested and experimented with, it has become possible for the very first time to follow the (generative) path of this (genetic) process of stabilization and renewal.

Moreover, regarding style, these tests develop and transform the debate on style in art history. Style has always been considered as rooted in a specific time and space, even when it has been theorized as hybrid and changing, as in Warburg and Focillon’s work. Visual style is currently theorized in art history, after Nelson Goodman, as something that cannot be detached from its content because “manner” and themes come together as the result of the same enunciation act.[46] Tests with generative models show us that styles have become not only something detachable from the time-space of their instantiation but also from the image’s theme, that is, what in art history reflection is named “content” (see Ruta et al. 2023).

To account for such detachable and moving styles requires, however, an expanded theoretical apparatus; in particular, and coming from a continental semiotics approach to enunciative praxis, it has been crucial to integrate notions from linguistic anthropology such as enregisterment and the entextualization. These concepts have allowed me to finetune and further specify the particular semiotic process and relations that are central to what Paris School semiotics has referred to as “modes of existence.” Potentialization has been aspectualized in two moments, an inchoative and a terminative one: the inchoative one depends on entextualization, the emergence of a cohering whole (a visual composition) following the event-occurrence of a prompt launch. The terminative moment of potentialization in the process of image generation coincides with the time during which these images have to find a place in the movement of visual culture: they can survive or not. If the user-selected images are published and implemented in public spaces, they may begin a path that will lead them to be recognizable and so enregistered in a reservoir of forms. Moreover, I have differentiated between images that are realizations, that is, paintings as works of art—see my comments on Ghirlandaio’s painting, an exemplification of the realization as a result of a stabilized combination of different modes of presence and what I call “general realizations”, that is, visualizations that are optimizations of prompts. These visualizations are not meaningful per se as accomplished, definitive, and isolated images. They are situated midway between actualization as a competence to draw and inscribe forms evoking enregistered visual styles, and singular realizations. In the case of outputs generated for the sake of scientific observation, as in this article, the actualization is accomplished through the mobilization and recall of various fragments and textures of enregistered discursive forms. In this specific context, the invocation of different extensions and intensities of visual forms passes through the enregistered filter of statistical probabilities, which selects among competing enregistered forms to be combined. Actualization, in scientific status, can be seen as a multitude of predictions, that is, probable and diverse associations between embeddings (of image and captions) that are crucial as means to explore databases and their characteristics. Realization, in the domain of generative models, is limited to the aesthetic status of images, when a visualization that is stratified through multiple manipulations has been selected as the unique image by the user (e.g., to be implemented in public spaces such as galleries, museums, social media, and so on).[47] In this sense, the intertwining between enunciative praxis, entextualization, and enregisterment has made it possible to specify different statuses of images and different domains of production and circulation.

The question of the virtualization of our tests remains to explore. Technically speaking, Midjourney, whose database is secret and determined by choices of the enterprise, prevent the virtualization of our image-explorations, in the sense that our images remain in a space that will not be re-employed by the model (these experiments are not, in any direct way, fodder for further model training). On the contrary, Stable Diffusion, having an

open-source and an ever-changing database, and having been trained through large datasets such as LAION-5B (an open-source collections of annotated images from Internet and social media), can be fed with all the images produced by the users. Contrary to what happens with Midjourney, where we have to become recognized artists to be able to take part in its database and to participate in the transformation of what is now sedimented, thence becoming virtual possibilities in the generation of new images, with Stable Diffusion we can enter the database through our generated images and so are able to make the dynamic of enunciative visual praxis begin again.

Acknowledgments. The author is indebted to the colleagues and friends who contributed to enhance this text thanks to their precious insights: Constantine Nakassis, who generously and constantly offers precious suggestions to his colleagues to enhance the potentialities of their work, and to Aurora Donzelli, who helped me to more deeply reflect on the notion of style. The experience in the "Enunciation/Entextualization" working group was very useful for my work and I especially thank the two main organizers, Constantine Nakassis and Tatsuma Padoan as well as the other members of this year-long seminar held at the Paris Center of the University of Chicago in 2023-2024. I'd also like to acknowledge three colleagues with whom I always fruitfully discuss my ideas: Enzo D'Armenio, Pierluigi Basso, and Adrien Deliège. The latter taught me a great deal about generative models, and I especially thank him here for the sections devoted to the different types of randomness on which image generation depends.

Endnotes

1. Realization is represented schematically by a dot, as it is the only pole within enunciative praxis that has a concrete manifestation on a material substrate, and the syntax of inscriptions is fixed within the surface of that substrate. The other three poles are processes that, from a semiotic point of view, can only be apprehended through comparisons between different discursive realizations (artifacts) produced in different moments of time. The use of a circle in the schema is understood when one considers that actualization and potentialization are two types of processes that strictly depend on the manifested textualization that is realization. Actualization is a process of organizing arguments into a form of schematization that bridges the general forms of virtualization and the specific, singular forms of each realization. In contrast, potentialization refers to the making-available of forms for the circulation; once potentialized, these forms may be adopted, transformed, or discarded. Virtualization is positioned outside the circle because it is more detached from particular realizations. It can be described as a kind of diagrammatic form in which realizations lose their syntagmatic singularity and instead assume the status of schematic formulas awaiting specification.↩

