

"Attention, Aesthetic Appraisal and Semantic Noise
while Reading of a Literary Text"

THESIS SUBMITTED FOR THE DEGREE "DOCTOR OF
PHILOSOPHY"

by

Amir Harash

Supervisor

Prof. Yeshayahu Shen

SUBMITTED TO THE SENATE OF TEL AVIV UNIVERSITY

May 2019

TEL AVIV UNIVERSITY

THE LESTER AND SALLY ENTIN
FACULTY OF HUMANITIES
THE SHIRLEY AND LESLIE PORTER
SCHOOL OF CULTURAL STUDIES



אוניברסיטת תל-אביב

הפקולטה למדעי הרוח
ע"ש לסטר וסאלי אנטין
בית הספר למדעי התרבות
ע"ש שירלי ולסלי פורטר

"Attention, Aesthetic Appraisal and Semantic Noise
while Reading of a Literary Text"

THESIS SUBMITTED FOR THE DEGREE "DOCTOR OF
PHILOSOPHY"

by

Amir Harash

Supervisor

Prof. Yeshayahu Shen

SUBMITTED TO THE SENATE OF TEL AVIV UNIVERSITY

March 2019

TEL AVIV UNIVERSITY

THE LESTER AND SALLY ENTIN
FACULTY OF HUMANITIES
THE SHIRLEY AND LESLIE PORTER
SCHOOL OF CULTURAL STUDIES



אוניברסיטת תל-אביב

הפקולטה למדעי הרוח
ע"ש לסטר וסאלי אנטין
בית הספר למדעי התרבות
ע"ש שירלי ולסלי פורטר

This work was carried out under the supervision of

Prof. Yeshayahu Shen

Acknowledgements

I owe a debt of enormous gratitude to my supervisor, Prof. Yeshayahu Shen, who has accompanied the growth of this work with complete faith in my methodology, and who provided me with intellectual, moral and, just as important, material support.

I also want to thank the Department of Literature and the School of Cultural Studies at the Tel Aviv University (TAU) for having enabled this dissertation and supported it in its various stages. I am grateful to the Attention Laboratory of the TAU School of Education and particularly Prof. Lilach Shalev-Mevorach for generously allowing a PhD student of literature to use the eye movement monitor. Finally, I am deeply grateful to the researchers and students of the International Society for the Empirical Study of Literature (IGEL) for accepting me with open arms and giving me encouragement. Thanks to the IGEL conferences and periodical, I have found an intellectual network where I felt at ease and could receive updates and meet the leading researchers in the area. Despite obvious gaps of age, status and knowledge, they have treated me like their equal.

Abstract

This dissertation deals with the relationship between the cognitive processing of a literary text and the literary-aesthetic experience of real-life readers. It presents several kinds of failures in the reading process, based on the foregrounding model. Foregrounding is the process whereby a textual element becomes salient relative to others, part of the text's "forefront", and therefore more available for literary interpretation. Foregrounding theory is one of the leading theories in the empirical study of literature, with considerable literature to support it.

The empirical investigation of foregrounding theory represents the most systematic and comprehensive attempt hitherto to empirically examine a model of literature reading (Van Peer, Zyngier & Hakemulder, 2007) – a model that deals with literariness itself, that is, with the key textual characteristics that differentiate a literary from a non-literary text. Yet, some of the classical experiments usually considered supportive of the theory have reported mixed findings (e.g. Emmott, Sanford, & Morrow, 2006; Miall and Kuiken, 1994). The foregrounding devices in the text have not always attracted the readers' attention, and have not always facilitated aesthetic effects. Recently, failure to replicate previous findings has led researchers to suggest that changes in the literary field such as the lack of a literary canon lead to a reading process that diverges from the predictions of the foregrounding theory (Van Peer & Chesnokova 2017). Accordingly, in what follows, I propose a model informed by the standard model of the foregrounding process that examines a possibility that has not yet been studied: that the foregrounding process

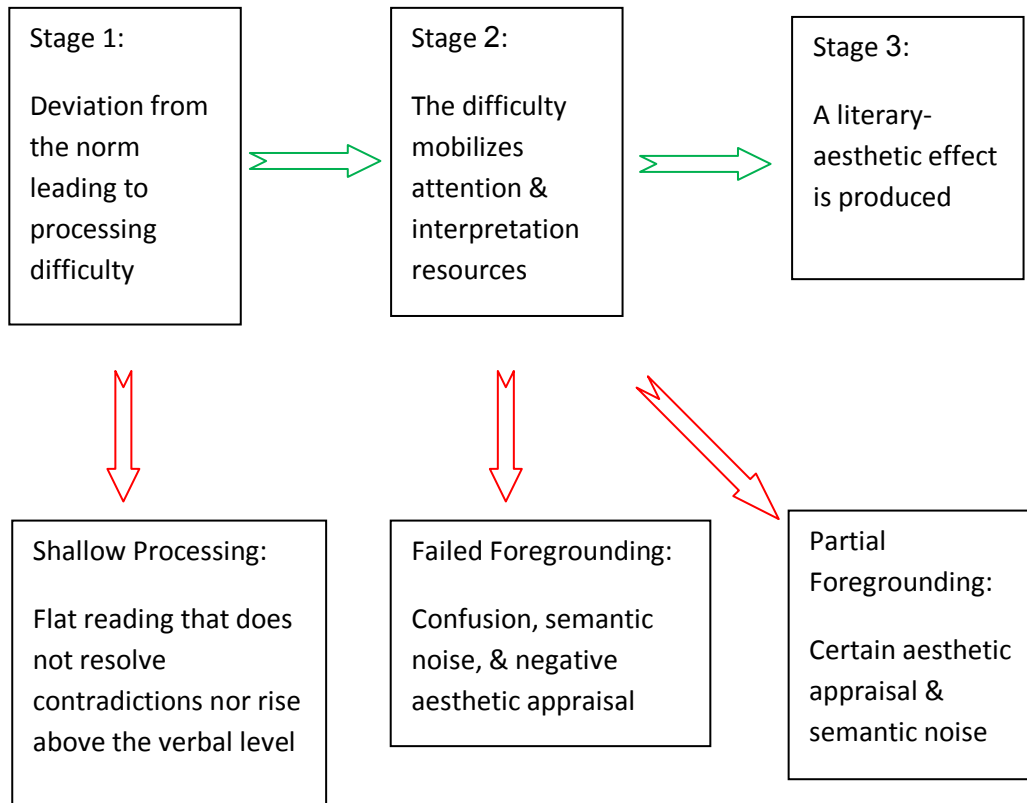
may fail and that this failure is not the exception but rather integral to actual reading of literature by real-life readers.

Incorporating the option of failure can help reconcile the contradictory findings mentioned above: If failure becomes a potential outcome of the foregrounding process, then more findings can be incorporated into the theory. Instead of looking for confirmations that the process exists, researchers can explore its boundaries. Importantly, addressing the possibility of failure is of literary importance for several reasons. First, this enables to provide a fuller and more adequate account of real readers' reading processes. Reading failure is common and familiar. Many become stuck in reading poetry and prose, do not understand the text well enough, become confused, fail to get to the bottom of the meaning, skip the problematic passage, read shallowly, or simply do not enjoy the text. This phenomenon is not limited to students and is not just an intermediate stage on the way to becoming proficient readers of literature – even the most skillful readers sometimes have difficulty with a complex literary text. Second, failure may also have (positive or negative) influence on the reader's aesthetic appraisal. Failures can sabotage the reading experience, or alternatively, precipitate a particular, more radical form of drawing pleasure from the text. It is also possible for different kinds of failures to have different aesthetic outcomes. Thus, by investigating the influence of failures on readers' aesthetic appraisals, more light may be shed on the consequences of difficulty for real readers.

The standard foregrounding model is composed of three stages. First, there is a deviation from the linguistic norm that causes processing difficulty. Second, the

difficulty mobilizes the reader's attention and interpretation resources. Third, this mobilization produces an effect of literary importance, a positive aesthetic appraisal. The model presented here derives from, expands and complements the standard foregrounding model. Two different types of failure are described: (1) in the transition from the first to the second foregrounding stage, leading to shallow processing; and (2) from the second to the third, leading to failed or partial foregrounding. The implications of these failures are described in terms of aesthetic appraisal and semantic noise experienced by the readers.

The differences between the various stages may be described using the analogy of a fruit. We can liken the difficulty or disruption to a nucleus around which a fruit grows. In *full foregrounding*, the fruit is comprised mainly of the positive aesthetic experience – the flesh – and within it all that remains is a tiny nucleus of semantic noise and difficulty. In the case of *partial foregrounding*, the fruit is made up mainly of the semantic noise and difficulty, surrounded by a thin layer of positive aesthetic experience. In the case of *failed foregrounding*, even that thin layer is nonexistent. The reader remains with only a difficulty and disruption, which have grown to the size of a whole fruit. Finally, in *shallow processing*, there is only the nucleus of initial difficulty, and since it has not been processed in depth, it has neither grown to a size that disrupts the reader significantly, nor borne fruit in the form of any aesthetic appraisal.



The failed foregrounding model: Sometimes the foregrounding process does not reach completion, resulting in partial or failed foregrounding – the processing difficulty is not converted into an aesthetic experience, resulting in confusion and semantic noise. The upper sequence represents a successful process, while the lower branches depict types of failures.

An eye-tracking experiment was conducted to evaluate the model. Forty-two volunteers read “The Chamber of Statues”, a short story by Jorge Luis Borges (1935), and completed a semantic noise and an aesthetic appraisal scales and an author recognition test (which served as an indication of readers’ experience or expertise in literature). They were also interviewed using a retrospective think aloud (RTA)

protocol based on the results of eye movement monitoring. RTA combines “soft” and “hard” evidence – verbal reports and eye movement patterns, and it is rich in terms of the information provided on the reader’s conscious experience.

Analysis of the interviews suggests that in 36% of the cases, readers did not even initiate the foregrounding process, and that they completed it successfully in only 21% of the cases. These rates varied significantly with the readers’ experience, aesthetic appraisal, reading strategy and textual passage.

The findings support the failed foregrounding model in several respects. First, in its ability to make statistically significant distinctions. Second, in that its predictions gain more support than those of its two rivals: the standard model and the radical aesthetician position.

The proposed model accounts for a series of observations. Positive aesthetic appraisal for the whole story was found related to full foregrounding in the key points while negative appraisal was found related to failed foregrounding. This is in line with the model’s prediction that failed foregrounding would have a more negative effect on readers’ aesthetic experience than shallow processing, and that full foregrounding would have a more positive influence than partial foregrounding.

We have also found that experienced readers often attain full foregrounding while inexperienced ones often opt for shallow processing. In addition, in-depth examination of reading strategies has revealed preference for strategies in which shallow processing or full foregrounding are central. Thus, it appears that in general,

the participants have preferred the two polar positions of the model – shallow processing and full foregrounding.

The failed foregrounding model may also be used to examine the effectiveness of various stylistic devices. Namely, one can examine which stylistic devices often lead to a full foreground effect, and which often lead to failures in the process. An examination of three stylistic devices – figurative descriptions, author comments and linguistic difficulty – found that the distribution of foregrounding profiles was not independent of these devices. Each had a significantly different foregrounding distribution, also affected by the reader's experience.

Figurative descriptions were the most effective in reaching full foregrounding. The author comments' literary importance was the easiest for the readers to recognize, and were very low on shallow processing. Nevertheless, they made it difficult for readers to complete the process with many halting in failed foregrounding. Linguistic difficulty was the least effective stylistic device: the readers found it difficult to both start the foregrounding process and complete it. Moreover, linguistic difficulty was the stylistic devices least sensitive to the reader's experience.

The new model suggests a general division into two types of effectiveness: (1) Effectiveness in introducing the reader into a foregrounding process; and (2) Effectiveness in bringing the reader to successful conclusion of the process. It was hypothesized that the layer where the initial difficulty is experienced has an important role in determining the devices' effectiveness in initiating the foregrounding process. When the difficulty was in the basic layer of linguistic processing, most readers tended to resolve it within the confines of that layer, with few going into literary

interpretation. When they came across unfamiliar foreign words, most readers struggled to understand “what” was said, rather than “why”.

The model is grounded in the assumption that full foregrounding is the key element in positive aesthetic appraisal. Nevertheless, it is flexible enough to allow the examination of other hypotheses – in particular, the approach I term the *radical aesthetician position*. This argues that failure does not ill affect the aesthetic experience – on the contrary, it is its core. The radical aesthetician believes that literary-aesthetic experience derives directly from the difficulty, from the inability to understand, from the very grappling with an insubordinate text, from recurrent knocking on the closed door of the uninterpretable. According to this approach, positive aesthetic appraisal is not the product of difficulty and its resolution, of defamiliarization and refamiliarization, but rather of dwelling on the difficulty, on the unfamiliar, on the incomprehensible.

Since the radical aesthetician position emphasizes effort and difficulty as leading to aesthetic appraisal, in terms of the failed foregrounding model the most important stages in radical aesthetic appraisal would be failed and partial foregrounding. Thus, should that position prove more correct, the essence of the literary encounter needs to be associated with both failed and partial foregrounding. Namely, it is not the completion of foregrounding that produces literary experience. Rather, whoever stops in the middle of the process, whoever responds to the invitation to interpret, but has failed to show up to the interpretation party, is the one who experiences the radical aesthetic experience. These predictions are completely the opposite of those of the failed foregrounding model.

The radical aesthetician position was also examined using the new model. According to this approach, failed and partial foregrounding should play a key role in aesthetic appraisal or at the very least, their frequency would increase with the reader's experience, and that the frequency of full foregrounding should drop. None of these predictions was supported. Even the search for readers who adopted reading strategies where those two elements are central was fruitless. Only four readers opted for a reading strategy where failed foregrounding was central, and in-depth examination of their interviews did not support the idea that theirs was a "radical" literary experience, but rather that they mainly stumbled across semantic noises of various kinds. It was not my impression that they experienced "bliss" while reading the text – precisely the opposite: they were highly frustrated by it.

The final section of the dissertation presents the main methodology: the RTA method combined with eye-movement monitoring. This method mitigates several key problems in both collecting verbal information and analyzing eye movements. First, it reduces the reactivity and verticality problems of collecting verbal information. Namely, it does not disrupt the reading and thinking process as it occurs, and it is relatively reliable in terms of recollection, since the eye movement patterns remind the participant of her reading process. It was also found to have a low likelihood of fabrication but a high likelihood of omissions. That is, not all the information in the eye movement findings is explained. These omissions, however, help reduce the amount of information and thus mitigate the big data problem. Finally, the method helps reduce the huge amount of data produced in eye movement studies in two additional ways. First, instead of referring to all words, the interviewer refers only to

those areas in the text where many of the readers have dwelled. Second, the method reduces the relevant data to those places in the text that the participants remember and have something to say about dwelling in them. Thus, RTA focuses researchers on specific, yet significant, phenomena that occur in the reading process, those that leave a lasting impression in the reader's memory.

The method proposed aligns two different types of responses by the same participant: verbal and physiological. It is verbally rich as well as localized spatially (or textually) and temporally. Moreover, it provides rich cognitive evidence, both according to previous usability studies and according to the analysis of common word combinations in the present experiment. The participant's reading experience was found correlated with interview length, meaning that this method has a relative advantage in studies on experienced readers, since they do better in verbalizing their reading process retrospectively. Note that the method's effectiveness is not limited to readers who have had a positive experience, since readers with negative aesthetic appraisal of the text as well as readers who have experienced semantic noises provided the same amount of data. Thus, the method is suitable for a wide range of literature reading experiences, and particularly for learning about comprehension difficulties and communication problems the reader experiences.

Note that the model presented here is not applicable to all types of literary experience. It is not designed to describe the only artistic technique, or to offer a "theory of everything". There may certainly be groups of readers or types of texts it does not describe well. Just as recurring readings may produce other effects, there may be

literary texts that lack foregrounding devices and provide a fluent reading experience, where literary effects are derived from other factors.

The findings lead to the conclusion that the failed foregrounding model is influenced by parameters that characterize the reader, the text and to a certain extent also the interaction between them. This sensitivity of the foregrounding profile to reader and text parameters suggests that the foregrounding process itself is more “fragile” than usually thought – with many factors affecting its effectiveness.

To conclude, the three main innovations in this dissertation are:

1. Developing the failed foregrounding model and validating it by analyzing interviews with readers.
2. Applying the retrospective think-aloud technique guided by eye movement patterns and validating it as an appropriate instrument for the study of literature reading.
3. Applying the concept of semantic noise to explain reading failures and developing methods to measure it.

Content

Acknowledgements	e
Abstract	f
Introduction	1
Overview	2
SECTION 1: THE PROPOSED MODEL	4
Motivation	4
Full, Partial and Failed Foregrounding	7
Breakdown between the First and Second Stages	8
Breakdown between the Second and Third Stages	9
The Radical Aesthetician Position	11
Empirical Evidence for Failed Foregrounding	13
Failed Foregrounding	13
Shallow Processing	15
SECTION 2: FAILED FOREGROUNDING AND THE READER	17
Studying Failed Foregrounding Using Questionnaires	17
Questionnaires	17
Results and Discussion	19
Theories Supported by the Findings	25
Studying Failed Foregrounding Using Reader Interviews	26
Main Findings	31
Foregrounding Profiles According to Aesthetic Appraisal	32
Theories Supported by the Findings	33
The Distribution of Foregrounding Profiles by Reading Experience	36
Reading Strategies	38
Summary of Section 2	41
SECTION 3: FAILED FOREGROUNDING AND THE TEXT	43
Foregrounding Distributions According to Stylistic Devices	43
Figurative Descriptions	46
Author Comments	47
Linguistic Difficulty	48
Interaction with Reading Experience	50

Summary and Discussion.....	50
Effectiveness of Stylistic Devices.....	50
The Limitations of Analysis by Stylistic Devices	53
SECTION 4: METHODOLOGY – RETROSPECTIVE THINK-ALOUD	55
Key Issues in Collecting Verbal Information: Reactivity and Verticality	56
Two Difficulties in Analyzing Eye Movements: Interpretation and Big Data.....	58
The Advantages of Retrospective Reflection Combined with Eye Movement Data.....	60
RTA in Usability Studies	60
An Eye Movement-Based Retrospection Experiment.....	63
Reliability.....	67
Discussion	73
Limitations.....	74
GENERAL DISCUSSION	75
Methodology	77
Is Foregrounding the Product of Controllable Behavior?.....	78
Effectiveness.....	79
Relation to Previous Foregrounding Models	80
Limitations of the Failed Foregrounding Model	81
Why Has a Similar Model Not Been Developed Hitherto?	84
References.....	86
Appendix A:	90
Amir Harash and Yeshayahu Shen	90
Appendix B: Semantic Noise Questionnaire	112
Appendix C: Aesthetic Appraisal Questionnaire	114
Appendix D: Names of Authors and Distractors in the Hebrew Version of the Author Recognition Test.....	116
Appendix E:.....	119
THE CHAMBER OF STATUES.....	119
חדר הפסלים.....	122
La cámara de las estatuas.....	123
"חיבור לשם קבלת התואר "דוקטור לפילוסופיה".....	א
מאת: אמיר חרש.....	א

TEL AVIV UNIVERSITY

THE LESTER AND SALLY ENTIN
FACULTY OF HUMANITIES
THE SHIRLEY AND LESLIE PORTER
SCHOOL OF CULTURAL STUDIES



אוניברסיטת תל-אביב

הפקולטה למדעי הרוח
ע"ש לסטר וסאלי אנטין
בית הספר למדעי התרבות
ע"ש שירלי ולסלי פורטר

Introduction

This dissertation deals with the relationship between the cognitive processing of a literary text and the literary-aesthetic experience of real-life readers. Its point of departure is that real-life readers are not ideal, and that therefore their reading processes are not perfect, but laden with difficulties, confusion and errors. Since these difficulties are highly common, interpreting them is necessary for gaining a full understanding of how real readers read literature.

In this dissertation, I point to a blind spot in the field of empirical study of literature, which is ignoring failures in reading processes. The empirical study of literature is a growing field that applies tools borrowed mostly from the social and cognitive sciences to describe and explain the way real-life readers read literature. The focus of this discipline is not on professional interpretation or ideal reading processes, but on the reading of normal individuals, who are not necessarily experienced in reading literature.

Nevertheless, the empirical study of literature has hitherto practically ignored reading failures. These may be thought of as failures in the foregrounding process, or as failures on the “poetic function”. Foregrounding is the process whereby a textual element becomes salient relative to others, becomes part of the text’s “forefront”, and therefore more available for literary interpretation. Foregrounding theory is one of the leading theories in the empirical study of literature, and there is already considerable research that substantiates it (Leech and Short 2007; Miall and Kuiken, 1994; Van Peer, 1986). Much is known about what makes a given element become foregrounded, as well as about the effects experienced by readers engaged in a successful foregrounding process. However, to the best of my knowledge, researchers have hitherto not asked when foregrounding succeeds, when it fails, what types of foregrounding strategies there are, and what their implications may be. Given the considerable importance of the foregrounding theory in the study of literature, there is particular interest in thorough examination of the foregrounding process, the likelihood of its success or failure and the reasons for that.

The present work provides several ways of compensating for this lacuna, the most important of which is using the failed foregrounding model. This model is derived from the standard foregrounding model, expanding it to cases of failure where the process does not complete its course. The failed foregrounding model is the core of this dissertation and spans over three sections. The last section complements them by discussing some of the steps necessary in order to develop and validate the model. It presents the experimental technique on which the model is based (thinking aloud with reference to eye movement monitoring findings). Finally, an appendix elaborates on the term semantic noise that is central to the model and has been crucial for its development. The dissertation sections are summarized below.

Overview

Section 1: describes a model with two types of potential failures in the foregrounding process. A full and successful foregrounding process involves the unsettling of the information processing system, as well as the attempt to overcome it, leading to the desirable literary-aesthetic effect. One type of potential failure is “shallow processing”, where the reader does not even initiate the foregrounding process, and the other is failure that occurs after an interpretive move has already begun, and is called “failed foregrounding”. Findings of two major experiments supporting the two failure types are reviewed, as are assessments regarding the effect of each on the reading experience. The model is presented vis-à-vis two alternatives: the standard foregrounding model and what is called here the radical aesthetician position, which argues that the pleasure in reading literature derives from the difficulty itself, from the inability to fathom the text.

Section 2: addresses the failed foregrounding model and the reader. It presents evidence supporting the model based on a reading experiment. Two types of evidence are presented: questionnaires that enable a general inquiry into the reading process, and interviews based on eye movement patterns, that allow a more local analysis based on key points in the text. Analyzing the interviews enables to create a foregrounding profile, or a distribution of various foregrounding conditions that apply to a given case. The section examines the foregrounding profiles of readers with high and low aesthetic appraisal, and with high and low experience in reading literature. In addition, the profiles are used to characterize different reading strategies. Finally, the predictions of the two alternative approaches described above are examined against the predictions of the present model.

Section 3: discusses the failed foregrounding model and the text. It examines the foregrounding profiles of three types of stylistic devices: author comments, figurative descriptions and linguistic difficulties. It finds that each device has a different foregrounding profile, a finding that is interpreted with respect to the concept of differences in the effectiveness of stylistic devices in encouraging foregrounding. Two different types of effectiveness are described, and hypotheses regarding factors encouraging each are suggested.

Section 4: The main technique in which the findings of this dissertation have been collected is RTA. In the past, researchers relied heavily on reader introspective, but for various reasons, they became more suspicious of introspective verbal information. Hence, there was a need to expand on this methodology, substantiate it and justify the validity of its use. For the sake of simplicity and in order not to distract the readers from the main line of argument presented in Sections 1-3, the methodology is presented in a separate section.

Appendix A: Semantic Noise: Here, I provide an extensive review of the concept of semantic noise in the context of empirical literature research. This self-standing section is attached as an appendix, as it was already published (Harash & Shen 2016). Semantic noise has played a key role in the failed foregrounding model, and has been a significant catalyst in its development. The appendix examines the relation between the difficulty in cognitive processing of a literary text and the aesthetic experience, and presents three competing theories: foregrounding, optimal innovation and fluent processing, each of which offers an essentially different explanation supported in all three cases by an extensive empirical literature. The concept of semantic noise is presented as a bridge that can span their contradictory findings.

To conclude, the three main innovations in this dissertation are:

4. Developing the failed foregrounding model and validating it by analyzing interviews with readers.
5. Applying the retrospective think-aloud technique guided by eye movement patterns and validating it as an appropriate instrument for the study of literature reading.
6. Applying the concept of semantic noise to explain reading failures and developing methods to measure it.

This dissertation is based on an experiment that combines various measurements, physiological, psychometric and verbal. Nevertheless, its focus is theoretical. The importance of the experimental approach presented here lies above all in that it has facilitated theory development. In a certain sense, this theory developed bottom-up – from the field, from the laboratory, out of conversations with the readers, and out of the basic findings about eye movements to which I was exposed in the process.

SECTION 1: THE PROPOSED MODEL

Motivation

Complex literary texts usually contain multiple elements that create difficulties for the reader. Literature readers apply interpretive processes designed to deal with those elements. Key conceptions in literary research (e.g. Russian formalism, Czech structuralism, foregrounding theory, and the new criticism) have made these difficulties the hallmark of the literary text. Towards the late 1980s, attempts have begun to empirically examine one of those approaches, foregrounding theory, based on real reader responses (Van Peer, 1986).

The study of literature is thousands of years old, but only in recent decades have the various theories begun to be tested in a way that allows their confirmation or refutation. While the mainstream of literature researchers continues to be interested in the text, its interpretation and its political, sociological and historical aspects, since the 1980s there has been a trend of literature researchers mainly interested in the interaction between actual readers and texts. The focus on the reader is not new, but even the reader-response criticism school (e.g. Fish 1970, 1980; Holland, 1968; Iser, 1978; Jauss, 1982) have hardly tried to test their claims using empirical tools, and settled for theoretical or interpretive claims regarding the way literature is read. Conversely, empirical researchers try to describe and explain the way real-life readers read literature, and their research is often conducted using tools borrowed from the social sciences and the cognitive and brain sciences. Within that growing body of knowledge, a special place is reserved for foregrounding theory.

Foregrounding theory is a productive starting point for a model of literary reading for several reasons: it is a major and well-established theory – its origins can be found in Aristotle's *Poetics* – yet it is updated with current developments and integrated in theories such as the neurocognitive poetics model (Jacobs, 2015). It is concrete enough and formulated in a way that enables researchers to derive confirmable or refutable hypotheses, and still general enough to be useful for stylistic and interpretative analysis of literature.

The empirical investigation of foregrounding theory represents the most systematic and comprehensive attempt hitherto to empirically examine a model of literature reading (Van Peer, Zyngier & Hakemulder, 2007) – a model that deals with literariness itself, that is, with the key textual characteristics that differentiate a literary from a non-literary text. It is against this background that we need to understand the following statement by Van Peer et al.: "We know of no single literary theory for which there is such a modest but yet convincing body of empirical evidence that has been accumulated over the past decades, no single theory that has withstood so many rigorous tests." (2007. p. 8).

Nevertheless, some of the classical experiments usually considered supportive of the theory have reported mixed findings (e.g. Emmott, Sanford, & Morrow, 2006; Miall and Kuiken, 1994). The foregrounding devices in the text have not always attracted the readers' attention, and have not always facilitated aesthetic effects. Recently, failure to replicate previous findings has led researchers to suggest that changes in the literary field such as the lack of a literary canon lead to a reading process that diverges from the predictions of the foregrounding theory (Van Peer & Chesnokova 2017). Accordingly, in what follows, I propose a model informed by the standard model of the foregrounding process that examines a possibility that has not yet been studied: that the foregrounding process may fail and that this failure is not the exception but rather integral to actual reading of literature by real readers.

A complete and successful foregrounding process involves the unsettling of the information processing system, as well as the attempt to overcome it, leading to the desirable literary-aesthetic effect. The process may also be seen as defamiliarization followed by refamiliarization. Conventionally, foregrounding begins with deviation from the normal usage of language, in the form of either parallelism or departure from the norm: grammatical irregularity, departure from the general linguistic norm of the period in which the text has been written, or departure from a local norm established within the given text. This deviation makes the reader delay and allocate additional attention in an attempt to interpret it, resulting in a literary-aesthetic experience (Leech and Short 2007; Miall and Kuiken, 1994; Van Peer, 1986; Van Peer at al. 2007). Foregrounding may also be thought of as a means of indirect communication between the author and reader. Through the difficulty, the author marks the key points, where the reader is encouraged to dwell and assign special importance. In this sense, difficulty and deviation serve as a kind of implicit invitation for interpretation sent by the author to the reader (Van Peer, 1986).

One possibility derived from the basic foregrounding model has hardly been discussed or studied – that foregrounding would sometimes fail in mid-process. In other words, it is possible that the linguistic deviation supposed to kick-start the foregrounding process would not lead to greater insight nor to a literary-aesthetic experience. In this case, the seed of defamiliarization does not bear an aesthetic fruit but remains a mere disturbance. Another, related possibility is for the foregrounding process never to begin in the first place: the reader does not respond to the invitation for interpretation by deepening his reading, but continues with relatively shallow reading, if not skips the difficulty entirely.

Incorporating the option of failure can help reconcile the contradictory findings mentioned above: If failure becomes a legitimate outcome of the foregrounding process, then more findings can be incorporated into the theory. Instead of looking for confirmations that the process exists, researchers can explore its boundaries. In addition, addressing the possibility of failure is of literary importance for several reasons. First, this enables to provide a fuller and more adequate account of real readers' reading processes. Reading failure is common and familiar. Many become

stuck in reading poetry and prose, many do not understand the text well enough, become confused, fail to get to the bottom of the textual meaning, skip the problematic passage, read shallowly, or simply do not enjoy the text. This phenomenon is not limited to students and is not just an intermediate stage on the way to becoming proficient readers of literature – even the most skillful readers sometimes have difficulty with a complex literary text.

Failure may also have (positive or negative) influence on the reader's aesthetic appraisal. Failures can sabotage the reading experience, or alternatively, catalyze a particular, more radical form of drawing pleasure from the text. It is also possible for different kinds of failures to have different aesthetic outcomes. Thus, by investigating the influence of failures on readers' aesthetic appraisals, better light may be shed on the consequences of difficulty for real readers. Additionally, detection and classification of these failures can even be used as a way to identify the more radical readers.

Examining reading failures is particularly important these days, when reading habits seem to change. Reading out of a printed book is only one option, and no longer an obvious one. Much of the current reading occurs in digital media. The physical and attentional conditions in the various reading contexts affect reading habits for good and bad. These changes may specifically affect the ways people read literature, with shallower processing affecting the foregrounding process, resulting in its less frequent completion. Thus, the future of the foregrounding process may no longer be taken for granted. A better understanding is needed of the new reading environments that are far from ideal for reading literature and of the costs and benefits of the alternatives to full foregrounding.

In this dissertation, instead of ignoring the difficulties in interpreting the literary text, I take them into account and integrate them into the foregrounding model. This can improve our understanding of literariness as it unfolds in the real life of every literature reader – of the non-ideal reader (that is, any one of us) who operates in suboptimal circumstances.

The remainder of this dissertation proceeds as follows. Section 1 presents the failed foregrounding model in detail, including its implications for readers' aesthetic experience and a review of previous research findings that support it. Section 2 deals with failed foregrounding and the reader, presenting a method appropriate for the study of the model's application to readings by real readers, the experiment conducted and its findings with respect to reading styles and strategies of various readers. Section 3 deals with failed foregrounding and the text, examining how different stylistic devices lead to different failures in the foregrounding process in an attempt to understand the factors involved in the effectiveness of various stylistic devices. Finally, Section 4 elaborates on the methodology required for developing and substantiating the model.

Full, Partial and Failed Foregrounding

This section presents the new model, which describes a full, partial and failed foregrounding process. It is based on existing foregrounding models (e.g. Leech & Short, 2007), which are basically three-stage models. First, there is the deviation from the linguistic norm that causes processing difficulty. Next, the difficulty mobilizes the reader's attention and interpretation resources. Finally, this mobilization produces an effect of literary importance (see Figure 1).

This standard model was empirically supported by Van Peer's (1986) study, which found that foregrounded textual elements lead to a sense of strikingness among poetry readers, as well as by Miall and Kuiken (1994), who showed that foregrounded textual elements lead, in addition to a strikingness affect, to slower reading speed and affective response among prose readers. The problem with this standard model is the assumption that the transition across the stages is smooth, that the process does not fail in the passage between the stages. At the very least, the fact that foregrounding theorists ignore the possibility of process failure reveals an implicit assumption: that this phenomenon has no literary interest for them. My own model, presented in Figure 2, is innovative in that it describes not only successful transitions from one stage to another, but also various possibilities for failure in the process. The model describes two types of breakdowns: in the transition between the first and second stages, and in the transition between the second and third stages.

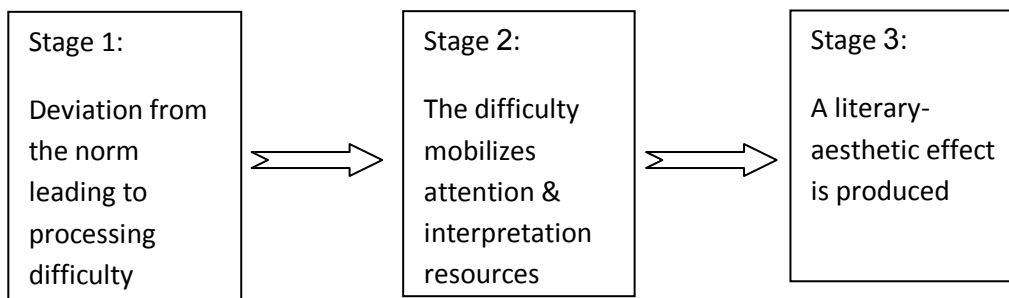


Figure 1: The standard foregrounding model: Linguistic deviation leads to mobilization of attention and interpretation resources, leading in turn to a literary-aesthetic effect. The model implicitly assumes smooth transition across the stages.

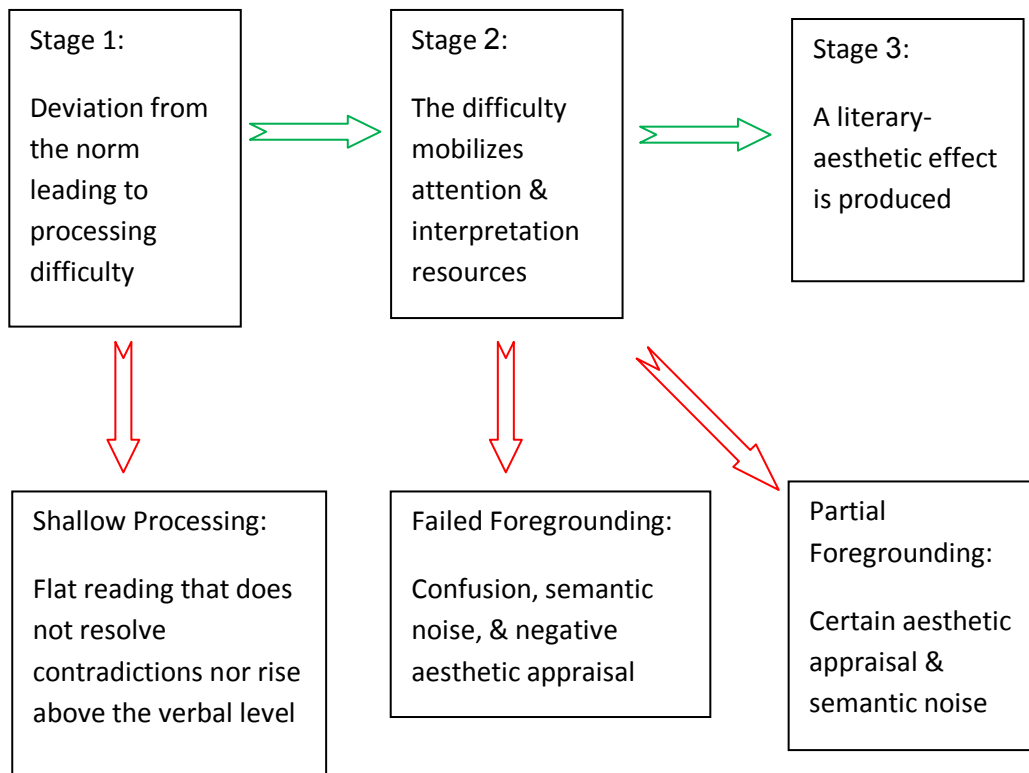


Figure 2: The failed foregrounding model: Sometimes the foregrounding process does not reach completion, resulting in partial or failed foregrounding – the processing difficulty is not converted into an aesthetic experience, resulting in confusion and semantic noise. The upper sequence represents a successful process, while the lower branches depict types of failures.

Breakdown between the First and Second Stages

Processing difficulty is supposed to lead to psychological prominence: mobilization of attention resources, extra processing, or an interpretive attempt to overcome the difficulty. But this is not inevitable: the reader may choose not to delve into the difficulty more than is necessary for his purposes, not to accept it as an invitation for interpretation. There may be different types of shallow processes, or a range of phenomena that can all fall under this title. On one end, readers can skip or browse through a textual passage, and on the other, they can attempt to understand it without shifting to levels “higher” than the verbal one, such as the plot, the characters, the author or any other “higher” or literary level of understanding.

The shallow processing and good-enough representation paradigms (Ferreira, Ferraro, & Bailey, 2002; Ferreira & Patson, 2007; Sanford, Sanford, Molle, & Emmott, 2006; Sanford & Sturt, 2002) show that in many cases, language users do not process the text in depth, but only to the limited extent that satisfies their immediate objectives. Linguistic information can be quite complex, and usually, in daily life, it is processed

under conditions of overload and limited time, without any social need to demonstrate deep and full understanding. Therefore, to save cognitive resources, language users ignore anomalies, do not notice errors, and fail to detect changes in the text. Researchers have demonstrated that even in relatively trivial cases of textual complexity, as in garden-path sentences (e.g. “The old man the boat”) or passive sentences, the readers do not make the effort required to fully understand the text, but rather use heuristics to simplify information processing. The result is partial representation: the linguistic information is represented non-specifically, in very general categories, and includes contradictory details. Accordingly, the findings of the shallow processing paradigm provide an alternative to the basic assumption of foregrounding theory, that deviation from the linguistic norm leads to deepened attention and interpretation. For more on the shallow processing paradigm, see the subsection "shallow processing" on page 15.

Breakdown between the Second and Third Stages

Smooth transition between the second and third stages indicates that the attentive or interpretive effort invested in the difficulty area has led the reader to a literary-aesthetic experience. However, it is also possible that the extra effort of the second stage has failed, leading the reader to experience nothing but confusion, irritation, indifference or even negative appreciation. In this case, foregrounding will be considered a failure. Another possibility is for the process neither to fail nor to succeed completely. This leads to partial foregrounding. *Partial foregrounding* involves some aesthetic appraisal, but it is raw, limited and underdeveloped in interpretive terms. This limited aesthetic experience can be accompanied by considerable misunderstanding or confusion.

