**THEORETICAL IMPLICATIONS OF FIGURE8**

Ainhoa SÁENZ DE ZAITEGUI TEJERO  
Universidad de Salamanca  
zaitegui@usal.es

A machine desires instructions as a garden desires discipline.  
Jimmie Durham

**Introduction**

Innovation in fiction has been a non-negotiable priority for literary theory ever since Russian formalism commenced work on contemporary poetics in the early 20th century. What constitutes novelty and how to achieve it technically became of paramount importance as creative processes increased in scope and velocity from modernism onwards. The integration of new media in art meant a new scenario for preexisting forms of fiction, regardless of genre: more complexity, more competitiveness. Decidedly inclined to the ideological side of things for the past fifty years, literary theory has, however, kept its friends close, and its enemies even closer: a formal approach to textual analysis has never been abandoned, possibly because Aristotle’s *Poetics* remains pivotal in the development of contemporary theory, thanks to the much-needed revision it was put through by the Neo-Aristotelianism of Chicago after the Renaissance debacle and subsequent chaos. Despite favoring a politically oriented pathway—as is only fit in a politically oriented context—, literary theory shares concerns with new rhetoric and discourse analysis, in which it looks back and ahead simultaneously. Of all the issues tackled by this interdisciplinary trifecta, metaphor reigns supreme.

The evolution of figurative language reveals the human struggle against the constraints imposed by language, perception, and cognition. As artificial intelligence collaborates ever closer with human intelligence in each and every field of knowledge, it is only natural that that cooperation expands to the creative aspects of the human experience, including the naming of things and its impact on the mental processing of them. Devised in the University of California, Santa Cruz, FIGURE8 is a system that generates similes and metaphors that are semantically valid and aesthetically innovative, targeting a human audience yet itself qualified to evaluate its own output autonomously. In the pages that follow, I will discuss some—of the many possible—strategies to insert computer-originated tropes within literary theory. Extensive research has been done in other areas of computational creation and literary theory. For instance, Pablo Gervás at the Universidad Complutense of Madrid studies the convergence of computational story generation and narratology within the frame of Vladimir Propp’s formalist
model. Propp’s methodology is particularly suitable for the purpose, since it proposes an attitude to genre—the folktales, in this case—as a predetermined set of instructions to be executed by an agent—the storyteller—with extremely specific prerogatives and restrictions—but mostly restrictions.

The search for image

To a great extent, contemporary literary theory results from a dispute over image. Viktor Shklovsky’s «Art as Technique» proposes the concept of defamiliarization as part of his argumentation against Alexander Potebnja, who stated that the image should be clearer than the item it represented. Shklovsky’s ostranenie is creation rising from destruction: the last blow to an already dying world. Potebnja—and the 19th century that came with—and operated under the assumption that literature was a matter of imagery, and that literary imagery had an in-built epistemological structure. It was all part of the Romantic doxa, inherited from Plato, surprisingly unaware of Aristotle, and—as usual with 19th-century ideologies—resilient beyond its natural chronologic boundaries. Misconceptions about poetry such as its peculiar link to truth or expression could be understood within the frame of rampant idealism, but it had already become sheer superstition ten years before the deconstruction officially began, by the time Käte Hamburger declared that lyrical poetry was non-fiction or was not at all.

Forty years before The Logic of Literature took up where Hegel left off, Shklovsky opposed the essentialist perception of art as a path to knowledge, if not knowledge itself. Following on Aristotle’s steps, the formalist considered art a device, a thing. As such, it has to serve a purpose. Finding out that purpose is the reason why the Poetics exists in the first place: Aristotle’s theory of catharsis corrects Plato’s error of expelling—and censoring, and politically exploiting—the poet on grounds that what he does is a lie. Aristotle understands that art is neither true nor false. It is something else, a self-contained system of instructions to make an object with language. It is important that these particular linguistic things are made correctly, as they are means to an end, and the end has strict requirements that must be met for the object to do its job.