2. According to Goodwin's definition (1994:606), professional competences, what he calls "professional vision," are "socially organized ways of seeing and understanding events that are answerable to the

distinctive interests of a particular social group."↩

3. The theory advanced by Agha concerns more generally the circuit of typification and singularization and change of discursive forms via voicing phenomena, as masterfully expressed in the following question: "In the course of any process of social dissemination, register models undergo various forms of revalorization, retypification, and change. What role do voicing phenomena play in encounters with registers and in the maintenance and transformation of register values over time?" (2005:38) In this sense, role alignment appears as an unstable moment of the process in which enregisterment occurs. Here the concept of voice is inspired by Bakhtin's work (1981, 1984) and intended as "ways in which utterances index typifiable speaking personae; similarly, many registers index social attributes of speaker such as gender, class, caste, and profession" (Agha 2005:39). Agha proceeds nevertheless to some adjustments and specify these voices as "enregistered voices," a sort of "socially performable figures," in contrast to entextualized voices (or voicing contrasts) that singularly emerge in particular texts-in-context. Enregistered voices are conventionalized entextualized voices. Yet not all entextualized voices end up becoming enregistered necessarily.↩

4. This process of stabilization of knowledge in contemporary scientific discourse is described in Dondero and Fontanille 2014[2012].↩

5. On the topic of de- and re- contextualization, see Nakassis's (2025a) paper in this issue. In a personal communication, Nakassis writes: "A text is a type (hence it is virtual, repeatable). A text is thus not an artifact, and is contrasted to a text-artifact, which is a material instantiantion of a text (every text must be manifested in some text-artifact, but is not reducible to it); this is like W. J. T. Mitchell's distinction of an image and a picture. What kinds of texts are there? How is a text 'made', i.e., what makes it come into existence, perdure, or dissolve? What is the text's relationship to what is thus constituted as its co- or con-text(s)? Entextualization, thus, is the process through which an emergent array of sign-tokens in time and space (that thus have some "texture") become a text." Nakassis explains that *Hamlet* can be considered a kind of complex text and states that *Hamlet*, the play, is not the same as the text-artifact like a DVD or a particular performance of Hamlet, or a book-artifact of *Hamlet*. "Yet *Hamlet* doesn't exist independent of such performances; such performances happen in contexts. And every such recontextualization can change the text (potentially)" (e-mail exchange, December 8, 2023).↩

6. For a contemporary semiotic analysis of the epistemological and methodological differences between textuality, interaction, practice, textualization and the notation of interactions—based on professional exchanges among participants in scientific interactions in architectural engineering—see Dondero 2014, 2017b.↩

7. The theory of figurative and plastic semiotics, as well as “semi-symbolism,” are based on the paintings of great artists such as Paul Klee, Andrea Mantegna, and Wassily Kandinsky. On this topic, see, for instance, Floch 1985. On the topic of religious iconography focused on its evolution (typification and change) from painting to fine art photography, see Dondero 2009[2007]; on Greek literature and Bacon’s painting see Basso Fossali 2013. More generally speaking, continental semiotics, and notably visual semiotics, has always had a close relationship with art history, as Marin’s (1994) work on uttered enunciation in paintings testifies. For some methodological affinities between art history and continental semiotics (in particular, on Victor Stoichita’s works on metavisuality), see Dondero 2020.↩

8. A crucial example of the singularization of a style is Francis Bacon, as described in *Francis Bacon: The Logic of Sensation* by Gilles Deleuze (2003[1981]). In this seminal book, the French philosopher explains how Francis Bacon found his specific way of painting by processes of exclusion, subtraction, inclusion, and differentiation from other painters.↩

9. Following the glossematics of Hjelmslev (1943), in Greimasean semiotics, an utterance, or text, is composed of a plane of expression and a plane of content, each of which are composed of a form and a substance. In Hjelmslev’s theory, the purport (which is abstract and external to the utterance) becomes substance at both planes, as the result of the work of the schematic form operating on the purport. In this sense, Cigana (2025) states: “To conceive of form as *transposable* is not to render it impermeable to substance: on the contrary, it is to make it accessible *through* a multiplicity of different substances, not only of expression (phonic substance vs. graphic substance, etc.) but also of content (affective substance, conceptual substance, volitional substance).” For instance, in a painting, the form of expression can be identified with visual forms such as oppositions and arrangements of lines, colors and positions of eidetic outlines; the substance of expression coincides with the canvas and the oil colors (as substrate and application respectively), which depend in turn on the intensity and manner of the painter’s marks on the canvas. The form of content can be conceived of as the genre of the painting (portrait, landscape, still life, etc.) and the substance of the plane of content as the contents and signifiers that are operational in a certain culture.↩

10. If an enregistered voice is a “way of speaking,” here a style is a “way of forming” (e.g., a way of painting); in this sense, just as, following Bakhtin, voices are always in dialogue within some discursive text (including in social interactions, including within a “word,” as Bakhtin puts it), so too are styles within a painting.↩

11. As explained in Dondero 2020, discussing the similarities between Focillon’s and Warburg’s works, an important notion examined by both scholars is the interval between the manifestation of one form and the manifestation of another which takes that form and transforms it. Concerning the interval, Focillon states that:

> It is not sufficient to know simply that events follow one upon the other; what is important is that *they follow at stated intervals.* And these intervals themselves authorize not only the classification of events, but even (with, of course, certain restrictions) the interpretation of them. The relationship between two facts in time differs according to the distance that separates them; *it is somewhat analogous to the relationship of objects in space and in light, to their relative sizes and to the projection of their shadows.* (Focillon 1992:138, my emphasis)