Both failed and partial foregrounding involves significant semantic noise. *Semantic noise* (Harash & Shen 2016; Shannon, & Weaver, 1963) refers to communication difficulties, disruptions resulting not from the texts itself, but from its interpretation process. This happens when the reading produces meanings that do not contribute to the communication: phrasings seen as cumbersome, distractive multiplicity of meanings, confusion between characters, associations that disrupt the reading, things that are perceived as errors, etc. In fact, the very linguistic deviation that initiates the foregrounding process produces a core of semantic noise. The failed attempt to understand and interpret the difficult passage is liable to produce additional semantic noise, which can be more disruptive than the immediate response to the linguistic deviation itself. It is therefore expected that a breakdown in the transition between the second and third foregrounding stages would be more noisy, disruptive and confusing than breakdown that leads to shallow processing. See Figure 3 for a graphic illustration of the possible results of the foregrounding process described in the model.

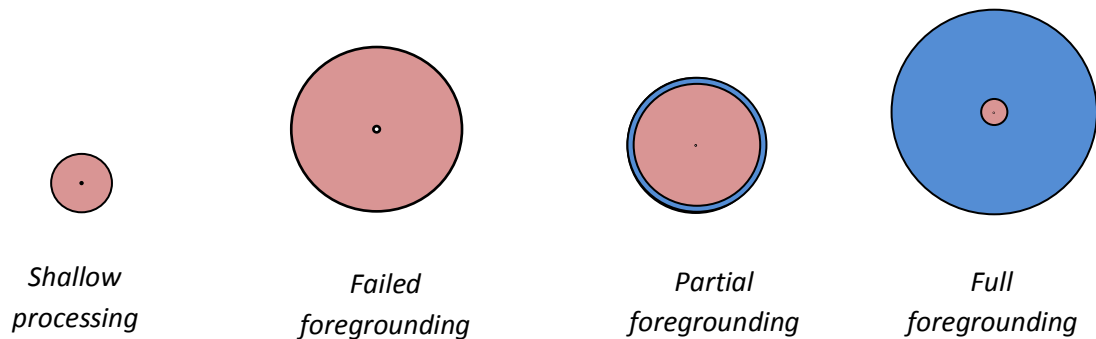


Figure 3: Scheme of the relation between positive aesthetic appraisal and semantic noise in all possible endings of the foregrounding process (blue represents positive appraisal; red represents semantic noise).

The differences between the various stages may be described using the analogy of a fruit. We can liken the difficulty or disruption to a nucleus around which a fruit grows. In *full foregrounding*, the fruit is comprised mainly of the positive aesthetic experience – the flesh – and within it all that remains is a tiny nucleus of semantic noise and difficulty. In the case of *partial foregrounding*, the fruit is made up mainly of the semantic noise and difficulty, surrounded by a thin layer of positive aesthetic experience. In the case of *failed foregrounding*, even that thin layer is nonexistent. The reader remains with only a difficulty and disruption, which have grown to the size of a whole fruit. Finally, in *shallow processing*, there is only the nucleus of initial difficulty, and since it has not been processed in depth, it has neither grown to a size that disrupts the reader significantly, nor borne fruit in the form of any aesthetic appraisal.

Note that there is no necessary relation between full foregrounding and correct, full or accurate understanding of the text. In particular, full foregrounding does not mean that the researchers agree with the reader’s conclusions and interpretation. It only means that a process that begins with a difficulty has led the reader to positive aesthetic appraisal or interpretive insight that is significant for him. Similarly, when describing a “failure” in the foregrounding process, I do not consider it as miscomprehension. Neither do I compare the concrete reader to an ideal reader or professional interpreter. The failure is primarily aesthetic. The reader does not fail because his interpretation is inaccurate, but because that interpretation does not serve him well enough. He or she has failed to convert the difficulty inherent in the text into something else that is more literary.

The Radical Aesthetician Position

The model is grounded in the assumption that full foregrounding is the key element in positive aesthetic appraisal. Nevertheless, it is flexible enough to allow the examination of other hypotheses – in particular, the approach I term the *radical aesthetician position*. This argues that failure does not ill affect the aesthetic experience – on the contrary, it is its core. The radical aesthetician believes that literary-aesthetic experience derives directly from the difficulty, from the inability to understand, from the very grappling with an insubordinate text, from recurrent knocking on the closed door of the uninterpretable. According to this approach, positive aesthetic appraisal is not the product of difficulty and its resolution, of defamiliarization and refamiliarization, but rather of dwelling on the difficulty, on the unfamiliar, on the incomprehensible.

It is difficult to find an explicit formulation of the radical aesthetician position in its pure form, but it is easy to find positions that overlap with it at least partly. This approach is common in literature as well as in music and the arts. Among the poststructuralists, Roland Barthes (1975) expresses views close to the radical aesthetician position. He distinguishes between two types of textual enjoyment: *plaisir* ("pleasure") and *jouissance* ("bliss"). The first is the more common, the more bourgeois, and the second is the more radical: an unpleasant experience, restless like boiling metal. See the subsection on "The radical aesthetician position" on page 35 for more on Barthes' *plaisir* and *jouissance*.

A similar approach may be found already among the artistic avant-garde of the early 20th century, particularly Dada and Surrealism, where the concept of meaningful poetry was attacked, as was the very meaning of words themselves as fundamental to the communication between the artist and audience. The Dadaist practices of simultaneous, phonetic and polyglot poetry were used to elicit language and meaning out of poetry. In simultaneous poetry, concurrent reading did not allow the listeners to hear the words. Phonetic poetry was made up of syllables and sounds that did not join into meaningful words and sentences and polyglot poetry included words from multiple languages, so that almost no listener could understand it fully (Hopkins, 2004). Hugo Ball's "Dada Manifesto" (1916) expresses the wish for a private language, unintelligible by others:

I don't want words that other people have invented. All the words are other people's inventions. I want my own stuff, my own rhythm, and vowels and consonants too, matching the rhythm and all my own. If this pulsation is seven yards long, I want words for it that are seven yards long (p. 2).

In the poetry field, starting from the modern era, difficulty and aesthetic quality have become intertwined. Difficult poetry researcher Iris Yaron (2010) explains:

In poetry, unlike in other forms of discourse, obscurity might be an aesthetic principle; indeed, poetic discourse enjoys a special privilege: it may run counter to the fundamental requirement of language, namely communicability, and may infringe some of the basic rules of language. It is free to disobey the rules of syntax, grammar or lexis. It is able to depart from the requirements of coherence, cohesion [...] And despite all this, simply because it is a poem, it will be perceived as a significant text (p. 3).

There is some relation between the radical aesthetician position and the aesthetics of noise music, which, in its most extreme manifestations, does not allow the listener any peace of mind, or any sense of musicality. In “The Paradoxical role of Noise in Music”, Cempsato (2011) describes the aesthetic experience of listening to noise music:

During the experience of a concert of *noise*, this process of musicalization becomes turbulent, creating an environment of constant annoyance. The situation remains disturbing from beginning to end, creating listening relationship that is close to a fight for survival, bordering the limits of the body. In this context noise remains a result of a specific situation, which resists generalizations, abstractions and analysis. It is the experience of a contingency. In this case noise is sublimated, augmented, almost as if it were to be worshiped (p. 86).

There is something in common between the musical noises of noise music and the semantic noises experienced while reading a difficult literary text. In both cases, these elements are undesirable and would be considered communication failures under normal circumstances, whose sensory quality is unpleasant, and deviates from that which is considered aesthetic. Indeed, and hence the radicalness of the radical aestheticians, they derive pleasure precisely from those difficulties and irritations, those noisy and bothersome elements, considered a disturbance for the ordinary reader or listener. Note also that while the radical aesthetician position is not completely contradictory to foregrounding theory, it highlights only one stage out of the entire process, that of the deviation, defamiliarization, confusion, disruption of sense-making, and prefers it to the complete process which includes defamiliarization but also refamiliaization, deviation but also interpretation that relies on and clarifies it.

Since the radical aesthetician position emphasizes effort and difficulty as leading to aesthetic appraisal, in terms of the failed foregrounding model the most important stages in radical aesthetic appraisal would be failed and partial foregrounding. Thus, should that position prove more correct, the essence of the literary encounter needs to be associated with failed and partial foregrounding. Namely, it is not the completion of foregrounding that produces literary experience. Rather, whoever stops in the middle of the process, whoever responds to the invitation to interpret, but has failed to show up to the interpretation party, is the one who experiences the radical aesthetic experience. These predictions are completely the opposite of those of the failed

foregrounding model. I know of no empirical researchers of literature who espouse the radical aesthetician position. Therefore, I see particular interest in examining these predictions against those of my own model as competitive approaches to explaining the behavior of real-life readers. This issue will be further elaborated on in the findings and discussion sections.

In the following sections, I review evidence supporting the failed foregrounding model from previous studies and my own research. Then, I demonstrate how the model and the observations it enables can contribute to the empirical study of literature reading, and how it compares to either the standard model or to the radical aesthetician position.

Empirical Evidence for Failed Foregrounding

Failed Foregrounding

One of the earliest findings which may be interpreted as supporting the failed foregrounding model have been provided by Miall and Kuiken's (1994) classical experiment, one of the studies most frequently quoted in support of the foregrounding theory. The experiment includes several important methodological innovations following Van Peer's (1986) pioneering experiments. For the first time, Miall and Kuiken have shown that foregrounding characteristics in a text are positively correlated not only with strikingness, as shown by Van Peer, but also with emotion ratings and slower reading. Note, however, that while the experiment's general findings support the foregrounding theory, some of them may be reinterpreted as supporting the failed foregrounding hypothesis.

Miall and Kuiken (1994) had their participants read three short stories disjointedly – one sentence after the other – and rank the sentences for strikingness and affect. The researchers analyzed the text based on three different kinds of foregrounding characteristics: phonetic, grammatical and semantic. They presented the findings as correlations between each of these foregrounding characteristics and the speed of reading, strikingness and the readers' affect. In two of the stories, there was support for the foregrounding theory, but in one of them – Virginia Wolf's "A Summing Up" (1973) – the pattern of the findings proved difficult to explain using the theory. In terms of reading speed, there was an effect, but only for grammatical foregrounding: the readers slowed down wherever there was deviation from the grammatical norm or rules. In the strikingness index, however, the story evidenced a significant effect for

Illustration of grammatical foregrounding in the opening of Virginia Wolf's "A Summing Up" (Wolf, 1972)

Since it had grown hot and crowded indoors, since there could be no danger on a night like this of damp, since the Chinese lanterns seemed hung red and green fruit in the depths of an enchanted forest, Mr. Bertram Pritchard led Mrs. Latham into the garden (p. 121)

The short story's opening sentence is characterized by considerable grammatical complexity that deviates from the literary norm. It contains three long causal clauses (since... since... since...) that push the main clause back to the end. Consequently, at first reading, it is difficult for the reader to follow the (purported) causal relationship between the clauses.

semantic and phonological foregrounding characteristics, but not for grammatical characteristics. Thus, while only the grammatical foregrounding characteristics in the text affected reading speed, they had no strikingness effect. In the additional, affective index, grammatical foregrounding devices also had no effect. These results were remarkable, particularly given that this was a story by Woolf, whose writing is typically rich in deviations from the grammatical norm – in fact, this is one of the hallmarks of her style. Why, then, did this particular characteristic of her writing, that made the readers slow down, not produce either strikingness or affective effect?

I suggest understanding Miall and Kuiken's (1994) finding as follows: the readers have reached the second stage in the foregrounding model, but not the third. Their slowed-down reading in the passages containing the grammatical foregrounding devices indicate that this literary device had an effect: it attracted attention and processing resources. The fact that precisely in these places there was no strikingness or affective effect suggests that the dwelling, in this case, did not lead the readers to a literary-aesthetic effect. Thus, the slowed reading indicates cognitive effort – the price of dealing with difficulty – that had no aesthetic reward.

It is not clear why the grammatical foregrounding failed in this case. Perhaps the challenge was too complex for the participants, students in an introductory psychology course. Had they been literature students, they may have been affected otherwise by the grammatical foregrounding devices. This may be also a byproduct of Miall and Kuiken's specific experimental conditions, since the participants read the text only once, sentence by sentence, and could not go back. Such disjointed reading may be particularly obstructive in dealing with Virginia Woolf's writing. Moreover, this may not necessarily be a problem unique to this readers' group or to these experimental conditions. It may be possible to generalize from the findings to the general population after all, and assume that most readers in most cases would act similarly. To understand the factors involved, additional experiments are required.

In any case, something can be learned from the type of failure revealed here. This case offers evidence to the effect that foregrounding devices have managed to attract the readers' attention to the text, but have failed in producing a literary effect of strikingness or of affect. This is a particular type of foregrounding failure, which occurs despite the fact that additional processing was given to a passage with foregrounding devices – i.e. despite the fact that the reader's attention was attracted to the text. In terms of the aforementioned model, this is evidence of failure in the transition between the second and third stages in the foregrounding process.

Before concluding this subsection and moving on to discuss shallow processing, note that I consider both the affective and the strikingness index as measuring part of the aesthetic experience. With regard to affect, this follows upon Van Peer, Hakemulder and Singer's (2007) indexes, as well as Menninghaus' (2015) work. The strikingness case is a bit more complicated. It seems that this effect is both cognitive and aesthetic. The early Van Peer (1986) presents it as evidence for attracting attention, and in

Leech and Short's (2007) terms, this is supposed to be an example for psychological prominence, that is, the second rather than the third stage of the model. However, in Van Peer's subsequent work with Hakemulder and Singer, when they developed six general indices for foregrounding effect, they includes strikingness in the aesthetic appreciation aspect. There, strikingness is presented next to the "beauty" and "musicality" of the text. Moreover, this index was found to be highly reliable (Cronbach's alpha = 0.67), suggesting that these three terms largely measure the same thing. Van Peer et al.'s (2007) aesthetic index reads as follows:

Aesthetic appreciation

- I think this line is musical
- I think the sentence is beautiful
- I found it striking

Shallow Processing

Another possibility for a failed foregrounding process is when the reader's attention is not at all attracted to a specific passage that deviates from the linguistic norm. A case in point may be found in the experiment of Emmott and her colleagues (Emmott et al. 2006). Combining two different research traditions, they added to the conventional foregrounding paradigm linguistic research tools borrowed from the shallow processing paradigm. The main concept of the latter paradigm is that linguistic processing is not perfect, but varies in depth according to the characteristics of the text and the reader's task requirements. Shallow processing researchers describe a series of effects from a variety of linguistic situations where readers demonstrate shallow processing. A known example is the so-called Moses Effect, based on questions such as "How many animals of each species did Moses take on the Ark?" A surprising number of participants, more than half, answer the question without noticing that it is Noah who took the animals on the Ark, according to the story of the Flood (Bredart & Modolo, 1988).

Shallow processing does not only prevent readers from noticing errors, but also makes it difficult for them to notice changes in the text. One of the main research methods in this area examines participants' ability to notice small changes between two versions of a sentence or a short paragraph. In this method, the researcher changes one word and lets participants read the two versions in order to determine whether they are identical or different (Sanford, & Sturt, 2002). Using this method, it was found that the degree of semantic detail in the representation is a function of linguistic focus.

Emmott, Sanford & Dawydiak (2007) used the finding of these experiments – it is easier to detect change when it is in the location under linguistic focus – to explore whether various stylistic devices do emphasize the text and attracts reader attention.

In a series of experiments, an unexpected result was found: an entire series of devices considered by the researchers to attract attention and lead to literary effect did not help participant detect change.

More formal ways of emphasis such as using italics, clefting, short sentences or independent clauses did lead participant to notice change. But emphatic devices defined by the researchers as content-related did not lead participant to detect change. These were indications of surprise, affect and importance. For example, the statement “What happened next made me furious” did not improve participants’ ability to notice changes in the following sentence (Emmott 2007, p. 214). Not only did these content-based emphatic devices not improve the ability to detect change, but one device, a statement of surprise, led to significantly lower detection rates compared to the control group.

The importance of this finding exceeds the specific question of whether the content-based emphatic devices examined in Emmott et al.’s (2007) experiments attract attention.¹ The important point here is the *very possibility* that this kind of failure may occur in the foregrounding process. This indicates a foregrounding failure different from that seen in Wolf’s story. According to the model presented above, this is failure in the transition between the first and second stages. The failure is in what is considered a preliminary stage in creating an aesthetic-literary effect using foregrounding.

The evidence offered above further support the model – particularly the existence of two types of failures in the foregrounding process. Another conclusion arising from these two experiments is that various foregrounding devices differ in their degree of effectiveness. This issue will be examined in length in the subsection on failed foregrounding and the text. Prior to that, however, in the section on failed foregrounding and the reader, I will present evidence that support the model based on a reading experiment. Two types of evidence will be presented: questionnaires that allow for a general examination of the reading process and interviews based on eye movement patterns that allow for a more local analysis according to the key points in the texts.

¹ The reason for failure in this case is unclear. It is too early to say whether the devices described in Emmott et al.’s (2007) experiments would normally fail to attract attention, or this is an artefact of the experimental conditions. It may be that, as Emmott suggests, this is due to limitations of the method of measuring the detection of change. This method is more sensitive to visual than to semantic or narrative changes. Another possibility is that the artefact is due to the artificiality of the texts in the experiment. Since they were not literary texts, the readers were not truly curious or expectant, hence their low attention levels. Further research is required to determine which explanation is more adequate.

SECTION 2: FAILED FOREGROUNDING AND THE READER

Studying Failed Foregrounding Using Questionnaires

In the experiment reported here, readers of a complex literary text were required to provide self-reports of aesthetic appraisal, semantic noise and experience in reading prose, measured by the author recognition test. I developed three Hebrew questionnaires for that purpose. The selected text was Borges' "The Chamber of Statues". It was considered suitable because most readers were unfamiliar with it, and because of its complexity, despite being very short (513 words in Hebrew).

The participants were 42 native Hebrew readers without any reading or learning disabilities. They included 25 women and 17 men, 88% of whom between 20 and 30 years of age. They were recruited through ads posted at the Tel Aviv University that read: "Wanted: Experimental subjects who love books". Prior to the experiment, they were asked to complete a personal details form in order to make sure they met the criteria for participation, as well as obtain an initial idea of their experience in reading literature.

The story was presented on a computer screen. Due to technical reasons, it was restructured into five instead of four paragraphs. To move to the next paragraph, participants had to press the space button; after doing so, it was impossible for them to return to the previous paragraph. As they read, the participants' head rested on a chin rest to minimize their head movements. The instructions for the readers were: "You are about to read a story written by a well-known author. Read it like you usually read literature. Concentrate, relax, and try to enjoy". After reading, the participants completed the Semantic Noise Questionnaire and the Aesthetic Appraisal Questionnaire. This was followed by an interview, at the end of which the participants completed the Author Recognition Test. At the end of the experiment, they were paid the equivalent of 15 USD for their participation. The experiment also included monitoring of eye movements that were presented to the readers during the retrospective think aloud protocol (as explained in the last section).

Questionnaires

The questionnaires were developed in a three-stage process. For each, a pilot version was developed and presented to 3-5 participants. The experimenter discussed the questionnaire with them to make sure the questions were understood. Following this pilot, several revisions were made. The process was reiterated with an additional pilot group to produce a third and final version.

Author recognition test

Since authors' names represent culture- and language-dependent knowledge, the task must be redeveloped for each language considering the local canon of literature. To the best of my knowledge, the task has not yet been adapted into the Hebrew language and culture. Therefore, I created a Hebrew questionnaire with 120 names. Sixty of these were names of authors specialized in the following genres: world literature, Hebrew literature (or Jewish authors in other languages), science fiction and fantasy, suspense and detective stories, and children's literature. The distractor names included politicians, painters and made-up names. The idea behind using politicians and painters' names was to create a sense of familiarity, so that the participants would not be able to identify the authors as such based only on a general sense of familiarity.

The participants were instructed to identify the authors, and warned not to guess, since they would receive one point for each name recognized and lose one for any misrecognition. They were asked to mark a name only if they were more than 90% certain. Their possible scores ranged between 0 and 60. Next, in order to enable comparison with the rest of the questionnaires, where a 1-7 scale was used, the scores were multiplied by 7/60, to obtain a maximum score of 7. The mean score was 2.75 and the standard deviation was 1.6. The score distribution of the participants was normal (Shapiro-Wilk's $W=.95622$, $p=.10807$), without a group of "experts" as opposed to "laypersons", but a range of experience level that covered almost the entire scale. The false recognition rate was rather low ($m=1.4\%$), which means the participants rarely guessed a name. See Appendix D for the full questionnaire.

Semantic noise questionnaire

The questionnaire included 15 statements related to the story. The participants were asked to indicate their level of agreement with each on a 1-7 rising Likert scale; six items were reversely scored, indicating not a noisy but rather a smooth and noise-free reading experience. The statements referred to various literary aspects: the author, the writing, the text, the story, the sentence and the word. They also referred to a variety of sensations related to semantic noise: confusion, comprehension difficulty, cumbrousness, complication, lack of clarity, problematic phrasing, stuckedness, and unfamiliarity. Opposite terms were also used: easy reading, smooth reading, well-paced story, as well as the author being able to convey his message to the reader.

The questionnaire's internal reliability was found to be Cronbach's alpha = .91, a satisfactory value. An examination of the internal correlations between each of the items and the questionnaire's overall score revealed the five most typical items: 14. The text was written in a way that made reading easier (reverse scored); 10. This kind of writing makes it difficult for the reader to understand; 3. In my opinion it was written cumbrously; 13. The author expressed himself awkwardly and confusingly;

11. The story was well-paced (reverse scored). See Appendix B for the full questionnaire.

Aesthetic appraisal questionnaire

I believe that since the aesthetic experience is so broad and complex, a permissive attitude is in order, of measuring as many aspects of that experience in one experiment; this approach guided the development of the questionnaire. It included 25 items, four of which were reverse-scored. Some of them were variations on existing questionnaires, particularly those used by Dixon et al. (1993) and van Peer et al. (2007), as well as by Menninghaus et al. (2015) and Knoop et al. (2016). Other items were deemed suitable by the researchers to describe the aesthetic experience of reading a short and poetic piece of prose. The participants were asked to indicate their level of agreement with each item, with reference to the story just read, on a rising 1-7 Likert scale.

The items referred to three spheres of the aesthetic experience: cognitive, affective and behavioral. Cognitive terms in the questionnaire included: style, beauty, surprise, and thought provoking. Affective terms included laughter, sadness, excitement and fear. Behavioral terms included recommending to a friend, reading more by the same author and wanting to read the same story again. The questionnaire's alpha Cronbach was satisfactory: .93. An examination of the internal correlations between each of the items and the questionnaire's overall score revealed the five most typical items: 20. The story is well written; 23. Reading this was a waste of time (reverse scored); 11. The story bored me (reverse scored); 24. I would like to read more stories by the same author; 9. The story was interesting. See Appendix C for the full questionnaire.

Results and Discussion

Correlations between questionnaires

No linear relationship was found between experience in reading literature as measured by the Author Recognition Test (ART) and the two other questionnaires: ART with the Aesthetic Appraisal Questionnaire (AAQ), $r=.22$; $p=.12$; ART with the Semantic Noise Questionnaire (SNQ), $r=-.14$; $p=.35$). Namely, literary experience did not predict the semantic noises revealed while reading the story or its aesthetic appraisal.

Nevertheless, AAQ and SNQ were correlated ($r=-.6$; $p<.0001$). Namely, the higher the semantic noise the lower the aesthetic appraisal and vice versa, the greater the aesthetic appreciation the lower the semantic noise (see Figure 4). This finding indicates either that semantic noise affects the aesthetic appraisal of literature, or that positive aesthetic appraisal weakens subjective semantic noise. The strength of the relationship suggests that a large percentage (36) of the variance in the aesthetic

experience may be explained by semantic noise or vice versa. Therefore, it is an important variable whose measurement can help researchers understand the nature of the literary aesthetic experience. Note that it is also possible that some third or fourth variable may determine the two others.

This finding may seem counterintuitive to literature researchers who assume that a literary text makes the reception process deliberately complex (e.g. Shklovsky, 1965/1917). They would have expected the opposite trend that semantic noises would lead precisely to positive aesthetic appraisal, if not by all readers than at least among experts. The finding is highly consistent with the fluent processing theory (see Appendix A on semantic noise for elaboration), but less consistent with the standard foregrounding model or the radical aesthetician position. Note, however, that additional analysis reveal a more complex picture. What follows may shed more light on this issue.

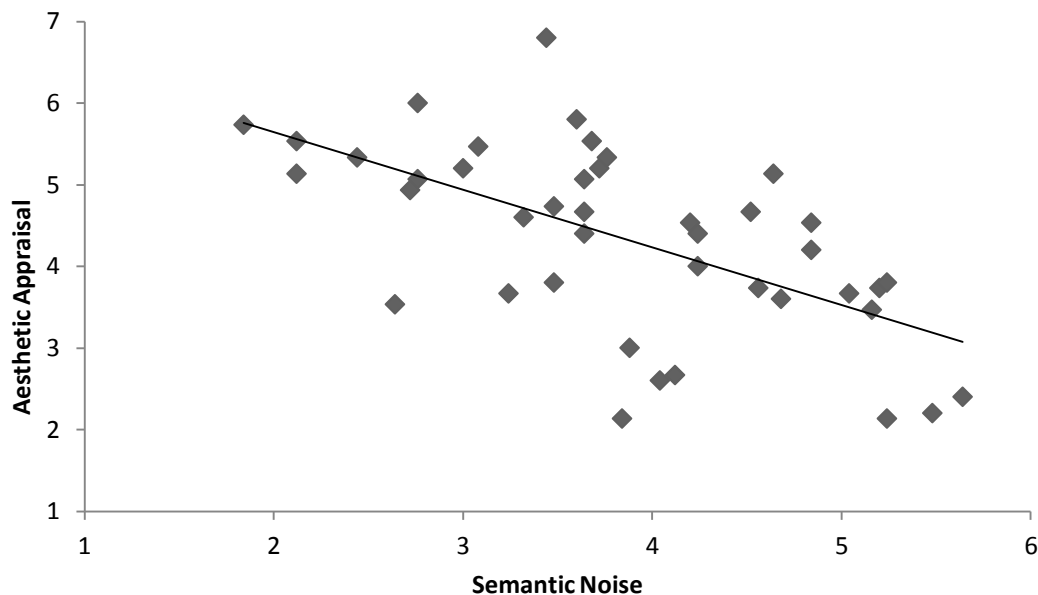


Figure 4: Correlation between semantic noise and aesthetic appraisal as measured in the respective questionnaires.

Cluster analysis

The machine-learning analysis method was selected since it allows detecting data trends that are difficult to locate otherwise, particularly when there is no clear-cut linear relationship. The type of machine learning performed was a k-means cluster analysis (see subsection "Cluster analysis" p. 38 for further details).

The findings presented below are based on three clusters. Since the data could be clustered in different ways, the decision to group the participants in three clusters should be justified. Two clusters were not sensitive enough, while the three clusters produced interesting and meaningful results. The division into four groups was identical to the three, apart for two participants who formed a fourth cluster of their own. Divisions into five groups and above were senseless given the fact that there were only 42 participants, resulting in clusters too small to have any statistical significance.

Although the *two-cluster structure* (see Figure 5) was not sensitive enough, it was important since it created a pattern that replicated some of the previous studies showing that experienced readers had greater appreciation for more complex literature, whether prose (Dixon, Bortolussi, Twilley & Leung, 1993) or poetry (Peskin 1998). This division formed one cluster of highly experienced readers ($n=13$; $m=4.67$; $SD=1.0$), with relatively high aesthetic appraisal ($m=4.4$; $SD=1.1$) and relatively low semantic noise ($m=3.86$; $SD=1.3$). The second cluster included participants with little literary experience ($n=29$; $m=1.9$; $SD=.9$), with relatively low aesthetic appraisal ($m=3.6$; $SD=.8$) and relatively high semantic noise ($m=4.54$; $SD=1.0$). The difference between the two clusters was significant: $F(3,38)=30.012$, $p<0.00001$; see Figure 5.

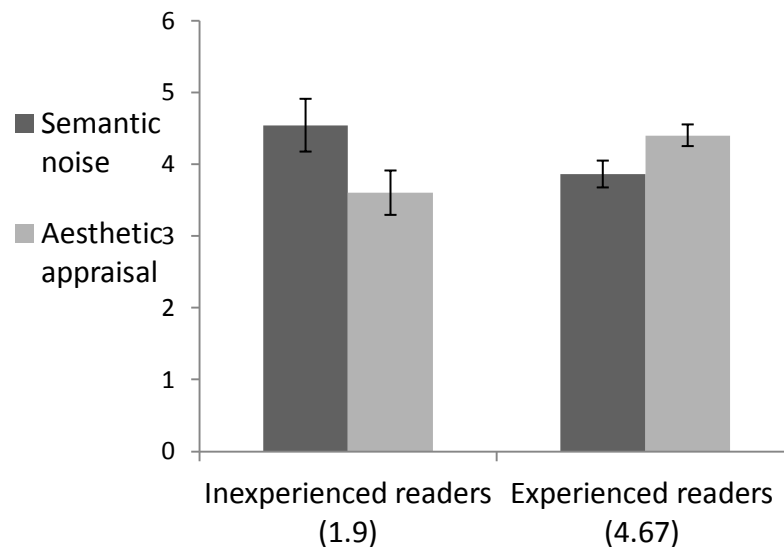


Figure 5: Grouping of participants into two clusters: Means and standard errors for semantic noise and aesthetic appraisal. In parentheses, the mean score in the author recognition test.

Using a *three-cluster structure*, however, a more complex picture is revealed (see Figure 6). One cluster ($n=19$) is made up of participants with little reading experience ($m=1.5$; $SD=.8$), and neutral scores in both aesthetic appraisal ($m=3.93$; $SD=.7$) and semantic noise ($m=4.0$; $SD=.8$). The second group ($n=13$) includes participants with

medium reading experience ($m=3.1$; $SD=1.1$), characterized by high noise ($m=5.51$; $SD=.5$) and low appraisal ($m=2.94$; $SD=.7$). The third cluster ($n=10$) included highly experienced readers ($m=4.63$; $SD=1.1$) with low noise ($m=3.36$; $SD=1.0$) and high appraisal ($m=4.86$; $SD=.7$). The difference between the clusters is significant, $F(6,74)=22.915$, $p<0.00001$).

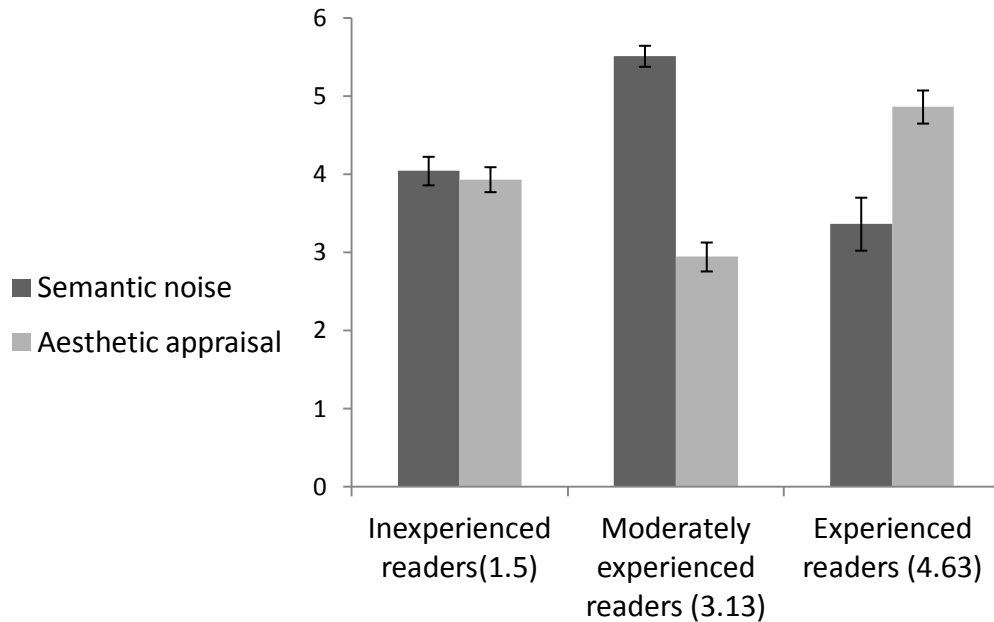


Figure 6: Grouping of participants into three clusters: Means and standard errors for semantic noise and aesthetic appraisal. In parentheses, the mean score in the author recognition test.

The first finding is the lack of any linear relationship between experience in reading literature and aesthetic appraisal or semantic noise. This finding does not mean there is no relationship between these variables, but that the relationship is not simple. Cluster analysis provides an in-depth view on the data, revealing what is hiding behind the lack of direct relationship. The clustering into two groups reported above reveals the classical finding, replicating previous studies showing that greater experience is related to higher aesthetic appraisal of a complex text.

Another, deeper pattern hidden in the data may be exposed only by a three-cluster structure. Indeed, the major innovation in this finding is revealed this way. People with very little experience did not rate the text as highly aesthetic, but also not as semantically noisy – their ratings were relatively neutral. Comparatively, among more experienced readers, the ratings tended to extremes. Participants with medium experience reported the least “literary” reading, with low aesthetic appraisal and high semantic noise. Those with the greatest experience had the most “literary” reading, with high appraisal and low noise.

Interpreting the findings in view of the model

The present findings may be interpreted in the spirit of the failed foregrounding model. It may be that inexperienced readers used the shallow processing strategy more often than the rest. They noticed less deviation in the text, and thus also sensed less semantic noise, but their potential of reaching significant literary insights was limited, and thus their aesthetic appraisal was also medium. Similarly, it could be that readers with medium experience reached failed foregrounding more often than the rest. These participants notice the deviations in the text, but had difficulty addressing these issues effectively due to their relative inexperience. Finally, the participants with the richest experience who reported high appraisal and low noise reached full foregrounding more often than the rest. They noticed deviations in the text and also dealt with them effectively, leading to an improved and more complete interpretation, as well as positive aesthetic appraisal.

The finding that experts experienced a low degree of semantic noise shows that successful interpretation was not difficult or cumbersome for them, and did not make them feel confused or stuck while reading. Locating and interpreting deviation in a way that contributed to aesthetic appraisal was almost effortless for them. This finding is consistent with Peskin (1998), who found that although literature experts applied cognitive strategies of greater complexity while reading poetry, their subjective experience was of pleasantness and effortlessness, whereas it was the laypersons who often expressed frustration while reading.

The difference between the three groups can also be explained by two variables: the ability to identify deviations in the text and the ability to deal with them. Inexperienced readers are low on both variables; those with medium experience are good in identifying deviation, but lack the ability to deal with them successfully; while experts are good in both. These may be two different components or aspects of literary expertise. Although literary experts have good strategies for dealing with deviation in the text, they find many more of these deviations that need to be explained and solved in a given text. And just as the large number of deviations detected by the expert reader can lead to a deeper and more satisfying reading, it can also lead to greater confusion, difficulty, stuntedness and miscomprehension.

The idea that these two factors are differentially sensitive to experience in reading literature is consistent with Hanauer's (1999) model for the development of literary knowledge. According to this model, literary knowledge develops from implicit to explicit knowledge that is accessible to reflection with the mediation of the attention system. Attention plays a role in that the reader learns to locate patterns in the text, detect textual deviations, as well as attend to specific literary devices. This ability is acquired either directly or indirectly. Directly, through explicit instructions that increase the general attention to the text or increase the attention directed at specific

textual features. And indirectly, by virtue of experience in reading literature. Equipped with this attentive discerning ability, the reader selects the information that enters the cognitive system as a basis for further processing and deeper interpretation. Therefore, it stands to reason that in the gradual development of literary knowledge, the ability to discern textual characteristics would precede the ability to provide a comprehensive and successful interpretation of the text.

Examples from the interview materials

These findings may be illustrated anecdotally using the interview materials. While reading, participants often lingered on the narrator's comments. For example when Solomon Son of David is mentioned, the narrator declares, "salvation be with both of them!" (See Appendix E). Three of the participants who lingered on those words provided the following typical explanations:

- *Inexperienced participant (shallow processing)*: Here it is again, that side comment. – Interviewer: And what did you think about it? – Here also, that it is kinda nice.
- *Participant with medium experience (failed foregrounding)*: Solomon Son of David – salvation be with both of them! That weird comment – who is the narrator and what is his ideology? Is it a story from the Jewish sources? Also, I really didn't understand what on Earth David Son of Solomon is doing here in this story, that seems to be for the general reader [rather than specifically Jewish].
- *Highly experienced participant (full foregrounding)*: All these interjections made me curious. They also reminded me of things I know about Islam. It made me think about how a certain phrase needs to be said after mentioning Muhammad's name, "May God honor him and grant him peace". So maybe I lingered there because of that.

The inexperienced participant had a relatively neutral experience ("kinda nice"), and provided a response that did not indicate any particular difficulty, interpretive literary insight or profound aesthetic experience. The deviation from the plotline was not experienced as a "problem" that had to be solved or reflected upon. The sentence remained disconnected and no attempt was made to integrate it with other issues raised by the story, as the two other participants have made. This response shows a typical *shallow processing* pattern.

The participant with medium experience thought the comment was odd, did not understand its role, and therefore wondered about the source of the story and the narrator's ideology. Clearly, that little authorial comment caused significant problems in understanding the story, that were in turn expressed on two levels of abstraction, in

an attempt to reintegrate the deviation (1) with the genre (a story for general readers? A story from Jewish sources?); and (2) with the narrator's perspective (What is his ideology?). Thus, issues were raised, questions were asked, and an attempt at integration was made, but no answers were found and the aesthetic experience was not satisfactory. Rather, the reader experienced difficulty and confusion, leading to a "weird" feeling that was not gratifying. This response shows a typical failed foregrounding pattern.

Conversely, the experienced participant described the narrator's comment as curious. He used the word "interjections" (referring to comments from the audience that disrupt the continuity of a discussion or a theater play), indicating his understanding of the passage as a deviation from the normal plotline that could be experienced as an interruption. In order to interpret that deviation, he too is forced to look outwards. Like the participant with medium experience, he too connects the comment to broader knowledge "located outside" the passage, in this case not to the story plane, but to general knowledge associated with Islam. Just as in Islam, the Prophet's name must be followed by a certain formula, so in the story, King Solomon's name must be followed by a formula. Thus, in this case, the experienced reader managed to solve the problem in a manner that pleased him, made sense given the overall Arab style of the story, intrigued him and even contributed to his aesthetic experience. This response shows a typical full foregrounding pattern.

Theories Supported by the Findings

These findings challenge approaches that highlight difficulty as a key factor in reading literature. Harash and Shen (2016; see Appendix A) present three main types of aesthetic theories that relate aesthetic appraisal and cognitive effort. At the extreme end of the cognitive difficulty continuum, we have theories that argue that interpretive difficulty is the heart of aesthetic experience, somewhat like an extreme version of the foregrounding theory that is similar to what was named here the radical aesthetician position. At the other end, we have the fluent processing hypothesis that argues that smooth and effortless processing is a prerequisite of an optimal aesthetic experience.

Fluent processing hypothesis

The negative correlation between the Aesthetic Appraisal Questionnaire and the Semantic Noise Questionnaire supports the fluent processing hypothesis. Globally speaking, in appreciating the entire story, the semantic noises did not contribute to a positive aesthetic appraisal. On the contrary, the noisier the reading the less positive the appraisal. More than a third (36%) of the aesthetic experience measured in the questionnaire could be explained by semantic noise. Nevertheless, as the questionnaires measure only the global effect rather than local effects within specific

points in the text, it may be argued that the global effect is a combination of both the background and the foreground of the text, and is therefore sensitive to both foregrounding and backgrounding processes. For the foreground to become salient, it needs a background, and it is better for the reader to process the background fluently, so that the slowdown in the foregrounded areas contrasts the smooth processing in the remainder of the text. Accordingly, support for fluent processing on the global level is consistent with foregrounding theory as I understand it.