What catharsis is to Aristotle, defamiliarization is to Shklovsky. When he objects to Potebnja’s assertion that the image must be clearer than that it represents, he is resisting the possibility that art is meaningful but useless. Art is technique. The product of art is an artefact. Far from revolutionary, this is the core of the Aristotelian philosophy of how to be human. If it is man-made, it is a tool. But as anyone who has ever come into contact with a metaphor knows, an image that is clearer than that it represents is hard to find, and probably not worth even seeking in the first place. Like Saussure, Shklovsky assigns opacity to the linguistic sign when codified in an artistic medium. The image stands between the brain and the concept as a wall, a bump in the road that demands taking a detour: rereading, rethinking. It stops the decoding process from happening automatically: the mind is not on autopilot anymore. It becomes aware. Instead of recognizing, it sees. Ultimately, art is a way to live better by making life harder.
As is often the case with human-generated products, the main challenge literature faces is the poor quality of the materials it is made of. Literary language—or the literary use of language, to put it properly—is constricted by the double layer of law that is grammar and rhetoric, the recte dicendi and the bene dicendi. When used conventionally, language does not need to abide by rhetorical rules. In fact, it does not even have to comply to grammatical rectitude, as long as it is comprehensible and conveys the message intended. In literary use, however, language undergoes a scrutiny as severe as crucial for the literary object to be sanctioned as literary at all. The code reaches its highest standards and becomes normative itself. At some point, it ceases to evolve so it can set a precedent. Literary genres are born as a result.

There is an irony to it that is not lost on Shklovsky: the instrument to drag our perception out of its comfort zone is prone to comfort zones itself. If images lean to stagnation, due to overuse or just because they overstayed their welcome, the incessant renewal of the repertoire must become a priority. In 1917, Russian formalism states that images are opaque linguistic objects that tend to get more and more transparent with use and overexposure, which renders them unable to perform as defamiliarizing agents. One century later, we have reinvented rhetoric and placed metaphor right at the heart of it, but images—the invention of them, in the etymological sense of the word—remain an issue nonetheless.

The strategies to provide fresh ways to say the same about the same topics struggle to keep up with the demand for artistic innovation, a process that has dramatically accelerated since modernism. Traditionally, literature has resorted to the manipulation of language as an element of surprise: hexameters like «quadrupedante putrem sonitu quatit ungula campum» or «ibant obscuri sola sub nocte per umbram» appeal to human curiosity for sound and meaning. But the unexpected has modest resilience, and it becomes lackluster long before rhetoric decides to add it to its oxymoronic catalog of predictable resources to surprise, in this case under the names onomatopoeia and hypallage. It seems safe to assume that when the ruler exerts regulation on rule breaking, the rule breaking is not exactly groundbreaking. It is only after modernism that the linguistic norm is challenged in a non-normative manner. A line like «through sames of am through haves of give» is idiosyncratic in ways only comparable to Finnegans Wake, but it is doubtful whether it defamiliarizes anything or just calls attention upon the apparently inexhaustible human determination to derive meaning from any semiotic construct that comes our way. Texts such as Pigmy or «Eleanor» take advantage of the arbitrary nature of the linguistic sign to reveal the arbitrary nature of the mind that encodes and decodes it. Do they serve a defamiliarizing purpose? Is our perception altered by them? Probably not. Not to mention that the use of images is not confined within the boundaries of art. As Lakoff demonstrated, conventional linguistic interaction is governed by figurative language, as is our conceptualization of the stimuli we perceive and process.

Images on repeat used to be regarded as valuable goods, essential even. Horace built his reputation on loci communes and misconstrued the Aristotelian mimesis as write like the Greeks and you’ll be fine. The sonnet heavily relies on the cumulative effect of cliches to conjure up a sense of philosophical depth and consistency. There was a time—and let’s be clear, it was literary history in its
entirety, not just a brief spell—when a used up image was an image that came with a guarantee. But then something changed. Then art became ambitious.