In this sense, when confronted with a composition of forms originating from different periods, we must study the temporal latency—the distance between the "original" form and its reshaping. This reshaping of an "original" form occurs within a process that also engages other forms from the past, whether from more distant or closer times. In this way, a picture can be described as an oscillation of forms from different times and latencies, seeking equilibrium.↩

12. This question is linked to the hypothesis Basso Fossali (2017:ch. 2) has made about the fact that the unequal distribution of modes of existence on the surface of the image produces, despite of the static character of the plane of expression of the image, a temporal effect of dynamicity on the spectator. For a study of the dynamism in still images from the double perspective of visual semiotics and image processing, see Deliège, Dondero, and D'Armenio 2024.↩

13. These different logics can be addressed as metrical contrasts that allow a contrastive individuation of voices according to the description of Agha (2005).↩

14. For a methodological tool in tensive semiotics for the analysis of literary texts, see Zilberberg 2012. For describing the rhythm in which a figure is inserted in an image (disruptively, slowly, regularly, irregularly, iteratively, punctually) and for a tensive analysis of the figure-ground dynamic in the portrait genre, see Dondero 2020, 2024a.↩

15. On this topic, see Didi-Huberman and Mannoni 2004. On the relation between movement depicted in images and empathy, Pinotti writes: "Motor empathy was particularly apt to explore the intensification of movement in static images conveyed by mobile accessory forms such as garments and hair, as symptoms of a renewed sensibility of the Renaissance for the antique" (Pinotti 2021:116).↩

16. Warburg's vision of the constitution of an image is comparable to Focillon's. The fabrication of an image constitutes a potential energy which is stored, and which may be reactivated and discharged by other images in succeeding periods. This process may, for instance, result in the revival of a motif, for example, a motif from Greek paganism in the Renaissance. In other words, images are considered to be *dynamograms*, which are forms having recorded a dynamic force. The aesthetics of dynamograms which Warburg aims to

establish can be described through three phases of a process explaining the survival of the patterns: polarization, depolarization, and repolarization. See Hagelstein 2014.↩

17. In linguistic anthropology, an important contribution to this question is Kockelman 2024, which highlights the standardized responses offered by models like ChatGPT and proposes strategies to misalign the values of the companies that created these models with the parameters governing ChatGPT and our interactions with it. The response we, as users, receive from ChatGPT is the one that best aligns with the criteria of the metaprompt—the rules governing the selection among various possible responses—and, more specifically, the one that achieves the closest alignment between the model's training data (refined through feedback) and the user's queries. On this subject, Kockelman reminds us that the judgments or feedback provided during the machine's training is, for now, conformed to the instructions of the companies managing these models. For a different perspective on alignment and disalignment of values during complex and iterated interactions with ChatGPT, see Basso Fossali 2025.↩

18. By contrast, a GenAI like Stable Diffusion represents a more open system (virtualization) insofar as its database was continuously fed with a public dataset of images and captions freely scraped from the Internet. Moreover, Stable Diffusion has openly released its code and model weights. While Midjourney and DALL•E (OpenAI) have not released information about which datasets were used to train them, all the information about Stable Diffusion is available: the latter was trained with the Large-scale Artificial Intelligence Open Network 5B (LAION-5B), released in March 2022. For more on the training of these models, see Somaini 2023.↩

19. Plastic analysis, in relation to figurative analysis, was first described by Greimas (1989). See also the perspective of Groupe Mu (1992) on figurative and iconic analysis. ↩

20. According to Benveniste, systems which do not have units and rules to govern their relations would be dependent upon natural language for any description or for any "interpretance." As Benveniste (1981[1969]) stated, verbal language is the system for interpreting other signs—and society more generally. For Benveniste, the handicap of non-verbal signs is indeed *the lack of distinct constitutive units* and rules which manage the linguistic system, such as the *rules of selectivity* (on the paradigmatic axis) and *recurrence* (on the syntagmatic axis). The problem, for theorists of the relations between the verbal and the visual such as Benveniste, is the presumed *liberty of non-verbal signs*. Gestures, sounds, and images would all lack a *repertoire*, a system of distinct signs, and syntactic rules to govern their syntagmatic dimension—there would be no grammatical rules that would guarantee the intelligibility of gestural or pictorial statements. Non-verbal signs would be unable to ensure the predictability of their occurrences or their transmissibility through a universally accepted system of notation. This problem also concerns plastic arts: "Therefore, the meaning of art may never be reduced to a convention accepted by two partners. New terms must always be