Standard foregrounding theory: Failed foregrounding model

The negative linear relationship between semantic noise and aesthetic appraisal does not support but also does not contradict the standard foregrounding theory, to the extent that these findings are interpreted as relevant to the background more than to the foreground of the text. Nevertheless, the division into three clusters paints a different picture. The failed foregrounding model is the only one that could explain the findings given that division. The pattern emerging among the participants with medium experience is consistent with the possibility that failed foregrounding leads to multiple semantic noises and negative aesthetic appraisal. On the other hand, the inexperienced participants showed a pattern more consistent with shallow processing. Thus, these findings support the existence of the two main types of failure described in the previous section: shallow processing and failed foregrounding. The interview-based analysis of the key points in the text, presented below, will allow to characterize the effects of these failures with greater accuracy, as well as the types of stylistic devices related to them.

Radical aesthetician position

This is the main approach not supported by the questionnaire findings. The difficulties and confusion arising during the reading, as measured by the semantic noise index, did not contribute to the story's aesthetic appraisal but vice versa.

Studying Failed Foregrounding Using Reader Interviews

In order to further investigate the failed foregrounding model, I am interested in gathering more detailed evidence that would enable more subtle distinctions: How to describe in detail what happens in a certain point in the text on which the readers dwell? What kinds of failures occur? How are the various readers' responses distributed at that point? And does that distribution vary according to the types of readers and the type of textual deviation to which they respond?

To examine these issues, I turned to interviews with readers. I found that if readers are interviewed in a certain way, they provide detailed evidence on foregrounding failures. The interviews focused on certain points in the text that make the readers dwell, and these provided a particularly rich type of verbal information regarding their conscious experience of the foregrounding process and its failures. As mentioned above, in the experiment, participants read a Hebrew translation of Jorge Luis Borges' "The Chamber of Statues" (1935). Subsequently, they completed an aesthetic appraisal and a semantic noise questionnaire. Afterwards, they took part in interviews based on heat maps of their eye movements: immediately after reading the text, I showed them graphic visualizations of their eye movements marking, in red, places where they delayed in the text. I asked the participants whether they remembered, or could assume, why they had dwelled precisely on these spots² (see the methodological section on p. 55 for further details on the interview technique).

During the interviews, I emphasized eight particular passages, or key points, identified in the pilot stage. Most pilot participants had dwelled on these passages and the verbal responses to them were rich and diverse (the key points are marked in the text of the story in Appendix E). The selected key points were identified as having the potential for both a full foregrounding process – realized in some of the cases by the participants – and failures of various types. Points characterized by semantic noise alone, without the potential for an aesthetic experience were not selected as key points. All interviews were recorded and transcribed and the responses to those eight passages were evaluated by external reviewers.

These interviews included many cases where the failed attempt to understand and interpret led to a disturbance that detracted from the reading experience. For example, the following sentence is taken from Borges' story: "and the front hooves of their horses did not touch the ground yet they did not fall, as though the mounts were rearing". This enigmatic figurative sentence led to a variety of participant responses; for example,

I didn't understand what it meant for the horses' hooves not to touch the ground and yet they didn't fall. If they didn't touch the ground, then how did they stand – on their hind legs? It's like I remember myself thinking about it, and reading the last sentence at least twice.

Interviewer: Did it confuse you?

I guess it did, and again, because I know myself, I guess it annoyed me a little.

² I explained to them that they had to try to recall what happened as they were reading and that if they could not remember, they should make an assumption based on their self-knowledge. In providing their answers, they were invited to use certain terms suggested in the questionnaires (such as beautiful, interesting, confusing and difficult), but did not have to do so. They were also explained that sometimes delay was not related to the word itself, but to the sentence or the passage. If their answer was not clear enough, I asked clarification questions.

This reader's response offers several indications of failed foregrounding. First, it appears she stumbled upon a textual problem. She identified a logical contradiction in the sentence. On the one hand, the horses' hooves do not touch the ground, and on the other, the text explicitly states that they did not fall. The reader found it difficult to reconcile the two contradictory pieces of information. Second, the reader clearly delved into the problem, reporting having thought about it and reading the sentence twice in an attempt to do so. Third, it appears the reader did not have a positive aesthetic experience nor reached interpretive insight, but was only confused and annoyed. Accordingly, the foregrounding process began – a deviation and difficulty was identified (stage 1), and the reader dwelled on the problematic passage in an attempt to understand it (stage 2), but this did not lead to affect, interpretive insight or positive aesthetic experience. Thus, it appears that this reader's response can be interpreted according to the model presented above as indicating failed foregrounding.

The following reader quotes are typical of the various model stages. The examples are taken from the interviews, but edited so that they can be understood even without knowing to which part of the story they refer:

Shallow Processing

- These are just words that were less familiar to me.
- I dwelled on it to make sure I'm reading correctly.
- I don't remember whether I read it correctly before, but now it confused me.
- I have no idea what this means – I simply saw it and did not attach any importance to it.

Failed Foregrounding

- There it is again, it already appeared before in the story. I tried to connect the two passages and came up with complete confusion. I couldn't understand how it all connects together. It really confused me, like crazy.
- I also didn't understand who's writing this and about what. Because on the one hand first they cursed the lead character, and now it seems the narrator is on his side. So I got stuck for a moment. I didn't understand the contexts.
- OK, and how does it help me? I mean, at the end I didn't understand what kind of hero he is in the story, how he contributes to the plot. What his function is – is that the word, what his function is?
- It's a kind of inner talk, so it felt to me, or like the narrator says his thing in the story. I concentrated in order to understand why it's important for him to have that kind of talk. I concentrated to understand why it's relevant.

Interviewer: And did you find it beautiful or interesting?

No, I found nothing in it.

Partial Foregrounding

- Part of it was very hard to understand, highbrow literature, highbrow language. The language was not uniform across the story.
Interviewer: And did you find this language also beautiful, or did it only make life difficult for you?
It was also kind of beautiful, it sort of created an interest.
- I remember stopping there because I kind of like wondered about that piece of information, but what does it give me? It mystified me, maybe confused me.
Interviewer: But did it seem like a poetic or intriguing detail to you?
Yes, it added something to the sort of materiality of the text... Yes, it did add something.
- Here I dwelled, I remember, because I tried to visualize it in my head and couldn't. Maybe because the phrasing was awkward for me. I repeated it several times. Such phrasing, a little bit awkward but also beautiful.
- I dwelled on it, I didn't understand it. Now I do, but before I didn't. It was also a bit funny. I don't know why. It felt like a kind of joke to me.

Full Foregrounding

- *Sensory experience*: Here specifically when I read it I saw it in my imagination. I really imagined the situation here.
- *Affective arousal*: It made me laugh, yes, it made me laugh. 'Cause it's nice, this tone.
- *Interpretive move*: There is an additional layer here, a kind of irony.
- *Literariness*: It's a poetic combination. I wouldn't have chosen to describe it this way myself. You have to be a writer for that.
- *Immersion*: It puts you in a certain atmosphere. That you're inside the story and really enter the spirit of the times.

Classifying these quotes involves a certain subjective element, since the borderlines between the categories are not always clear-cut. Therefore, I had to rely on external reviewers. Three reviewers analyzed the readers' responses according to the various model stages. They were personally known to me as experienced in literature. I had them read "The Chamber of Statues" and explained about the experiment and the model. I then presented them with a scheme of the model described here and explained each of the stages and the expectable failures. An additional tool for their judgments was the decision tree presented in Figure 7. Finally, I presented them with several examples, and they practiced their decision-making on ten cases, explaining their considerations. While reviewing the rest of the cases, I was available to the reviewers for consultation.

The decision tree that was presented to the reviewers (see figure 7) is made up of two stages, in each of which the reviewer is asked a binary question. The first question is: Was there a positive aesthetic experience? Since I believe in a broad definition of

aesthetic experience, I preferred not to narrow its boundaries in advance. Accordingly, the reviewers did not receive a restrictive definition, but rather offered several examples, leaving the issue open to their discretion. If the answer to the first question is negative, the reviewer must ask: Did the attempt to deal with the difficulty transcend the straightforward verbal level? If the answer is negative, then we have a negative or neutral aesthetic experience without an attempt to transcend the verbal level, hence shallow processing. If the answer is positive, we have an attempt to delve into the text that did not produce a positive aesthetic experience, hence failed foregrounding.

And if the answer to the first question is positive, the reviewer must ask: Was there an interpretive move? And was the processing relatively noise-free? If the answer to both of them is negative, we have partial foregrounding. If the answer to both or even one of those questions is positive, we have full foregrounding – a positive aesthetic experience, relatively free of noises or having an interpretive element.

Similarly to the aesthetic question, the question “Does the attempt to deal with the difficulty transcend the straightforward verbal level?” had no single, clear answer. If the reader said, for example, “I delayed in order to read correctly, because it’s important for me to read like they want me to”, it is not clear whether this was simply an attempt to overcome a difficulty as the most basic level of reading, just an attempt to “read correctly”, making this shallow processing, or whether the phrase “read like they want me to” suggests that the reader had an internalized model of communication with the author(s), and consequently, that his very effort to read correctly was part of a discursive move that transcended the straightforward verbal level.

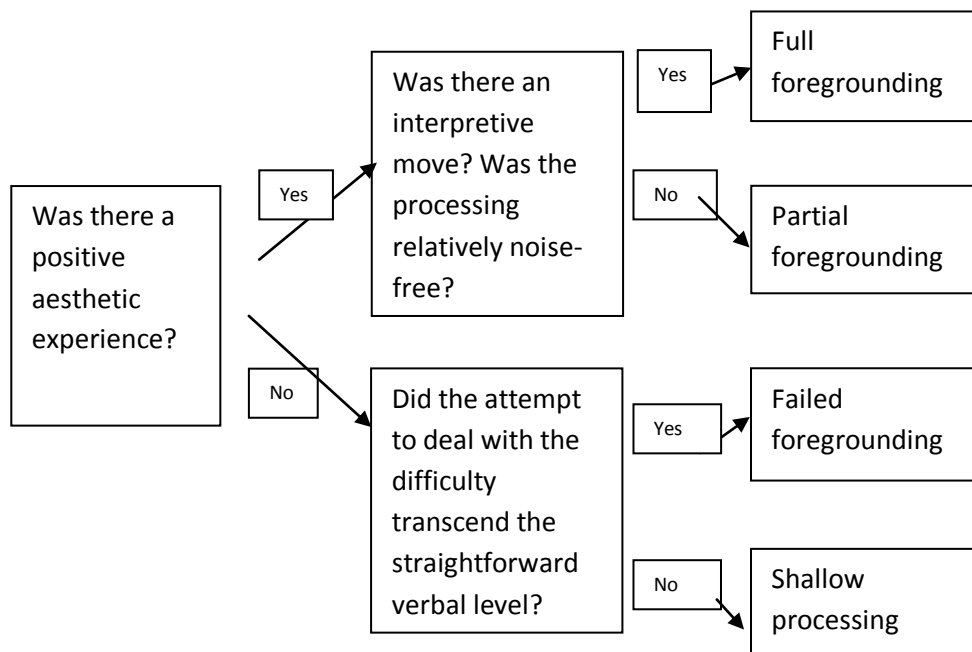


Figure 7: The decision tree used by the reviewers to classify readers' responses according to the failed foregrounding model. Thus, since any decision regarding such statements was highly dependent on their context and their wording, I preferred to trust the reviewers' judgment, rather than restricting them with rigid rules. Fortunately, although the reviewers had to make complex decisions and did not even consult one another to reach consensus, interrater agreement was rather high. In an interrater reliability index called Krippendorff's alpha (for ordinal variables) that ranges between 0 and 1, the result was 0.74.

Main Findings

This section will demonstrate how the analysis of the interviews can characterize the frequency of the various model stages. It will examine, among other things, the aesthetic implications of the model stages and the relationship between them and experience in reading literature. The first and most general finding is that shallow processing is the most common strategy, as demonstrated in Figure 8.

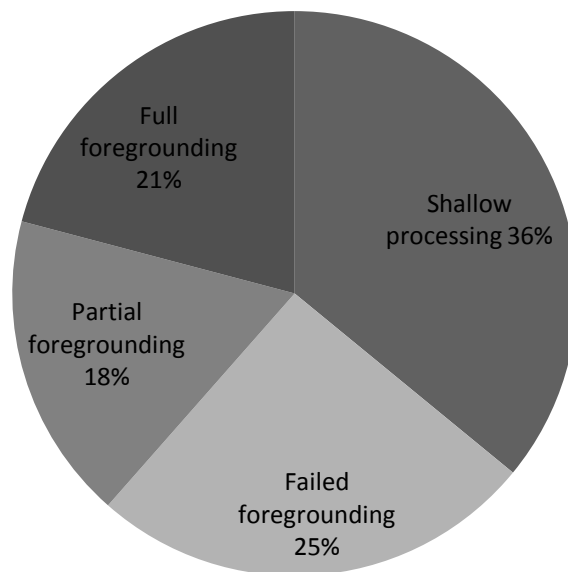


Figure 8: Pie chart of the foregrounding stages in all of the experimental observations. There is a significant difference in stage frequency ($F(3,1140)=14.77$; $p<0.000001$), due mainly to the higher frequency of shallow processing.

In 36% of the cases, the foregrounding process did not start at all. Instead, the readers settled for shallow processing of the difficulty before moving on with their reading. In other words, in more than a third of the cases the readers did not accept the difficulty as an “invitation for interpretation”. Either they did not identify the invitation as such, or they identified it and preferred to remain with shallow processing for other reasons.

When the readers did choose to continue dwelling on the difficulty, there was a more or less similar likelihood for the foregrounding process to fail, be partial or succeed. Full foregrounding occurred in only 21% of the cases. Thus, the likelihood for the foregrounding process to be complete was not significantly higher than for the other possibilities, and was significantly lower than the likelihood of shallow processing.

These findings indicate that the proportion of cases explained by the standard foregrounding model is rather low. In fact, the standard model ignores the absolute majority of cases of real participant readings. The experiment indicates that smooth transition across foregrounding stages occurs only in about one-fifth of the cases. The standard model's disregard of cases of misunderstanding, confusion and shallow processing means that it ignores most of what goes on when real readers approach the text. Not only was there no higher frequency of full foregrounding among the participants, but they evidenced a general tendency not to rush into a foregrounding process when coming across a linguistic deviation that would facilitate it. As you can see below, various factors moderate that tendency. For example, experienced readers opt for shallow processing less frequently than do inexperienced ones. Nevertheless, even for them this is a quite common choice.

Foregrounding Profiles According to Aesthetic Appraisal

The previous section presented the general frequency of the various foregrounding stages, but examining frequency alone is not enough to learn about how each stage affects the reading experience. The results presented here concern the relationship between the foregrounding profile and general aesthetic appraisal in reading the story. The participants were divided into two groups of high and low aesthetic appraisal, respectively, according to the aesthetic appraisal questionnaire. The relationship between aesthetic appraisal and the foregrounding profile is presented in Figure 9. Note that the aesthetic index is a general one that applies to the reading of the entire story, while the foregrounding profile is based on local testimonies referring only to the eight key points. Thus, the relationship found in this comparison is one between a general effect relevant to the entire reading experience and local effects in key points that have attracted the readers' attention.

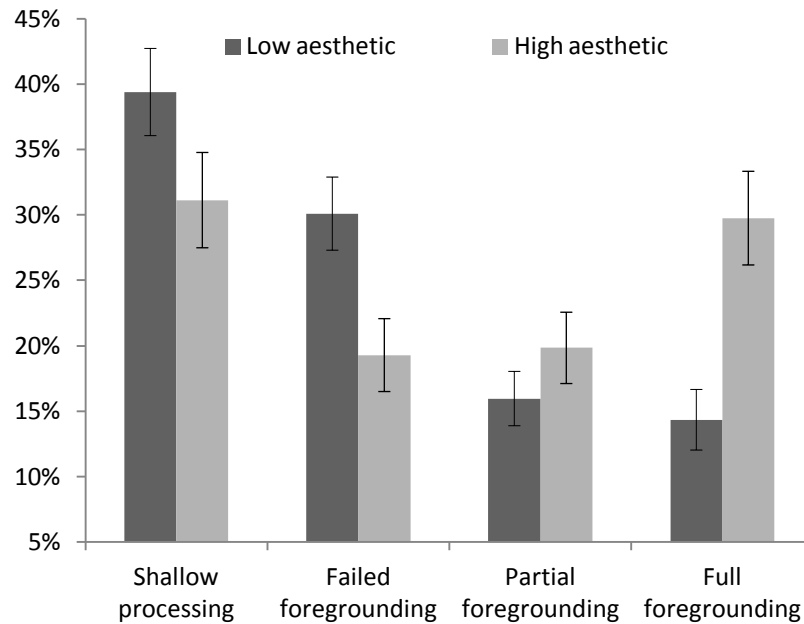


Figure 9. The distribution of foregrounding profiles by global aesthetic appraisal, measured in a questionnaire after reading the story (see Appendix C). The Y axis represents the percentage of cases where reviewers have determined that a participant's response to a certain textual passage indicates a given foregrounding stage. The vertical lines represent standard errors.

There is a significant difference in the foregrounding profiles between high and low aesthetic appraisal ($F(4,279)=3.6325$, $p=.00664$). A t-test showed that the difference was due to the fact that more participants with high aesthetic appraisal achieved full foregrounding ($t=-3.365$ $p=0.0009$), and more participants with low aesthetic appraisal achieved failed foregrounding ($t=2.555$, $p=0.011$).

Theories Supported by the Findings

The failed foregrounding model

These findings support the proposed model for the following reasons. *First*, the very relationship between the distribution of the foregrounding profiles and the aesthetic appraisal rankings shows that this distribution is not random, but related to the readers' aesthetic appraisal for the whole story. This means that the issue of failed foregrounding is not related exclusively to reading comprehension, but has literary-aesthetic implications. Since there is significant difference between the patterns, it is possible to characterize the foregrounding elements associated with either high or low aesthetic appraisal.

Second, readers who reported high aesthetic appraisal for the whole story differed from readers who reported low appraisal in that they achieved full foregrounding more frequently in the key points. This can mean one of two things: that full foregrounding on the local level contributes to positive aesthetic appraisal on the global level, or that positive appraisal on the global level contributes to the ability of completing the foregrounding process successfully on the local level.

Note that it is too early to determine whether something special in the key points makes them affect the global aesthetic appraisal more than other points in the text; this issue requires further study. If, however, it turns out that key points have greater effect on overall appraisal, this would support one of the claims of foregrounding theory regarding the distinction between the foreground and background of the text. If the key points represents the foreground, they may be expected to have a great effect on the interpretive move of the story as well as on its global aesthetic appraisal.

In addition, it appears that partial foregrounding plays little role in producing global positive appraisal. Although partial foregrounding in the key points was higher also among those with global positive appraisal, this was to a small and non-significant extent. Accordingly, it appears that full foregrounding is the main element that contributes to high global aesthetic appraisal.

Third, readers reporting low aesthetic appraisal for the whole story differed from readers with high appraisal in achieving failed foregrounding more frequently. This means that failed foregrounding in the key points disrupted them from forming high appraisal for the whole story, or low appraisal disrupted them from successfully completing foregrounding processes in the key points. Conversely, the rate of shallow processing did not differ significantly between the groups. This may suggest that failed foregrounding is associated with global negative aesthetic appraisal much more than shallow processing, a finding consistent with the model described above. As you may recall, shallow processing is a case where the reader ignores the difficulty and does not try to delve into the disturbance, whereas failed foregrounding is a failed attempt to delve into the problem, leading to greater semantic noise than the initial disturbance. Thus, failed foregrounding is predicted to disrupt aesthetic appraisal more than shallow processing.

The standard model

The standard model is partly supported here, in that when the full foregrounding process does occur in the key points, it is related to global positive aesthetic appraisal. For readers high in aesthetic appraisal, it was quite common, representing almost 30% of the cases – more than twice the frequency among those low in aesthetic appraisal.

The radical aesthetician position

Recall that the most important stages in radical aesthetic appreciation are failed and partial foregrounding. Thus, should that position prove more correct, the essence of the literary encounter seems to be associated with failed and partial foregrounding. In the case of the present comparison, these two stages should be linked with positive aesthetic appraisal for the entire story, and this is true particularly of failed foregrounding, that reflects this position more purely.

The radical aesthetician would argue that the local difficulty experienced by the reader in the story's key points needs to lead precisely to an effect of positive appraisal of the story as a whole. This position is not supported by this pattern of findings, in two respects. First, in that partial foregrounding is not significantly related to positive aesthetic appraisal. Second – and this is the more significant finding from the point of view of the radical aesthetician – in that failed foregrounding is related to negative global aesthetic appraisal. In other words, cases of comprehension difficulty, confusion and futile interpretive effort were not related to positive global aesthetic appraisal, as predicted by the radical aesthetician position– in fact, the opposite was found.

Despite these findings, the radical position may still be defended. It is possible be that the effect of radical reading is not aesthetic in the same sense as measured in the aesthetic appraisal questionnaire. Although the questionnaire is rather broad, it may have been ill suited for measuring the type of pleasure derived from radical reading. As mentioned in Section 1, in *The Pleasure of the Text*, Roland Barthes (1975) distinguishes between two types of textual effects: *plaisir* (“pleasure”) and *jouissance* (“bliss”). The first represent the “bourgeois” reading experience – convenient and satisfactory reading of a text derived from and connected to culture. The second, however, is a different type of experience, closer to the Lacanian *jouissance* – a more radical reading, perverse and even masochistic; this reading is not pleasant, but rather unsettling, like “hot metal”. It is beyond any communication, shakes the very foundations of the reader’s psyche, transforming his attitude to language. Similarly to Barthes’ distinction, one can argue that the aesthetic questionnaire used in this experiment measures “pleasure” from text, but not “bliss”. If so, then the experiment still has a way of testing the claims of the radical aesthetician. In this case, experience in reading literature would be the index that could support it, since it may be expected that beginner readers would only tend to derive pleasure from the text, whereas more experienced readers have already developed the ability to derive bliss. Only the latter can dwell on the area of confusion and misunderstanding without hastening to flee, either by reaching some kind of conclusion or by giving up in advance and selecting shallow processing. Accordingly, the prediction of the radical aesthetician could be as follows: readers low on literary experience would be higher in full foregrounding, as it measures simple and less sophisticated “pleasure”. Conversely, experienced readers would have a stronger tendency for failed and partial foregrounding, as they are mature enough as readers to experience bliss in a way that involves a sense of

misunderstanding and a struggle to create meaning. The next section will examine whether these predictions have been substantiated by the current research.

The Distribution of Foregrounding Profiles by Reading Experience

Comparing foregrounding profiles between readers with high vs. low reading experience shows a significant effect of experience ($F(4,279)=3.392, p=0.0099$). A t-test indicates that the difference is due to the fact that experienced readers have frequently attained full foregrounding, while inexperienced ones have frequently attained shallow processing ($t=2.09, p=0.0071$; $t=-2.651, p=0.0085$, respectively).

These findings, presented in Figure 10, support the failed foregrounding model in several respects. First, the very fact that there is a relationship between the distribution of the foregrounding profiles and experience in reading literature serves as further indication that this distribution is not random and is meaningful in literature research.

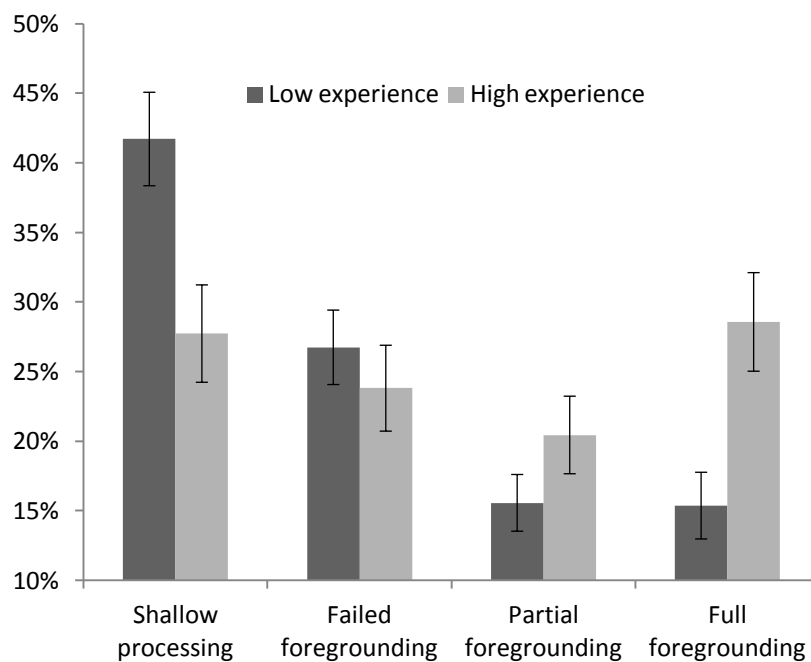


Figure 10: The distribution of foregrounding profiles according to experience in reading literature. Experience was measured using an author name recognition questionnaire (see Appendix D). The Y axis represents the percentage of cases where reviewers have determined that a participant's response to a certain textual passage indicates a given foregrounding stage. The vertical lines represent standard errors.

Second, the differences between experienced and inexperienced readers were found on both poles of the foregrounding process. Experienced readers attained full foregrounding more often, while inexperienced readers initiated foregrounding less frequently, opting for a shallow processing strategy. This may be interpreted as

follows: inexperienced readers feared to start the foregrounding “adventure” as it required effort and there was no way of telling whether it would lead to high or low aesthetic appraisal. Apparently, the inexperienced readers followed a conservative but safe strategy: refusing to accept the “invitation to interpret”. Conversely, the experienced readers that took a chance more often and risked starting the foregrounding process frequently managed to reach its successful ending. Thus, their experience apparently enabled both the “courage” to start the foregrounding process and the required interpretive skill to reach its successful end.

The standard model

These findings also point to the superiority of the failed foregrounding model over the standard one. It appears the standard model is more suitable to experienced readers, but less so to readers with little literary experience, who, in more than 40% of the cases, preferred not to initiate foregrounding and settled for shallow processing of the problematic passage.

The radical aesthetician position

This position would have predicted that highly experienced readers be higher in failed and partial foregrounding, as their experience allows them to dwell on areas of miscomprehension, confusion and difficulty. However, there was no significant difference between experienced and inexperienced readers in failed and partial foregrounding. Moreover, when difference was found in full foregrounding, it was in the direction opposite to that predicted by the radical aesthetician position. Recall that the radical aesthetician would argue that textual enjoyment that is not bliss characterizes the unskilled reader, and would therefore expect unskilled readers to reach full foregrounding often. Nevertheless, exactly the opposite was found - that experienced readers are the ones that manage to complete the foregrounding process successfully. This indicates that what experience in literature enabled in this case was not the ability to dwell on the difficulty, but the ability to resolve it with relative ease.

Nevertheless, as in the previous subsection, the radical aesthetician position can still be defended. It may be that radical aesthetic reading is independent of reading experience. Rather, it is a position that develops among certain readers out of a certain predilection that cannot be explained in terms of expertise. Let us try, therefore, to think of the radical aesthetician’s reading as a strategy followed by certain readers regardless of their experience. To test this possibility, we must search for readers who opt for a reading strategy where failed or partial foregrounding is dominant. If such readers can be found, their interviews can be analyzed to determine whether their reading comments are consistent with the radical aesthetician position.

Reading Strategies

The concept of reading strategies may be useful in organizing the findings in another way. The personal foregrounding profile can be used to characterize the participant's preferable reading strategy. There is a certain similarity between what is presented below and the reading strategies described by Vipond and Hunt (1984). Vipond and Hunt describe three reading strategies: information-driven, story-driven and point-driven. Each focuses on a different textual layer. The *information-driven* strategy is designed to reach only basic understanding and to extract all the relevant information from the text. This strategy is common in reading textbooks, newspapers or train schedules, for example. The *story-driven* strategy operates on the plot layer and focuses on constructing the narrative world, characters and their motives, etc. Finally, the *point-driven* strategy operates on the discursive layer in an attempt to understand why the narrator or the author describe the events they describe. The readers using this strategy ask themselves what is the point the narrator or author tries to make. Vipond and Hunt describe several cases in which using the point-driven strategy leads to failure, whereby the reader does not comprehend the narrator or author's intent, experiences the story as pointless and asks, "so what?" In their experiment, where participants read a short story by John Updike, many of the participants experienced such failure, whereas only 5% managed to attribute meaning to the author's intention.

Despite the differences between the two, both the foregrounding profile and Vipond and Hunt's (1984) strategies seek answering the same questions: which readers preferred to remain on the basic level of understanding the story, and which tried to approach it at a higher analytic level?

Cluster analysis

The reading strategies were not predetermined, but arose from the findings using cluster analysis, an exploratory technique that tries to identify structures within the data. It identifies homogenous groups of cases when the grouping is not previously known. The advantage of a cluster analysis is that it can group observations into a series of clusters and help build a taxonomy of groups regardless of prior expectations. Hence, clustering is useful in that it can lead to the discovery of previously unknown groups within the data. This technique is often used in biology, climate studies, medicine, information retrieval, web search, image pattern recognition, etc. (Rajaraman & Ullman, 2011).

Clustering was an adequate solution in this case because there was no specific hypothesis on which kinds of foregrounding profiles would consolidate in to reading strategies. The clustering algorithm used to classify the data point is K-means cluster – a method to quickly cluster large datasets in which the researcher defines the number of clusters in advance. The number of clusters selected was four, because of the four stages in the failed foregrounding model. This way, it was possible for each

stage to be a key factor in a specific reading strategy: a strategy of the group of readers with much shallow processing, a strategy of the group that had much failed foregrounding, and so on. However this was not the case, as shown below: the strategies found did not exactly overlap with the four stages. No strategy was found where partial foregrounding was a central component, and one strategy was found that had no stage as a central component; instead, the stage distribution within it was similar to the general mean.

Strategy A included participants who opted mainly for shallow processing (n=12). Strategy B included participants who opted mainly for failed foregrounding (n=4). Strategy C included participants who acted according to the general average (n=14). Finally, Strategy D included participants who opted mainly for full foregrounding (n=11). The difference between the four groups was significant in all measures: shallow processing [F(3,37)=29.3, p<0.00001]; failed foregrounding [F(3,37)=23.7, p<0.00001]; partial foregrounding [F(3,37)=3, p<0.05]; and full foregrounding [F(3,37)=31.2, p<0.00001]. The different strategies are presented in Figure 11 below.

Reviewing those four strategies suggests several basic observations. First, three out of the four strategies have a dominant element that is much more common than the others are. In A, this is shallow processing (56%); in B, failed foregrounding (54%); and in D, full foregrounding (43%). Second, partial foregrounding is not a dominant element in any strategy. Third, C has no dominant element, but it is rather distributed very similarly to the average distribution in the population (compare to Figure 8). Finally, only four participants used a strategy where the dominant element was failed foregrounding.

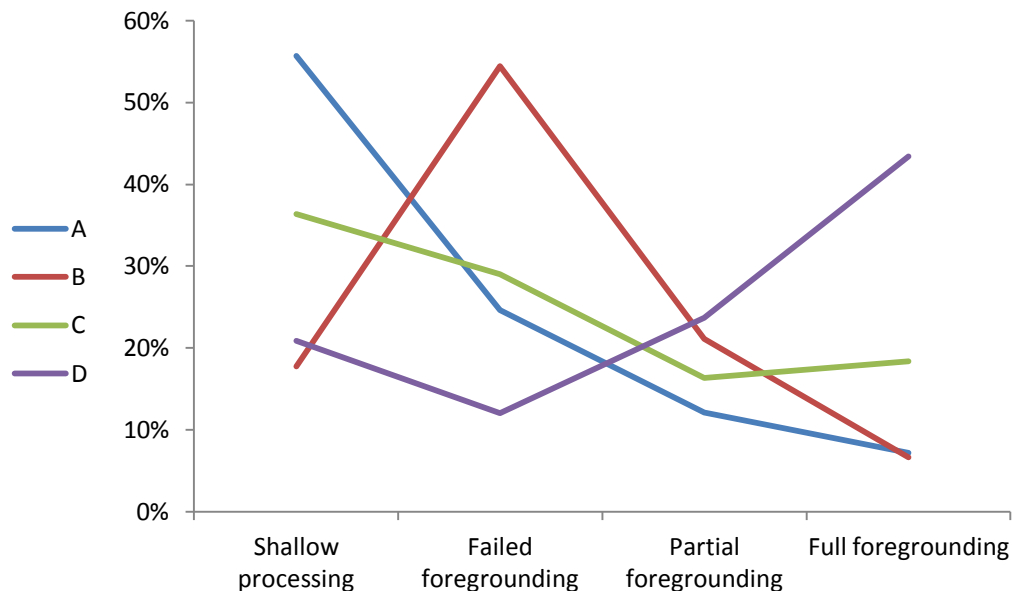


Figure 11: The four strategies found in a cluster analysis of the data according to participants. The Y axis represents the percentage of cases where reviewers have

determined that a participant's response to a certain textual passage indicates a given foregrounding stage.

The participants have shown a general preference for strategies where one of the two polar conditions dominates: shallow processing or full foregrounding. This behavior can be considered as logical given the findings reported above with regard to aesthetic appraisal. Recall that partial foregrounding was not related to positive aesthetic appraisal, whereas failed foregrounding was related to negative appraisal. Therefore, it is reasonable to assume that participants avoided reading strategies where these two elements of the foregrounding process played a central role. Namely, they preferred strategies where they rarely started the foregrounding process, or completed it often, while the rest of the participants, who had no particular preference for any of the stages, acted according to the general average.

The extent to which the strategies described here are controllable is unclear, since part of the ability to end a successful foregrounding process has to do with the experience in reading literature and the difficulty of the text itself. Nevertheless, this pattern of findings suggests that at least the ability to initiate a foregrounding process could be partly controllable. One reason for assuming that is the small number of failed foregrounding participants. This may be due to participants preferring shallow processing in order to avoid the frustration and semantic noise involved in recurring failure in the foregrounding process. In the Discussion, I will further elaborate on this issue.

The radical aesthetician position

The fact that only few readers opted for strategies where failed or partial foregrounding were a central element does not support that position. No reader opted for a strategy where partial foregrounding was a central element, and only four opted for a strategy where failed foregrounding was central. On the other hand, the existence of these four participants allows us to take a closer look at the failed foregrounding experience. In-depth analysis of the interviews held with them could shed light on whether theirs was a more radical literary-aesthetic experience – whether they experienced “bliss”.

An examination of the interviews with those four participants paints a picture of a restless reading. They described the story as difficult and pointless. They experienced an impressive variety of semantic noises. In their descriptions, they often referred to confusion, misunderstanding and frustration, stating that many passages in the story “say nothing” and that “the words are not connected”. Some of their comments attested to failures in the point-driven reading strategy, such as “I don’t understand what they want to say here”, or “who cares?” They also made frequent associations to books and films that only made it more difficult for them to understand the story. Some of these were quite random and due to alliteration. One reader kept reading

multi-meaningful words in their wrong meaning. Another found it difficult to concentrate because she often felt a sense of familiarity and tried but could not remember why things in the story were familiar to her. Still another reader kept applying interpretive principles that only made it difficult for her to understand the story. This graduate literature student had recently participated in a course about folktales. Her reading was affected by her search for typological numbers, in an attempt to understand “the narrator’s ideological position”, a pointless search for allegorical meanings, and what seemed to her as unjustified deviations of the text from the folktale norms. She was also frustrated by the lack of cultural knowledge that prevented her from understanding the story and said: “I feel I had lots of hypotheses and they didn’t help me with anything”.

The interviews show that the four readers with the failed foregrounding strategy have experienced multiple disturbances and noises in their reading, which has proven dissatisfactory despite the considerable effort. Accordingly, it is easy to understand why so few readers have adopted this strategy, as it seems to have been highly frustrating. It is doubtful whether this was the reading experience Barthes referred to when he wrote about “bliss”.

Summary of Section 2

The failed foregrounding model. The findings hitherto support the failed foregrounding model in several respects. First, in its ability to make statistically significant distinctions. Second, in that its predictions gain more support than those of its two rivals: the standard model and the radical aesthetician position.

We have shown that the proposed model accounts for a series of observations. Positive aesthetic appraisal for the whole story was found related to full foregrounding in the key points and negative appraisal was found related to failed foregrounding. This is in line with the model’s prediction that failed foregrounding would have a more negative effect on readers’ aesthetic experience than shallow processing, and that full foregrounding would have a more positive influence than partial foregrounding.

We have also found that experienced readers often attain full foregrounding while inexperienced ones often opt for shallow processing. In addition, in-depth examination of reading strategies has revealed preference for strategies in which shallow processing or full foregrounding are central. Thus, it appears that in general, the participants have preferred the two polar positions of the model – shallow processing and full foregrounding. These findings have implications for both the standard model and radical aesthetician position,

The standard model proved successful in that full foregrounding in the key points did predict positive aesthetic appraisal for the whole story. Its weakness was,

however, in that the case it described was relatively rare. Even among the participant with the highest full foregrounding scores, its frequency was around 30-40%. In other words, even there most of the cases were of various failures in the foregrounding process. Thus, it seems that failures in the foregrounding process, particularly shallow processing, are not the exception, but are rather integral to the process of reading a literary text by real-life readers. The standard model assumes smooth passage through the three foregrounding stages, but these findings show that successful conclusion of the process is not the common case. It would therefore be more accurate to think of foregrounding not as a situation but as a distribution of situations. This distribution was found sensitive to the level of aesthetic appraisal, the readers' experience and their reading strategies.

The radical aesthetician position was also examined using the new model. This position argues that literary aesthetic experience is the result of a failed struggle against the text's incomprehensibility. According to this approach, failed and partial foregrounding should play a key role in aesthetic appraisal or at the very least, their frequency would increase with the reader's experience, and that the frequency of full foregrounding should drop. None of these predictions was supported. Even the search for readers who adopt reading strategies where those two elements are central was fruitless. Only four readers were found who opted for a reading strategy where failed foregrounding was central, and in-depth examination of their interviews did not support the idea that theirs was a "radical" literary experience, but rather that they mainly stumbled across semantic noises of various kinds. It was not my impression that they experienced "bliss" while reading the text – precisely the opposite: they were highly frustrated by it.

This section focused mainly on the readers, their experience while reading the text, their expertise in reading literature, their reading strategies, and how all these related to the foregrounding process. The following will emphasize the aspects of the text. I will present an analysis of the distribution of foregrounding according to various stylistic devices. It will be a relatively preliminary analysis, but will indicate the potential contribution of the model to the study of literary style. This will be followed by a methodological section about RTA: the main technique in which the findings of this dissertation have been collected. After that I will then summarize and discuss the model's implications, inquire as to the degree it reflects controllable behavior, and the factors that encourage initiating the foregrounding process and its successful conclusions. Finally, I will try to explain why a similar model has not been developed in the past, and detail some of the model's limitations.