Deconstructing creativity

In his analysis of M. Boden’s creativity model, G. A. Wiggins (2006: 2) defines computational creativity as «[t]he study and support through computational means and methods, of behaviour exhibited by natural and artificial systems, that would be deemed creative if exhibited by humans». The problem with this approach is that theoretically speaking, it attempts to answer questions no one is asking. Can machines be creative? It depends on whether human beings can be creative or not. There seems to be a double standard: when applied to machines, creativity means creation ex nihilo, whereas human creativity is merely a matter of choice. Human beings are creative because they are able to select known elements and combine them in arguably unexpected ways. Moreover, new combinations are enabled not by new rules, but by the breaking of old rules. These acts against the laws of art are in turn assimilated as laws themselves sooner rather than later. There is no clear taxonomy or hierarchy regulating the old and the new sets of rules either. For all the innovative techniques introduced in poetry during the past hundred years, most readers still associate the genre with rhyme, tropes, and introspection. There is only so much imagism can do to overthrow Wordsworth.

More specifically, Wiggins’ premise is hard to reconcile with the death of the author, by virtue of which the human origin of a text is denied the right to a behavior. The insistence of computational creativity theorists on measuring AI-originated art by human standards is just unnecessary, as literary theory does not care for human origins. In true deconstructionist fashion, Barthes killed the author, but he did so only because biographism forced him to, much like Shklovsky was compelled to rethink image only because Potebnja thought it wrong. Barthes’ author dies not by murder, let alone assassination: «the sway of the Author» (Barthes, 1968: 313) was too short-lived and ideologically dependent to accommodate such extreme measures. Quite the opposite, the author dies of dehumanization. Strip them of the features that make them human, and they are gone. Since those features used to be era-appropriate during the birth and rising of capitalism, but fail to meet the requirements of postmodern philosophy, Barthes is confident a bullet has been dodged without exposing ourselves to a missile in the process.

Tirelessly does Barthes highlight the purely instrumental role of the author in the creative procedure. References to «the body writing» (Barthes, 1068: 313) or «the hand, cut off from any voice, borne by a pure gesture of inscription (and not of expression)» (Barthes, 1968: 315) depict the author as an empty vessel that ideology invariably occupies, leaving no room for a supposed individuality that Barthes deems yet another construct of ideology anyway. Significantly, he calls it «the inner “thing”» (Barthes, 1968: 315). That «inner “thing”» comprises everything the citizens of the Romantic and post-Romantic world hold dear about themselves: thoughts, emotions, passions, the delusion of uniqueness. Properly, the «thing» cannot be expressed, it can only be «translated», as «[it] is itself only
a ready-formed dictionary» (Barthes, 1968: 315), dictionary being used here in its semiotic technical capacity. Since the dictionary cannot be owned, only used – it belongs to the community, just like language –, expression is not an option, only inscription is. And to inscribe preexisting content, you don’t need individuality. You only need a body and a socially integrated mind –a mind that has been provided with a dictionary. Have one of each? Congratulations: you are an author.

The depletion of individual resources elevates the text and reshapes the reader as an entity no longer human, but spatial. The text needs to happen some time, somewhere. The reader, «[who] is without history, biography, psychology» (Barthes, 1968: 316), is that time, that place. A consciousness for the phenomenological process to occur, as Iser would put it. We exist so everything else can exist.

Unlike the reader, the author does not need a consciousness, only a dictionary. The literary artefact is always borrowed, never created. It can do without all the human drama. Aristotle could not care less about Sophocles’ life, or what his authorial intention was, or whether he sympathized with Oedipus or not. In fact, the Poetics urges tragedy to stir clear of any human input and adhere to abstract systems whenever possible. The author simply does not factor in. Creativity is restricted to elocutio, and only because like Shklovsky, Aristotle understood that in order to trigger catharsis –or defamiliarization– the audience had to see as new a sequence of events that was anything but. The myth is above the woman and the man. The woman and the man have been trained by ideology to make cultural devices that contain and spread ideology. They have been given the frames, the rules, the materials. Ideology abhors chance, so it leaves no room for it. But –as Aristotle and Shklovsky warn– an art stale is an art rendered inoperative. In order to keep it effective, a heuristic approach to creativity must be enabled.