found, since they are unlimited in number and unpredictable in nature; thus, they must be redevised for each work and, in short, prove unsuitable as an institution. On the other hand, the meaning of language is meaning itself, establishing the possibility of all exchange and of all communication, and thus of all culture (Benveniste 1981[1969]:16). On Barthes' translinguistics and critiques on his natural language model to study visual texts, see Dondero 2020 and Lagopoulos et al. 2025. A discussion on visual *langue* according to the problems raised by Benveniste and Barthes can be found in Dondero 2020, which states that the problem of a visual language can be methodologically solved assuming a perspective going from the global to the local dimension of signs, through the study of the relationship between *image social status* (artistic, scientific, religious, etc.), *visual genres* (portrait, landscape, etc.), and *plastic analysis* (eidetic, chromatic and topological categories), the latter reconfigured through tensive semiotic approaches. Concretely, this means that in an image, the direction of a glance, of a raised hand, of a pointed finger, but also the formal/plastic characteristics such as the change of the luminosity within a painting, the change in saturation and so on are crucial to produce meaning such as, for example, "elevation" or "fall," "terrestrial" or "transcendental." These couplings between plastic oppositions present on the plane of expression of paintings and oppositions on the plane of content are determined by a methodology called "semi-symbolism" in Paris School semiotics (Greimas 1989). More recently, these couplings have been studied according to their genre and their field of affiliation/social domain/status (Basso Fossali and Dondero 2011). The geometry of a gesture also counts: for instance, in Renaissance and Baroque painting, a gesture composing a circular figure on the plane of expression reflects stability and calmness on the plane of content; on the contrary, a gesture composing an irregular triangle will reflect disrupting directions, and a conflict in semantic tension on the plane of content.↩

21. I'm convinced that the most useful work on this direction has been made by a painter, and not by a scholar: Paul Klee. In his *Pedagogical Sketchbook* (1953) and *Notebooks* (1961[1956], 1973[1970]), Klee built a reservoir of lines and color distributions, classifying them from the simplest to the most complex, and then exploring their combinations (lines with lines and lines with colors). As for the content plane, he studied measures for lines as well as their active/passive agency in respect to planarity, and weight for colors. Subsequently, Klee experimented with these lines and colors in his paintings and drawings to test the various possible combinations.↩

22. Generally speaking, big data have transformed the practices of artists and the theoretical approaches in many disciplines specialized in the study of images, such as art history, philosophy, aesthetics, and semiotics. In recent works (Dondero 2025a), I considered two manners of developing the current availability of visual big data and their constitution into large collections (for example, large collections of paintings): image analysis (in which databases play a crucial role in studying the evolution of styles in the history of painting and in retracing new genealogies of forms) and image generation, as discussed in the main text.↩

23. As explained by Somaini (2023), the diffusion model operates in two phases: “forward diffusion” and “reverse diffusion.” The first phase, forward diffusion, turns the images of the dataset used for training into indistinguishable “noise images” (i.e., random pixels distributed within the grid-like image space). It does so by recursively adding a bit of “noise” to the image (i.e., the image becomes more and more noisy as noise continues to be added). During the second phase, the phase of reverse diffusion, starting from the “noise images,” the model learns how to recover the initial images of the training set. It does this through a “denoising process” that learns to remove progressively the noise added to the initial images. This allows the model to subtract, for each given “noise image,” the layers of noise that were added until it finds the initial image. This double process trains the machine to overcome some obstacles (the noise) in order to learn how to reconstruct the composition of an image through prompts that guide the generation process.↩

24. For a general discussion about originality, fake and machinic creativity, see Leone 2023.↩

25. We could state that the machine operates by averages accompanied by the aleatoric while a painter uses a diagrammatic device to produce something new out of past forms. On diagrammatic production, see Deleuze (2003) on Bacon’s paintings and Dondero 2023 for a comparison between the diagrammatic thinking in the works of Peirce, Deleuze, and Goodman.↩

26. I’d like to make it clear that I use the term average for the sake of convenience. Indeed, I should be talking about “distribution” rather than average. The notion of distribution encompasses the random sampling of visual features characterizing the average, because within a distribution these characteristics are unequally divided, whereas the average is a concept that brings to mind something “fixed” that gives rise only to “a single version” of a set of features. To be more precise, the images generated use features that do, in fact, generally revolve around the mean. I am indebted with Adrien Deliège for our discussions on this subject.↩

27. In addition, the users/experimenters, if they are programmers, can fine-tune an existing neural network through annotations, building finer correspondences between the lists of numbers that identify natural language descriptions and the lists of numbers that identify images. To make the production closer to one’s wishes and thus minimize the bias or noise produced by overly generic databases, it’s also possible to refine the prompt in different ways, for example, by explicitly indicating to the machine the technique to be used, such as “chalk drawing,” “oil painting,” “fresco,” and so on. Additionally, Midjourney has recently introduced a tool for refining a part of the image already produced, allowing the experimenter to select a part of the image through the “Vary Region” command. This function allows the user to circle/highlight the part to be modified and to enter a prompt corresponding to what the user wants to see appear for just this part of the image. For instance, the user can select an empty part of the image and request the addition of a visual object in order to alter the overall composition. If the user aims at localized modifications in an image, this

kind of instruction is much more efficient than modifying a prompt. It is also a tool which serves to minimize the statistical bias or the aleatory process at the basis of every computational modification within Midjourney.↩

28. We could even argue that the realization process is achieved only in a minor way in scientific practice—specifically, when considering the processes of article publication and dissemination/popularization. In fact, it is rare for a scientific object to be represented by a definitive iconography or an image capable of fully capturing its processual character. ↩

29. I relay here a crucial distinction between *type* and *class* by Aurora Donzelli who, in comments on an earlier draft in May 2024, writes: "Type—an abstract individual category resulting from processes of induction made from its concrete occurrences (tokens)—is used to deduct/produce more concrete instantiations: from the concrete/individual to the abstract individual to, again, more concrete individual tokens. In a sense, types can be understood as generative entities in that they operate like a stylistic matrix that determines/shapes the concrete instantiations emanating from it. To the contrary, class is not an individual abstract entity. Rather, it is a collective inventory of one or more properties, optionally supplemented by rules for determining elements' membership within it. In spite of their similarities, type and class operate quite differently. While class encompasses, gathers, and organizes its elements, type generates its tokens."↩