SECTION 3: FAILED FOREGROUNDING AND THE TEXT

The previous section described the failed foregrounding model using an analysis of readers' responses to Jorge Luis Borges' "The Chamber of Statues". The comparisons were according to reader characteristics: aesthetic appraisal, reading experience, or reading strategy. Such comparisons are important as they can be used to develop a general literature reading model or to compare the predictions of different models.

Another type of comparison is possible, based on the text rather than the reader. Readers' responses to a specific passage may be examined, or several passages may be analyzed jointly and compared to others that differ in a certain key aspect. It would seem that the potential number of such comparisons is huge, limited only by the researchers' creativity.

This section presents an analysis of the foregrounding characteristics of the eight key points in the text, based on a stylistic classification, with emphasis on effectiveness in completing the foregrounding process.

Foregrounding Distributions According to Stylistic Devices

The failed foregrounding model may be used to examine the effectiveness of various stylistic devices. Namely, one can examine which stylistic device often lead to a full foreground effect, and which often lead to failures in the process. Previous studies offer some findings that support this line of inquiry. For example, Emmott et al. (2006) found that formal devices such as italics, clefting, short sentences or independent clauses attract the reader's attention, while content-based devices do not. Emmott's findings, however, are limited to the model's first stage, and it is impossible to tell, without additional data, whether the formal stylistic devices only attract attention or also produce a literary aesthetic effect.

Conversely, Miall and Kuiken's (1994) findings can be used to examine the completing of the foregrounding process. Miall and Kuiken themselves did not try to answer the effectiveness question, since they sought to prove the aggregate effect of all foregrounding devices. Alternative analysis of their findings, however, provides some insight into this question as well. Miall and Kuiken performed four experiments, and in each presented correlations between foregrounding characteristics and affect, strikingness, and reading speed. Some of these correlations were statistically significant and some were not. Table 1 below presents the number of significant correlations found in all four experiments. This table summarizes the strength of the evidence found for each of the various foregrounding devices. This will reveal an interesting trend regarding the effectiveness of foregrounding devices.

	Semantic Foregrounding	Grammatical Foregrounding	Phonetic Foregrounding
Affect	4	0	6
Strikingness	6	0	7
Speed	6	2	4
<i>Total</i>	16	2	17

Table 1: The number of significant correlation found in all experiments by Miall and Kuiken (1994), according to three types of foregrounding devices. This comparison helps find differences in the effectiveness of various foregrounding devices.

Table 1 indicates that grammatical foregrounding is different from both semantic and phonetic foregrounding in terms of effectiveness. While the latter two attained a similar number of significant correlations in Miall and Kuiken’s study (1994) – 16 and 17, respectively – the relation between grammatical foregrounding and the experimental indicators was significant in only two cases, and was thus much weaker than either semantic or phonetic foregrounding in its effect on affect, strikingness or reading speed.

Other studies also found different degrees of effectiveness for different foregrounding devices. Sopčák (2007) studied readers’ responses to foregrounding devices in Joyce’s drafts, and found a general effect on strikingness and affect. Analysis according to the various levels, however, revealed that semantic foregrounding was the most effective, since only it had a significant effect on *both* indices. Grammatical foregrounding did not affect strikingness and phonological foregrounding did not affect affect. This finding regarding the centrality of semantic foregrounding is also consistent with Van Peer (1986), who found that semantic devices played a much stronger role than grammatical ones, which in turned produced stronger effects than phonological ones. Thus, although foregrounding researchers have long been aware of the differential effectiveness of foregrounding devices, this finding has not received comprehensive theoretical attention.

Like Emmott et al. (2006) and Miall and Kuiken (1994), the findings of this experiment also indicate differential effectiveness of stylistic devices. The graph below presents the foregrounding profiles for three types of stylistic devices: linguistic difficulty, author comments, and figurative descriptions (see Appendix E for a detailed listing of stylistic devices). I focused on those three devices since they were common in the key points in the text on which many of the readers dwelled and regarding to which they had important things to say in order to explain that dwelling.³

³ Some of the stylistic devices examined recurred elsewhere in the text. Not every author comment or figurative descriptions became a key point in the text. The comment “salvation be with both of them”

In retrospect, it seems that each of these devices also presents a different textual layer or reading strategy based on Vipond and Hunt's (1984) strategy classification. Linguistic difficulty encourages an information-driven strategy. Figurative description requires the reader to deal with the story world layer. While author comments direct readers to the discursive layer, where they are the author's interlocutors. The author's comments can raise the question, what does the author want to say, or what is the point she is trying to make.

Since reading experience has a significant effect on the foregrounding profile, I will present the foregrounding distribution of the three stylistic devices in two complementary ways: for all participants (Figure 12), and according to the participants' reading experience (Figures 13-15).

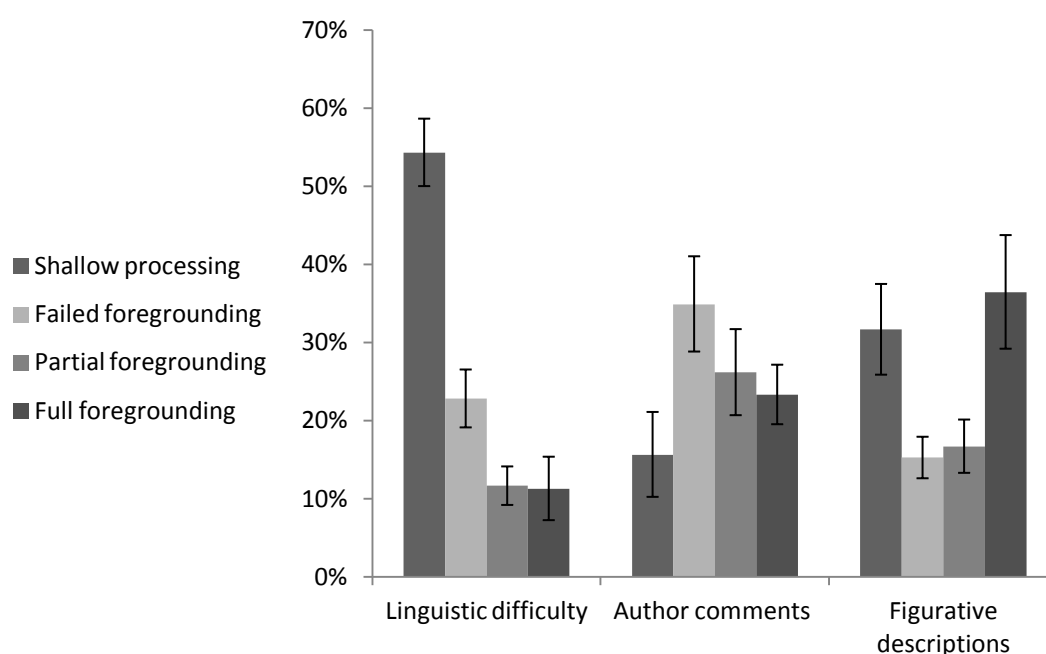


Figure 12: Foregrounding profiles for three types of stylistic devices. The overall difference between the three groups is significant ($F(6,16)=7.5, p=0.0006$). Local difference exists in three out of the four profiles: shallow processing (0.0001), failed foregrounding ($p=0.03$), and full foregrounding ($p=0.01$). The vertical lines represent standard errors.

was repeated twice, but most participants dwelled only on the first comment. Repetition may have dulled the sting of this innovation.

One figurative description also did not become a key point, that of the silence in the chamber: “while not a single human voice or clarion sound was heard”. It may be that the description’s negative character made it less striking, since in this case the figure does not create an image of something but of its lack. These are but conjectures, since we did not examine when or why a certain point in the text became a key point, but rather examined the three stylistic devices found within the key points themselves.

Figurative Descriptions

Figurative descriptions appear in three passages, where a certain physical element is described (long room, warrior statues, and a marvelous mirror), and the description includes action by a character that makes it highly visual. I called this group “figurative descriptions”, since they create secondary figures through which something is said about the key object. For example, the story mentions a long room described as follows: “The seventh appeared empty, and it was so long that the ablest of archers, had he loosed an arrow from its doorway, would not have hit its end”. Here, the figure – the able archer shooting an arrow from the doorway to the distant wall – is used to describe the room and emphasize its length. Similarly to this example, other figurative descriptions also involve a lifeless element described through the action of a live human. Just as the room is described through the archer, the statues are described through a blind man’s touch, and the mirror by what “whoso looked might see”.

Figure 12 presents the foregrounding profiles for figurative descriptions for all readers. It was found that passages including figurative descriptions were the highest in full foregrounding (36%). The foregrounding profiles formed a U shape, indicating that these passages were high in both shallow processing and full foregrounding, and that the intermediate stages were infrequent. A strong experience effect can be seen in Figure 13, when distinguishing between experienced and inexperienced readers, the U shape almost disappears.

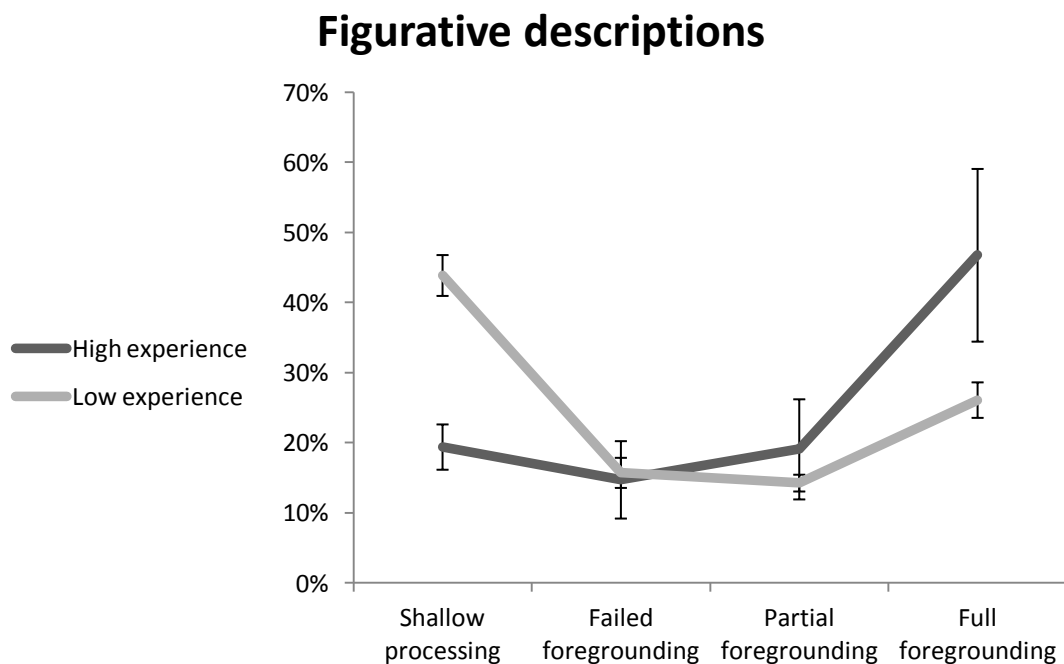


Figure 13: Foregrounding profiles for figurative descriptions by reader experience. The vertical lines represent standard errors.

Figure 13 presents the foregrounding profiles for figurative descriptions by reader experience. When distinguishing between experienced and inexperienced readers, it becomes clear that the tendency for full foregrounding is typical of experienced readers, while the tendency for shallow processing is typical of inexperienced ones. It was easy for experienced readers to initiate a foregrounding process in these descriptions (only 19% shallow processing) and complete it successfully (46.7%). Namely, for experienced readers it was both easy to identify the literary importance of the figurative descriptions and interpret them. Inexperienced readers, however, found it much more difficult to initiate foregrounding (44% shallow processing) and to complete it (only 26% attained full foregrounding). In summary, the figurative descriptions were the most effective passages in terms of foregrounding, but this was highly affected by experience.

Author Comments

Author comments appear in two passages in the story, where the narrative sequence is disrupted and the author comments about the characters, expressing a personal opinion – a blessing or a curse. These comments are part of Borges' attempt to create an *Arabian Nights*-style narrator. For example, when we are told that the king opened the forbidden gate with his right hand, the author wishes parenthetically that the hand may “burn through all eternity”. Comments of this kind made the readers think, among other things, of the author as a folk storyteller, about the text's authenticity, the relation between the author and the characters, and about irony.

Figure 12 presents the foregrounding profiles for author comments for all readers. Passages that included such comments were higher in failed foregrounding than others (35% for author comments, compared to 23% for linguistic difficulty and 15% for figurative descriptions). It was also found that the degree of shallow processing was the lowest of the other types of passages (16%). This pattern of results shows that it was easy for the participants to identify that these passages were literarily important and that they therefore tried to delve into them and moved to the second foregrounding stage. Completing the foregrounding process was relatively challenging, however, and relatively few readers managed it (26%).

Figure 14 presents the foregrounding profiles for author comments by reading experience. Readers' experience was found to have a non-significant effect on the rate of failed foregrounding, although experience did slightly change the overall trend reported earlier. As mentioned, in general shallow processing was relatively low here; it was lower than the general mean for inexperienced readers (24%), and quite rare for experienced ones (7%). Experienced readers were higher than inexperienced ones in their ability to complete the foregrounding process (29% vs. 18%). Thus, although it was easier for experienced readers to initiate and complete the foregrounding process,

the overall trend was identical, and the unique characteristic of author comments – the high degree of failed foregrounding – was experience independent.

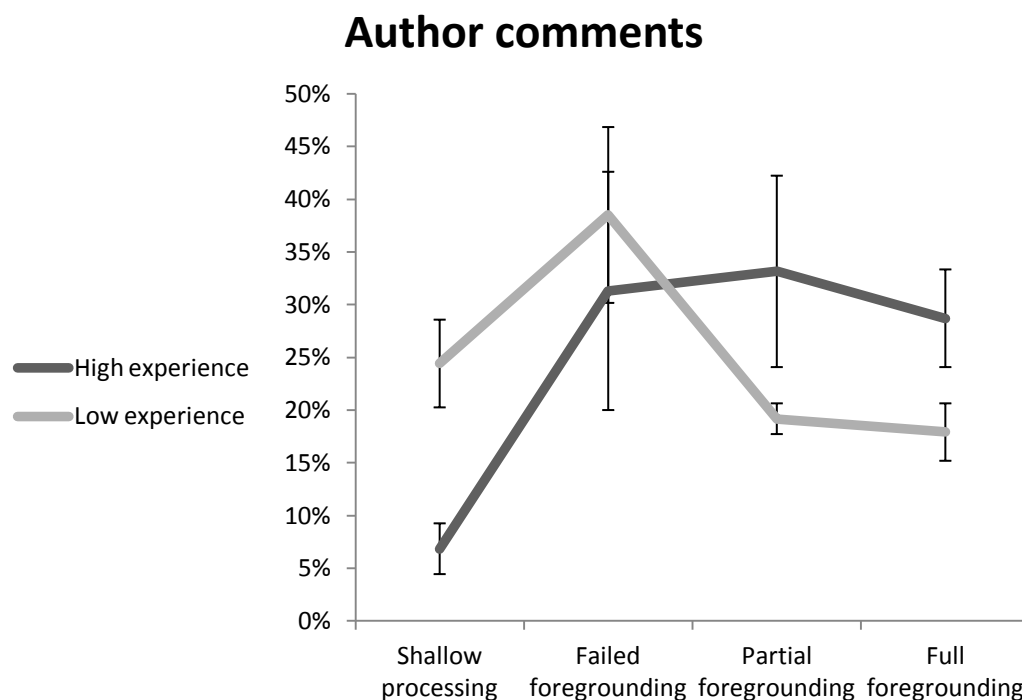


Figure 14: Foregrounding profiles for author comments by reading experience. The vertical lines represent standard errors.

Linguistic Difficulty

Linguistic difficulty appears in three passages: a foreign word whose meaning was not apparent to most readers; a very long sentence with multiple commas whose beginning and end could not be easily connected; and a particularly complex passage with three foreign words and a somewhat unusual syntactic structure that raised pragmatic difficulty. Appearing in the beginning of the story, the latter case involved particularly great difficulty, as attested by the eye movements. When the city where the plot unfolds is mentioned, its name is said to be “Labtayt, or Ceuta or Jaén”. This sentence proved difficult for the readers both due to the foreignness of the sounds to the Hebrew ear, and because of the unusual syntactic structure of using the word “or” twice in a row in reference to a place name. Thus, the readers found it difficult to understand both how to read the cities’ names and to decide which of the possibilities was indeed its name.

Figure 12 presents the foregrounding profiles for linguistic difficulty for all readers. It was found that passages with linguistic difficulty were higher in shallow processing than others (54.3%). It appears that this stylistic device made it difficult for most

readers to initiate the foregrounding process, and only few readers managed to attain full foregrounding (11.2%).

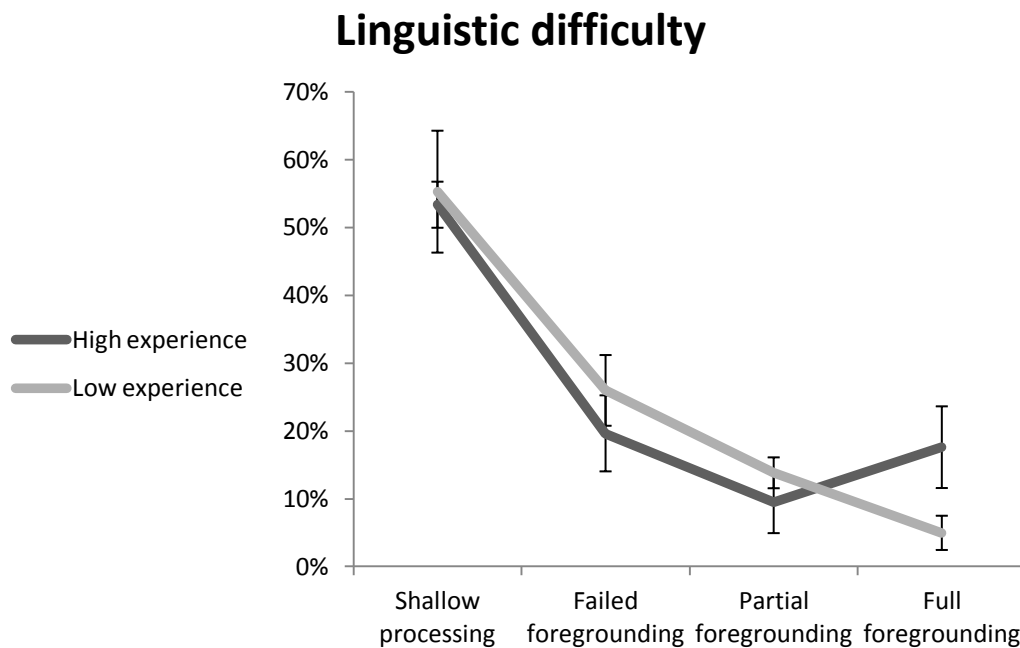


Figure 15: Foregrounding profiles for linguistic difficulty by reading experience. The vertical lines represent standard errors.

Figure 15 presents foregrounding profiles for linguistic difficulty by reading experience. It was found that the effect of experience on passages with linguistic difficulties was small. Both experienced and inexperienced readers found it very difficult to initiate the foregrounding process and had more than 50% shallow processing. A significant difference was found in the ability to complete the foregrounding process, however. Inexperienced readers found it very difficult to resolve the linguistic difficulty and achieve positive aesthetic appraisal (5% full foregrounding). Experienced readers, on the other hand, managed to do so at a much higher frequency (17.5%). Nevertheless, both still represent low frequency of full foregrounding compared to the other stylistic devices. In summary, linguistic difficulty was the least effective, and the effect of reading experience was limited to the completion of the foregrounding process, without significantly affecting the other stages.

It appears that there is a connection between the finding that linguistic difficulty encourages shallow processing and Miall and Kuiken's (1994) finding regarding the ineffectiveness of grammatical foregrounding. Not because these are the same stylistic devices, but because in both cases the ineffectiveness of the foregrounding process is related to a difficulty in the relatively basic layer of linguistic processing, one where the very beginning of the sense-making process is disrupted. I will elaborate on a possible implication of this finding in the Discussion.

Interaction with Reading Experience

The interaction between experience in reading literature and the stylistic characteristics was not significant ($F(6, 16)=1, p=0.44$): the same general trend observed for experience emerged with regard to stylistic devices as well: experienced readers both entered the foregrounding process and completed it more often. There was only one case of deviation from this trend – that of linguistic difficulty. Linguistic difficulty was the only stylistic device where experience in reading literature did not affect the tendency to enter the foregrounding process. The percentage of shallow processing was independent of experience – experience contributed only to completing, but not to initiating the foregrounding process.

This finding may suggest that linguistic difficulty, at least of the type examined here, could have a special status that differentiates it from the other two stylistic devices. Is it the case that linguistic difficulty makes the entry into the foregrounding process particularly unlikely? Is this difficulty more resilient to experience in reading literature than other stylistic devices? To determine whether this is the case, further research is required. This is the first indication of potential interaction between textual and reader-driven factors. Such interaction, if found, will show that the foregrounding profile is sensitive not only to either text- or reader-related parameters, but also to their interaction.

Summary and Discussion

An examination of three stylistic devices – figurative descriptions, author comments and linguistic difficulty – found that the distribution of foregrounding profiles was not independent of these devices. Each had a significantly different foregrounding distribution, also affected by the reader's experience.

Figurative descriptions were the most effective in reaching full foregrounding. The author comments' literary importance was the easiest for the readers to recognize, and were very low on shallow processing. Nevertheless, they made it difficult for readers to complete the process with many halting in failed foregrounding. Linguistic difficulty was the least effective stylistic device: the readers found it difficult to both start the foregrounding process and complete it. Moreover, linguistic difficulty was the stylistic devices least sensitive to the reader's experience.

Effectiveness of Stylistic Devices

The three stylistic devices differed in terms of their effectiveness in initiating foregrounding. The question when a text would lead to a positive aesthetic experience and when to semantic noise is, in a certain sense, the proverbial holy grail of the failed foregrounding research agenda. It would be tempting to approach this issue as a

natural scientist would: to create a taxonomy of all known stylistic devices – not only the three studied here – and examine, for each of them, how effective they are in initiating successful foregrounding. This would be a huge task requiring enormous resources, but even if completed, its theoretical importance would be limited. Since the most important questions are what this effectiveness is made of, and what factors affect it.

The new model suggests a general division into two types of effectiveness: (1) Effectiveness in introducing the reader into a foregrounding process; and (2) Effectiveness in bringing the reader to successful conclusion of the process. This division raises a series of questions. First, what factors affect each type of effectiveness? Second, are the two types related or independent? Third, how do those factors interact with additional, non-textual parameters such as reading experience and strategy? Below, I suggest several directions for finding the answers.

Textual elements that promote the initiation of foregrounding

I hypothesize that the layer where the initial difficulty is experienced has an important role in determining the devices' effectiveness. When the difficulty was in the basic layer of linguistic processing, most readers tended to resolve it within the confines of that layer, with few going into literary interpretation. When they came across unfamiliar foreign words, most readers struggled to understand “what” was said, rather than “why”. This notion is consistent with Miall and Kuiken’s finding (1994) that grammatical foregrounding is ineffective in causing strikingness effects. Grammatical foregrounding also occurs in the basic layer of linguistic processing, and makes it difficult, primarily, to interpret the very “surface” of the text. Or, in the terms of Vipond and Hunt’s reading strategies (1984), linguistic difficulty sensitizes the readers to an information-driven strategy.

Conversely, in the case of an author comment it is easier for readers to identify literary importance. Author comments sensitizes the readers to a point-driven reading strategy that makes them think about the discursive layer, about the author or the narrator and what they want to say. Thus, deviations occurring in the discursive level are seen by the readers as invitations to interpretation, leading them to initiate a foregrounding process, albeit often ending in failure.

Textual elements that promote successful foregrounding

A major factor promoting the successful conclusion of the foregrounding process appears to be the ease in which the difficulty may be resolved. Often, readers who achieved full foregrounding did so by relying on extra-textual knowledge: familiarity with literary genres and conventions, historical background, other literary works, etc. Some deviations were difficult to resolve – riddles that required rare literary

knowledge, such as acquaintance with Borges or Italo Calvino, while others could be interpreted by referring to familiar texts such as Harry Potter, the Bible, or Disney movies.

Nevertheless, familiarity with the required texts is no guarantee of successful interpretation. It is an essential but not a sufficient condition. Another potentially relevant factor is the degree of textual integration required to signify the textual deviation. The greater the integration required, the more challenging the completion of the foregrounding process. Some passages require integration with yet-unread passages to be understood. Such integration is only possible in a second reading; therefore, in these cases it would be difficult to achieve full foregrounding upon first reading, even if the literary importance of the given passage is clear to the reader. Accordingly, the reader can enter a foregrounding process one cannot complete upon first reading.

Figurative descriptions were characterized by a high degree of full foregrounding. Perhaps the very use of a figure for descriptive purposes constitutes a clear invitation for entering an interpretive process. The relative ease of completing the process could be due to that it doesn't demand too much from the reader. No integration of far away text segments is needed, because the source and the figure are close to each other. No point driven strategy is needed because the difficulty can be solved within the story-world. It may also be that something in the sensual quality of the visual image makes completing the process easier for the readers. A review of the interviews indicates that this also applies to passages leading to emotional response or even laughter. It may be that the emotional and sensory response is more immediate, more accessible to the readers, therefore leading without considerable effort to full foregrounding.

The reader's influence on initiating and completing foregrounding

General positive aesthetic appraisal was correlated with successful foregrounding in the key points. The causal direction of this finding is unclear. The more obvious possibility is that full foregrounding in the key points contributes to global aesthetic appraisal, but the opposite direction is also conceivable. It may also be that the causal relation is circular – that is, that general positive appraisal of the story helped readers address the difficulty in the specific passage by helping them understand that it has some literary importance. Conversely, the successful experience in that specific passage may have contributed to general appreciation of the story, affecting in turn the ability to successfully complete a foregrounding process when struggling with the difficulty in the next key point.

Experience in reading literature had a significant effect on the tendency to enter the foregrounding process, as well as to complete it successfully. This finding was obtained both out of questionnaire analysis and out of interview analysis – i.e., on both the global level of reading the story and on the local level, the key points. There

was only one exception to this rule: in cases of linguistic difficulty, experience did not affect the initiation of foregrounding (see Figure 15). The significance of this finding is yet unclear, nor is it certain whether it may be generalized to other cases of different texts and readers. It is possible that this is due to the layer of difficulty. The linguistic difficulty may have attracted readers' attention to the basic layer of linguistic processing. Thus, even though they did identify the deviation itself, they failed to identify it as having literary significance. To the extent that they did identify that the difficulty had literary significance, experience certainly helped them resolve it.

The Limitations of Analysis by Stylistic Devices

The distinction made here between different stylistic devices was relatively crude. These are not predetermined categories nor independent variables that the experiment had been especially designed to examine. "Linguistic difficulty", for example, is a category that can include a much broader range of devices than those found in this specific text. Had the experiment been designed especially to examine cases of linguistic difficulty, it would have been constructed otherwise, so as to examine this variable more systematically. The experiment was not focused on the effect of various linguistic difficulties on the reader, however, but was rather an experiment in the first reading of a short story, examining readers' explanations for their attention foci. Thus, the analysis presented here is a post-hoc move of following the data, namely, following the interaction between Borges' particular text and the experiment's particular participants. Hence, the ability to generalize from these findings to all cases of "linguistic difficulty", or "figurative description" or "author comments", for that matter, is relatively limited. On the other hand, these preliminary findings can certainly form the basis of more detailed future studies.

A productive guideline for a more systematic study of stylistic devices and the types of failure they may cause the reader may be found in Castiglione (2017). His linguistic-aspects-of-difficulty (LAD) model describes 20 different aspects of linguistic difficulty occurring in complex poetry that may cause the reader to delay in his reading. Castiglione distinguishes between five levels of processing that could be disrupted by linguistic difficulty, at a rising order of disruptiveness:

1. **Word recognition:** Interpreting a string of letters as an existing word.
2. **Decoding:** Assigning meanings to words.
3. **Parsing:** Assigning thematic roles and grammatical class to word.
4. **Integrating:** Building a global mental representation for the text.
5. **Inferencing:** Making sense of aesthetically foregrounded profiles through an interpretive act.

The LAD model predicts a relationship between specific aspects of linguistic difficulties and their effects on specific processing levels. For example, rare words, first names or words unfamiliar to the reader may affect the second level, decoding, whereas a change of tone or perspective may affect the fourth, integrating. Based on these predictions, it is possible to surmise which types of linguistic difficulty would be more effective as invitations for interpretation: those that cause difficulty at higher processing levels.

The findings of this experiment may be interpreted as consistent with the LAD model's predictions. The linguistic difficulty described in the experiment is expected to affect the second level, and therefore not to serve as an effective invitation for interpretation. While the author comments and figurative descriptions were expected to cause difficulty on the fourth or fifth levels, and were therefore effective invitations for interpretations. The LAD model, or other linguistic models like it, can serve as a basis for more subtle distinctions than those made here, or those usually made in foregrounding studies.

Accordingly, the analytical method presented in this section marks the target of describing the degrees of effectiveness of various stylistic devices and provides with the tool with which to perform such an inquiry: the failed foregrounding model. This is a highly ambitious target: in order to properly carry out this project, future studies would have to be more focused on the issue of the effectiveness of stylistic devices, and be informed by more detailed linguistic models.

SECTION 4: METHODOLOGY – RETROSPECTIVE THINK-ALLOUD

This section is dedicated to the methodology that enabled the development of the model presented in the previous sections. I preferred separating the presentation of the model and the methodology for several reasons.

The first reason is rhetorical. The main innovation of this dissertation is theoretical: the failed foregrounding model. It was therefore presented early on, without burdening the readers with methodological details that could distract them from the main point. The division into two sections was therefore designed to enable the readers to focus on each of these issues separately and to devote to each the attention they require.

Second, the use of a retrospective interview informed by the eye movement findings is relatively unusual in the current empirical research of literature. Introspection used to be a highly common method in the field, but for various reasons researchers have become increasingly suspicious of introspective verbal information, hence the need to expand on that method, substantiate it and justify it as reliable.

The main suggestion of this section is that a technique that combines an interview with the eye-movement presentation is an effective way of obtaining rich and accurate verbal information about the conscious aspect of the reading experience. This technique has several names. Other names include stimulated retrospective think-aloud, cued retrospective reporting, retrospective testing, retrospective protocol, retrospective report, think after, post-experience eye tracking protocol (PEEP), and post-task testing (Hyrskykari, Ovaska, Majaranta, Rähkä, & Lehtinen, 2008). In what follows, it is referred to mainly as retrospective think-aloud (RTA).

Compared to other methods of collecting verbal information, RTA is unique in that it uses eye movements as recall cues for participants. It is commonly applied in website usability⁴ or user interface studies, and several studies have established its validity in that area (Guan, Lee, Cuddihy & Ramey, 2006; Mayhew, 2017). To the best of my knowledge, this is the first time it is used in an empirical study of a literary text.

Part of the uniqueness of this research method lies in the way it integrates two very different types of information: verbal information, which is essentially “soft”, very rich in meaning and difficult to quantify; and eye movements – a “hard” physiological, quantitative indicator, highly precise in time and space but indifferent to the reader’s experience. This section presents how this method combines these two types of information and in doing so, solves some problems for collecting or analyzing both types of data. The combination proposed here is also one between methodological “fashions” from different periods, since the use of verbal information

⁴ Usability labs improve ease-of-use during the design process of products, and are especially common in developing the user interface of websites.

was more typical of the early days of empirical research of literature, becoming less common in the recent decades, whereas eye-movement monitoring has become more popular recently thanks to technological advances that have made it simpler to apply.

Verbal information was the key source of data for the early empirical researchers of literature. Steen (1991), for example, reviewed ten volumes of *Spiel*, a major German media culture journal at that time, and found that out of 33 methods, 13 were based on verbal information provided by thinking aloud, interviews and group discussions. An informal survey of the *Scientific Study of Literature* in 2014-2017, shows that only two out of 42 studies used equivalent techniques; both used interviews, and no study was found that used thinking aloud or group discussion.

Some of the reasons for this trend shift may be methodological. Verbal information collected in an interview or by thinking aloud leaves the freedom to the participant, rather than the researcher. This is opposed to multiple-choice questionnaires or Likert scales, where the participants' responses are channeled in advance into a narrow range. In the past, the participant's freedom was considered an advantage for researchers, most of whom were from the humanities. As time went by, this freedom came to be perceived as a disadvantage, since the richer and more diverse the information collected, the more difficult it is to subject it to reliable quantitative analysis. Moreover, the very collection of the verbal information raises reliability issues (Nisbett & Wilson, 1977).

Recent years have seen a flourishing of the empirical study of literature, also related to greater methodological caution. Today's researchers ask narrow questions and rely on methods that are more careful. Giving up on the readers' verbal reports is part of this trend, exacting a heavy price in richness and variety – themselves key characteristics of the literary experience. I therefore believe that suggesting a synthesis of rich verbal information and more accurate indicators is optimal as it has the potential of combining hard and soft science and overcoming the weaknesses of both. Precisely now that the pendulum tends towards the more stringent quantitative direction is the right time to prepare for it to swing back – and connect the two edges.

Key Issues in Collecting Verbal Information: Reactivity and Verticality

The two main issues involved in collecting verbal information are reactivity and verticality. *Reactivity* is the way speech, designed to report on thought processes, affects and modifies them. In other words, this is a cognitive disturbance due to the verbal reporting in real time. *Verticality* refers to the partiality of verbal information due to the participants' forgetfulness or unawareness of their own cognitive processes. These two factors are related and there is a tradeoff between them. If researchers ask a reader to report about her experiences while reading literature, verticality would be very low, since those experiences would be fresh in her memory; reactivity however would be high, since reporting would disrupt the reading flow and possibly influence

some ongoing cognitive processes. Conversely, if we ask the reader to report on her reading retrospectively, there would obviously be no reactivity effect, but verticality would be high, and grow higher the longer the time gap between the reading and the interview.

Ericsson and Simon (1980) discovered that the type of task assigned to the participant significantly affects the severity of reactivity. Considered trailblazers in the methodology of thinking-aloud studies, they reviewed dozens of studies and concluded that if the reported information is verbal information available in the participant's short-term memory, the reactivity problem becomes minimal. Ericsson and Simon (p. 227) described three levels of verbalization, with decreasing reliability of verbalization and increasing reactivity with the move from one level to the next:

- › **Level 1:** Direct articulation of information stored in a language (verbal) code.
- › **Level 2:** Articulation or verbal recoding of non-propositional information without additional processing.
- › **Level 3:** Articulation after scanning, filtering, inference, or generative processes have modified the information available.

According to the researchers, level-1 verbalization is the most reliable, since the participant is not required to perform any additional processing of his experience, but only report or reiterate the information that is available in his short-term memory and already stored as lingual data, requiring no further conversion. Level-2 verbalization involves only a simple conversion, for example from the visual to the verbal. Level-3 verbalization, however, requires significant processing before it is possible to verbalize the information; it is the least recommended, since the need to process, filter and summarize the information significantly disrupt the process on which the participant has to report. Retrospective reports are free of such influences, but they too are problematic because of verticality issues. In particular, verticality issues become exacerbated if the participant is required to summarize his experiences rather than report them as they occurred in given points in time.

Based on Ericsson and Simon (1980), there are several good reasons to suspect serious reactivity problems with thinking aloud about reading literature. Although reading is fundamentally verbal, much of what interests literature researchers requires level-3 verbalization. Only in the narrow case where the researcher is interested in the verbal decoding itself are we talking about level-1 verbalization, where reactivity is minimal.⁵ But when the researcher is interested in emotions arising while reading, in intertextual associations, thoughts about the author, reasons for reading difficulties,

⁵ Yet even in that minimal case, the pace of reading would slow down, due to the difference between the tasks. In silent reading, the reader is not required to complete his linguistic processing, and can settle for under-specification. Full linguistic realization as required when reading aloud, however, involves full vocal representation.

and any other subject requiring reflection, level-3 verbalization would be required, where reactivity is high. Such verbalization disrupts the very process it is supposed to describe and may distort or improve it.

There is another reason to suspect that reactivity problems would be particularly severe when studying literature reading. According to the foregrounding theory, the mechanism whereby literature operates is by slowing down the reading, causing the reader to devote increased attention to the text, thereby encouraging deeper interpretive processes, which in turn lead to positive aesthetic appraisal. If the very requirement to pause may improve aesthetic appraisal, then when thinking aloud the slowdown due to the verbal report may lead to deeper reflection that would not have occurred without the need to report, thereby producing aesthetic effects. Thinking aloud is thus exposed to the danger of inflating literary-aesthetic effects. Some support for this claim may be found in Tartakovsky and Shen (forthcoming), who found that participants asked to interpret non-standard similes rated them as more aesthetic than participants who rated them without having explained them first: indicating that the extra attention allocated during the explanations affected the aesthetic appraisal. For these reasons, the think-aloud method may be an effective tool for raising the reader's awareness of processes occurring while reading, and as such can be effective in teaching literature, but empirically it is not optimal, since it is difficult to accept its findings as representative of real-life reading.

Two Difficulties in Analyzing Eye Movements: Interpretation and Big Data

Recent years have seen increased use of eye movements in the empirical study of literature. Previously costly and cumbersome, this measurement technique became accurate, low-cost and user-friendly. In most universities today, there is at least one eye movement monitor in the psychology, linguistic or education departments, allowing easy access to literature researchers. In Germany, in particular, there are two large laboratories devoted exclusively to empirical aesthetics studies that have the instruments and qualified personnel required (one at the Free University of Berlin and the other at the Max Planck Institute for Empirical Aesthetics in Frankfurt). Most recently, the 2018 Conference of the International Society for the Empirical Study of Literature (IGEL, Norway) dedicated – probably for the first time in the history of the field – a panel to the study of eye movements in reading long literary text. There are therefore good reasons to expect this research direction to gain further momentum.

Despite the great enthusiasm surrounding this research tool, its use in the study of literature is still in its infancy. Although the resolution of the information obtained from eye movement is very high in space and time, its meaning requires *interpretation*. The same physiological result can suggest opposite conclusions. For example, regression of the gaze can attest to great difficulty in understanding, but also to rereading to enjoy the text again. Slow reading can indicate pleasure or confusion.

A dilating pupil can be evidence of cognitive difficulty or emotional arousal, etc. Eye movements tell the researcher that something important occurred in terms of cognition or reading experience, but do not reveal exactly what. This, then, is the interpretation problem.

There is extensive linguistic research on eye movements and their cognitive significance (Rayner, 1998). Linguistic research has resolved the interpretation problem using carefully controlled research conditions; for example, multiple versions of the text and maximal control of all parameters that differentiate them. This solution requires short, simple and artificial texts. Obviously, this is irrelevant for literature researchers, as a literary text is usually too long and complex for the researcher to isolate the variables. Replacing the literary text with a simple one-liner means throwing the baby out with the bathwater – missing the elusive object of study – literariness.⁶

In addition, the interpretation problem is exacerbated due to the nature of the literary ploy. If, as foregrounding theory assumes, the literary ploy is aesthetically appraised through difficulty or even confusion, it would be impossible to judge on the basis of the text alone whether readers dwell due to a ploy that worked and led to a literary experience, or due to one that failed and left the reader with only the difficulty and confusion. A researcher faced with the interpretation problem would find it very difficult to resolve without some direct information from the reader.