It can be taught, so it must be able to learn

There are three points where computational creativity theory and literary theory converge. Firstly, both agree that innovation is a requisite for an artefact to be socially adopted as art, regardless of the nature –human or artificial– of its originator. Secondly, they concede that innovative artefacts are possible, as long as innovative is used strictly as a synonym for unexpected, or even unexpectable. Thirdly, the originator must have self-evaluating resources that make it more productive and better at production, allowing it to validate the good outcome and discard the fails:

[It is perfectly possible for an exploratorily creative system to generate real novelty (because it can produce results which are unpredictable) and in some circumstances for those results to be valued, even if it works by brute enumeration, and for the system to be capable of judging the value of the artefact for itself. These are the criteria for “creativity” in humans –why, objectively, should machine creativity be different? (Wiggins, 2006: 15. Italics in original).

The fact that the creative system is able to assess its own product does not make its assessment socially binding. Cervantes believed La Galatea to be his masterpiece, and look where that got him. Self-evaluation is rather a making-it-up-as-I-go strategy than a verdict on the value of the final product. To express it in polysystem theory form, the future of the artefact is the responsibility of the institution,
not the producer. The heuristic component to creation remains nonetheless, and applies both to human and machine:

 […] a human artist who produces valued outputs by exhaustive enumeration is generally less likely to be heralded as an important artist than an artist who is capable of using his or her own heuristics to arrive directly at a valued artefact. This is equally true of a creative computer system: a system which is capable of using heuristics to navigate its conceptual space directly to a valued solution is likely to be deemed more successful that one which is not (Wiggins, 2006: 15).

In «Ten Questions Concerning Generative Computer Art», McCormack, Bown, Dorin, McCabe, Monro, and Whitelaw pose as the first of their debating points, «Can a machine originate anything?» (McCormack, Bown, Dorin, McCabe, Monro, and Whitelaw, 2012: 3) If a system is programmed to be and act in a specific way, what room is there for an existence and a behavior outside the programmer’s plan? The authors address G.H.R. Parkinson’s thesis that the room is none: according to Parkinson, the system executes instructions and cannot escape neither the instructions nor the execution because it has no concept of option outside either. The role of the system is purely instrumental. In opposition to this, McCormack, Bown, Dorin, McCabe, Monro, and Whitelaw (2012: 3-4) support the heuristic approach, arguing that absolute control over a system is as unattainable as absolute knowledge of it:

Program behaviour, while defined by the program (and created by the programmer), typically has a large, sometimes vast, number of executable pathways. This makes it impossible for the programmer to completely understand and predict the outcome of all but the most trivial programs – one reason why software has «bugs». […] Computer programs, like people, can be adaptive, they can learn, and so initiate new and potentially creative behaviours.

An AI creative program thus functions in ways similar to Barthes’ author. Both are equipped with a set of instructions – code or ideology – they are required to execute. The requirement implies performance according to rules and standards as much as the act of performing itself. Whether you are human or machine, software is to be installed in your brain or hardware. Barthes’ author can dream themselves to be as original as they wish to. The harsh reality is, their only chance to become an author is compliance with literary genres, linguistic laws, canon criteria, market regulations. There is no author outside authorial frames. You can opt out and do your thing, but then you will be doing your thing, not art. You will be unique, just not an author.