30. In the case of DALL•E, the results show a gap between natural language requests, which are hyper-politically correct due to the prompt revised by ChatGPT, and the visual output, which does not adhere to the instructions that the machine itself provided, particularly regarding the skin color of the depicted figures or their attributes (characteristics on which the machine revision had specifically worked). This means that certain semantic characteristics available in natural language, made relevant by the automatic revision of the prompts, are not directly made operational in the context of visualization. It also means that the image generations certainly follow the instructions, but they also follow internal rules of the visual composition itself that are independent of the natural language instructions. On these compositional rules, see D'Armenio et a. 2024, and D'Armenio et al. 2025. ↩

31. In the case of a model like DALL•E, which is known to revise the prompt, there is also a degree of randomness at the natural language level. Indeed, the model that revises the prompt is a language model. As such, it functions like all LLMs—that is, probabilistically; it formulates the revised prompt based on the most probable tokens given the initial prompt. There is a kind of random selection within the token distribution, which means that the revised prompt is not always the same for a given initial prompt. Therefore, the prompt actually submitted to the generative model—the revised prompt—already contains an element of randomness that can influence the final image.↩

32. One could even argue that randomness is embedded in the very way models learn—whether in learning to create embeddings, to form text-image associations, or to generate images. This is because, during training, a model's set of parameters or weights is initially randomized. Different initializations can lead to different final models, and consequently to different embeddings and associations. Furthermore, training involves processing large datasets, which are often presented to the model in a randomized order. This introduces additional variability into how the model learns and internalizes concepts. There is also the consideration that core mathematical operations within models—such as addition and multiplication—are performed on GPUs, which may exhibit slight non-determinism. As a result, even identical inputs and computations can yield small differences in outcomes.↩

33. J. Porquet (2025) comments on Wilf's text and proposes another vision on styles. In his view, generative models, rather than reproducing Peircean Thirds, are more adept at extracting Firsts—abstract qualities such as textures, colors, and so forth—from a fixed dataset, and aligning them cotextually in the production of new images. This means that these chains of Firsts, even if they may emulate the style of certain artists to some observers, never exceed the sum of their parts. In other words, the production of a sequence of visual qualities does not imply the capacity to constitute a style.↩

34. Nakassis thus analyzes "the ways in which verbal discourse, in its emergent, evenemential poetics, *functions as an image*" (2023a:77, my emphasis), the term image reminding to a complex of social, ideological, etc. perspectives indexed by a discursive event. In this article, Nakassis also suggests studying the relation between the image-text and pictures made of forms and colors in the following terms: "Images, thus, are not in opposition to language; nor are they the special preserve of our prototypes for images: paintings, film, and so on. Images are in discourse, *they are its shape* (that is, *the contours of discourse cast an image*); they are the basis of our feel for language and of its pragmatic effectivity" (2023a:79, my emphasis).↩

35. For the difference between autographic arts such as painting (one-phase art) and allographic arts (two-phase art, composed of a text of instruction and multiple executions) see Goodman 1976. For an understanding of the status of the original in paintings in the sense of Goodman, in contrast with Walter Benjamin's theory, see Dondero 2009. Regarding the difference between painting and photographic reproduction in respect to aura, see Basso Fossali and Dondero 2011. In the latter, photography is considered as an exemplification of a hybrid art, that is "autography with multiple objects," an autographic mode built on two steps, in which the executions may be differentiated as originals or as reproductions thanks to their degrees of legitimacy (e.g., vintage prints). For some considerations about reproducibility in digital world and in NFT in relation to aura, see Dondero 2025b.↩

36. In this conception, the image functions as a singular and non-repeatable complex of forms, and as a whole structured by a point of view (or a relation of points of view) which makes each painting a reference for thinking an optic or a way of seeing (as in the case of Wölfflin's [1915] work on classic and baroque optics).↩

37. Paris School semiotics has studied the image not only as a text (a structured surface) but also as an object (an inscribed matter), on the basis of the propositions contained in Fontanille 2008 on substrate (*support*), application (*apport*), and gestures of inscription. For a discussion of the materiality of images and their production process, see Donzelli 2025 in this issue; for a discussion on the distinction between genetic and generative approaches to images, see chapter 9 in Lagopoulos et al. 2025.↩

38. In this sense, the artistic image is, in contemporary continental semiotics, considered as a unique object. The difference between the analytical parameter of a text and of an object is that the first one considers an artifact as a closed texture of relations and oppositions on the plane of expression and a complex of relations and oppositions on the plane of content. The approach to the artifact as an "object" highlights the materiality of images and the work needed to produce them. In this direction, these artistic images are not considered as composed solely of relations and oppositions (as in a structural organization) but also of unique, tangible substrates that are non-repeatable and not structurally organized. The substance of the plane of expression is in this way emphasized, in contrast to the early visual semiotics represented by the studies of Greimas and Floch.↩

39. On this topic, Nakassis writes: "the database organizes its output as denotational texts — strings of propositional information — that is then translated into an output that maps such information to perceptible qualia (colors, lines, etc.) organized in a diagrammatic way, that is, as image-texts" (personal communication, December 15, 2024).↩