Additional problems in analyzing eye movements of literature readers have to do with *big data*. Since today's researchers can obtain split-second information about each word separately, with regard to a large number of indicators (duration of fixation, duration of first fixation on a word, inward and outward regressions, total dwelling time, etc.),⁷ researchers find themselves flooded with data.

Extracting the knowledge hidden in the information requires non-linear models or machine learning and information mining methods that are usually not available to literature researchers. It is difficult to approach this task without programming knowledge. Even a detailed quantitative analysis of each word based on multiple stylometric parameters requires considerable expertise, and is possible only for a few European languages for which a dedicated stylistic analysis software has been developed. Thus, analyzing eye movement findings has become a project that literature researchers find hard to complete without the help of statisticians and programmers.

⁶ Exceptions to this rule are studies on extremely short literary texts, such as aphorisms, metaphors or poem lines. These studies are important, but can only capture limited aspects of the literary experience. For example, in their study on reading single poem lines, Van Peer, Hakemulder and Zyngier (2007) reported difficulty in emotional arousal, and inferred that the reading time was too short for that.

⁷ The software used to analyze eye movement findings, EyeLink Data Viewer 2.6, provides more than sixty different indicators for each word.

The Advantages of Retrospective Reflection Combined with Eye Movement Data

Combining retrospective reflection with eye movement findings goes a long way to solving the problems discussed above. Instead of the participant thinking aloud and disrupting her own task, the task is performed in silence. Afterwards, she is shown a record of her eye movements and asked to explain her thoughts and actions retrospectively. The eye movement findings serve as highly effective retrieval cues, making the explanations reliable.

Specifically, the combined RTA method addresses the following issues presented above:

- **Reactivity:** This problem is almost completely resolved, since reading is not disrupted by requiring the reader to report her thoughts. Nevertheless, the awareness of the eye movement measurement may slightly affect the reading process.
- **Verticality:** This problem is mitigated. Although time passes between reading and reporting, so some forgetting is inevitable, eye movements serve as reliable retrieval cues that awaken the participants' memory.
- **Interpretation:** This issue is largely resolved since the participant serves as the interpreter of her own eye movements. The participant has privileged knowledge on her own mental states (even though this knowledge is limited to conscious processes). Therefore, the participant's help can resolve much of the interpretation problem. Nevertheless, as the information becomes verbal, researchers have a different interpretation problem, as now they are required to interpret not the eye movements but the explanations provided by the participants.
- **Big data:** The proposed method reduces the excessive eye movement data in two ways. First, not every data point is checked (each fixation or word), but only the salient ones, those on which the participants dwelled more during their reading. Moreover, the readers do not necessarily have anything to say about each such point, but only on those points on which they dwelled consciously, and which were significant enough to leave a trace in their memory. This significantly reduces the amount of information, and converts it from numerical into verbal form.

I know of no applications of this method in the study of literature. But it has been used in the usability area for over a decade, and several methodological studies have examined its strengths and weaknesses, as presented next.

RTA in Usability Studies

Several methodological studies demonstrate the advantages of retrospection combined with eye movements. This method has been used for several years in the usability area

and because of its intuitive advantages in solving methodological problems involved in thinking aloud and in eye movement measurements, usability researchers have begun to use it spontaneously at the same time in different laboratories (Hyrskykari et al. 2008).

Hyrskykari et al. (2008) examined three conditions of verbal reporting: thinking aloud, retrospective thinking aloud based on eye movements, and the combination of the two. They compared the three conditions in terms of both the amount of verbal information received and its type. The task given was to search for a car to buy on a used car website. It was found that retrospective thinking produced more verbal information: the mean length of a thinking-aloud interview was 1148 words, while retrospective thinking produced 3309 or 4236 words (depending on the experimental condition).

The verbal content was divided into three types of comments: *manipulation* those describing the basic manipulations the participant was doing at the moment (e.g. “I write the name into this field”); *visual* comments (e.g. “Then I look for a picture of the car”); and *cognitive* comments (e.g. “I remember seeing it before). Hyrskykari et al. (2008) found that when thinking aloud, most (82%) of the comments described the manipulation performed by the participants, while only 4% were cognitive. Conversely, in the retrospective thinking condition, the rate of cognitive comments increased to 33% or 43% (depending on the condition). Thus, the retrospective method is not only more reliable in that it does not disrupt the thinking process but it also provides more and cognitively richer information.

The logic behind these findings is straightforward. Manipulation comments require a level-2 verbalization, that is, simple conversion of a non-verbal action into verbal reporting. This verbalization is easier to provide while performing the task. Conversely, reporting cognitive information requires level-3 verbalization, which is more disruptive of the task. Therefore, the participants who concentrated on performing the task found it difficult to provide that information in real time, but easier to do so retrospectively.

Other studies lent support to these findings. Van Gog, Pass, van Merriënboer and Witte (2005) examined thinking aloud during problem solving and found that in a guided retrospection combined with eye movements condition, more metacognitive information was obtained than in thinking aloud. Similarly, Altabour, Alhadreti and Mayhew (2017) found the eye movement RTA produces more cognitive information, better identifies comprehension problems and improves the ability to recall behavior details, compared to retrospection guided only by a video documentation of the screen and mouse movements.

Further evidence on the kinds of information obtained using the different methods was found by Eger, Ball, Stevens and Dodd (2007). They gave their participants internet searching tasks and compared three methods of verbal reporting: thinking

aloud; retrospection with eye movements; and retrospection without eye movements, but with documentation of the computer screen and mouse movements. They found that retrospection with eye movements found more usability issues in the website than thinking aloud, and particularly more problems related to the user's lack of comprehension and the feedback received from the website.

It was also found that task difficulty affected the effectiveness of the type of verbal reporting. Eger et al. compared two different search engines: a familiar and user-friendly one (Google), and an unfamiliar and complex one (Infomagnet). Retrospection with eye movement exposed many more usability issues in Infomagnet than retrospection without eye movements. This indicates that retrospection with eye movements is better suited for complex and cognitively challenging tasks, as well as for detecting miscomprehension and communication issues between the user and the website. It was also found that compared to thinking aloud, retrospective methods were more pleasant for the participants, did not disrupt task performance, and provided more complete and comprehensible statements.

Guan et al. (2006) examined the reliability of retrospection with eye movements by having their participants solve numerical or visual problems on two complexity levels. They compared the eye movements and the retrospective explanations provided by the participants, and found that in 88% of the cases the information provided was reliable, and matched the eye movements as recorded. Nevertheless, in 3-4% of the cases, there were fabrications, defined as participant reports on an area of the screen the participant never focused on. Surprisingly, they discovered that the rate of fabrication dropped significantly with task complexity: in complex tasks, there was only 1% fabrication. Nevertheless, Guan et al. found that a more significant issue in using this method is omission of information. This was defined as points on the screen on which the eyes did focus, but were not included in the verbal explanation. Unlike fabrications, the number of omissions grew with task complexity. According to the researchers, omissions are due, among other things, to the gap between the abstraction of the verbal information and the precision and density of eye movement data. In other words, this is a verticality problem not due to forgetfulness but due to the difference between the nature of verbal and physiological information. For example, instead of stating explicitly that the eye went right and left seven times, participants summarized and said their eyes shifted from side to side. Thus, one weakness of this method is that it does not provide information that is as detailed on each of the eye movements. This weakness is also a strength, however, in that it reduces the overall amount of information, thereby contributing to minimizing the big data issue.

These usability studies provide initial indications for the reliability of combining retrospection with eye movements and its advantages over other methods of verbal reporting. Note, however, that there are many differences between the tasks reported in these experiments and reading literary text. In usability studies, the task is to solve a problem, search for data or navigate in a virtual environment, not to read for pleasure.

Moreover, there are also differences in the purposes of researchers from the different disciplines. Usability researchers focus on improving websites and care mainly about the results – their motivation is usually practical rather than theoretical. They do not try to develop a general theory on the cognition of website users, but rather minimize the comprehension difficulty and user errors in their website, to make the user experience smoother and more friendly. On the other hand, literature researchers are more interested in theory, in the reading process as an objective in its own right, and as much as they are interested in problems encountered by the reader, they are more interested in the various facets of the aesthetic-literary experience.

The distance between the disciplines may not be as great as it appears, however. My own research combines these two research directions. The study of semantic noise and comprehension failures brings the literary discipline closer to that of usability researchers. Therefore, for the purposes of this study, RTA is important precisely for the potential of detecting comprehension problems. Thanks to this method's advantage in detecting comprehension problems and communication failures, it was possible to pay better attention to aspects of the literary reading process hitherto neglected. The failed foregrounding model highlights the difficulties and failures of the real reader, and thus this method was ideal for developing it.

An Eye Movement-Based Retrospection Experiment

This is the same experiment described in Section 2: reading Jorge Luis Borges' "The Chamber of Statues" (1935). A few minutes after the reading, the retrospective interview began. In the short recess between the reading and the interview, I produced the "heat maps" and the participants completed an aesthetic appraisal and semantic noise questionnaire (see Section 2 for the questionnaires and their analysis).

At the start of the interview, the participants received explanations about the heat maps, how they represented the eye movements, and how they should be read (see Figure 16). They were explained that they would be asked why they dwelled on particular points in the text and that they would have to try to remember what happened during their reading, and if they could not, then try to infer based on their knowledge about themselves. In their responses, they could borrow terms that appeared in the questionnaires (such as "beautiful", "interesting", "confusing", or "difficult"), but did not have to do so. Every answer that came to their mind was relevant. They were also told that sometimes their dwelling on a certain word was not related to the word itself, but to the sentence or the larger text area.

Next, during the interviews, the participants saw the heat maps that documented their eye movements as they read, and were asked about all the points on which they dwelled. If necessary, the eye movements were presented in additional ways in order to resolve ambiguities in the heat maps, such as using a film or a map of fixations and saccades. Particular attention was devoted to eight specific passages identified as key

points during the pilot stage: most of the pilot participants dwelled on these passages and provided rich and varied verbal responses to them. All interviews were recorded, and the comments on the eight passages were transcribed.

In what follows, several general analyses of the interview material are presented: the length of the interview as a function of the participants' reading experience; categorization of the 100 most common words in the interviews; and analysis of several eye movement indicators according to the participants' reading strategies. Based on these analyses, conclusions are suggested on the quality of the information collected using the RTA method.



Figure 16: A participant's heat map of the first paragraph. Red indicates dwelling (see legend on the right). Note that the text is justified to the right as Hebrew is read from right to left.

General description of the interview material

My first impression of the interviews is that the information obtained using this method is very rich. The explanations for the delays were varied and ranged across a broad spectrum. Some of these were related to processing difficulties, such as unfamiliar words, unusual syntax, miscomprehension, confusion, details that do not match, etc. Others were related to higher-level processes with obvious literary importance, such as the text's poetics, pleasure, intertextual associations, emerging emotions, visual similarity, thoughts about the genre or author, etc.

Most participants were curious about their eye movements – an aspect of their behavior that is usually hidden from them. My impression was that they enjoyed researching their own eye movements. In most cases, they found it easy to recall the reason for their delays, and when they didn't recall, they managed to infer based on their self-knowledge. It also seemed they did not find it difficult to separate their

experience during the first reading from their experience during the second reading, when they could see their eye movements superimposed on the text. A certain, albeit anecdotal evidence for the participants' substantial confidence in their memory may be found in the fact that they barely used the phrase "I don't remember". An examination of the interview transcripts showed that this phrase was used an average of 0.7 times (median = 0) per interview.

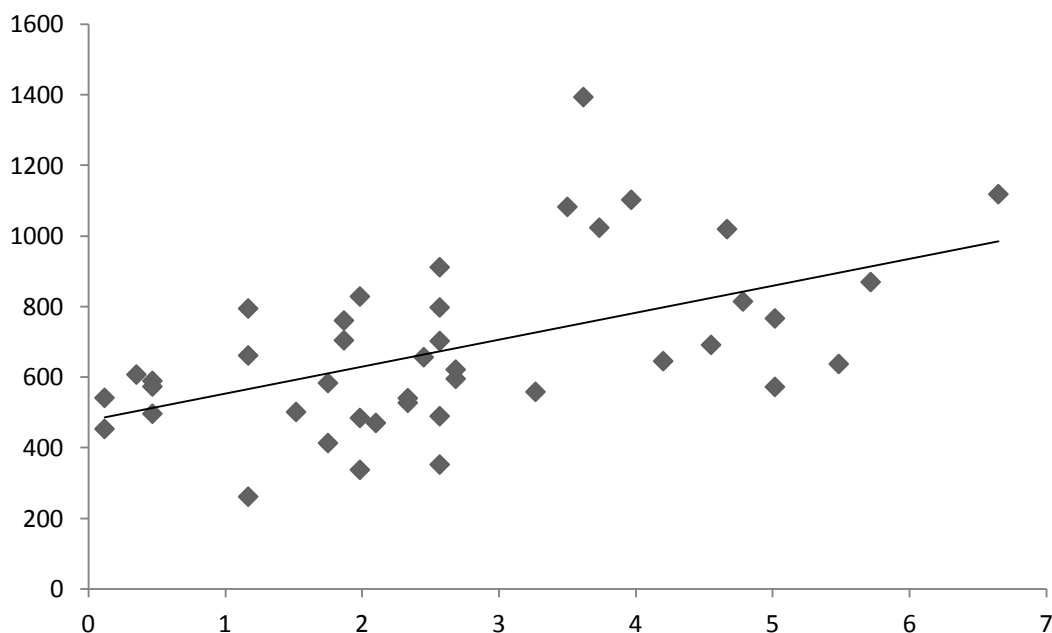


Figure 17: Interview lengths in words (Y axis) as a function of the participants' experience in literature reading as measured by the author recognition test (ART), normalized to a 1-7 scale

The variance in interview lengths was very high. A strong direct correlation was found between the length of the transcribed interviews (in words) and the participant's experience in reading literature as measured by the author recognition test (ART) (see Figure 17) ($R = 0.53$; $p = 0.0003$). This relationship can mean that (1) participants with reading experience are aware of more aspects of their subjective experience as they read literature; and/or (2) experienced participants are more verbal, and therefore provide more detailed explanations. Perhaps other factors are also at play here. Whatever the reason, experienced readers provided more detailed reports, which, moreover, impressed me as richer in terms of literary insights.

Interview length was not correlated with aesthetic appraisal nor with semantic noise. This finding suggests that the nature of the subjective experience of reading had no impact on the participants' ability to explain their reading processes. Namely, those who enjoyed the story did not remember more or was better aware of the reading process. And conversely, those who disliked the story or found it difficult to read did not recall less about their reasons for dwelling. This finding may be understood also

in light of the usability studies that showed that this method is particularly effective in collecting comments on incomprehension and communication problems. Thus, even those readers who did not have a positive aesthetic experience, or experienced high semantic noise while reading, had much to say when required to explain their reading delays, providing explanations about problems and difficulties that were no less detailed than the explanations about the aesthetic literary experiences.

Quantitative analysis of common word pairs

The section on the failed foregrounding model presented a detailed quantitative analysis of the interviews. Below, I only present a brief analysis of the most common word pairs, to demonstrate the type of information produced by this interview technique. This method allows the researcher to detect general, common trends in a highly rich and varied material. Using Primitive Word Counter (version 1,09; 2007-2009), a list of the most common 100 combinations of words used in conjunction in the interviews was produced. Most combinations were of two words but some were of several.

I divided these combinations into four categories. This division is inspired by the analytic methods used in the RTA usability studies reviewed above, but is different, among other things because of the content of the interviews is different. For example, the *procedural* category that is meaningful in browsing websites is irrelevant to the reading of literary texts. The *cognitive* category is shared both by this categorization and those commonly used in usability studies.

The first category of *cognitive* combinations includes a verb that suggests a cognitive action, such as “I remember”, “I did not understand”, “I don’t know”, “I think”, “I tried to understand”. *Quoting* combinations cite directly from the text or include words such as “the story”, “the sentence”, etc. *Referential* combinations include the word “it” or its equivalent, suggesting that the participant is referring to something specific, but there is no way of knowing exactly what without analyzing the interviews in greater depth. Finally, *discursive* combinations are common expressions that do not match any of the other categories, including negation and affirmation expressions, “yes”, “no”, “you don’t”, conditional phrases, adverbs and other common figures of speech: “If he”, “too much”, “so”, “possibly”, etc.

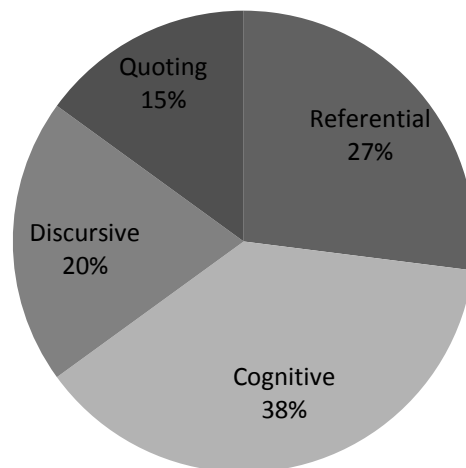


Figure 18: The commonest 100 word combinations in the interviews by categories.

Figure 18 shows that the commonest category is the cognitive one: 38% of the word combinations had a clear cognitive aspect. The second most common were the referential combinations, with 27%, followed by the discursive (20%) and quoting (15%) categories. The high frequency of cognitive word combination is similar to Hyrskykari's (2008) finding of 33% or 43%. It therefore seems that in reading literature as well, the procedure of retrospection combined with eye movements encourages essentially cognitive comments. Therefore, it is suitable for studies on readers' cognitive processes.

The high rate of referential comments may be explained either by the high frequency of the word "it" in Hebrew,⁸ or by the situation itself: the interviewees refer to the heat maps presented to them. Since they see the same thing as the interviewer, indication – whether physical (pointing the finger) or symbolic (using the word "it") – is highly common. The infrequent use of direct quotes can also be understood against this background. The participants have little use repeating the passages verbatim, since it is right there on the screen in front of the interviewer, so that referring to them indirectly was enough.

Reliability

An important issue for future study is the reliability of the information collected using this method. This section presented theoretical considerations pointing to the minimal extent of verticality and reactivity issues, and presented usability studies supporting

⁸ According to *Ma'agarim* (2019), the historical dictionary of the Academy of the Hebrew Language, the frequency of the word "it" in Hebrew is about one percent of *all* words in newspapers and periodicals, as well as in texts of all kinds written in the 20th century.

these claims. Future studies using this method in literature research, however, would do well to back them with empirical evidence specific to literature.

Specifically, I am not as concerned with forgetfulness as I am with the possibility of *fabrication*. That is, the possibility that when presented with their own heat map, the participants come up with an explanation that is fabricated, whether they are aware of it or not. Although Guan et al. (2006) suggested that the rate of fabrications is very low, the danger exists, and the extent of this phenomenon and its potential impact should be assessed in the literature reading area as well. It is plausible to assume that even when the participants do not truly recall what happened, their response remains far from random. Since the text is right in front of them, in the form of a heat map, whatever they infer or even invent is based on their second encounter with the text.

There are two basic possibilities regarding the potential impact of this second reading on participant reports. When participants report that the reading delay is due to *semantic noises*, and rely on semantic noises experienced in the second reading, it is reasonable to assume that semantic noises were experienced in the first reading as well. This is because it is to be expected that semantic noises would be reduced from reading to reading. Namely, it is more likely for RTA to inflate the number of positive aesthetic appraisals, and not of semantic noises. The analysis of the findings presented in the previous sections indeed shows that most interviewees refer to semantic noises, whereas noise-free aesthetic appraisals (“full foregrounding”) represent only about one-fifth (21%) of the cases. Thus, even if fabrication is feared, most of the findings remain valid, and if the full foregrounding category has been inflated, it could not have been inflated considerably, as this is a relatively small category to begin with.

Now I will examine the second scenario, where the participants report a *positive aesthetic experience*. My personal impression is that the full foregrounding reports are highly specific and sometimes describe detailed and even original interpretive moves. While it is impossible to ensure that none of these moves reflect processes that occurred during the second reading, the likelihood of such contamination is low, for several reasons. First, the second reading in this case is not ordinary, uninterrupted literary reading, but only a reviewing of a heat map during an interview. That is, the participants do not reread the entire text, but only reexamine specific sections of it, in order to provide localized explanations for their dwelling patterns. Thus, any second reading effects can be expected to be much smaller than in ordinary second reading.

Second, studies on second reading have not reached clear-cut conclusions regarding its effect. Even in studies where such an effect was found, it was not very strong. For example, Zyngier, Van Peer & Hakemulder (2007) found an effect only in the most complex text out of the three, and only in one out of three reader groups. The effect size was approximately an increase of 7% in the aesthetic measurements after the

second reading.⁹ This is a maximal estimate of the potential of the second reading to improve aesthetic appraisal. It is reasonable to believe that in the RTA condition, the effect is much lower, as this is not a full second reading, and that the 7% were found by Zyngier et al. (2007) in one condition out of nine.

A third way of estimating the effect size of fabrication in this method is based on Guan et al. (2006). As described above, they found that the cases of fabrication in retrospection with eye movements represent a mere 3-4% of all cases, and that when the task is complex, this rate drops drastically to 1%. Since the reading assignment in the experiment involved a complex text, it may be assumed that the rate of fabrication was extremely low in the present case as well.

Another approach to assessing the reliability of the information collected is not to rely exclusively on the verbal data, but to *cross-reference* it with other research instruments or data sources, such as questionnaires or eye movements. The previous sections presented two such cross-references, one with the aesthetic appraisal questionnaire and the other with the participants' reading experience. In brief, these comparisons were highly significant, supporting the method's reliability. While there is some concern that in the course of the interview, the aesthetic experience was somewhat inflated due to the second reading, there is no such fear when it comes to the aesthetic questionnaire, since it was collected immediately after the reading and before presenting the heat maps, i.e. with no further reading. Therefore, because the reports collected in the RTA method matches the aesthetic questionnaire findings, this supports the method's reliability.

Comparing with eye movement patterns reinforces RTA reliability. Based on the selection method presented in the previous part, the participants were divided into four groups based on their main reading strategy. It was found that these groups also differed in their eye movement patterns, as seen in the graphs below. The first two graphs represent indicators of early cognitive processes, that is, the initial processing of a word (duration of the first fixation on the word and the number of fixations in the first time a word was looked at, i.e. without backward regressions; Figures 19 & 20, respectively). The following two graphs represent subsequent processes.

⁹ The effect size is not provided directly in the article, only its significance. The estimated effect size presented here has been calculated out of the figure 2 on p. 671 of their article.

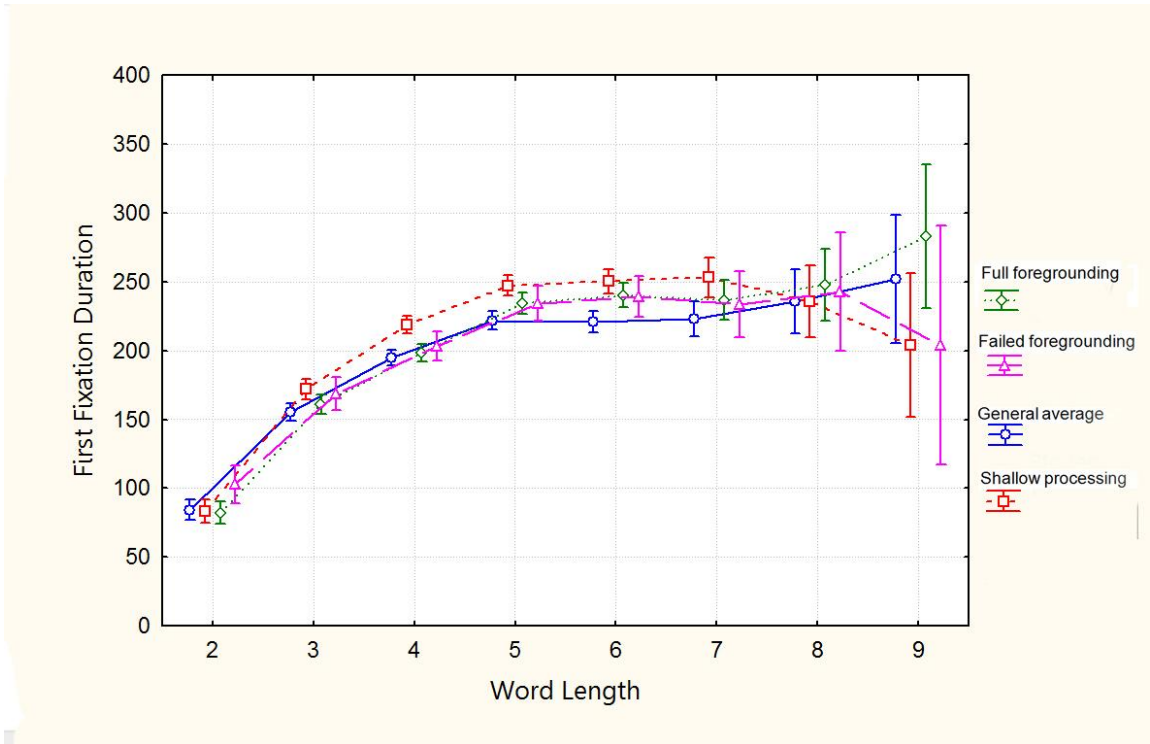


Figure 19: Duration of the first fixation (in milliseconds) by word length in letters and reading strategy. Vertical bars denote 0.95 confidence intervals. Current effect: $F(21, 20847)=1.8683, p=.00925$

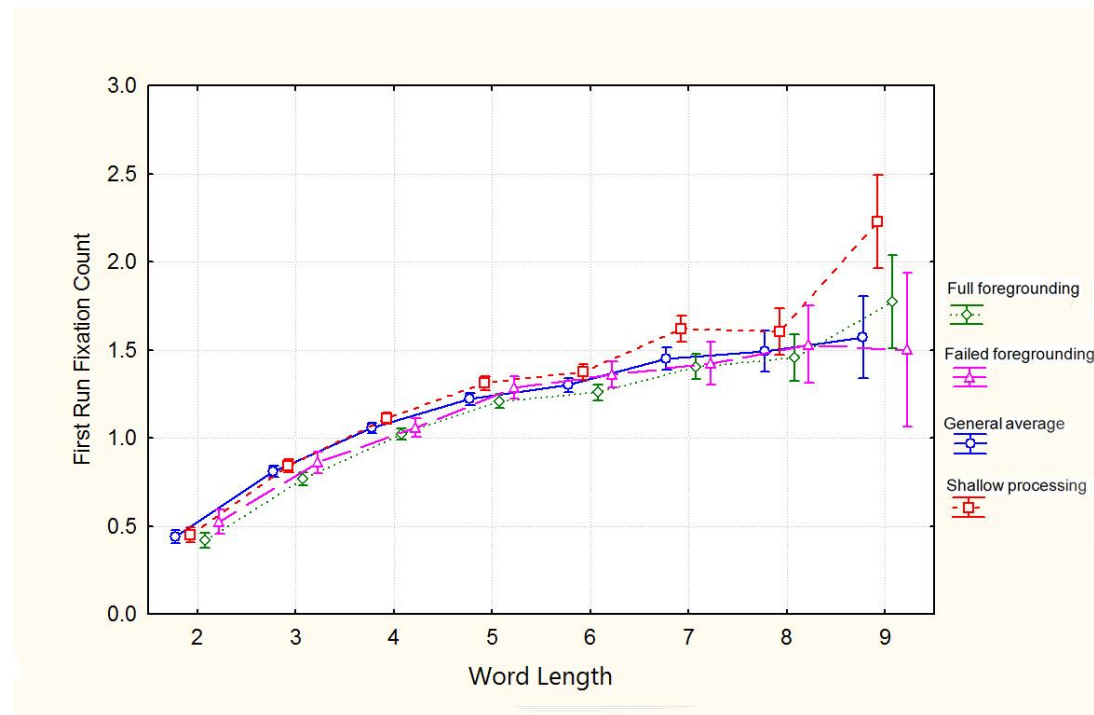


Figure 20: Number of fixations in the first time the word was looked at by word length in letters and reading strategy. Vertical bars denote 0.95 confidence intervals. Current effect: $F(21, 20847)=1.7166, p=.02168$

These two graphs show a clear difference between the shallow processing and the rest of the groups. It appears that the length of their first fixation was higher than that of the rest, for words of medium length. Moreover, the number of fixations in the first reading of the word was higher than that of the rest – for long words. These two findings suggest that readers of the shallow processing strategy had trouble upon their first encounter with the text – a difficulty in its initial decoding – as well as in basic reading processes. This finding is in line with the interview findings, which suggested that the difficulty of the readers in this group was not due to the attempt to perform an interpretive move but to more basic comprehension problems and to multiple semantic noises. These readers became stuck in the basic stages of processing, and failed to reach deeper processing that could enable literary interpretation.

Two other eye movement measures that examine subsequent processing reveal a different picture, however. Figure 21 presents the “run count” measure, or the number of times the gaze enters and exits a given word “area”, followed in Figure 22 by “dwell time”, or the total duration of the eye’s dwelling on a given word.

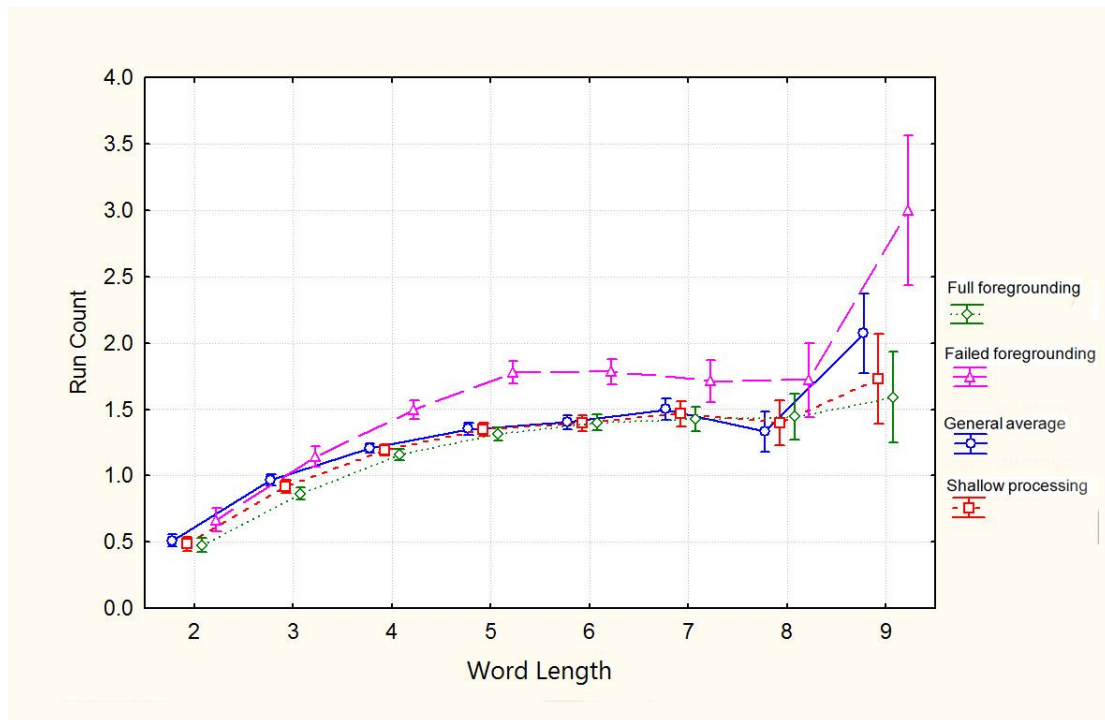


Figure 21: Number of times the gaze enters and exits a given word “area” by word length in letters and reading strategy. Vertical bars denote 0.95 confidence intervals. Current effect: $F(21, 20847)=2.0599, p=.00293$

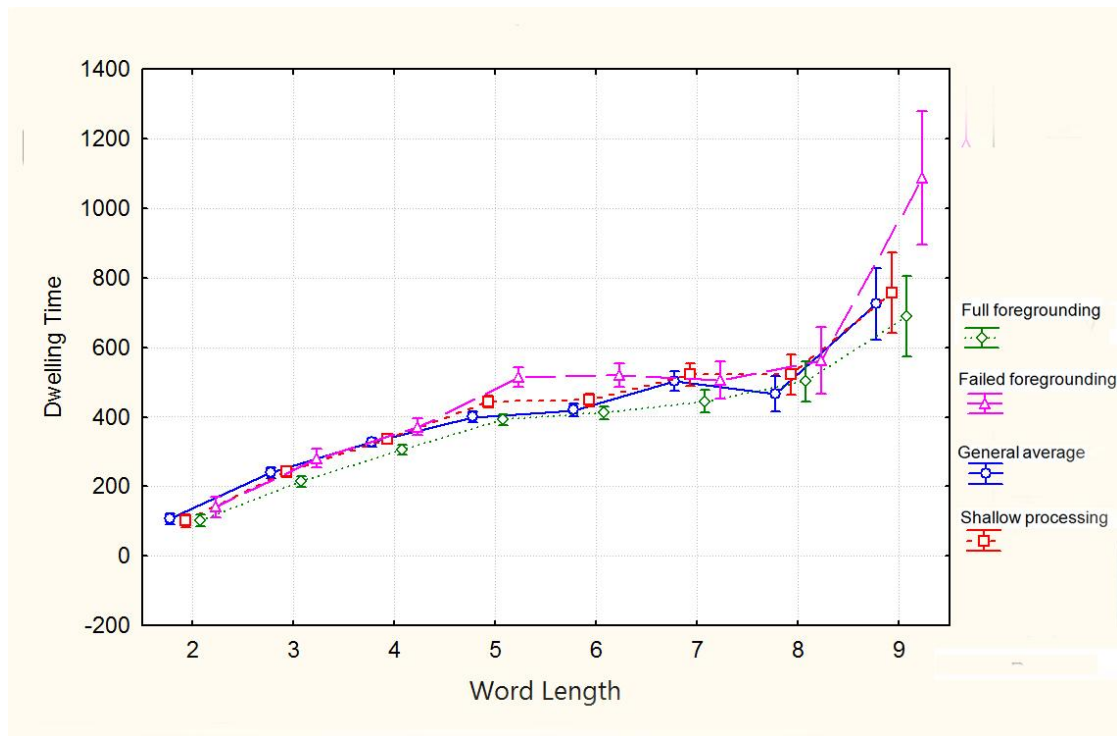


Figure 22: The total dwelling time on a given word (in milliseconds) by word length in letters and reading strategy. Vertical bars denote 0.95 confidence intervals. Current effect: $F(21, 20847)=2.3366, p=.00050$

These last two graphs indicate that the failed foregrounding strategy is exceptional. These participants reread many more words than members of the other groups, in almost all word lengths, and their dwelling time on letters with five or more words was longer than the rest of the groups. This suggests that they invested more, or had much more difficulty in the later stages of reading – a finding that is in line with the interview finding that their reading was fraught with foregrounding failures. In other words, they tried to interpret the text but failed, entered an interpretive process, were involved in the text, but their efforts failed to mature into a positive aesthetic experience.

Another finding arising from these graphs has to do with readers whose main strategy was full foregrounding. These readers dwelled the least, both in terms of the total duration devoted to a word and in terms of rereading words. This suggests a smoother processing, an effortless reading experience relative to the other groups. These readers had much less difficulties. The fact that participants reporting full foregrounding had less difficulty already during the first reading dose no support the claim that the second reading (during the interview) was responsible for their positive aesthetic appraisal. In general, all these findings mitigate the suspicion for belated influence of the interview stage, since they indicate a difference between the reader groups already during the first reading, and that the differences in eye movements match the patterns arising from the interviews. This finding also supports the questionnaire findings, in

that there is a general effect of fluent processing of the text among participants experiencing positive aesthetic appraisal.

Given all the considerations presented here, it may be concluded that the reliability of the method used for collecting verbal information in the experiment is at least adequate. This conclusion is based both on theoretical analysis of previous studies and on empirical evidence showing that the differences in the interviews between various participants mirror differences found using other instruments and approaches, including physiological data collected during the reading. Therefore, this information may be relied upon for the analyses and conclusions reported in the previous parts.

Discussion

This section presented the RTA method – combining retrospective thinking aloud with eye movement monitoring. This method mitigates several key problems in both collecting verbal information and analyzing eye movements. First, it reduces the reactivity and verticality problems of collecting verbal information. Namely, it does not disrupt the reading and thinking process as it occurs, and it is relatively reliable in terms of recollection, since the eye movement patterns remind the participant of her reading process. It was also found to have a low likelihood of fabrication but a high likelihood of omissions. That is, not all the information in the eye movement findings is explained. These omissions, however, help reduce the amount of information and thus mitigate the big data problem. Finally, the method helps reduce the huge amount of data produced in eye movement studies in two different ways. First, instead of referring to all words, the interviewer refers only to those areas in the text where many of the readers have dwelled. Second, the method reduces the relevant data to those places in the text that the participants remember and have something to say about dwelling in them. Thus, RTA focuses Researchers on specific, yet significant, phenomena that occur in the reading process, those that leave a lasting impression in the reader's memory.

The method proposed aligns two different types of responses by the same participant: verbal and physiological. It is verbally rich as well as localized spatially (or textually) and temporally. Finally, it provides rich cognitive evidence, both according to previous usability studies and according to the analysis of common word combinations in the present experiment.

The participant's reading experience was found correlated with interview length, meaning that this method has a relative advantage in studies on experienced readers, since they do better in verbalizing their reading process retrospectively. Note that the method's effectiveness is not limited to readers who have had a positive experience, since readers with negative aesthetic appraisal of the text as well as readers who have experienced semantic noises provided the same amount of data. Thus, the method is suitable for a wide range of literature reading experiences, and particularly for

learning about comprehension difficulties and communication problems the reader experiences.

Limitations

First, the RTA method privileges relatively short texts: poems, fragments, short stories or novel chapters. When the text is long, participants may forget the details of their reading process, exacerbating the verticality issue. Second, the method does not enable examining phenomena of which the participant is unconscious. Third, it provides less information from inexperienced participants. Fourth, some participants felt uneasy with their eye movements being monitored and with being asked about them.

Moreover, the interviews, transcription and analysis require considerable time, and one of the most significant challenges the method poses to the researcher is to analyze the interviews reliably as well as allow for quantitative analysis. Even having managed that, the entire complexity of the interview materials would be beyond the researchers' reach. Therefore, complementary qualitative analysis is recommended.

Finally, the method does not enable the researcher to obtain the same amount of information on every point in the text. The information is concentrated in key points that seem important to the readers and who therefore have much to say about them. Conversely, some of the information recorded in the eye movement monitoring is not addressed by the participants, whether due to the higher abstraction level of verbal information or due to recall difficulties. If the researchers are interested in information about points omitted by the interviewees, this could be a serious limitation.

GENERAL DISCUSSION

This dissertation presented a model describing the various paths of failure in the foregrounding process – a model that derives from the standard foregrounding model, expands and complements it. The main innovations of this work are that it sheds light on failures in reading literature, provide a theoretical conceptualization of these failures, classify them under the failed foregrounding model, and develops a methodology designed to help study these failures. Two different types of failure were described according to the three foregrounding stages: failure in the transition from the first to the second stage, leading to shallow processing; and failure in the transition from the second to the third, leading to semantic noise. Another kind of failure in the transition from the second to the third stage is partial foregrounding: some raw aesthetic appraisal, limited and underdeveloped in interpretive terms, which can be accompanied by considerable misunderstanding or confusion. The implications of these failures were described in terms of aesthetic appraisal and semantic noise experienced by the readers.