It is in the narrow space between the frame and the unpredictability which makes intelligent systems intelligent that creativity lies. Freedom resides in learning. Intelligence means managing the code you are given – Barthes’ dictionary – in such a manner that the code is pushed to evolution while staying recognizable. It also means ability to calibrate. Self-assessment is not intended to lead to self-knowledge or self-discovery, as the concept of self is a construct itself. However, the products the system generates because it has no other choice can and must be evaluated by the system, in order to ensure the heuristic mechanism that gives art a future.
Images are meant to shock. Shklovsky’s choice of examples illuminates the strangeness—beyond unexpectedness— inherent to a defamiliarizing linguistic device: «It would be instructive to apply [Potebnya’s] principle to Tyutchev’s comparison of summer lightning to deaf and dumb demons or to Gogol’s comparison of the sky to the garment of God» (Shklovsky, 1917: 6). Images are a detour: the rhetoric counterpart to the formalist linguistic deviation. The politics of image making relies on a delicate balance: too conservative, and the image will be dead on arrival; too innovative, and no one will understand.

It is, however, the scarce resources for image renewal that slows down the semiotic evolution most. Literary fiction generators—commonly known as authors—have been reusing the same themes and motifs to convey the same meaning through the same formal structures since the Iliad. It is called culture. Shklovsky acknowledges that creation—a term only vindicated by Romanticism, but what a vindication that was—does not exist, and that poses no problem, as it would not matter anyway. Images are produced within cultural frames, and the frame never gives. Moreover, the author is in turn limited by their own personal boundaries, determined by their knowledge of Barthes’ dictionary, or lack thereof. Not that that is of import either: a stronger grasp on the code may lead to more static images, tightly clutched by the weight of tradition, whereas ignorance can be not only bliss, but also freedom. Even Bloom’s hyperaware poet can relate to this. According to Shklovsky, art is a combinatorial technique, just like to Aristotle’s mind, tragedy is a thing made of parts. The innovation is not in the parts. It is in the ensemble:

The more you understand an age, the more convinced you become that the images a given poet used and which you thought his own were taken almost unchanged from another poet. The works of poets are classified or grouped according to the new techniques that poets discover and share, and according to their arrangement and development of the resources of language; poets are much more concerned with arranging images than with creating them. Images are given to poets; the ability to remember them is far more important than the ability to create them (Shklovsky, 1917: 7).

Art needs a more comprehensive search engine, a faster processor. Barthes’ author is dead because they cannot operate in the postmodern world. Art accepts repetition in themes and core values, as one cannot help but accept the inevitable. What art does not tolerate is repetition in form. Incest has been frowned upon for millennia, and it seems safe to assume that it will continue to be advised against for millennia to come. There are only so many elements that can be juggled in the deconstruction and reassembling of the myth. But the material structure must rely upon fresh joints. Art demands yet unseen, unheard, unimagined images that can support sustainable defamiliarization.

In her 2015 paper, «FIGURE8: A Novel System for Generating and Evaluating Figurative Language», Sarah Harmon introduced what can well be regarded as the only author Barthes would never feel tempted to kill: an originator of écriture devoid of human baggage. FIGURE8 is a system on a mission: to put concepts and words together in order to produce images. At its disposal it has the most comprehensive dictionary ever compiled: online databases. Like Barthes’ reader, it has no history, no biography, no psychology. It does have self-assessment skills. As Harmon describes it:
FIGURE8 is a system that uses a web-driven approach to form a preliminary knowledge base of nouns and their properties. The system is provided with a model of the current world and an entity in the world to be described. A suitable vehicle is selected from the knowledge base, and the comparison between the two nouns is clarified by obtaining an understanding via corpora search of what these nouns can do and how they can be described. Sentence completion occurs by intelligent adaptation of a case library of valid grammar constructions. Finally, the comparison is ranked by the system based on semantic, prosodic, and knowledge-based qualities. In this way, FIGURE8 simulates the human-authoring process of revision by generating many vehicle choices and linguistic variations for a single tenor, and choosing the best among them as its favorite. While FIGURE8 does not claim to have a comprehensive set of rules—for example, it does not consider phonetics in its evaluation of description quality—it provides a novel foundation for an intelligent figurative language generation and assessment system (Harmon, 2015: 72).