40. https://www.moma.org/collection/works/79802↩

41. https://en.wikipedia.org/wiki/Orchard_with_Cypresses ↩

42. https://www.metmuseum.org/art/collection/search/436529↩

43. Meyer notes important characteristics of current generative models that are able to detach the styles as patterns of visual information which can travel through different semantic zones and cease to be rooted in a specific time-and-space relation:

> Historical as well as contemporary, collective as well as individual forms of representation can seemingly be detached at will from their time and place of origin and the work of their authors. It is not least for this reason that O'Reilly and others speak of plagiarism. More importantly, the logic of the prompt radically expands and de-hierarchizes the notion of style […] *Thus 'style' ceases to be a historical category and becomes a pattern of visual information to be extracted and monetized.*" (2023:106–7, my emphasis)

Cf. Nakassis's (2025b) discussion of the dialectic traffic between what he calls "person-" and "persona-indexing" registers.↩

44. Of course, I do not mean to suggest that artificial styles have the same status as historical styles. Nor do I wish to engage in discussions about beauty, authorship, aesthetic intentions, and related topics. For more on this issue, see Porquet, Wang, and Chilton 2025. These tests, involving combinations of various historical visual syntax of elements, aim to highlight the machinic process of reconfiguring well-known, stabilized forms. Historical styles serve as the necessary and unique foundation for assessing this "machinic gaze" in comparison to other forms of perceiving the world.↩

45. I thank Constantine Nakassis for this reflection on visual units.↩

46. On this issue, see the important reflections of Pinotti 2012.↩

47. On the signature as a form of accomplishment and election in the art world of paintings, see Stoichita 1992.↩

References

Agha, Asif. 2003.The Social Life of Cultural Value. *Language & Communication* 23(3–4):231–73.

Agha, Asif. 2005. Voicing, Footing, Enregisterment. *Journal of Linguistic Anthropology* 15(1):38–59.

Agha, Asif. 2011. Commodity Registers. *Journal of Linguistic Anthropology* 21(1):22– 53. https://doi.org/10.1111/j.1548-1395.2011.01081.x

Bakhtin, Mikhail M. 1981. Discourse in the Novel. In *The Dialogic Imagination: Four Essays*, pp. 259–422. Austin: University of Texas Press.

Bakhtin, Mikhail M. 1984. *Problems of Dostoevsky's Poetics*, edited and translated by Caryl Emerson. Minneapolis: University of Minneapolis Press.

Basso Fossali, Pierluigi. 2013. *Il Trittico 1976 di Francis Bacon. Con "Note sulla semiotica della pittura."* Pise, ETS.

Basso Fossali, Pierluigi. 2017. *Vers une écologie sémiotique de la culture. Perception, gestion et réappropriation du sens.* Limoges: Lambert-Lucas.

Basso Fossali, Pierluigi. 2025. Rationalités correctives et intelligence artificielle assistée. Les doubles contraintes des humanités numériques, *Semiotica* 2025(262):71–109. https://doi.org/10.1515/sem-2024-0189.

Basso Fossali, Pierluigi and Maria Giulia Dondero. 2011. *Sémiotique de la photographie.* Limoges: Pulim.

Benveniste, Émile. 1971. *Problems in General Linguistics*. Coral Gables, FL: University of Miami Press.

Benveniste, Émile. 1981. The Semiology of Language. *Semiotica* 37:5–23.

Beyaert-Geslin, A. 2017. *Sémiotique du portrait. De Dibutade au selfie.* Louvain-La-Neuve. De Boeck Supérieur.

Bordron, Jean-François. 2013. *Image et vérité - Essais sur les dimensions iconiques de la connaissance.* Liège: Presses Universitaires de Liège.

Bourdieu, Pierre. 1984[1979] *Distinction: A Social Critique of the Judgement of Taste,* translated by Richard Nice. Cambridge, MA: Harvard University Press.

Briggs Charles and Richard Bauman. 1992. Genre, Intertextuality, and Social Power. *Journal of Linguistic Anthropology* 2(2):131–72.

Chumley, Lily. 2016. *Creativity Class: Art School and Culture Work in Postsocialist China*, Princeton University Press.

Cigana, L. 2025. Forme, structure, affectivité. La logique des sentiments en sémiotique structurale. *Signata* 16.

D'Armenio, Enzo, Maria Giulia, Dondero, Adriene Deliège, and Alessandro Sarti. 2025. For a Semiotic Approach to Generative Image AI: On Compositional Criteria. *Semiotic Review*, 9. https://doi.org/10.71743/ee5nrx33.

D'Armenio, Enzo, Adrien Deliège, and Maria Giulia Dondero. 2024. Semiotics of Machinic Co-Enunciation. About Generative Models (Midjourney and DALL•E). *Signata* 15. https://doi.org/10.4000/127x4.

Deleuze, Gilles. 2003[1981]. *Francis Bacon: The Logic of Sensation*. London: Continuum.

Deliège, Adrien, Maria Giulia Dondero, Enzo D'Armenio. 2024. On the Dynamism of Paintings Through the Distribution of Edge Directions. *Journal of Imaging* 10(11):276. https://doi.org/10.3390/jimaging10110276.

Didi-Huberman, G. and L. Mannoni. 2004. *Mouvements de l'air. Etienne-Jules Marey, photographe des fluides.* Paris: Gallimard.