I then provided several types of support for the model, based on previous foregrounding experiments and the findings of my own research. Questionnaires were used for global examination of reading processes, and interviews were used for local examination. The analysis of the aesthetic appraisal, semantic noise and reading experience questionnaires supported the fluent processing hypothesis. However, cluster analysis that divided the participants into three groups revealed a pattern that is more adequately accounted for by the failed foregrounding model. The interviews enabled a more fine-tuned observation of local effects, and the classification of foregrounding profiles, which are the distributions of foregrounding failures and successes in a given case.

Foregrounding profiles did not distribute randomly, but according to the participants' experience in reading literature and their global aesthetic appraisal. The model also managed to characterize the participants' reading strategies, as well as the nature of their responses to various stylistic devices. This led to the conclusion that the failed foregrounding model is rather sensitive and influenced by parameters that characterize the reader, the text and to a certain extent also the interaction between them. In 36% of the cases, readers didn't even initiate the foregrounding process, and they completed it successfully in only 21% of the cases. These rates varied significantly with the readers' experience, aesthetic appraisal, reading strategy and stylistic device. This sensitivity of the foregrounding profile suggests that the foregrounding process itself is more "fragile" than usually thought – with many factors affecting its effectiveness. This sensitivity of the foregrounding process is consistent with the struggle of experiments in this area to produce broad and consistent findings.

The model enables a series of distinctions regarding the various stages of the process. It was found that full foregrounding contributes to positive aesthetic appraisal while failed foregrounding detracts from it. It was also found that readers inexperienced in reading literature tend to settle for shallow processing, while experienced readers complete the foregrounding process more often. In general, it was found that participants prefer a “polar” strategy: there is almost no preference for reading strategies where failed or partial foregrounding are a major element.

It was also found that different stylistic devices yielded different foregrounding profiles and effectiveness levels. Figurative descriptions were the most effective in reaching full foregrounding. The author’s comments were the easiest for the readers to identify as having literary importance, and were very low on shallow processing. Nevertheless, they were not conducive of ending the process, with many readers being stuck in failed foregrounding. Least effective was linguistic difficulty: the readers found it difficult to start the foregrounding process as well as complete it. Moreover, linguistic difficulty was also the stylistic device least sensitive to the reader’s experience.

The main findings regarding the various foregrounding stages are summarized in Table 2 below. Note that the findings presented in this work and the hypotheses they inform do not represent a comprehensive summary of what may be discovered using the model, but more of a starting point. They mark the direction, the type of discoveries that can be made using this model and methodology, the questions that may be asked and the hypotheses that may be raised in turn.

Full Foregrounding	Partial Foregrounding	Failed Foregrounding	Shallow Processing
Typical of high aesthetic appraisal & reading experience	Does not contribute significantly to aesthetic appraisal	Typical of low aesthetic appraisal	Typical of low reading experience
A major reading strategy of 11 participants	Not a reading strategy of any participant	A major reading strategy of only four participants	A major reading strategy of 12 participants
Typical of passages with figurative descriptions	Not typical of any stylistic device	Typical of passages with author comments	Typical of passages with linguistic difficulty

Table 2: Characterization of the various stages of the failed foregrounding model based on the findings readers interviews.

The standard model proved successful in that full foregrounding in the key points did predict positive aesthetic appraisal for the whole story. Its weakness was, however, in that the case it described was relatively rare. Even among the participant with the highest full foregrounding scores, its frequency was around 30-40%. In other words, even there most of the cases were of various failures in the foregrounding process. Thus, it seems that failures in the foregrounding process, particularly shallow processing, are not the exception, but are rather integral to the process of reading a literary text by real-life readers. The standard model assumes smooth passage through the three foregrounding stages, but these findings show that successful conclusion of the process is not the common case. It would therefore be more accurate to think of foregrounding not as a situation but as a distribution of situations. This distribution was found sensitive to the level of aesthetic appraisal, the readers' experience and their reading strategies.

The radical aesthetician position was also examined using the new model. This position argues that literary aesthetic experience is the result of a failed struggle against the text's incomprehensibility. According to this approach, failed and partial foregrounding should play a key role in aesthetic appraisal or at the very least, their frequency would increase with the reader's experience, and that the frequency of full foregrounding should drop. None of these predictions was supported. Even the search for readers who adopt reading strategies where those two elements are central was fruitless. Only four readers were found who opted for a reading strategy where failed foregrounding was central, and in-depth examination of their interviews did not support the idea that theirs was a "radical" literary experience, but rather that they mainly stumbled across semantic noises of various kinds. It was not my impression that they experienced "bliss" while reading the text – precisely the opposite: they were highly frustrated by it.

Methodology

The attempt to describe the entire foregrounding process, rather than just its successful completion, required the researcher and participant to be highly sensitive to the course of the reading itself. The experience of reading literature is fleeting, and any attempt to describe it may be likened to trying to paint the wing of a bird in flight. All those failures and partial successes pass by quickly and are usually filtered out by both the researcher and reader as irrelevant to literary reading. In order to capture those moments, those pieces of information before they are forgotten, a special methodology was required that combined both the high resolution in time and space enabled by eye movement tracking and the experiential wealth enabled only by the reader's introspection.

The main methodological innovation is the use of the retrospective think-aloud (RTA) technique in a literature study. It combines “soft” verbal with “hard” physiological methods, thereby bridging between traditional and current trends in the empirical study of literature. The integration it offers mitigates some of the problems associated with each of the individual methodologies: verticality, reactivity, interpretation and big data. It is particularly suitable for poems and short stories or novel chapters, for participants with reading experience and for researchers interested in a broad range of literary experiences, from semantic noises to literary interpretations. The proposed methodology provides rich cognitive information and allows the researcher to address reading comprehension difficulties, as well as failures to attain an aesthetic-literary experience. Nevertheless, like any other methodology, it is best not to rely on it exclusively, and there is some advantage to using additional approaches at the same time, including questionnaires, statistical analysis of eye movements, and qualitative analysis of the interviews. Finally, although the methodology has been extensively studied and validated in the usability area, studies specific to literature reading are required to assess the quality of the information provided and resolve issues related to reliability.

Beside the methodological advantages enumerated above, there are also ethical and ecological advantages in returning to methods of collecting freely volunteered verbal information. In this approach, the participant is not restricted to selecting a certain number on a Likert scale, but can speak and express herself freely. In this approach, the participant is not only an informant, but also a partner. She joins the researcher in the attempt to interpret her own reading patterns. This freedom allows the participant to contribute a greater part of her humanity, rather than functioning as an information-processing automaton.

Is Foregrounding the Product of Controllable Behavior?

The extent to which the foregrounding profile reflects the participant’s literary competence is still unclear, as is the degree to which it is the product of a strategy. It may also be that the ability to choose is greater in the transition between the first and second stages, that the reader’s freedom is expressed in the decision whether to initiate foregrounding, and that the ability to complete the process is more dependent on other factors, such as the reader's interpretive skills. Experienced participants initiate and complete foregrounding more often; It is less clear whether this can be attributed only to higher ability, to higher motivation, or to greater confidence that the investment would pay off.

Even if the readers do have greater influence over the initiation of foregrounding, this does not necessarily mean that they are aware of making a choice. Still, many participants have acted conservatively, and avoided embarking on a foregrounding

“adventure”. I use the word "adventure" on purpose, since a foregrounding process may be a fruitful investment, but also a risky one. One never knows in advance where the positive process might lead: aesthetic appraisal or semantic noise. Venturing out on this voyage is not only risky, but also costly in terms of cognitive resources. Thus, if shallow processing is a strategy, it is one that attempts to maximize the aesthetic experience for minimal cognitive effort.

Effectiveness

Another contribution of the new model to the stylistics area relates to the effectiveness question, both in formulating it directly and in proposing an approach for dealing with it. The issue at stake is to what extent and in what way different stylistic devices make the readers experience aesthetic effects. That question can be divided into two independent effectiveness types: (1) Effectiveness in introducing the reader to a foregrounding process; and (2) Effectiveness in taking the reader to its conclusion. The first is what we previously called “invitation for interpretation”: To what extent is this stylistic device interpreted by the reader as an invitation to interpret, and does the reader accept it? The second type has to do with the ability of the reader who has accepted the invitation to achieve an aesthetic experience.

Stylistic devices seem to vary in their effectiveness. Based on the present findings, figurative descriptions appear to be high in both effectiveness types, and linguistic difficulty to be low in both. Conversely, author comments are high in Type 1 and low in Type 2, meaning that they offer a good invitation for interpretation, but are very limited in enabling the readers to realize its potential. These preliminary findings have led me to hypothesize a relationship between the layer where processing failure or difficulty occur and Type 1 effectiveness. If the difficulty attracts the reader’s attention to the layer where it occurs, the obvious step would be to try to resolve the difficulty where it is found. Therefore, it stands to reason that difficulties in more basic layers of information processing, such as linguistic difficulty, would be less effective “invitations” for interpretation than difficulties on higher layers, such as the discursive layer, where difficulties raised by author comments occur.

Type 2 effectiveness depends on the ease in which the difficulty arising from the text may be resolved. Accordingly, it may also be related to textual factors such as the degree of integration required between various points in the text, as well as to reader-dependent factors such as interpretive ability and experience in reading literature. As can be seen in Figures 13-15, both effectiveness types seem to depend not only on the text, but also on the reader’s experience. Based on the findings presented hitherto, several hypotheses may be proposed for future studies with regard to what these effectiveness types require of the reader:

Type 1 effectiveness posits, at least, the following three requirements:

1. Identifying the difficult point in the text.
2. Understanding that the difficulty has literary significance, i.e. that the solution does not require (only) an information-driven strategy.
3. Motivation to enter a foregrounding process, or a reading strategy that prioritizes engagement in foregrounding.

Type 2 effectiveness requires at least three things:

1. The reader does not settle for shallow processing.
2. The textual difficulty is resolvable based on the reader's knowledge or ability (therefore, here too reading experience is advantageous).
3. The reader does not apply inappropriate interpretive strategies that will only confuse him.

Note, however, that these suggestions rely on very few stylistic devices and on a relatively crude classification, linguistically speaking. Future studies informed by more complex linguistic models, such as the LAD model (Castiglione 2017), may shed more light on this issue.

Relation to Previous Foregrounding Models

It appears most foregrounding theoreticians have not seriously considered the possibility of failure, or at the very least, have not found it theoretically interesting. However, there is at least one foregrounding model that does explicitly raise the possibility of failure – that of Leech and Short (2007). Leech and Short do not automatically assume that each stage leads to the next. Rather, they claim that writers' preferences could promote idiosyncrasies of style which draw readers' attention but have no discernible literary function. Leech and Short describe a logical relation between the stages from the type of A is a necessary but not sufficient condition for B. Every effect that has literary significance is due to psychological prominence that results from a deviation from a linguistic norm. The opposite is not necessarily true however: not every deviation from a linguistic norm leads to psychological prominence, and not every psychological prominence leads to literary meaning. In that, Leech and Short mark the two potential points of failure in the foregrounding process: linguistic deviation that does not lead to psychological prominence, and psychological prominence that does not lead to a literary effect. Despite having discerned the weak points in the process, Leech and Short have not developed this idea forward by addressing the failures themselves and putting their model to an

empirical test. What they wanted to emphasize was that because linguistic deviation is a necessary but insufficient condition for producing a literary effect, statistical linguistic analysis is not enough to determine whether the reader experiences foregrounding.

The current model, unlike that of Leech and Short (2007), does not only argue that such failure is possible, but explicitly identifies, classifies, and characterizes the types of failure. According to this model, foregrounding is not a linear process where each stage normally leads to the next. Rather, it is a “tree” of possibilities, which also represents what happens if the stages do not progress according to the optimal scenario.

Limitations of the Failed Foregrounding Model

A significant methodological weakness of this experiment lies in its sample size. Although 42 participants represent an adequate sample size for an RTA eye movement experiment, for a questionnaire study it is quite small, particularly if the sample is divided into three groups. This weakness is partly compensated for by the fact that the research involves several types of measures, and the fact that the findings seem to converge: The results of the different measures may be explained using the same model – the failed foregrounding model.

The classification system used in the model and the analysis presented here were intentionally broad. Aesthetic appraisal, shallow processing, and interpretation all capture a wide range of phenomena. The use of these generalizations has allowed the model to be relatively simple and elegant. Other researchers may choose to narrow or deconstruct the model into more specific types – according to affect, sensory experience, level of interpretation, etc. Following Miall and Kuiken (1994), it is possible to create a model based on the importance of affect rather than interpretation. Such a model may be more suitable to other stories where emotions are more salient. For Borges’ relatively dry, rational and enigmatic writing, however, the model was particularly suitable.

It is also possible to measure semantic noise, interpretation and aesthetic appraisal separately, and define different types of full, partial, and failed foregrounding according to different combinations of these three elements. In order to analyze the evidence collected from the interviewees, my classification approach and assumptions were good enough. It is conceivable, however that other classification systems be appropriate as well, or that other texts would have different effects requiring more appropriate classification systems.

The distinction made in this dissertation between different stylistic devices was relatively crude. These were not predetermined categories nor independent variables that the experiment had been especially designed to examine. “Linguistic difficulty”,

for example, is a category that can include a much broader range of devices than those found in this specific case. The LAD model, or other linguistic models like it, can serve, in future experiments as a basis for more subtle distinctions.

Note also that the method presented here for measuring the effectiveness of foregrounding focuses on the quantitative aspect, although the process clearly has a qualitative aspect as well. Obviously, not everyone who attains full foregrounding undergoes the same aesthetic experience or reaches the same interpretive insight. Let us take grammatical difficulties as an example. Although only few readers managed to decrypt them, the investment paid. The interviews showed that the participants who reached full foregrounding in these passages experienced impressive literary insights that contributed considerably to their interpretation of the story and overall aesthetic experience. Hence, effectiveness of the type presented here is only one aspect of foregrounding. Future studies may need to describe qualitative differences within that stage in the model as well.

Another limitation has to do with the partial adoption of the fluent processing hypothesis. The model assumes that high semantic noise (almost) never produces high aesthetic appraisal. This assumption may be contested by researchers supporting the radical aesthetician position, who believe literary experience to be directly derived from difficulty, miscomprehension, and repeated struggle against a recalcitrant text. To these researchers I can respond that at least in this case – first reading of a short story by Borges – such a phenomenon was not found. Instances of semantic noise were not associated with positive aesthetic appraisal, but only affected it negatively. “Failed foregrounding” was associated with negative appraisal, and an in-depth review of the interviews with the four readers who had adopted a strategy with failed foregrounding presented frustrated, confused and restless readers, far from experiencing “bliss”. The effect of experience was also incompatible with the predictions of the radical aesthetician position: experienced readers did not show higher frequency of failed and partial foregrounding, but of full foregrounding.

Nevertheless, the model does include an element that mixes semantic noise and positive aesthetic appraisal – “partial foregrounding”. It appears there is a certain affinity between partial foregrounding and a positive literary experience that mainly involves difficulty. They both seem to be located on the same spectrum, with partial foregrounding being a light case of Sisyphean but pleasurable reading. Nevertheless, the findings with regard to that element did not support it as a lead player in the field of literariness: its role in producing aesthetic appraisal was rather negligible; it was not a key characteristic of any of the stylistic devices checked; and was not a preferable reading strategy. Thus, the findings do not support the struggle against a recalcitrant text as an essential aspect of literary aesthetic experience.

Nevertheless, one cannot reject the possibility that in other situations, with different texts or readers, struggling against textual incomprehensibility may play a central or even positive role. This experience may be much more common among literary

scholars than among novices, with the former having been "trained" to derive satisfaction from it. For a reader of literary classics such as Goethe's *Faust* (particularly Part B) or Joyce's *Finnegan's Wake* it may be impossible not to adopt such a strategy. Readings of these kinds of texts lie beyond the scope of the model.

Alternatively, it may be that the difference here is not so large. Perhaps even experienced students of literature who come across a highly enigmatic text would read it for the first time in a way that could be described using the failed foregrounding model. They too may experience semantic noise and frustration. The difference is, however, that instead of giving up during or after the first reading, they would read again and again. Their training, their awareness of facing a masterpiece, their reference to representation levels distant from the story world, and other factors all combine to give them the power to move on. Subsequent readings will exceed the scope of the failed foregrounding model. First, as it is expected that the semantic noise would decrease gradually from reading to reading. And second, as this reading would be more similar to hermeneutic practices of the study of a holy text or the reading of a philosophical treatise, than to a first reading of a short story for pleasure.

Thus, it may be that the very areas in the text that create confusion and miscomprehension upon first reading would be discovered as particularly conducive to the dwelling effects of recurrent readings. The tight knots that resist opening in the first reading loosen up as the text is reread. Because the research presented here refers only to a first reading, it does not refute the possibility that additional reading would have revealed a more complex pattern. Although the findings support a relatively fluent processing, it is impossible to reject the possibility that higher semantic noise has belated effects in repeated reading, or after reflecting on and discussing the text further. It may be that precisely under such conditions, readers who had initially experienced higher semantic noise would be able to take the additional step towards significant interpretation and a positive aesthetic experience. Since the present study did not examine such texts or such readers, these are at present only hypotheses; additional research is required to characterize such readings.

The model presented here is not applicable to all types of literary experience. It is not designed to describe the only artistic technique, or to offer a "theory of everything". There may certainly be groups of readers or types of texts it does not describe well. Just as recurring readings may produce other effects, there may be literary texts that lack foregrounding devices and provide a fluent reading experience, where literary effects are derived from other factors. There may even be cases where writers produce certain aesthetic effects by deliberately encouraging shallow processing, whether by effects of suspense or by using punctuation or specific editing techniques. Finally, it may be that writers deliberately induce failed foregrounding, whether as a price they are willing to pay for text complexity or in order to produce an unusual aesthetic experience.

Why Has a Similar Model Not Been Developed Hitherto?

As shown above, potential failures in the foregrounding process were identified by Leech and Short (2007), and specific cases of such failures were reported by Miall and Kuiken (1994), and more recently by Emmott et al. (2006). Apparently, ignoring the failed foregrounding phenomenon has not been coincidental, but part of a theoretical approach.

The implied readers of foregrounding theory as optimal readers

The implied readers of the foregrounding theory can be described as optimal, at least in the following narrow sense: they make an effort to understand the text. When the implied readers of foregrounding theory run into some kind of difficulty, they do not quit nor skip the difficulty, but dwell on it to interpret it. Not only do they want to understand in depth, to interpret the linguistic deviations in the text, but they also have the cognitive resources and literary skills required for it.

The implied readers of foregrounding theory are no exception in a world of literary theories about the reader. Many theories assume some kind of reader optimality. According to Menachem Perry (1977), the readers raise hypotheses as they read, which they then test and update continuously, always ensuring that they explain a maximal number of detail most accurately. Such readers act like scientists by following the rules of confirming and refuting scientific theories. Similarly, the reader of Wolfgang Iser's reading theory (1978) follows an optimal and exhaustive reading strategy by not letting the spots (or places) of indeterminacy in the text remain indefinite, but takes the effort to "fill" them, so to speak, to produce a coherent and significant text.

There is no a priori problem in assuming optimal literary reading. It may be that in certain contexts, actual readers do act optimally by raising and testing hypotheses, investing maximal interpretive efforts to achieve the fullest understanding and carefully following interpretation rules aligned with the requirements of the specific text. However, there are reasons to suspect that optimal theories do not optimally describe the choices and conduct of most readers. This is not a blind spot of those theories, but the result of conscious choice. Researchers from the reader-response criticism school such as Perry and Iser had little interest in real-life readers. What they described was a kind of optimal interpretive move. It may be that this general tendency in literature studies has also affected the empirical study of foregrounding. Researchers focused more on interpretive success – or on the various affective and aesthetic effects – and less on interpretive and aesthetic failures. Hence, looking for the optimal readers and neglecting the rest.

Foregrounding and the humanities

Another reason for the lack of similar models may be found in the relation between foregrounding theory and the heritage of the humanities. By assuming a committed reader who makes a successful effort to overcome textual difficulties, foregrounding theory is not only an optimal theory of literature reading, but also an optimistic one. This optimism is part of the humanities tradition, which postulates a special place for the act of reading, viewing it as transformative and consciousness shaping, giving the readers wisdom, making them more moral, and enabling them to realize more of their humanity (Hakemulder 2018).

The findings presented here do not undermine these assumptions, but only limit them to a relatively small portion of readings. If the transformative potential of literature lies in deeper, full foregrounding readings, it is important to understand the conditions where they occur and the factors supporting them. It appears that writers, editors, researchers and teachers who want to encourage such readings face an uphill task. First, because the frequency of full foregrounding, at least in this sample, is low. Second, because shallow processing is a less adventurous strategy in terms of the reader's investment than starting a foregrounding process without any guaranty of completing it. Therefore, to achieve full foregrounding the reader needs to make a leap of faith. In the long run, failure and the semantic noise it involves may be worthwhile, as after gaining enough experience, readers may complete the process more often. Crossing this threshold, however, requires persistence. Perhaps training helps readers believe the effort is worthwhile. This challenge is exacerbated by the high frequency of shallow processing precisely among inexperienced readers, since many literature teachers and writers want to reach out not only to the experienced readers, but to the entire public.

References

- Ball, H. (1916) Dada Manifesto. *The Anarchist Library*. Retrieved from <https://theanarchistlibrary.org/library/hugo-ball-dada-manifesto.a4.pdf> on March 17 2019.
- Barthes, R. (1975). *The pleasure of the text*. Trans. R. Miller. New York: Macmillan.
- Borges, J. L. (1935). *Historia universal de la infamia*. Buenos Aires: Editorial Tor.
- Borges, J. L. (1998). *Collected Fictions*./Jorge Luis Borges/Translated by A. Hurley–New York: Penguin Group.
- Bredart, S., & Modolo, K. (1988). Moses strikes again: Focalization effect on a semantic illusion. *Acta Psychologica*, 67(2), 135-144.
- Campeato, L. (2011). The paradoxical role of noise in music. In Fourth International Conference of Students of Systematic Musicology, University of Cologne, Germany. October 5-7, 2011, (pp. 1-8)
- Castiglione, D. (2017). Difficult poetry processing: Reading times and the narrativity hypothesis. *Language and Literature*, 26(2), 99-121.
- Dixon, P., Bortolussi, M., Twilley, L. C., & Leung, A. (1993). Literary processing and interpretation: Towards empirical foundations. *Poetics*, 22(1-2), 5-33.
- Eger, N., Ball, L. J., Stevens, R., & Dodd, J. (2007, September). Cueing retrospective verbal reports in usability testing through eye-movement replay. *In Proceedings of the 21st British HCI Group annual conference on people and computers: HCI... but not as we know it-Volume 1* (pp. 129-137). British Computer Society.
- Elbabour, F., Alhadreti, O., & Mayhew, P. (2017). Eye tracking in retrospective think-aloud usability testing: Is there added value? *Journal of Usability Studies*, 12(3), 95-110.
- Emmott, C., Sanford, A. J., & Dawydiak, E. J. (2007). Stylistics meets cognitive science: Studying style in fiction and readers' attention from an interdisciplinary perspective. *Style*, 41(2), 204-225.
- Emmott, C., Sanford, A. J., & Morrow, L. I. (2006). Capturing the attention of readers? Stylistic and psychological perspectives on the use and effect of text fragmentation in narratives. *Journal of Literary Semantics*, 35(1), 1-30.

- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215-251.
- Ferreira, F., & Patson, N. (2007). The good enough approach to language comprehension. *Language and Linguistics Compass*, 1, 71–83. doi: 10.1111/j.1749-818X.2007.00007.
- Ferreira, F., Ferraro, V., & Bailey, K. G. D. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11(1), 11-15.
- Fish, S. (1980/1970). Literature in the reader: Affective stylistics. In *Is there a text in this class? The authority of interpretive communities* (pp. 21–67). Cambridge, MA: Harvard University Press.
- Guan, Z., Lee, S., Cuddihy, E., & Ramey, J. (2006, April). The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1253-1262). ACM.
- Hakemulder, F. (2018). The importance of confounds in research, The 16th IGEL Conference, Stavanger, Norway 25-28 July 2018.
- Hanauer, D. (1999). Attention and literary education: A model of literary knowledge development. *Language Awareness*, 8(1), 15-29.
- Harash, A., & Shen, Y. (2016). Reading as an obstacle race: Processing difficulty, semantic noise and the aesthetic experience. *Versus*, 45(1), 71-92.
- Holland, N. N. (1968). *The dynamics of literary response*. New York: W. W. Norton.
- Hopkins, D. (2004). *Dada and surrealism: A very short introduction*. Oxford: Oxford University Press.
- Hyrskykari, A., Ovaska, S., Majaranta, P., Rähkä, K. J., & Lehtinen, M. (2008). Gaze path stimulation in retrospective think-aloud. *Journal of Eye Movement Research*, 2(4), 1-18.
- Iser, W. (1978). *The act of reading: A theory of aesthetic response*. Baltimore: Johns Hopkins University Press.
- Jacobs, A. M. (2015). Towards a neurocognitive poetics model of literary reading. *Cognitive neuroscience of natural language use*, 135-59.

- Jauss, H. R. (1982). *Toward an aesthetic of reception* (T. Bahti, Trans.). Brighton, UK: Harvester Press.
- Knoop, C. A., Wagner, V., Jacobsen, T., & Menninghaus, W. (2016). Mapping the aesthetic space of literature “from below”. *Poetics*, *56*, 35-49.
- Leech, G. N., & Short, M. (2007). *Style in fiction: A linguistic introduction to English fictional prose*. Harlow, UK: Pearson Education.
- Ma'agatim (2019) the Historical Dictionary Project of the Academy of the Hebrew Language. search results for the word "הו" (it\this) by historical period, the years after 1919.
- Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky, E., Kuehnast, M., & Jacobsen, T. (2015). Towards a psychological construct of being moved. *PloS ONE*, *10*(6), e0128451.
- Miall, D. S., & Kuiken, D. (1994). Foregrounding, defamiliarization, and affect: Response to literary stories. *Poetics*, *22*(5), 389-407.
- Moles, A. (1966). *Information theory and esthetic perception*. Trans. J. E. Cohen. Urbana: University of Illinois Press.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231-259.
- Peskin, J. (1998). Constructing meaning when reading poetry: An expert-novice study. *Cognition and Instruction*, *16*(3), 235-263.
- Rajaraman, A., & Ullman, J. D. (2011). *Mining of massive datasets*. Cambridge: Cambridge University Press.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372-422.
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, *6*, 382–386. doi: 10.1016/S1364-6613(02)01958-7
- Sanford, A. J., Sanford, A., Molle, J., & Emmott, C. (2006). Shallow processing and attention capture in written and spoken discourse. *Discourse Processes*, *42*(2), 109-130.
- Shannon, C. E., & Weaver, W. (1963). *The mathematical theory of communication*. Champaign: University of Illinois Press.

- Shklovsky, V. (1965/1917). Art as technique. In L.T. Lemon and M. Reis (Eds.), *Russian formalist criticism: Four essays* (pp. 17-23). Lincoln: University of Nebraska Press.
- Sopčák, P. (2007). Creation from nothing': A foregrounding study of James Joyce's drafts for *Ulysses*. *Language and Literature*, 16(2), 183-196
- Steen, G. J. (1991). The empirical study of literary reading: Methods of data collection. *Poetics*, 20(5-6), 559-575.
- Tartakovsky, R. & Shen, Y. Flat as a wall or flat as a butcher's hand: The poetic simile and second-order semantic foregrounding. *Scientific Study of Literature* (forthcoming).
- Van Gog, T., Paas, F., Van Merriënboer, J. J., & Witte, P. (2005). Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied*, 11(4), 237.
- Van peer, W. (1986). *Stylistics and psychology: Investigations of foregrounding*. London: Croom Helm
- Van Peer, W. & Chesnokova, A. (2017). The Invisible Net, IGEL Paris Conference: What is Literariness? October 10-13 2017.
- Van Peer, W., Hakemulder, J., & Zyngier, S. (2007). Lines on feeling: Foregrounding, aesthetics and meaning. *Language and Literature*, 16(2), 197-213.
- Van Peer, W., Zyngier, S., & Hakemulder, J. (2007). Foregrounding: Past, present, future. In *Stylistics* (pp. 1-22). Leiden: Brill Rodopi.
- Vipond, D., & Hunt, R. A. (1984). Point-driven understanding: Pragmatic and cognitive dimensions of literary reading. *Poetics*, 13(3), 261-277.
- Woolf, V. (1972). *A haunted house and other short stories*. New York: Houghton Mifflin Harcourt.
- Yaron, I. (2010). Obscurity and Dylan Thomas's early poetry. Retrieved from: <http://cle.ens-lyon.fr/anglais/obscurity-and-dylan-thomas-s-early-poetry-106097>.
- Zyngier, S., Van Peer, W., & Hakemulder, J. (2007). Complexity and foregrounding: In the eye of the beholder? *Poetics Today*, 28(4), 653-682.

Appendix A:

Reading as an Obstacle Race:

Processing Difficulty, Semantic Noise and the Aesthetic Experience¹⁰

Amir Harash¹¹ and Yeshayahu Shen^{12 13}

Abstract

Three theories describe the relation between the relative difficulty of stimuli processing by readers and their aesthetic experience: foregrounding, the fluent processing hypothesis, and optimal arousal theories. The first argues that aesthetic pleasure derives from processing difficulty and the resulting delay; the second – from fluent and accurate processing; while the third suggests that there is an optimal level of difficulty where aesthetic experience is maximized. The three theories are contradictory both in their arguments and in their findings, which may be explained by the fact that each is informed by a different field. New studies, however, compare them within the same field (poetry or proverbs) and report mixed findings, which lend support to more than one explanation. The present article proposes “semantic noise” as a new variable that has not been taken into account in previous experiments, and that may explain some of the contradictory findings. Semantic noise is a communication failure not caused by a problem in the communication channel. The term refers to the interaction between certain textual characteristics and the reader who experiences them as undesirable (e.g. perceived errors, multiple and distracting meanings, or confusion between story characters). According to this concept we propose two hypotheses: (1) Literature experts will be more sensitive to semantic noise; and (2) A minimal difference between two stimuli can produce greater semantic noise than a big difference. These hypotheses generate predictions that differ from those of the foregrounding, fluent processing, and optimal arousal theories.

Keywords: aesthetic experience; foregrounding; semantic noise, fluent processing, optimal arousal .

1. Foreword

In the diverse and growing field of empirical aesthetics, the answer to one fundamental question remains controversial: How does the relative difficulty of processing artwork affect the reader’s (or viewer, listener, etc.) aesthetic experience? Three competing

¹⁰ *Versus*, 45(1), 71-92.

¹¹ Tel Aviv University, Israel, amirhara@yahoo.com

¹² Tel Aviv University, Israel, yshen@post.tau.ac.il

¹³ We thank the Porter Institute for Poetics and Semiotics and the Shirley and Leslie Porter School of Cultural Studies at Tel Aviv University for supporting Amir Harash's PhD. project, as well as The Israel Science Foundation administered by The Israel Academy of Sciences and Humanities, grant no. 119612 for Yeshayahu Shen. This text was translated from Hebrew by Ami Asher.

theories, which developed from diverse backgrounds, view the aesthetic experience as resulting from different causes and predict different outcomes. Foregrounding theory claims that aesthetic experience is caused by difficulty in processing and the resulting delay in reading. The fluent processing hypothesis claims the exact opposite: that aesthetic experience is facilitated by ease in processing the stimulus. Finally, optimal arousal theories, such as Daniel Berlyne's (1971) and his followers', claim that aesthetic experience culminates at a certain point where the artistic stimulus is not too familiar and not too unfamiliar, not too simple and not too complex.

All three theories are well-founded and successful, and all have managed to produce specific predictions and support them in a variety of areas, through prolonged research endeavors. And yet they all contradict each other in their claims and findings. Following a brief explanation of each, we will try to understand the reasons for this phenomenon. Then, we will propose a key variable excluded from existing theories that may explain some of their contradictory findings – semantic noise.

2. Foregrounding and Defamiliarization

One of the major directions in the empirical study of literature deals with textual characteristics that cause difficulty and delay. Originating in Russian Formalism (Shklovsky, 1965 [1917]), key concepts in those experiments and related theory include deautomatization, defamiliarization, salience and foregrounding (Van Peer, 1986; Miall and Kuiken, 1994; Emmott, Sanford and Morrow, 2006). According to Viktor Shklovsky's 1917 *Art as Technique*, art deautomatizes the perception of objects, and slows the processing down, turning it into an end in itself. Thus, for example, artistic usage of language is one that defamiliarizes the words. According to Shklovsky:

In studying poetic speech... we find material obviously created to remove the automatism of perception; the author's purpose is to create the vision which results from that deautomatized perception. A work is created 'artistically' so that its perception is impeded and the greatest possible effect is produced through the slowness of the perception. (1965 [1917]: 21)

In 1932 Jan Mukařovský coined in *The Esthetics of Language* the term "foregrounding" (in Czech: *aktualisace*) to refer to making a textual element stand out, or throwing it into relief against the background of the norms of ordinary language.

The first comprehensive empirical study designed to validate the theory was conducted by Willie van Peer in 1986. Van Peer describes two main types of foregrounding: deviation and parallelism. Each can operate on the semantic, syntactic or phonological level. Examples include rhyming, alliteration, neologisms, metaphors, irony, syntactically irregular structures, and syntactic inversions. Initially, studies were based mostly on self-reports, but in recent years Catherine Emmott and others (e.g. Emmott, Sanford and Morrow, 2006) began

using advanced attention research techniques, informed by the change blindness paradigm. They present subjects with two versions of the same text that differ only in one word; if subjects are blind to the change, then it is safe to conclude that the change locus has not been foregrounded.

The aesthetic experience evoked by the foregrounding process has been variously described and measured. In *Stylistics and Psychology* (1986), van Peer found that the main feature of highly foregrounded passages is strikingness, that is, readers found these passages to be more striking, salient, sharp, lively, and impactful. According to David S. Miall and Don Kuiken's *Foregrounding, Defamiliarization, and Affect* (1994), the main effect of foregrounding is affective response. According to this theory, textual deviation evokes a certain affective response in the reader, which in turn leads to reinterpretation. As the researchers put it, "the novelty of an unusual linguistic variation is defamiliarizing, defamiliarization evokes feelings, and feelings guide 'refamiliarizing' interpretative efforts" (Miall and Kuiken 1994: 392). They found that highly foregrounded passages were read more slowly and rated as more evocative of feeling.

In 2007, van Peer, Jèmeljan Hakemulder, and Sonia Zyngier proposed six scales of aesthetic experience: (1) aesthetic appreciation; (2) aesthetic structure; (3) cognitive aspects; (4) emotive aspects; (5) social context; and (6) attitudinal aspects. They found correlation between foregrounding devices in a text and between the structural, cognitive and attitudinal scales.

Another aesthetic experiential effect related to foregrounding is depth of appreciation, that is, improved evaluation ratings following the second reading of a given passage. This effect was found in studies of Jorge Luis Borges' "Emma Zunz" and Salman Rushdie's *The Satanic Verses* (Dixon, Bortolussi, Twilley and Leung, 1993; Hakemulder, 2004). A depth-of-appreciation effect was found only in versions with foregrounded passages; and in Dixon et al.'s study, it was found only among those with more prose reading experience. Hence, it was not mere repetition and in-depth reading that produced the aesthetic experience, since the effect was not found in versions from which foregrounding devices had been removed.

3. Processing Fluency

A considerable body of literature supports the theory that associates aesthetic judgment with processing fluency (e.g. Topolinski and Strack 2009; Reber, Schwarz and Winkielman 2004; Whiettlesea, Jacoby and Girard 1990; Zajonc 1968). Fluency is a function of the speed and accuracy of stimulus processing. Manipulations that increase processing speed without detracting from its accuracy – such as subliminal priming (prior unconscious exposure to the stimulus) or even a more legible handwriting – increase processing fluency.

This line of research may be called the hedonic fluency hypothesis (Bohrn, Altmann, Lubrich, Menninghaus, and Jacobs 2012) or hedonic marking of processing fluency (Winkielman, Schwarz, Fazendeiro, and Reber 2003). According to Rolf Reber and colleagues (2004: 365), “Aesthetic experience is a function of the perceiver's processing dynamics: The more fluently the perceiver can process an object, the more positive is his or her aesthetic response”. They argue that fluent processing is hedonically biased, in that it indicates a positive state of affairs in the world or in the cognitive system, and that this positive feeling is interpreted by the participant as related to the object viewed (or read), who in turn attributes an aesthetic value to it. This line of study also argues that processing fluency can explain multiple parameters linked in previous studies to the beauty of objects, such as symmetry, Gestaltian “good forms” and the clear differentiation between figure and ground.

In *The Pleasure of Reading*, Kringelbach, Vuust, and Geake (2008) argue that the pleasure of language and music is derived from the ability to foresee the next move. They rely on a cerebral model according to which there is feedback between low reception brain areas and higher areas that predict incoming information. When the prediction and the information actually received by the brain are incompatible the brain registers an error, requiring it to change its model. While listening to music the brain produces a range of expectations related to rhythm, tonality, harmony and melody. When these are satisfied, listeners experience internal reward, and when they are not, listeners experience tension. Accordingly, in *Predictive Coding of Music* (Vuust, Ostergaard, Pallesen, Bailey and Roepstorff 2009) researchers had participants listen to drum beats that involved rhythmic disorders, and located rapid brain responses that suggested inconsistency between prediction and incoming data, as well as a more delayed response of higher recognition of the error. These responses were heightened among jazz musicians as opposed to listeners without musical training. This line of argument is supported by an evolutionary reasoning: an organism able to predict the near future at a greater precision level will have higher chances of survival, and is therefore rewarded by internal pleasure for the ability to predict information received from the environment.

Isabel Bohrn et al. (2012) rely on findings in the visual perception area, as well as findings from the linguistic research area to apply this hypothesis to the study of literary reading. According to this hypothesis, an easily read text (low cognitive processing demands) will be preferred to a difficult one, with high demands, all other things being equal.

4. Optimal Arousal

Berlyne (1971) developed the research program known as psychobiological aesthetics, and grounded it in fundamental nervous-system characteristics. He argues that the optimal hedonic tone is a function of arousal. Based on neurobiological findings, he argues

for a reward system and an aversion system whose activation depends on the organism's level of arousal. The arousal potential of a given stimulus, in turn, depends on its properties, including novelty, surprise, complexity, ambiguity, and asymmetry. According to this theory, low arousal level will produce a positive hedonic experience, which will intensify up to a certain level where high arousal will also activate the aversion system. The greater the arousal, the more the negative response will grow, until eventually turning the experience into a distressful one. However, before the hedonic tone begins to fall, the function will reach a local maximum. This point is the optimum in an inverse U-shaped hedonic function. From this point on, any increase or decrease of arousal will negatively affect the hedonic experience. Berlyne himself tried to find support for his theory by studying the observation of simple geometric forms, as well as artworks. This theory formed the basis for an extensive empirical study, and is considered fundamental to the empirical esthetics area. It was also adjusted to the narrative structure and linguistic fields.