FIGURE8 operates on the premise that an image—simile or metaphor—must be constructed and evaluated according to a six-property set of criteria: clarity, novelty, aptness, unpredictability, prosody, and sentence completion. It searches for linguistic and conceptual compounds that bring together elements belonging to ontological realms as far apart as possible, whilst retaining a semantic connection that is solid enough to produce recognizable meaning. Surprise is key, but so is preservation of the sign. For example, «The turtle darkened like a blue ocean» (Harmon, 2015: 72) is deemed invalid, as it fails to meet the clarity standard despite combining two conceptually close items—«turtle» and «ocean»—, which, in itself, is considered a flaw too.

Submitted to human perusal, FIGURE8’s images are accepted as comprehensible both semantically and aesthetically. The human evaluators do not question the origin of the artefact, assuming it is human, not machine: the Turing test is passed by FIGURE8. Furthermore, human readers and artificial author both concur in their image ranking. «The queen stands like a strong castle» (Harmon, 2015: 75) is regarded as successful by its originator and highly appreciated by the human subjects alike.

The obvious—and therefore unnecessary—question that usually arises at this point is, Who cares what a machine has to say? Apparently, people do—at least as long as they are unaware of being engaged in a communicative interaction with a machine. The representational side to art—Aristotle’s mimesis—falls outside the scope of our paper, as does the distinction between combinatorial creativity and emergent creativity (McCormack, Bown, Dorin, McCabe, Monro, Whitelaw, 2012: 10). FIGURE8 does not claim to generate poetry, but images. The probability of poetry being but a sequence of images remains high, nonetheless. And in this, imagism is not the peak, just the unmasking. Once again, it is Barthes that allows to leave those anthropocentric considerations aside. The question is not whether or not a program can be creative ex nihilo or conceive a humanly relevant, mimesis-based model of reality. Rather, the question is whether a human being can. Reality itself being a construct, it is

1 Narratology shares similar concerns as to how to maintain a balance between patterns semiotically recognizable and the inescapable urge to step outside the confort zone: «One [issue regarding the creative potential of the proposed framework] is the need to avoid repetition, so that once a particular sequence of character functions has been produced by the system, the following attempts by the system try to avoid using the same character functions in the same order. Another is the possibility to break away from the rigid procedure for constructing stories, so that good stories beyond the set defined by the given rules can be achieved» (Gervás, 2014: 7). As a matter of fact, The Making of Americans remains the only occurrence where repetition in literature has ever been a good idea.
language that dictates normative structures. And as Barthes points out, «language knows a “subject”, not a “person”» (Barthes, 1968: 314).

Conclusion

Although the differences between human and artificial creativity are frequently the focal point in the debate, contemporary literary theory suggests that the similarities are just as critical—if not more. Only an essentialist model of the artistic process can sustain predeconstructionist misconceptions such as imagination as creation or expression as inherent to individuality. The creativity verb par excellence is not create, but make—or invent, once again in the etymological sense of the word. From Shklovsky’s teleological take on art to Foucault’s purely pragmatic consideration of the human role in literature—the dead author as a social commodity—, theorists systematically have proceeded to dehumanize word-based fiction, in an attempt to keep it relevant and in agreement with the postmodern Zeitgeist. Not only computer-generated literature can be assimilated to the theoretical models already existing, it can also contribute to the validation of some abstract systems, or to disprove and discard them altogether. More importantly, computational creativity can help literary theory redefine its identity and goals, and stay relevant as an interdisciplinary meeting point for scientific discussion:

Some […] have argued that theory has been discredited; some that it has simply grow old and outdated; some that it has completed its task, that theory has now vanished into new, and better critical practice; others that it is impossible to talk of the end of a body of thought that itself does so much to problematize notions of historical linearity. In addition, there are those who point out that the word «after» can mean not only «following in time» but also «in pursuit of» or even «in imitation of» (Payne and Schad, 2003: xi).

Maybe it is high time that literary theory simply ignored the walls it has worked so hard to tear down. We human beings have already proven quite good at making things out of words. Maybe it is time for us to start listening to—and reading—something other than ourselves.

Quoted works (Academia only)