Dondero, Maria Giulia. 2009[2007]. *Le sacré dans l'image photographique. Études sémiotiques.* Paris: Hermès.

Dondero, Maria Giulia. 2014. Sémiotique de l'action. Textualisation et notation. *CASA: Cadernos de Semiótica Aplicada* 12(1):15–47.

Dondero, Maria Giulia. 2017. Du texte à la pratique: Pour une sémiotique expérimentale. Semiotica 2017(219):335–56. https://doi.org/10.1515/sem-2017-0081

Dondero, Maria Giulia. 2019. Visual Semiotics and Automatic Analysis of Images from the Cultural Analytics Lab: How Can Quantitative and Qualitative Analysis Be Combined? *Semiotica* 230:121–42. https://doi.org/10.1515/sem-2018-0104

Dondero, Maria Giulia. 2020. *The Language of Images. The Forms and the Forces.* Cham, Switzerland: Springer.

Dondero, Maria Giulia. 2023. The Experimental Space of the Diagram According to Peirce, Deleuze and Goodman. *Semiotic Review*, 9. https://doi.org/10.71743/g5pn4v44.

Dondero, Maria Giulia. 2024a. The Face: Between background, enunciative temporality and status. *Reti, linguaggi, saperi. The Italian Journal of Cognitive Sciences,* 1/2024:49–70, doi: 10.12832/113797.

Dondero, Maria Giulia. 2024b. The Relations between Charles Goodwin's Interactional Linguistics and Semiotics: Some Reflections on Multimodality and on the Diagram in Scientific Discourse, *Signs and Society* 12(3):257–75 .

Dondero, Maria Giulia. 2025a. Semiotics of artificial intelligence: enunciative praxis in image analysis and generation. *Semiotica* 2025(262):111–46. https://doi.org/10.1515/sem-2024-0195.

Dondero, Maria Giulia. 2025b. The Aura of the Original and Serial Reproduction. The Cases of Painting, Photography and the Digital. In C. Repapis and S. Myrogiannis, eds. *Economics and Semiotics*, pp. 75-89. London: Routledge.

Dondero, Maria Giulia and Jacques Fontanille. 2014[2012]. *The Challenge of Scientific Images. A Test Case for Visual Meaning.* Ottawa: Legas.

Donzelli, Aurora. 2025. On the Making of Signboards: Corporeal Inscriptions and Material Transpositions of the Writing into the Written. *Semiotic Review*, 12. https://doi.org/10.71743/eq18s297.

Drucker, Johanna. 2019. *Visualization and Interpretation: Humanistic Approaches to Display.* Cambridge, MA: MIT Press.

Fabbri, Paolo. 1998. *La svolta semiotica.* Rome: Laterza.

Floch, Jean Marie. 1985. *Petites mythologies de l'œil et de l'esprit. Pour une sémiotique plastique.* Paris: Hadès-Benjamins.

Focillon, Henri. 1992[1934]. *The Life of Forms in Art.* New York: Zone Books.

Fontanille, Jacques. 1989. *Les espaces subjectifs. Introduction à la sémiotique de l'observateur.* Paris: Hachette.

Fontanille, Jacques. 2006[1998]. *Semiotics of Discourse.* New York: Peter Lang.

Fontanille, Jacques. and Zilberberg, C. 1998. *Tension et signification.* Liège: Mardaga.

Goodman, Nelson. 1976. *Languages of Art: An Approach to a Theory of Symbols.* London: Bobbs Merrill.

Goodwin, Charles. 1994. Professional Vision. *American Anthropologist* 96(3):606–33.

Goodwin, Charles. 2018. *Co-operative Action.* Cambridge: Cambridge University Press.

Greimas, A. J. 1989. Figurative Semiotics and the Semiotics of the Plastic Arts. *New Literary History* 20(3):627–49.

Groupe Mu. 1992. *Traité du signe visuel. Pour une rhétorique de l'image.* Paris: Seuil.

Hagelstein, M. 2014. *Origine et survivances des symbols. Warburg, Cassirer, Panofsky.* Hidelsteim: OLMS.

Impett, L. and F. Offert. 2025. *Vector Media.* Minneapolis: Meson Press & University of Minnesota Press.

Keane, Webb. 2007. *Christian Moderns: Freedom and Fetish in the Mission Encounter.* Berkeley: University of California Press.

Klee, Paul. 1953. *Pedagogical Sketchbook*, translated by S. Moholy-Nagy. New York: Frederik A. Praeger Publishers.

Klee, Paul. 1961[1956]. *Notebooks. Vol. 1. The Thinking Eye*. London: Percy Lund, Humphries & Co., Ltd.

Klee, Paul. 1973[1970]. *Notebooks. Vol. 2. The Nature of Nature*. London: Lund Humphries.

Kockelman, Paul. 2024. *Last Words: Large Language Models and the AI Apocalypse*. Chicago: Prickly Paradigm Press.

Lagopoulos, Alexandros, Karin Boklund-Lagopoulou, Maris Giulia Dondero, Jacques Fontanille, Maria Ilia Katsaridou, and Rea Walldén. 2025. *Semiotics of Images. The Analysis of Pictorial Texts*. Berlin: De Gruyter.

Leone, Massimo. 2011. (In)efficacy of Words and Images in 16th-century Franciscan Missions in Mesoamerica: Semiotic Features and Cultural Consequences. In V. Plesch, C. MacLeod, and J. Baetens, eds. *Efficacy: How To Do Things With Words and Images?*, pp. 57–70. Amsterdam: Rodopi.