William Brewer and Edward Lichtenstein (1981, 1982) proposed the structural affect theory of stories. They argue that stories are a subgroup of narrative structures: those that arouse pleasure in the readers. They showed that readers tend to enjoy narratives that have a particular structure that increases arousal and then reduces it. On the other hand, stories which only increase arousal without reducing it in the end, or those that do not arouse at all, will be rated by readers as less enjoyable and also less "story-like" (1981). They describe three types of stories which have this general structure: stories that evoke tension, surprise or curiosity. As with Berlyne (1971), the arousal curve that represents the structural affect theory's predictions is shaped like an inverted U, with one difference: the horizontal axis represents the time of reading the story, rather than the stimulus's complexity or intensity.

In the figurative language area, Shen (2008) proposed a similar concept, according to which the semantic structure of various poetic figures of speech (simile, synthetic metaphor, oxymoron, etc.) reflects a kind of "compromise" between two constraints: on the one hand, the figurativeness interferes with cognitive conceptual structures by "conflating" two common conceptual domains. On the other, this very interference conforms to cognitive constraints. Thus for example, poetic similes conform to the constraint that the source term is, typically, more concrete than the target (as in: "Emptiness is like a weight, heavy on the heart"), rather than the other way around. Based on a comparison of hundreds of similes used in Hebrew, Arabic and Russian poems, Shen found that in almost all cases, novel similes appearing in these poetic corpuses conform to this cognitive constraint. In a series of experiments, Shen (1997) found that these similes were judged as more natural, easier to understand and more memorable than their inverted counterparts (in which a concrete concept was compared with an abstract one). Hence, similes used in poetry represent a kind of cognitive optimum: on the one hand they compare concepts belonging with different

categories (hence their innovation and aesthetic effect); on the other hand, this very interference with common categorization is subject to cognitive constraints that facilitate its communicativeness.

Influenced by Berlyne's theory, psycholinguist Rachel Giora formulated the Optimal Innovation Hypothesis (Giora et al. 2004). According to this hypothesis, a sentence will be optimally pleasurable if the response to it combines the activation of new meanings with the retention of familiar ones. Innovation should lead to a response that differs not only quantitatively but also qualitatively from the original response to the stimulus, but at the same time retain the latter, so that both will make sense and be identifiable and comparable. For example, the phrase "a peace of paper" retains the original meaning of "a piece of paper", but also adds a new one, suggesting that the peace agreement in question is worthless. On the other hand, variants such as "a single piece of paper" or "a sheet of paper" will not add new meanings and will therefore not represent optimal innovation. Conversely, if the meaning continues to stray away from the source, as in "a pill of pepper", nothing remains of the original meaning, rendering the innovation too great to be optimal. Support for this theory was found in experiments that asked participants to rank four variations on the same proverb for familiarity and pleasurability; the version in which a novel meaning was combined with an old one was rated the most pleasurable.

5. Are the Three Theories Truly Contradictory?

The three theories presented here in general terms use slightly different terminologies, and the terms textual diversion, linguistic innovation, familiarity, predictability, arousal, fluent processing, defamiliarization and foregrounding are not completely equivalent. However, we believe that each emphasizes a slightly different aspect of processing difficulty. Each locates aesthetic pleasure elsewhere on the scale it conceives between new and familiar, complex and simple, etc. and even if those scales are not completely identical, they are comparable. It should be noted that foregrounding and fluent processing are straightforwardly comparable on the processing difficulty axis; the optimal arousal theory can also be included in the same conceptual framework though with some elaboration. Indeed, we can find studies that combine the three theories, or two of them, as providing different predictions for the same experiment (Bohrn et al. 2012; Van Peer et al. 2007)

How are we to understand the existence of three well-founded theories that contradict each other both in their key concept and in their findings? One possible explanation is that each theory is applicable to a different media or genre according to the stimulus's level of complexity or processing difficulty. Along the range of textual fields, for example, we could say that foregrounding optimally describes poetry reading; that optimal arousal is suitable for

prose and proverbs; and that fluent processing is most appropriate for news stories or commercials. However, it would seem that even such a distinction fails to resolve the issue, as new studies suggest that even within a narrow generic area such as proverbs or poems it is difficult to decide between the three.

Bohrn et al.'s fMRI study ("Old Proverbs in New Skins"; 2012) attempted to decide between the three theories in the domain of proverbs. The researchers presented to participants proverbs of four levels of familiarity, including highly familiar ones such as "Rome was not built in a day", highly unfamiliar ones, substitutions for familiar proverbs that retain their meaning ("Rome was not erected in a day"), and variations that transform the meaning ("Rome was not destroyed in a day"). They assumed that the more familiar the proverb the quicker and more fluent its processing. In this case, every theory would have a different prediction regarding beauty ratings as a function of proverb familiarity. The fluent processing theory would predict a decreasing monotonous function: the less familiar the proverb, the less it is pleasurable. Foregrounding would predict an increasing function. Finally, optimal arousal would predict an inverse U-shaped function. Since Giora documented a function shaped similarly to an inverted U, it was expected that proverbs would be conducive to the optimal innovation theory.

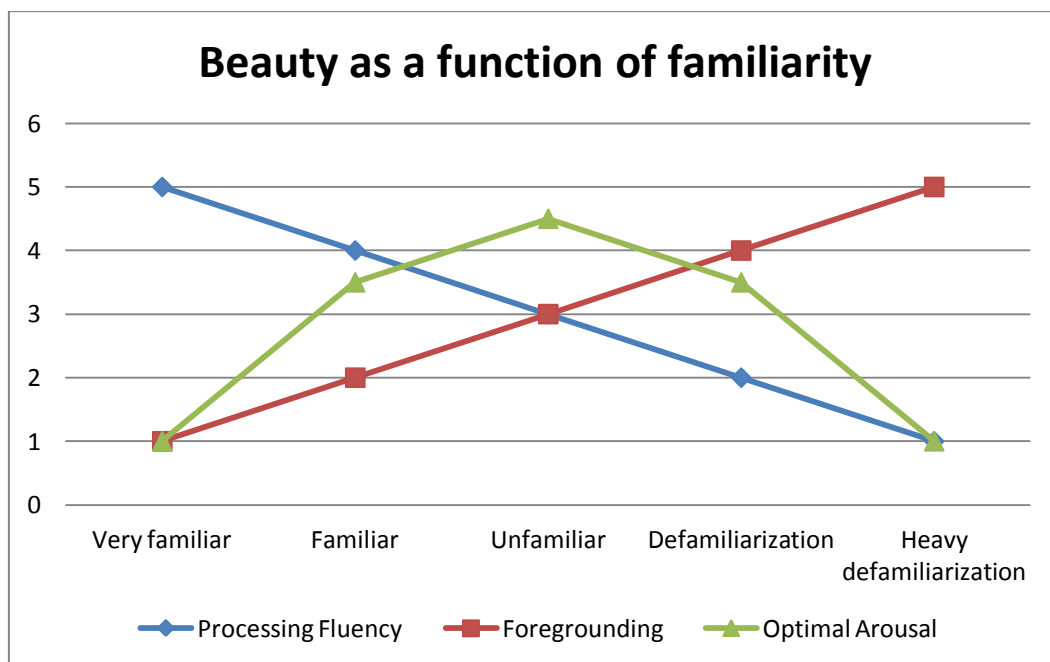


Figure 1. Predictions of the three theories regarding beauty as a function of familiarity

However, unlike Giora's studies, they found that the most familiar proverbs were rated as most beautiful by the participants. Moreover, they found no monotonous function, neither increasing nor decreasing, and no inverse U-shaped function either. Their function

does not support any of the theories. It begins as a decreasing function, but levels near the edge. The variation and substitution were rated without any significant difference between them and the control group of non-rhetorical sentences (however, note that the substitution proverb was rated a bit lower than the rest, albeit not significantly. We will elaborate on this finding below). The decreasing section of the graph would seem to support the fluent processing theory, but the cerebral findings shed a different light on the beauty ratings, and supported the foregrounding and optimal arousal theories.

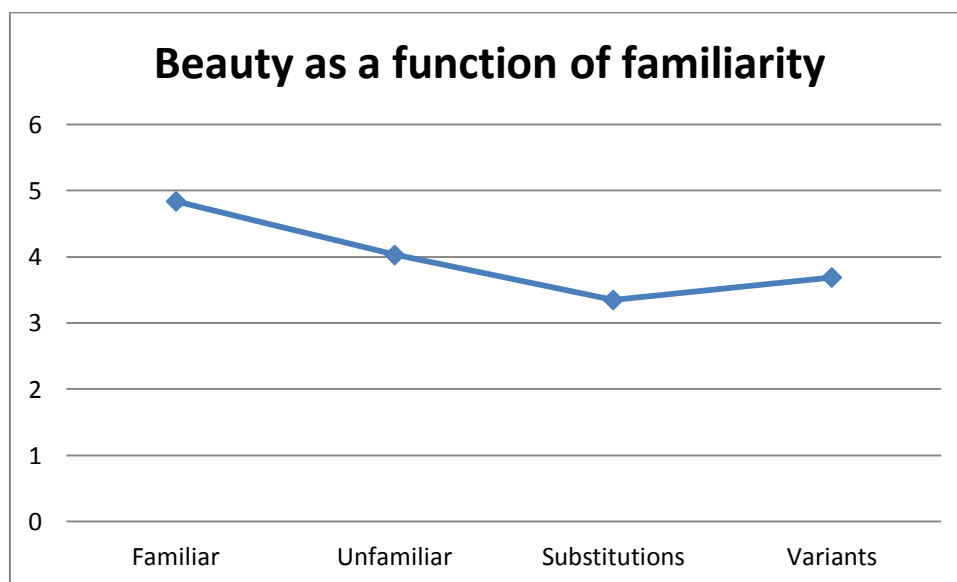


Figure 2. Beauty as a function of proverb familiarity (Bohrn et al. 2012)

Contrary to the predictions of the fluent processing hypothesis that argues for pleasant feelings as causing aesthetic appreciation, and in keeping with the foregrounding and optimal arousal theories, the unfamiliar proverbs activated emotional brain areas, while the familiar proverbs (rated, as seen in Figure 2 above, as the most beautiful) did not. Similarly, emotional areas were activated in a pattern supportive of foregrounding when comparing the variants with the substitutions. Thus, within the proverbs realm, and even within the same study, findings supported more than one theory. Below we will suggest a factor ignored by the current theories that may explain some of these contradictory findings.

Participants in a new study by Anna Chesnokova and Willie van Peer (2016, in this issue) read a little known poem by E. E. Cummings, with and without the original foregrounding devices that deviate from language norms. The poem was nine stanzas long and exceptionally rich in all kinds of deviations: syntactic, stylistic, grammatical, etc. For example, the “deviant” words were removed from the opening line, “anyone lived in a pretty how town”, to turn it into: “someone lived in a pretty old town”. Chesnokova and van Peer presented the poem in either of the two versions to three groups of readers with varying literary expertise and asked them to rate it on three different points on van Peer et al.’s (2007)

six-indicator scale. Significant differences between the versions were found in the emotive scales. Surprisingly, only the group with medium literary expertise rated the foregrounded text higher on the emotive scale, while the two other groups rated the un-foregrounded text higher. Thus, in terms of this indicator at least, this experiment supports the foregrounding theory only for one reader category, and supports other theories for inexperienced readers, as well as very experienced ones (literary department staff members).

The very difference between readers with varying levels of expertise is not new, however, as previous experiments found that the higher the expertise, the stronger the foregrounding effect. Here, however, the group that could have been expected to be most sensitive to foregrounding preferred the poem in its simplest form. This finding surprised Chesnokova and van Peer (2016, in this issue) to such a degree that they concluded their paper by suggesting "that the traditional and recent views on foregrounding may need certain revision".

A previous study by van Peer et al. (2007) also produced mixed results that supported both foregrounding and optimal arousal. The experiment involved six complexity variations on a single line from a poem by Portuguese poet Antero de Quental, the most complex of which was the original: "I feel shedding over my resting place icy tears of disbelief"; the simplest was: "I cry bitter tears of sadness on my bed". Out of the six indicators of esthetic experience, three were significantly correlated with complexity level, consistent with foregrounding, but one, the attitudinal indicator actually presented an inverted U-shaped curve, as predicted by optimal arousal.

Not only do those findings refute the idea that each theory is applicable to a separate area, but findings on the speed of reading literature point in a different direction. In *Lost in a Book*, Victor Nell (1988) found that those who read for the sake of pleasure read at a wide range of speeds, and sometimes their reading speed is much faster than usually found in laboratory experiments. Nell also found that the ratio between the highest and lowest speed ranged between 1.46 among slow readers and 7.79 among the fastest. The fast readers read at speeds of over 600-800 words per minute, which is considered as browsing or skimming. The very existence of this wide range, where the reader shifts between rapid and slow processing, suggests that in reading literature, there is an important role also for the faster reading, one which is beyond being just the background for the foreground. If the aesthetic experience relies mostly on the foreground, there is no point in so much background. On the contrary, it would seem that the same reader also derives pleasure from fluent, rapid processing, from passages in the book that are "page-turners" as well as from slow and careful processing, and also from passages read at medium speed. All are part of the aesthetic reading experience. If we liken literary reading to an obstacle course, it would seem that the reader derives pleasure both from the joy of fast running between the obstacles and from the effort of overcoming

them. In that sense, the optimal arousal principle is not a sufficient synthesis of the two other theories, because instead of describing the joy of both poles of the processing difficulty continuum, it places it in the middle, thus failing to explain cases in which aesthetic experience is indeed located in the poles.

6. Semantic Noise: The Missing Link?

One possible conclusion given the contradictory findings reviewed above is that there are additional variables that significantly affect the relationship between the esthetic experience and processing difficulty, which experiments have hitherto found difficult to hold constant. We propose one such variable – semantic noise. After defining the term, in the following section we will illustrate how some of the findings support its existence and effect. But first, a musical allegory and literary anecdote will help understand the importance of this missing link.

The term *semantic noise* is borrowed from information processing theories, but the word noise is actually a musical metaphor. The distinction between music and noise is a basic one that cuts across the entire musical field, one that was fiddled with and doubted during the 20th century, but nevertheless remained important and useful. In *Noise Water Meat: A History of Sound in the Arts*, Douglas Kahn (1999) describes the dynamics of the development of music as such, that when a new musical genre is formed, it is first perceived and branded by some of the public as “noise”. With time, the audience becomes used to the new genre or instrument, and experiences it as music, and when another innovation appears it is again seen as noise. Over the 20th century, this was true of jazz, rock-n-roll, and more.

This dynamics may be plotted on the time axis, but also divides the cultural space. Oriental music, for example, often sounds as no more than noise to listeners educated on western music, and vice versa. In a story called “Professor Leonardo”, Hebrew author Yaacov Churgin describes such a scene from the early 20th century, where classical music sounded like noise to a local audience in Palestine. In the scene, a musician called Professor Leonardo organizes a violin “concert” for an audience gathered in the sand in front of his house:

The bedsheet was down, in five minutes we would all find out the nature of that wondrous thing called “concert”. And the moment came – the two rang the bell. The curtain was moved all the way. This time, the professor showed up hatless, in the full glory of his locks, a violin and a bow in his hand. He pinched the strings several times, passed the bow over them once or twice, and then murmured something to the audience which nobody could figure out, and started.

We understood very little in music. All we could hear was a medley of whines and blasts and broken sounds. His head hither and thither in the storm, over

the violin, and his forest of locks rocked with him, crashing once against this ear and then against the other. When the sounds stopped and the noise ceased – a deep, embarrassed silence overcame the courtyard: it was the silence of disappointment. We all expected to see a “concert”, that is something surprising that we couldn’t even guess – and there it was, nothing more than plain playing! (Churgin 1990: 10)

This humoristic anecdote is illustrative of the idea that the same musical stimulus that the professor likens to a “very expensive phial of Parisian perfume” (8) may sound to listeners from another culture, or with little experience in the particular genre, as a “medley of whines and blasts”, or simply “noise”. The audience did expect something surprising, and was prepared for the new and unusual, but it seems they failed to interpret and conceive of the new and unusual in the violin concert as superb music, and could only experience it as an instrument played in the simplest way, just as people often say of modern artworks that they have been painted by three year-olds.

What is particularly important for the purpose of this article, however, is not the circumstances where the professor’s playing was considered as noise, but the listeners’ experience, particularly the proximity between the new and surprising on the one hand, and the grating and noisy, a proximity that demands our attention and understanding in the field of study we are dealing with.

7. Semantic Noise and Foregrounding

According to information theory models, communication is never an ideal process, since noises of various kinds disrupt effective communication. Shannon and Weaver’s (1963) communication model includes two types of noise that disrupt the information transfer between sender and receiver. The first is physical noise (also called channel or external noise) due to disturbances that distort the message: for instance, drilling sounds that disturb a conversation, or a blot of ink on a newspaper. The second kind is semantic – derived from the message itself and its decoding process. Semantic noise exists even when there is no distortion and the message received is technically identical to the message sent: when the message contains multiple meanings and ambiguities, and when the language in which it is coded is different than that in which it is decoded, when the signs are meaningless for the receiver, and in many other cases. “Noise” is sometimes defined in motivational terms, as part of the message the sender did not mean to send, or the receiver did not want to receive (Moles 1966). Throughout the years, the term semantic noise became foundational in the interpersonal communication area (Wood 2010), and has even been used in educational studies to explain misunderstanding and communication difficulties between teachers and students (Greene 1974; Pickett 1988).

Although there is no absolute physical or structural way of differentiating noises from signals, some features are commonly part of communication breakdown: mistakes, glitches, interruptions, etc. These features are not "noisy" by themselves, but only as they interact with the reader's actions, preferences or tasks. Frequent typos will bother those who read for pleasure, but if the task is proofreading, the errors will not be the "noise" but rather the "signal" and the literary plot or descriptions will only encumber the proofreader searching for errors. Thus, according to our understanding, noise will include congruence between certain characteristics of the text, or its processing, and the reader who experiences them as undesirable. These are processes where the reading produces meanings that do not contribute to the communication process: cumbersome turns of phrase, perceived errors, multiple and distracting meanings, confusion between characters, associations that disrupt the reading, and so on.

Although foregrounding theory does not refer directly to semantic noise, there seems to be an intimate link between the two concepts. A great many foregrounding theorists share a conception of information processing whereby structural deviation causes a kind of confusion, disorientation, unclear meaning or disruption in the readers' sense-making process (Van Peer, 1986; Van Peer et al. 2007; Miall and Kuiken, 1994; Leech and Short 2007). According to Mukařovský (2014 [1932]), the main function of language in daily life is communicative, but foregrounding exceeds this function, and is characterized by disturbing the communication process. Foregrounding crosses not only boundaries of ordinary language use but also those of the literary canon. Even some of the terms used by Mukařovský are reminiscent of noise. In his words, new poetic trends are perceived by the canon as a "distortion" of tradition and as "errors against the very essence of poetry" (2014 [1932]: 46). It is unclear, however, under which conditions the confusion, disturbance and discommunication will lead to an aesthetic experience and when they will have no added value for the reader and be seen as unnecessary in every sense, that is, semantic noise. According to Mukařovský ordinary language also has a low dosage of foregrounding. However, the difference between it and foregrounding in literature lies in that in literature, it is systematic and hierarchic. These factors may also be responsible for some of the differences between noise and foregrounding.

8. Semantic Noise in Experiments

We suggest that some of the difficulty in deciding among the competing theory lies in that not every difficult and faltering processing is worthwhile: sometimes it causes semantic noise, while in other cases the same textual stimulus leads to aesthetic, thrilling or pleasurable experience. Therefore, semantic noise is a variable that can affect experiments and must be taken into account and included in any theory that seeks to settle the current contradictions in

the literature. The question, when noise is experienced and when aesthetic pleasure is experienced, is complex and far beyond the scope of the present article. At the moment, however, we can only suggest two cases where semantic noise disrupted the experiments reviewed above, leading to predictions that conflict with those of the foregrounding, fluent processing hypothesis, and optimal arousal theories:

- (1) Literature experts will have heightened sensitivity to semantic noise.
- (2) A minimal difference between two stimuli can cause greater semantic noise than a big one.

(1) Heightened sensitivity among literary experts

As already mentioned, Chesnokova and van Peer's (2016, in this issue) experiment produced a surprising finding for foregrounding researchers: just as students with little literary experience, the literature department staff failed to appreciate the more complex form of Cummings' poem. But the surprise did not end here. Among the latter, the disagreement with the foregrounding theory was greater than among the former. Their evaluation gap between the two poem versions was larger. Can this finding be explained using the semantic noise concept?

Some of the foregrounding experiments found differences according to the participants' literary reading experience, but the underlying cause is not clear. Van Peer et al. (2007) suggest that the ability to detect the deviations on which the foregrounding process is dependent relies on profound familiarity with language norms, the result of frequent reading. If indeed, as they suggest, experienced readers detect such deviations with greater ease (just as jazz musicians are more sensitive to rhythmic deviations than unprofessional listeners), this still does not guarantee that the same readers experience these deviations as aesthetically pleasurable – it may be a necessary condition, but certainly not sufficient. It may be that they would experience them as semantic noise. Inasmuch as experienced readers identify more deviations in the text, they have a greater potential of having a full foregrounding experience, but also more semantic noise. Thus, rich literary experience suggests two divergent predictions: greater sensitivity to foregrounding and greater sensitivity to semantic noise. Even if we do not know yet when a certain passage would cause either, we may assume that for highly experienced readers, both experiences would be more extreme.

The graph below depicts the emotive level on three points in the text, and compares the original poem with the manipulated, simplified version; the curve indicates the effect size. The findings show that both the staff and the students experienced the linguistically

complicated version as less emotive, as well as that the trend among the staff is much stronger (and also statistically significant).

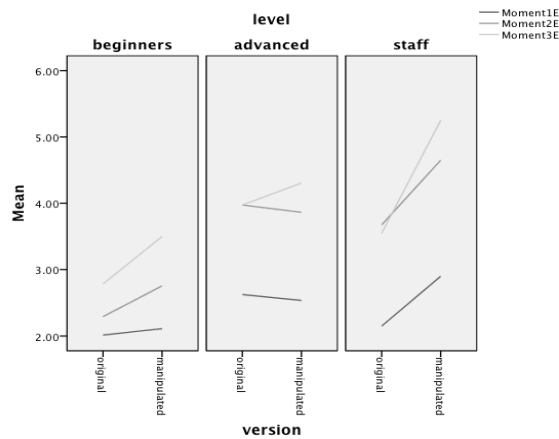


Figure 3. Emotive ratings based on poem version and literary expertise (Chesnokova and van Peer, 2016, in press)

The two groups differ in that having been exposed to many literary texts in the course of their career, the staff members may be expected to have developed aesthetic sensitivity. Part of that sensitivity is heightened sensitivity to noise. Therefore, had the staff members managed to experience the complication as aesthetic, they would probably have appreciated it more than all the other experimental groups, but otherwise, they should have experienced it as noise and consequently enjoy reading less than the other groups.



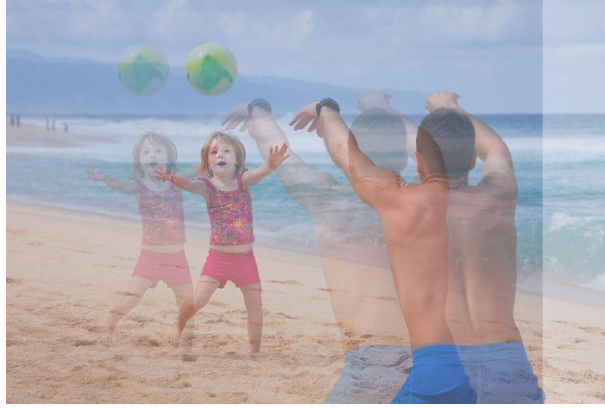


Figure 4. From top to bottom: the original photo; the same photo with doubling and a slight shift to the right; the same photo with a greater shift to the right

(2) Noise resulting from a minimal difference

Our hypothesis is that a minimal difference between two stimuli can cause greater confusion and semantic noise than a major one. If plotted on a graph, this prediction will look very different than all the theories reviewed above. As you may recall, in the graph of esthetic appreciation as a function of processing difficulty, the foreground and fluent processing theories predict a monotonous function, while the optimal arousal theory predicts an inverted U-shaped function. But in certain cases, as in the case of explicit or implicit comparison between two almost identical stimuli, the semantic noise concept predicts a U-shaped function. In this case, a minimal change may cause noisy confusion while a greater change may not be confusing because the two stimuli would be distinct enough. The following three photos demonstrate this phenomenon:

In the middle photo, you can see that when the image is doubled and there is a slight shift, you experience a blurring effect, as if the photographer's hand was shaking when he was using long exposure. This is visual noise where the deviation is not perceived as two superimposed images but as a disturbance within one. In the image below, you can see the two superimposed layers more clearly, and although you do experience visual noise, it is different than in the middle photo. To judge by our own experience, it appears that the middle photo is much more irritating than the lowest one. Thus, it is not always the greater distance from the original that creates an object more difficult to process. Certainly, a similar effect may be demonstrated in the music or sound area, whereby a minimal shift causes a sense of a disturbing echo or being off-scale, while a major shift is experienced as two clearly distinct acoustic objects.

In a different modality, a similar case may be found in Bohrn et al.'s (2012) experiment. The substitution proverb "Rome was not *erected* in a day" was considered closer

to the original and rated more familiar by the participants. Nevertheless, it was also rated as the least beautiful (compared to the original and semantic variation), albeit not significantly. This finding is predicted by none of the theories, as it creates a local U-shaped curve. Similarly, Giora et al. (2004) found that the proverb with the slight variation was considered the least beautiful (see Figure 5). Although in both cases the difference is not significant, but since it has been independently found in those two experiments, and contradicts every prediction, perhaps this tiny deviation that creates a kind of local U-shaped curve should be taken seriously.

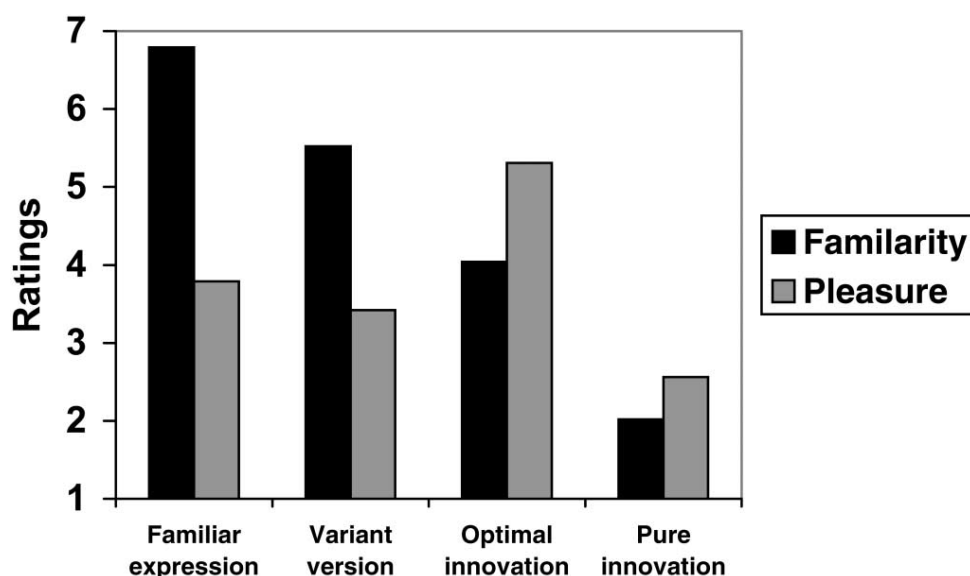


Figure 5. Familiarity and pleasure of four versions of the same proverb (Giora et al. 2004).

Note: What Giora calls a “variant” is what Bohrn et al. (2012) call “substitution”; and what she calls “optimal innovation” is what they call “variant”.

Bohrn et al.’s (2012) cerebral findings in the comparison between the slight variations on the same proverb (“information is power”) reveal that the more common ones activate cognitive areas related to error detection more intensely. They conclude:

Proverb-substitutions, which do not question the content of the corresponding familiar proverbs, did not recruit this moral emotion network; instead, they activate the right IPL, left fusiform gyrus, and the ACC which are associated with attention shifting, error detection, and conflict management... Functional data suggest that this less innovative condition may have been processed as containing errors (Bohrn et al. 2012: 10).

These findings are consistent with the explanation offered here, that the participants have seen thee alternative version of the proverb and compared it to the original, familiar one, experiencing it – at least at first, not as a distinctly different proverb, but as the same proverb with an error. This processing of the proverb as erroneous is an example of the effect of noise,

which in this case probably disturbed and detracted from the aesthetic experience. And since this phenomenon caused a disturbance in a direction not predicted by any of the theories reviewed here, it prevented the experimental findings from being unequivocal.

9. Conclusion

In this article we have presented the concept of *semantic noise* and suggested that it can explain inconsistent findings in studies of the relation between processing fluency and aesthetic experience. We have argued that this variable may account for the experimental findings that seem inconsistent, from the standpoint of the three theories presented: foregrounding, fluent processing hypothesis, and optimal arousal. Semantic noise is related to terms such as complexity, processing difficulty and comprehension difficulty, but is distinct from them. Not every case of high complexity or great processing difficulty leads to semantic noise, as evident in cases where the foregrounding process is successful and the reader has an aesthetic experience. We have also presented two hypotheses according to which taking semantic noise into account predicts novel hypotheses: heightened sensitivity to noise among experts, and a minimal change that causes confusion. Should these hypotheses be empirically supported, this would establish the importance of semantic noise as an independent variable.

Semantic noise, and noise in general, can contribute to the empirical study of esthetics because they are applicable to a variety of media and art forms other than literature, including music, cinema, and theater. Each medium is associated with different types of perceptual failures and disturbances, and the study of noise in art may reveal broad common denominators between media, as well as clearly distinguish between them on that basis.

We believe that the key question is what differentiates a sign that produces semantic noise from that which produces an esthetic experience. Further research is required to address this complex issue, but based on the studies and theories reviewed here, we can already suggest several directions or levels where the question may be answered: on the *sociological-cultural level*, culture-dependent listening habits, as in the case of western music in oriental ears demonstrated in Chorgin's story; *individual variables*, such as exposure to the text and literature reading experience in Dixon et al.'s and Chesnokova and van Peer's experiments; and variables related to the *framing and context* of the reading task, such as the experimenter's instructions affecting the degree and type of reader attention, or reading in on the web as opposed to printed format. Obviously, there are also variables related to the stimulus itself that help determine whether it is experienced aesthetically or as noise, such as hierarchy and internal logic in the use of linguistic deviations, as opposed to random use of such literary devices. Although existing theories take some of these factors into account, none of them considers the effects of semantic noise on the aesthetic experience.

If reading is an obstacle course, it is difficult to determine in advance whether the reader will experience a given obstacle as a challenge or a burden. Will overcoming it arouse pleasure or a sense of beauty, or will the same obstacle seem impassable, unnecessarily difficult or simply bothersome? Sometimes readers enjoy an almost obstacle-free course, where they can simply run through the text, while on other occasions they will enjoy the obstacles, so long as they are not impassable. In still other cases (or with still other readers) they want to conquer obstacles that seem impassable, try to figure out a way around them, and continue charging at them until they manage to reach a peak of interesting interpretation.

References

BERLYNE, D. E.

1971 *Aesthetics and Psychobiology*, New York, Appleton Century-Crofts.

BOHRN, I. C., ALTMANN, U., LUBRICH, O., MENNINGHAUS, W. and JACOBS, A. M.

2012 "Old Proverbs in New Skins—an fMRI Study on Defamiliarization", *Frontiers in Psychology*, n. 3, 204, pp. 1-18.

BORGES, J. L.

1964 "Emma Zunz", in *Labyrinths: selected stories and other writings*, Vol. 186, New York, New Directions Publishing.

BREWER, W. F. and LICHTENSTEIN, E. H.

1981 "Event Schemas, Story Schemas, and Story Grammars", *Attention and Performance*, n. IX, pp. 363-379.

1982 "Stories Are to Entertain: A Structural-Affect Theory of Stories", *Journal of Pragmatics*, n. 5, 6, pp. 473-486.

CHESNOKOVA, A., and VAN PEER, W.

2016 "Anyone Came to Live Here Some Time Ago: A Cognitive Semiotics Approach to Deviation as a Foregrounding Device", *Versus*, in press.

CHURGIN, Y.

1990 *Professor Leonardo*, Jerusalem, Keter.

DIXON, P., BORTOLUSSI, M., TWILLEY, L. C. and LEUNG, A.

1993 "Literary Processing and Interpretation: Towards Empirical Foundations", *Poetics*, n. 1, 22, pp. 5-33.

EMMOTT, C., SANFORD, A. J. and MORROW, L. I.

2006 "Capturing the Attention of Readers? Stylistic and Psychological Perspectives on the Use and Effect of Text Fragmentation in Narratives", *Journal of Literary Semantics*, n.1, 35, pp.1-30.

GIORA, R., FEIN, O., KRONROD, A., ELNATAN, I., SHUVAL, N. and ZUR, A.

2004 "Weapons of Mass Distraction: Optimal Innovation and Pleasure Ratings", *Metaphor and Symbol*, n. 2, 19, pp. 115-141.

GREENE, J. W.

1974 "On Rapping With the Inner-City School Dude", *Journal of Negro Education*, n. 43, pp. 310-319.

HAKEMULDER, J. F.

2004 "Foregrounding and its Effect on Readers' Perception", *Discourse Processes*, n. 2, 38, pp. 193-218.

KAHN, D.

1999 *Noise, Meat, Water: A History of Sound in the Arts*, Cambridge, MIT Press,.

KRINGELBACH, M. L., VUUST, P. and GEAKE, J.

2008 "The Pleasure of Reading", *Interdisciplinary Science Reviews*, n. 4, 33, pp. 321-335.

LEECH, G. N. and SHORT, M.

2007 *Style in Fiction: A Linguistic Introduction to English Fictional Prose*, Harlow, Pearson Education.

MIALL, D. S., and KUIKEN, D.

1994 "Foregrounding, Defamiliarization, and Affect: Response to Literary Stories", *Poetics*, n. 5, 22, pp. 389-407.

MOLES, A. A.

1966 *Information Theory and Esthetic Perception.*, Urbana, Univ. of Illinois Press.

MUKAŘOVSKÝ, J.

2014 "The Esthetics of Language" [1932], in *Chapters from the History of Czech Functional Linguistics*, J. Chovanec (ed.), Brno, Masarykova univerzita, pp. 41-52.

NELL, V.

1988 *Lost in a Book: The Psychology of Reading for Pleasure*, New Haven, Yale Univ. Press.

PICKETT, J. R.

1988 "Semantic Noise and the Classroom", *ETC: A Review of General Semantics*, n. 3, 45, pp. 278-280.

- REBER, R., SCHWARZ, N. and WINKIELMAN. P.
2004 "Processing Fluency and Aesthetic Pleasure: Is beauty in the Perceiver's Processing Experience?", *Personality and Social Psychology Review*, n.4, 8, pp. 364-382.
- RUSHDIE, S.
2011 *The Satanic Verses: A Novel*, New York, Random House,.
- SHANNON, C. E., and WEAVER, W.
1963 *The Mathematical Theory of Communication*, Champaign, Univ. of Illinois Press.
- SHEN, Y.
1997 "Cognitive Constraints on Poetic Figures", *Cognitive Linguistics*, 1, 8, pp. 33-72.
2008 "Metaphor and Poetic Figures." in *The Cambridge Handbook of Metaphor and Thought*, R. Gibbs Jr. (ed.), Cambridge, Cambridge Univ. Press, pp. 295-307.
- SHKLOVSKY, V.
1965 "Art as Technique" [1917], in *Russian Formalist Criticism: Four essays*, L. T. Lemon and M. Reis (eds.), Lincoln, Univ. of Nebraska Press, pp. 17-23.
- TOPOLINSKI, S. and STRACK, F.
2009 "The Analysis of Intuition: Processing Fluency and Affect in Judgements of Semantic Coherence", *Cognition and Emotion*, n.8, 23, pp. 1465-1503.
- VAN PEER, W.
1986 *Stylistics and Psychology: Investigations of Foregrounding*, London, Croom Helm.
- VAN PEER, W., HAKEMULDER, J. and ZYNGIER, S.
2007 "Lines on Feeling: Foregrounding, Aesthetics and Meaning", *Language and Literature*, n. 2, 16, pp. 197-213.
- VUUST, P., OSTERGAARD, L., PALLESEN, K. J., BAILEY, C. and ROEPSTORFF, A.
2009 "Predictive Coding of Music: Brain Responses to Rhythmic Incongruity", *Cortex*, n.1, 45, pp. 80-92.
- WHITTLESEA, B. W. A., JACOBY, L. L. and GIRARD, K.

1990 "Illusions of Immediate Memory: Evidence of an Attributional Basis for Feelings of Familiarity and Perceptual Quality", *Journal of Memory and Language*, n. 6, 29, pp. 716-732.

WINKIELMAN, P., SCHWARZ, N., FAZENDEIRO, T and REBER ,R.

2003 "The Hedonic Marking of Processing Fluency: Implications for Evaluative Judgment." in *The Psychology of Evaluation: Affective Processes in Cognition and Emotion*, J. Musch and K. C. Klauer (eds.), pp. 189-217

WOOD, J.

2010 *Interpersonal Communication: Everyday Encounters*, Boston, Cengage Learning.

ZAJONC, R. B.

1968 "Attitudinal Effects of Mere Exposure." *Journal of Personality and Social Psychology*, n. 2, part 2, 9, pp. 1-28.

Appendix B: Semantic Noise Questionnaire

1. The text confused me from time to time
2. I came across words or expressions in the story that were not familiar enough
3. In my opinion it was written cumbrously
4. Writing such as that helps the reader understand (reverse scored = RS)
5. Some of the sentences were phrased a bit problematically
6. The reading was smooth (RS)
7. I managed to be completely concentrated on my reading (RS)
8. There were sentences or words there that made me get stuck for a moment
9. Sometimes it wasn't all too clear to me what exactly it was all about
10. This kind of writing makes it difficult for the reader to understand
11. The story was well-paced (RS)
12. The text was complicated
13. The author expressed himself in a cumbersome and confusing manner
14. The text was written in a way that made reading easy (RS)
15. I think the author managed to deliver his intended message (RS)

שאלון רעש סמנטי

1. הטקסט בלבל אותי מידי פעם.
2. נתקלתי בסיפור במילים או ביטויים לא מספיק מוכרים.
3. לטעמי זה היה כתוב בצורה מסורבלת (שאלה הפוכה).
4. כתיבה כמו זאת עוזרת לקורא להבין.
5. לחלק מהמשפטים היה ניסוח בעייתי במקצת.
6. הקריאה זרמה חלק (שאלה הפוכה).
7. הצלחתי להיות מרוכז לגמרי בקריאה (שאלה הפוכה).
8. היו שם משפטים או מילים שגרמו לי להיתקע לרגע.
9. לפעמים לא היה לי ברור על מה בדיוק מדובר.
10. כתיבה כמו זאת מקשה על ההבנה של הקורא.
11. לסיפור היה קצב טוב (שאלה הפוכה).
12. הטקסט היה מסובך.
13. הכותב התבטא בצורה מגושמת ומבלבלת.
14. הטקסט היה כתוב באופן שמקל על הקריאה (שאלה הפוכה).
15. לדעתי הכותב הצליח להעביר את מה שהוא רצה והתכוון (שאלה הפוכה).