Leone, Massimo. 2023. The Spiral of Digital Falsehood in Deepfakes. *International Journal of Semiotics and Law* 36:385–405. https://doi.org/10.1007/s11196-023-09970-5

Manovich, Lev. 2025. From Representation to Prediction: Theorizing the AI Image. In P. Conte, A. C. Dalmasso, M. G. Dondero, and A. Pinotti, eds. *Algomedia: The Image at the Time of Artificial Intelligence*. Cham, Switzerland: Springer.

Manovich, Lev and Emanuele Arielli. 2021–2024. *Artificial Aesthetics: Generative AI, Art and Visual Media*. https://manovich.net/index.php/projects/artificial-aesthetics

Marin, Louis. 1994. *De la représentation.* Paris: Seuil.

Meyer, Roland. 2023. The New Value of the Archive: AI Image Generation and the Visual Economy of 'Style'. In: *IMAGE. Zeitschrift für interdisziplinäre Bildwissenschaft*, Jg. 19, Nr. 1, S. 100-111.http://dx.doi.org/10.25969/mediarep/22314

Murphy, Keith. 2015. *Swedish Design: An Ethnography*. Ithaca, NY: Cornell University Press.

Nakassis, Constantine V. 2019. Poetics of Praise and Image-texts of Cinematic Encompassment. *Journal of Linguistic Anthropology* 29(1):69–94.

Nakassis, Constantine V. 2023a. A Linguistic Anthropology of Images. *Annual Review of Anthropology* 52:73–91.

Nakassis, Constantine V. 2023b. *Onscreen/Offscreen.* Toronto: University of Toronto Press.

Nakassis, Constantine V. 2025a. (Im)personalizing Enunciation. *Semiotic Review*, 12. https://doi.org/10.71743/k99f5r76.

Nakassis, Constantine V. 2025b. Person-indexing Registers, Stardom, Auteurism. *Signata* 16.

Pinotti, A. 2012. Formalism and the History of Style. In M. Rampley, T. Lenain, H. Locher, A. Pinotti, C. Schoell-Glass, and C.J.M. Zijlmans, eds. *Art History and Visual Studies in Europe*, pp. 75–90. https://doi.org/10.1163/9789004231702_007

Pinotti, Andrea. 2021. Formalism and Kunstwissenschaft: The "How" of the Image. In K. Purgar, ed. *The Palgrave Handbook of Image Studies*, pp. 109–30. Cham, Switzerland: Palgrave Macmillian.

Porquet, Julien. 2025. Copier, extraire, générer: Réflexions sur le transfert de style par IA. Invited talk at the *Séminaire International de Sémiotique à Paris* "Intelligence artificielle

générative et nouveaux enjeux sémiotiques. Traduction et appropriations creatives," May 28. Recording available: https://www.youtube.com/watch?v=ICULM3J-Tmw&t=12s

Porquet, Julien, Sitong Wang, and Lydia Chilton. 2025. Copying style, Extracting value: Illustrators' Perception of AI Style Transfer and its Impact on Creative Labor. In *CHI Conference on Human Factors in Computing Systems* (CHI '25), April 26–May 1, 2025, Yokohama, Japan. ACM, New York, NY. https://doi.org/10.1145/3706598.3713854

Dan Ruta, Gemma Canet Tarrés, Andrew Gilbert, Eli Shechtman, Nicholas Kolkin, John Collomosse. 2023. DIFF-NST: Diffusion Interleaving for Deformable Neural Style Transfer. *arXiv preprint* arXiv:2307.04157.

Schneider, Britta. 2022. Multilingualism and AI: The Regimentation of Language in the Age of Digital Capitalism. *Signs and Society* 10(3):362–87.

Seguin, Benoît. 2018. *Making Large Art Historical Photo Archives Searchable*. PhD dissertation. EPFL Scientific Publications. https://infoscience.epfl.ch/record/261212

Silverstein, Michael and Greg Urban. 1996. The Natural History of Discourse. In M. Silverstein and G. Urban, eds. *Natural Histories of Discourse*, pp. 1–16. Chicago: University of Chicago Press.

Somaini, Antonio. 2023. Algorithmic Images: Artificial Intelligence and Visual Culture. *Grey Room* 93:74–115. https://doi.org/10.1162/grey_a_00383

Stoichita, Victor. 1992. Manet raconté par lui-même (II) : La Forme du nom. *Annales d'Histoire de l'Art et d'Archéologie* 14:95–110.

Stoichita, Victor. 1997[1993]. *The Self-Aware Image: An Insight into Early Modern Meta-Painting.* Cambridge: Cambridge University Press.

Thom, René. 1983. Local et global dans l'œuvre d'art. *Le Débat* 2(24):73–89.

Warburg, Aby. 2012[1924–1929]. *L'Atlas Mnémosyne.* Paris: Éditions Atelier de l'écarquillé.

Wölfflin, Heinrich. 1932[1915]. *Principles of Art History: The Problem of the Development of Style in Early Modern Art.* Los Angeles: Getty Publications.

Wilf, Eitan Y. 2013. From Media Technologies That Reproduce Seconds to Media Technologies That Reproduce Thirds: A Peircean Perspective on Stylistic Fidelity and

Style-Reproducing Computerized Algorithms. *Signs and Society* 1(2):185–211.