Appendix C: Aesthetic Appraisal Questionnaire

1. I enjoyed reading the story
2. The story excited me
3. The story was beautiful
4. Sometimes, the story made me laugh
5. The story was senseless (RS)
6. This story has value as a literary work
7. The story was scary
8. The story moved me
9. The story was interesting
10. The story has a poetic feel to it
11. The story bored me (RS)
12. Sometimes the story saddened me
13. The story was original
14. The plot was interesting/ special
15. The use of language was interesting/ special
16. I loved the style
17. The story has a profound meaning
18. The story felt too long to me (RS)
19. This was a thought-provoking story
20. The story is well-written
21. I'd recommend it to a friend who loves to read
22. I'd like to reread the story
23. Reading this was a waste of time (RS)
24. I'd like to read more stories by the same author
25. The story was surprising

שאלון הערכה אסתטית

1. נהייתי לקרוא את הסיפור.
2. הסיפור ריגש אותי.
3. הסיפור היה יפה.
4. לפעמים הסיפור הצחיק אותי.
5. הסיפור היה סתמי (שאלה הפוכה).
6. יש לסיפור הזה ערך כיצירה ספרותית.
7. הסיפור היה מפחיד.
8. הסיפור היה נוגע ללב.
9. הסיפור היה מעניין.
10. היתה לסיפור תחושה פיוטית.
11. הסיפור שעמם אותי (שאלה הפוכה).
12. לפעמים הסיפור העציב אותי.
13. הסיפור היה מקורי.
14. העלילה היתה מעניינת / מיוחדת.
15. השימוש בשפה היה מעניין / מיוחד.
16. אהבתי את הסגנון.
17. יש לסיפור משמעות עמוקה.
18. הסיפור הרגיש לי ארוך מידי (שאלה הפוכה).
19. זה היה סיפור מעורר מחשבה.
20. הסיפור כתוב טוב.
21. הייתי ממליץ עליו לחבר שאוהב לקרוא.
22. הייתי רוצה לקרוא שוב את הסיפור.
23. לקרוא את זה היה בזבוז זמן (שאלה הפוכה).
24. הייתי רוצה לקרוא עוד סיפורים של אותו סופר.
25. הסיפור היה מפתיע.

Appendix D: Names of Authors and Distractors in the Hebrew Version of the Author Recognition Test

Authors	סופרים
Marcel Proust	מרסל פרוסט
Irit Linur	עירית לינור
Amos Oz	עמוס עוז
Assaf Inbari	אסף ענברי
Damon Runyon	דיימון ראניון
Dr. Seuss	ד"ר סוס
Lewis Carroll	לואיס קארול
Liad Shoham	ליעד שוהם
Philip K. Dick	פיליפ ק דיק
Italo Calvino	איטאלו קאלוינו
Vladimir Nabokov	ולדימיר נבוקוב
Dror Burstein	דרור בורשטיין
Bruno Schulz	ברונו שולץ
Yossel Birstein	יוסל בירשטיין
Toni Morrison	טוני מוריסון
Shin Ben-Tzion	ש. בן ציון
Haim Sabato	חיים סבתו
Jorge Amado	ז'ורז' אמדו
Shel Silverstein	של סילברשטיין
Miriam Roth	מרים רות
Franz Kafka	פרנץ קפקא
Mikhail Bulgakov	מיכאל בולגקוב
Yehuda Burla	יהודה בורלא
Paulo Coelho	פאולו קואלו
George R. R. Martin	ג'ורג' ר ר מרטין
Ofir Touche Gafla	אופיר טושה גפלה
Ram Oren	רם אורן
Alona Frankel	אלונה פרנקל
Yehoshua Bar-Yosef	יהושוע בר יוסף
Richard Bach	ריצ'ארד באך
Ronit Matalon	רונית מטלון
Batya Gur	בתיה גור
Leo Tolstoy	לב טולסטוי
Orson Scott Card	אורסון סקוט קארד
Chinghiz Aitmatov	צ'ינגיס אייטמטוב
Harlan Coben	הרלן קובן
Uri Nisan Gnessin	אורי ניסן גנסיין
Albert Cohen	אלבר כהן
Kurt Vonnegut	קורט וונגוט
Yoel Hoffmann	יואל הופמן
H. G. Wells	ה. ג. וואלס
Dan Brown	דן בראון
Tamar Bornstein-Lazar	תמר בורנשטיין לזר
Agatha Christie	אגתה כריסטי
Antoine de Saint-Exupéry	אנטוואן דה סנט אכזופרי

Arthur Conan Doyle	ארתור קונן דויל
J. R. R. Tolkien	ג. ר. ר. טולקין
Jack London	ג'ק לונדון
Haruki Murakami	הארוקי מורקאמי
John Grisham	ג'ון גרישם
Michael Ende	מיכאל אנדה
Eduardo Mendoza	אדוארדו מנדוסה
Stieg Larsson	סטיג לרסון
Raymond Chandler	ריימונד צ'נדלר
Yigal Mossinson	יגאל מוסינזון
Wilhelm Busch	וילהלם בוש
Galila Ron-Feder Amit	גלילה רון פדר
Jo Nesbø	יו נסבו
Douglas Adams	דאגלס אדאמס
A. A. Milne	א א מילן

Distractors	מסיחים
Carol Popp de Szathmáry	קרול פופ סטמארי
Sliman Mansour	סולימאן מנסור
Calvin Coolidge	קלווין קולידג'
Mula Ben-Haim	מולה בן חיים
Binyamin Temkin	בנימין טמקין
Aharon April	אהרון אפריל
Peretz Bernstein	פרץ ברנשטיין
Justin Henry	ג'סטין הנרי
Rachel Cohen-Kagan	רחל כהן כגן
Andrew Jackson	אנדרו ג'קסון
Theodor L. Feininger	תיאודור ל פיינינגר
Fritz von Uhde	פריץ פון אוהדה
Joaquin Phoenix	חואקין פיניקס
Wassily Kandinsky	וסילי קנדינסקי
Rutherford Hayes	רוטפורד הייז
Otto Dix	אוטו דיקס
Jasper Johns	ג'ספר ג'ונס
Shmuel Ben David	שמואל בן דוד
James Monroe	ג'יימס מונרו
Eşref Armağan	אשרף ארמאן
Avraham Poraz	אברהם פורז
Felix Nussbaum	פליקס נוסבאום
Tilda Swinton	טילדה סווינטון
Melissa Leo	מליסה לאו
David Remez	דוד רמז
Yaniv Weizman	יניב ויצמן
Etai Pinkas	איתי פינקס
Pinchas Rosen	פנחס רוזן
Nissan Slomiansky	ניסן סלומינסקי
Franz Marc	פרנץ מארק
Viktor Vasnetsov	ויקטור וסנצוב
John Quincy Adams	ג'ון קווינסי אדמס

Olympia Dukakis	אולימפיה דוקאקיס
Randy Quaid	רנדי קווייד
Charlie Biton	צארלי ביטון
Richard S. Castellano	ריצ'ארד ס. קסטלנו
James Buchanan	גיימס ביוקנן
Boris Lekar	בוריס לקר
Woodrow Wilson	וודרו ווילסון
Zvi Lurie	צבי לוריא
Shlomo-Yisrael Ben-Meir	ישראל שלמה בן מאיר
Mercedes Ruehl	מרסדס רוהל
William McKinley	וויליאם מקינלי
Richard Farnsworth	ריצ'רד פרנסוורת'
Avraham Katznelson	אברהם כצלנסון
M. G. Tuchman	מ. ג. טוכמן
Ruth Gordon	רות גורדון
Constantin Daniel Rosenthal	קונסטנטין ד רוזנטל
Dr. Ticho	ד"ר טיכו
John H. Twachman	ג'ון ה טווצ'טמאן
Peretz Bernstein	פרץ ברנשטיין
Walid Haj Yahia	ואליד חאג'י-יחיא
Emil Nolde	אמיל נולדה
Gustave Courbet	גוסטב קורבה
Igael Tumarkin	יגאל תומרקין
Sergei Eisenstein	סרגי אייזנשטיין
Zvi Hendel	צבי הנדל
Paul Signac	פול סיניאק
Rachel Weisz	רייצ'ל וויס
Amelia Nolde	אמיליה נולדה

Appendix E:

THE CHAMBER OF STATUES

Jorge Luis Borges

In the early days, there was a city in the kingdom of the Andalusians where their monarchs lived and its **name was Labtayt, or Ceuta, or Jaén** [Linguistic Difficulty]. In that city, there was a strong castle whose double gate was designed neither for going in nor for coming out, but for keeping closed. And whenever a King died and another King took the Kingship after him, with his own hands, he set a new and strong lock to that gate, till there were four-and-twenty locks upon the castle, according to the number of Kings. After this time, an evil man, who was not of the old royal house, took the throne by force and instead of adding a lock, he had a mind to open these locks, that he might see what was within the castle.

The Vizier and Emirs implored him to desist; they hid from him the iron key ring and told him that it was much easier to add a new lock to the gate than to force four-and-twenty, but he cunningly repeated his words, saying, “It is my wish to examine the innards of this castle”. Then they offered him all that their hands possessed of monies and treasures and things of price, of flocks, of Christian idols, of gold and silver, if he would but refrain; still, he would not be baulked. So he pulled off the locks **with his right hand (may it burn through all eternity!)** [Author Comment] and entering, found within the castle figures of Arabs on their horses and camels, habited in turbands hanging down at the ends, with swords in baldrick-belts hanging down from their ears and bearing long lances in their hands. All these figures were sculpted and threw shadows on the ground, **and a blind man could identify them by the mere touch of the hand,** [Figurative Description] and the horses’ hooves did not touch the ground yet they did not fall aground, as though they were standing on their hind legs.

These exquisite figures filled the king with great amazement; even more wonderful was the excellent order and silence that one saw in them, for every figure's head was turned to the same side, the west, while not a single human voice or clarion sound was heard. Such was the first room in the castle. In the second, the king found the table that belonged to **Suleyman, son of David — salvation be with both of them!** —

[Author Comment] This table was carved from a single grass-green emerald, a stone whose occult properties are indescribable yet genuine, for it calms the tempest, preserves the chastity of its wearer, keeps off dysentery and evil spirits, brings favorable outcome to lawsuits, **and is of great relief in childbearing** [Linguistic Difficulty]. In the third room, two books were found: one was black and taught the virtues of each metal, each talisman, and each day, together with the preparation of poisons and antidotes; the other was white, and though the script was clear, its lesson could not be deciphered.

In the fourth room found he a mappamundi figuring the earth, the towns, the seas, the castles and the perils, each with its true name and exact shape. In the fifth, they found a marvelous mirror, great and round, of mixed metals, which had been made for Suleyman, son of David — salvation be with both of them!— **wherein whoso looked saw the faces of his fathers and the faces of his sons** [Figurative Description], from the first Adam to those whose ear shall hear the Doomsday Trumpets. The sixth room was filled with that hermetic powder, one drachma of which elixir can change three thousand drachms of silver into three thousand drachms of gold. The seventh appeared empty, and it was so long that **the ablest of archers, had he loosed an arrow from its doorway, would not have hit its end** [Figurative Description].

Carved on that far wall, they saw a terrible inscription. The king examined it, and understood it, and it spoke in this wise: "If any hand opens the gate of this castle, the warriors of flesh at the entrance, who resemble warriors of metal, shall take possession of the kingdom." These things occurred in the eighty-ninth year of the **Hegira** [Linguistic Difficulty]. Before the year reached its end, Tarkio would conquer that city and slay this King after the sorriest fashion and sack the city and make prisoners of the women and boys therein and get great loot. Thus it was that the Arabs spread all over the cities of Andalusia — a kingdom of fig trees and watered plains in which no man suffered thirst. As for the treasures, it is widely known that Tarik, son of Zayid, sent them to his lord, the caliph who entombed them in a pyramid.

(From the Book of the Thousand Nights and a Night, Night 272)

Note: The above text is based on Andrew Hurley's English translation (Borges, J. L. 1998) after having been adjusted according to the Hebrew version presented to the interviewees, which is closer to the Spanish original. The underscored sentences are the key points that the interviewees reported during the retrospective think aloud protocol.

חזר הפסלים

ח.ל בורחס

בשכבר הימים הייתה במלכות האנדלוסים עיר אשר בה ישבו מלכיה ושמה לבטיט או סוטה או חאן. וטירה מבוצרת היתה בעיר הזו, אשר שיערה הכפול לא נועד לבוא או לצאת בו, כי אם להינעל בלבד. כל אימת שנפטר מלך ואחר ירש את כסאו הרם, הוסיף זה במו ידיו מנעול חדש לשער, עד אשר הגיע מספר המנעולים לעשרים וארבעה, מנעול לכל מלך. ברבות הימים אירע שאיש רשע אחד, אשר מוצאו לא היה מבית המלכות, גזל את הכס בכוח, ותחת להוסיף מנעול, ביקש לפתוח את עשרים וארבעת המנעולים הקודמים, למען יראה את תוכן הטירה ההיא.

הנזיר והאמירים התחננו לפניו לבל יעשה כדבר הזה והחביאו מפניו את מחזיק המפתחות העשוי ברזל, באמרם כי נקל להוסיף מנעול אחד מלפרוץ עשרים וארבעה, אולם הלה חזר על דבריו בערמה: "חפץ אני לבחון את תוכן הטירה הזו". אז הציעו לו אוצרות ככול אשר השיגה ידם במקנה, בפסלי נוצרים, בכסף וזהב, אך מיאן להיכנע ופתח את השער ביד ימינו (שתבער לעולמי עולמים!) בפנים מצאו דמויות של ערבים במתכת ובעץ, על גבי גמלים וסוסים זריזים, ולהם צניפים המתבדרים על גבם וחרבות המשתלשלות מן האוזניים והחנית הישרה בימנים. כל הדמויות האלה היו מפוסלות והטילו צללים על הרצפה, ועיוור היה מכירם במגע יד בלבד, ורגלי הסוסים לא נגעו באדמה והמה לא צנחו ארצה, משל ניצבו על רגליהם האחוריות.

מורא גדול הטילו על המלך הדמויות הנהדרות הללו, ויותר מהן הסדר והשקט המופתי אשר ניפרו בהן, שכן כולן ניבטו לצד אחד, צד מערב, ולא נשמע לא קול אדם ולא צליל חצוצרה. זה היה בחדר הראשון בטירה. בשני עמד שולחנו של שלמה בן דוד – יזכו שניהם לגאולה שלמה! – חצוב באזמרגד יחיד, אשר צבעו, כידוע, ירוק הוא, ואשר את סגולותיו אין לתאר, אף כי ממשיות הן שהרי בכוחו לשפך את הסערות, לשמור על טהרת בעליו, למגר את חולי המעיים ואת הרוחות הרעות, להוציא משפט חסד ורב עזרו בלידות. בשלישי מצאו שני ספרים: האחד היה שחור ולימד את סגולות המתכות, הקמעות וימות השנה, וכן רקיחת הרעלים ומפיגיהם; והאחר היה לבן וְחֹכְמָתוֹ לֹא נִיתְנָה לַפְעֵנוֹחַ, אף על פי שכתבו היה ברור.

ברביעי מצאו מפה של העולם ובה הממלכות, הערים, הימים, הטירות והסכנות, כל דבר בשמו האמיתי ובדמותו המדויקת. בחמישי מצאו מראה מעוגלת, מעשה ידי שלמה בן דוד – יזכו שניהם לגאולה שלמה! – אשר מחירה רב מאוד, שכן עשויה היתה מתכות שונות והמתבונן בה ראה את פני אבותיו ופני בניו, למן אדם הראשון ועד לאלה אשר אוזנם תשמע את חצוצרות יום הדין. השישי היה מלא בשיקוי אשר די בטיפה אחת ממנו להפוך שלושת אלפי אונקיות כסף לשלושת אלפי אונקיות זהב. השביעי נראה להם ריק וכה ארוך היה, עד כי הטוב בקשתים, אם היה משלח חץ מפתחו, לא היה קולע לסופו.

על הקיר הקיצוני היתה חקוקה כתובת נוראה. המלך בחן אותה וירד לסוף כוונתה, וזה היה דברה: "אם תפתח יד אדם את שער הטירה הזו יבואו לוחמים בשר ודם הדומים בכל ללוחמי המתכת שבכניסה וישתלטו על הממלכה." כל הדברים האלה ארעו בשנת 89 להיג'רה. לפני תום השנה כבש טריק את המבצר הזה והביס את המלך הזה ומכר את נשותיו ובניו לעבדים והשם את אדמותיו. כך פשטו הערבים במלכות אנדלוסיה, על תאג'יה ושדותיה הרבים, אשר לא ידעו את החורב. ואשר לאוצרות, נודע כי טריק בן זיאד שלחם אל הח'ליף, אדונו, אשר שמר אותם בתוך פירמידה.

(מתוך אלף לילה ולילה, לילה 272)

La cámara de las estatuas

J. L. Borges

En los primeros días había en el reino de los andaluces una ciudad en la que residieron sus reyes y que tenía por nombre Lebtit o Ceuta, o Jaén. Había un fuerte castillo en esa ciudad, cuya puerta de dos batientes no era para entrar ni aun para salir, sino para que la tuvieran cerrada. Cada vez que un rey fallecía y otro rey heredaba su trono altísimo, éste añadía con sus manos una cerradura nueva a la puerta, hasta que fueron veinticuatro las cerraduras, una por cada rey. Entonces acaeció que un hombre malvado, que no era de la casa real, se adueñó del poder, y en lugar de añadir una cerradura quiso que las veinticuatro anteriores fueran abiertas para mirar el contenido de aquel castillo. El visir y los emires le suplicaron que no hiciera tal cosa y le escondieron el llavero de hierro y le dijeron que añadir una cerradura era más fácil que forzar veinticuatro, pero él repetía con astucia maravillosa: "Yo quiero examinar el contenido de este castillo". Entonces le ofrecieron cuantas riquezas podían acumular, en rebaños, en ídolos cristianos, en plata y oro, pero él no quiso desistir y abrió la puerta con su mano derecha (que arderá para siempre). Adentro estaban figurados los árabes en metal y en madera, sobre sus rápidos camellos y potros, con turbantes que ondeaban sobre la espalda y alfanjes suspendidos de talabartes y la derecha lanza en la diestra. Todas esas figuras eran de bulto y proyectaban sombras en el piso, y un ciego las podía reconocer mediante el solo tacto, y las patas delanteras de los caballos no tocaban el suelo y no se caían, como si se hubieran encabritado. Gran espanto causaron en el rey esas primorosas figuras, y aun más el orden y silencio excelente que se observaba en ellas, porque todas miraban a un mismo lado, que era el poniente, y no se oía ni una voz ni un clarín. Eso había en la primera cámara del castillo. En la segunda estaba la mesa de Solimán, hijo de David —¡sea para los dos la salvación!—, tallada en una sola piedra esmeralda, cuyo color, como se sabe, es el verde, y cuyas propiedades escondidas son indescriptibles y auténticas, porque serena las tempestades, mantiene la castidad de su portador, ahuyenta la disentería y los malos espíritus, decide favorablemente un litigio y es de gran socorro en los partos.

En la tercera hallaron dos libros: uno era negro y enseñaba las virtudes de los metales de los talismanes y de los días, así como la preparación de venenos y de contravenenos; otro era blanco y no se pudo descifrar su enseñanza, aunque la escritura era clara. En la cuarta encontraron un mapamundi, donde estaban los reinos, las ciudades, los mares, los castillos y los peligros, cada cual con su nombre verdadero y con su precisa figura.

En la quinta encontraron un espejo de forma circular, obra de Solimán, hijo de David —¡sea para los dos la salvación!—, cuyo precio era mucho, pues estaba hecho de diversos metales y el que se miraba en su luna veía las caras de sus padres y de sus hijos, desde el primer Adán hasta los que oirán la Trompeta. La sexta estaba llena de elixir, del que bastaba un solo

adarme para cambiar tres mil onzas de plata en tres mil onzas de oro. La séptima les pareció vacía y era tan larga que el más hábil de los arqueros hubiera disparado una flecha desde la puerta sin conseguir clavarla en el fondo. En la pared final vieron grabada una inscripción terrible. El rey la examinó y la comprendió, y decía de esta suerte: "Si alguna mano abre la puerta de este castillo, los guerreros de carne que se parecen a los guerreros de metal de la entrada se adueñarán del reino".

Estas cosas acontecieron el año 89 de la hégira. Antes que tocara a su fin, Tárik se apoderó de esa fortaleza y derrotó a ese rey y vendió a sus mujeres y a sus hijos y desoló sus tierras. Así se fueron dilatando los árabes por el reino de Andalucía, con sus higueras y praderas regadas en las que no se sufre de sed. En cuanto a los tesoros, es fama que Tárik, hijo de Zaid, los remitió al califa su señor, que los guardó en una pirámide.

(Del Libro de las 1001 Noches, noche 272)

TEL AVIV UNIVERSITY

THE LESTER AND SALLY ENTIN
FACULTY OF HUMANITIES
THE SHIRLEY AND LESLIE PORTER
SCHOOL OF CULTURAL STUDIES



אוניברסיטת תל-אביב

הפקולטה למדעי הרוח
ע"ש לסטר וסאלי אנטין
בית הספר למדעי התרבות
ע"ש שירלי ולסלי פורטר

קשב הערכה אסתטית ורעש סמנטי בקריאה של טקסט ספרותי.

חיבור לשם קבלת התואר "דוקטור לפילוסופיה"

מאת: אמיר חרש

מנחה: פרופ' ישעיהו שן

הוגש לסנאט של אוניברסיטת תל-אביב

מאי 2019

קשב הערכה אסתטית ורעש סמנטי בקריאה של טקסט ספרותי.

חיבור לשם קבלת התואר "דוקטור לפילוסופיה"

מאת: אמיר חרש

מנחה: פרופ' ישעיהו שן

הוגש לסנאט של אוניברסיטת תל-אביב

מאי 2019

TEL AVIV UNIVERSITY

THE LESTER AND SALLY ENTIN
FACULTY OF HUMANITIES
THE SHIRLEY AND LESLIE PORTER
SCHOOL OF CULTURAL STUDIES



אוניברסיטת תל-אביב

הפקולטה למדעי הרוח
ע"ש לסטר וסאלי אנטין
בית הספר למדעי התרבות
ע"ש שירלי ולסלי פורטר

עבודה זו נעשתה בהדרכת:

פרופ' ישעיהו שן

תקציר

עבודה זו עוסקת בקשר בין עיבוד קוגניטיבי של טקסט ספרותי והחוויה אסתטית-ספרותית של קוראים ממשיים. בעבודה זו אני מצביע על נקודה עיוורת שקיימת בשדה המחקר האמפירי של הספרות, וזו ההתעלמות מכשלים בתהליכי הקריאה. כיוון שקשיים אלו הם שכיחים מאוד הבנה שלהם חיונית לתמונה המלאה של האופן בו קוראים ממשיים קוראים ספרות. נקודת המוצא של העבודה היא כי קוראים ממשיים אינם קוראים אידיאליים, תהליכי הקריאה שלהם אינם מושלמים אלא משופעים בקשיים, בלבולים ותקלות. אפשר לחשוב על הכשלים האלה כעל תקלות במה שיעקובסון מכנה "הפונקציה הפואטית". העבודה מציגה מספר כישלונות בתהליך הקריאה ומסווגת אותם על בסיס מודל הפורגראונדינג. פורגראונדינג הוא התהליך שבעזרתו אלמנט טקסטואלי מסוים הופך בולט ביחס לאחרים - הופך לחלק מ"חזית" הטקסט, ובכך הופך זמין יותר לפרשנות ספרותית (Van peer, 1986). תיאוריית הפורגראונדינג היא אחת התיאוריות המובילות במחקר האמפירי של הספרות ומחקר רב תומך בה.

תיאוריית הפורגראונדינג עוסקת בספרותיות עצמה, כלומר במאפייני הטקסט שמבדילים בין טקסט ספרותי באופיו לכזה שאינו ספרותי. המחקר האמפירי של תיאוריית הפורגראונדינג מהווה הניסיון המקיף והשיטתי ביותר שנעשה עד כה לבחינה אמפירית של מודל של קריאת ספרות (Van Peer, Zyngier & Hakemulder, 2007). עם זאת, חלק מהניסויים הקלאסיים שנחשבים בדרך כלל כתומכים בתיאוריה מכילים ממצאים מעורבים (e.g. Emmott, Sanford, & Morrow, 2006; Miall and Kuiken, 1994). של הקוראים ולא תמיד הובילו לאפקטים אסתטיים. לאחרונה, לאור קושי בשחזור ממצאים קודמים, הציעו חוקרים כי שינויים בשדה הספרות, כמו העדר של קנון מוסכם, מובילים לתהליכי קריאה החורגים מתחזיותיה של תיאוריית הפורגראונדינג (Van Peer & Chesnokova 2017). בהתאם לכך, בעבודה זו אציע מודל חדש, המבוסס על מודל הפורגראונדינג המקובל אך יוצא ממנו ומפתח אפשרות שלא הוקדשה לה עד כה תשומת לב תיאורטית: האפשרות שתהליך הפורגראונדינג עלול להיכשל, ושהכישלונות האלה אינם היוצא מן הכלל, כי אם חלק בלתי נפרד מקריאת ספרות על ידי קוראים ממשיים.

שילוב אפשרות הכישלון בתוך המודל עשויה לעזור ליישב את הממצאים המעורבים שהוזכרו לעיל, שכן, אם כשלון הוא אחת התוצאות האפשריות של תהליך הפורגראונדינג אזי יותר ממצאים יכולים להיות מוכלים בתוך התיאוריה. זה מאפשר לחוקרים לא רק לחפש אישושים לכך שתהליך הפורגראונדינג אכן מתרחש, אלא גם לתור אחר גבולותיו: לתאר מתי הוא מצליח, מתי הוא נכשל, ובאילו אופנים. ישנה גם חשיבות ספרותית רחבה יותר בבדיקת מקרי הכישלון עצמם, וזאת ממספר סיבות. ראשית, כך ניתן יהיה לתת דין וחשבון מלא ומדויק יותר אודות תהליכי קריאת ספרות על ידי קוראים ממשיים, כלומר, קוראים שאינם אידיאלים. הרי כישלונות בקריאה הם שכיחים ומוכרים לכל, רבים נתקעים במהלך קריאת פרוזה או שירה, מתקשים להבין את הטקסט במידה מספקת, מתבלבלים, לא יורדים לסופם של הדברים, מדלגים על קטעים "בעייתיים", קוראים באופן רדוד, או פשוט לא נהנים מהקריאה. התופעות הללו אינן מוגבלות לקוראים מתחילים, לתלמידי תיכון או לסטודנטים בקורסי מבוא לתורת הספרות. הן אינן שלב ביניים בדרך להפיכה לקורא ספרות מנוסה - אפילו הקוראים המוכשרים והמנוסים ביותר מתקשים וכושלים לפעמים במפגש עם טקסט ספרותי מורכב. שנית, יתכן ולכישלונות הללו ישנן השפעה (חיובית או שלילית) על ההערכה האסתטית של הקוראים. כישלונות עלולים לחבל בחווית הקריאה, אך לחילופין, יתכן גם שהם מחוללים חווית קריאה רדיקאלית, מעיין סוג שונה, חריף יותר, של התענגות מן הטקסט. לכן, על ידי חקירה של הכישלונות והשפעתם על ההערכה האסתטית של הקוראים ניתן לשפוך אור נוסף על ההשלכות השונות, ואפילו הסותרות, של קושי בקריאה על קוראי ספרות ממשיים.

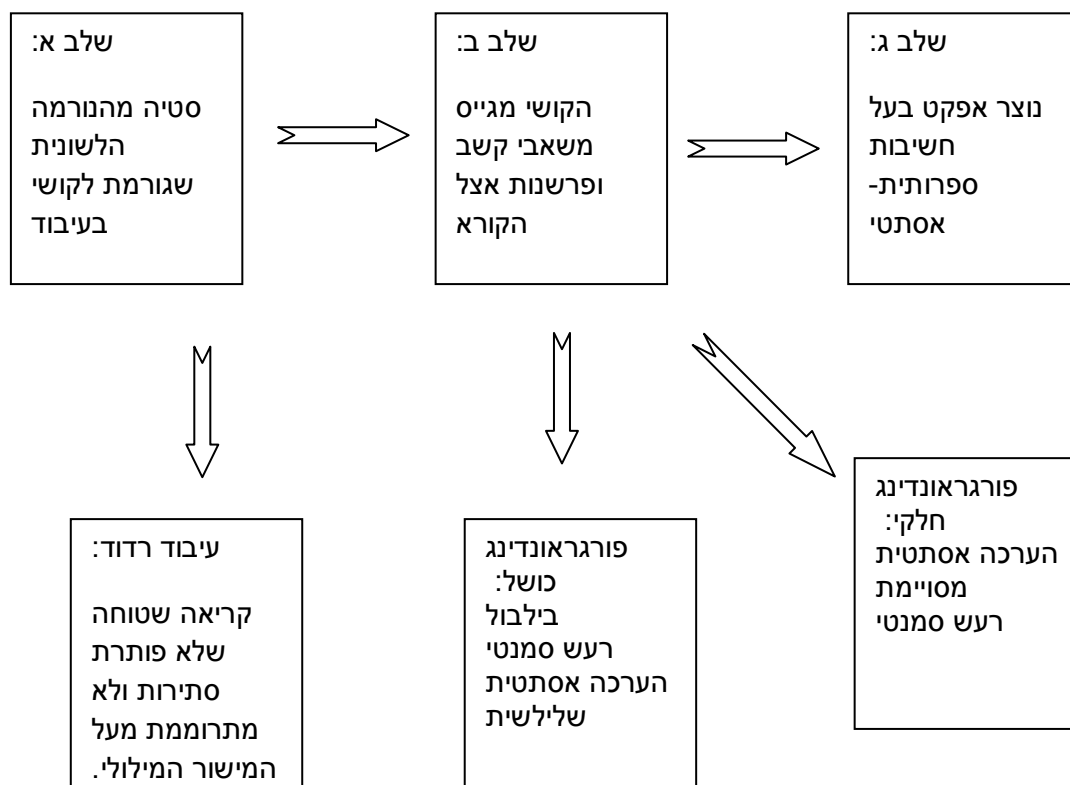
מודל הפורגראונדינג הסטנדרטי הוא מודל תלת שלבי: ראשית ישנה סטייה בטקסט מהנורמה הלשונית שגורמת לקושי בעיבוד, שנית, הקושי מגייס משאבי קשב ופרשנות אצל הקורא, ושלישית, הגיוס הקשבי מוביל לאפקט בעל חשיבות ספרותית-אסתטית. המודל המוצג בעבודה זו יוצא מתוך המודל הסטנדרטי הנ"ל ומרחיב אותו לאפשרויות נוספות, בהן המעבר בין שלבי המודל לא מתרחש בצורה חלקה.

מתוארות שני סוגים של תקלות:

תקלה במעבר בין השלב הראשון לשני - הקושי בעיבוד אמור להוביל לבולטות פסיכולוגית: לגיוס של משאבי קשב, לעיבוד נוסף, או לניסיון פרשני להתגבר על הקושי, אולם הוא לא חייב להוביל

אליו. הקורא יכול שלא להתעכב על הקושי יותר מהנחוץ למטרותיו, ולנקוט באסטרטגיה של עיבוד רדוד. יתכנו סוגים שונים של עיבוד רדוד, תחת הכותרת הזו יכול להיכנס טווח רחב של תופעות: בקצה האחד ישנה קריאה מרפרפת או אפילו דילוג על מקטע טקסט שלם, ובקצה השני ניסיון להבין שלא מתרומם מעל המישור הלשוני גופו, כלומר, לא פונה לעלילה, לדמויות, למחבר, או לכל מישור הבנה גבוה יותר.

תקלה במעבר בין השלב השני לשלישי - המאמץ הפרשני או הקשבי שהושקע באזור הקושי, אמור להוביל את הקורא אל החוויה האסתטית-הספרותית. אך ישנה גם האפשרות שהמאמץ הנוסף של השלב השני לא הוביל להערכה אסתטית חיובית אלא לבלבול, לרוגז, לאדישות או אפילו להערכה שלילית. במקרה כזה נתאר את הפורגראונדינג ככושל. אפשרות נוספת היא שהתהליך לא יכשל לחלוטין, אך גם לא יצליח לגמרי. במקרה כזה יוצר פורגראונדינג חלקי. פורגראונדינג חלקי כולל הערכה אסתטית גולמית, מוגבלת, לא ומפותחת מבחינה פרשנית.



מודל הפורגראוודינג הכושל. לפעמים תהליך הפורגראוודינג לא משלים את מהלכו ונוצר פורגראוודינג כושל או חלקי: הקושי בעיבוד לא מותמרים לחוויה אסתטית ונוצר בלבול ורעש סמנטי. המסלול העליון מתאר את התהליך המוצלח, והענפים התחתונים מתארים סוגי כשלים בתהליך.

אפשר לנסות לתאר את ההבדל בין השלבים השונים בעזרת דימוי של פרי, אם נחשוב על הקושי הראשוני בעיבוד כגרעין שסביבי צומח הפרי של חווית הקורא. בפורגראוודינג השלם עיקר הפרי הוא החוויה האסתטית החיובית, ובתוכו נותר גרעין קטן של רעש סמנטי וקושי. במקרה של פורגראוודינג חלקי עיקר בשר הפרי הוא הרעש הסמנטי והקושי, וסביבו מתוחה קליפה דקה של חוויה אסתטית חיובית. במקרה של פורגראוודינג כושל, לעומת זאת, אפילו הקליפה הדקה הזו אינה קיימת. הקורא נותר רק עם קושי והפרעה שגדלו למימדים של פרי שלם. ובעיבוד הרדוד

ישנו רק את גרעין הקושי הראשוני, וכיוון שלא התעמקו בו הוא לא גדל למימדים שמפריעים מאוד לקורא, מאידך הוא גם לא הצמיח סביבו הערכה אסתטית כלשהיא.

כדי לבחון את מודל הפורגראונדינג הכושל בוצע ניסוי קריאה הכולל מעקב אחרי תנועות עיניים. ארבעים ושניים נבדקים קראו את הסיפור הקצר "חדר הפסלים" מאת הסופר הארגנטינאי חורחה לואיס בורחס (1935), ומילאו שאלונים הבודקים רעש סמנטי והערכה אסתטית, כמו כן עברו מבדק של היכרות עם שמות סופרים ששימש כמדד לניסיון בקריאת ספרות. מיד לאחר מילוי השאלונים הם רואיינו בשיטה של חשיבה רטרוספקטיבית בקול רם, Retrospective Think Aloud - RTA. הוצגו לקוראים "מפות חום" המתארות את דפוסי תנועות העיניים שלהם בעת הקריאה והם התבקשו להסביר מדוע התעכבו במקומות בטקסט עליהם עיניהם השתהו או חזרו אחורנית במהלך הקריאה. בשיטה הזו, המשלבת עדויות "רכות" ו"קשות", דיווח מילולי יחד עם מידע פיזיולוגי מדויק, מתקבל מידע עשיר אודות החוויה המודעת של הקוראים.

הממצאים תומכים במודל הפורגראונדינג הכושל מכמה בחינות. ראשית, ביכולת של המודל לבצע הבחנות מובהקות מבחינה סטטיסטית. שנית, בכך שהתחזיות שלו התגלו כתקפות יותר מאלו של שתי התיאוריות המתחרות שנבדקו: מודל הפורגראונדינג הסטנדרטי ועמדת האסתטיקן הרדיקאלי.

מתוך ניתוח של הראיונות נמצא כי רק ב-36 אחוז מהמקרים הקוראים לא התחילו את תהליך הפורגראונדינג אלא נשארו בעיבוד רדוד, ורק ב-21 הם אחוזים סיימו אותו בהצלחה. האחוזים האלה השתנו באופן מובהק בהתאם למידת הניסיון של הקורא, ההערכה האסתטית שלו, אסטרטגיית הקריאה, ולמקטע הטקסט. קוראים שדיווחו על הערכה אסתטית גבוה נבדלו מקוראים עם הערכה אסתטית נמוכה בכך שהם הגיעו בשכיחות גבוהה יותר לפורגראונדינג מלא. מכאן ניתן להסיק כי הפורגראונדינג המלא תורם להערכה אסתטית חיובית, או שהערכה אסתטית חיובית כללית תורמת ליכולת לסיים תהליך פורגראונדינג בהצלחה. בנוסף, נראה שלפורגראונדינג החלקי אין תפקיד מרכזי ביצירה של הערכה אסתטית חיובית. אומנם גם הפורגראונדינג החלקי היה שכיח יותר בקרב בעלי ההערכה האסתטית החיובית, אולם במידה קטנה ולא מובהקת. לאור זאת נראה כי הפורגראונדינג המלא הוא הרכיב המרכזי שתורם להערכה אסתטית חיובית. קוראים

שדיווחו על הערכה אסתטית נמוכה נבדלו מקוראים עם הערכה אסתטית גבוהה בכך שהם הגיעו בשכיחות גבוהה יותר לפורגראודינג כושל; או שהפורגראודינג הכושל הפריע לגיבוש הערכה גבוהה, או שההערכה הנמוכה הקשתה על סיום מוצלח של תהליכי פורגראודינג והביאה לכישלונם. לעומת זאת העיבוד הרדוד לא היה שונה באופן מובהק בין הקבוצות. מכך ניתן להסיק שהפורגראודינג הכושל מקושר הרבה יותר להערכה אסתטית שלילית מאשר העיבוד רדוד. הממצאים הללו מתאימים למודל המתואר למעלה.

ההבדלים בין בעלי הניסיון בקריאת ספרות לחסרי הניסיון נמצאו בשני הקצוות של תהליך הפורגראודינג. בעלי הניסיון הגיעו לעיתים קרובות יותר לפורגראודינג מלא, וחסרי הניסיון נמנעו מלהיכנס לתהליך הפורגראודינג לעיתים קרובות יותר, ונקטו באסטרטגיה של עיבוד רדוד לעיתים קרובות יותר. אפשר לפרש זאת כך: חסרי הניסיון חששו (במודע או שלא במודע) להיכנס ל"הרפתקת" הפורגראודינג, שדורשת מאמץ ואין לדעת אם היא תוביל להערכה אסתטית גבוהה או נמוכה. במובן הזה נראה שהקוראים חסרי הניסיון נקטו באסטרטגיה שמרנית אך בטוחה. לעומתם בעלי הניסיון, שבחרו לעיתים תכופות יותר להסתכן ולהתחיל תהליך פורגראודינג, הצליחו להגיע בחלק גדול מהמקרים לסופו המוצלח. כך שנראה שהניסיון אפשר להם גם את ה"אומץ" הדרוש כדי להתחיל את תהליך הפורגראודינג, וגם את המיומנות הפרשנית הדרושה כדי להגיע לסופו המוצלח.

בדיקה של התפלגות פרופילי פורגראודינג לפי שלושה אמצעים סגנוניים שונים גילתה את הדפוסים הבאים:

מקטעי טקסט שכללו תיאור פיגורטיבי היו הגבוהים ביותר בפורגראודינג מלא ובעיבוד רדוד. דפוס הפורגראודינג יצר צורה של U המעידה כי שלבי הביניים בהם נמוכים בשכיחותם. לעומת זאת, נמצא כי מקטעי טקסט שכללו הערות מחבר היו גבוהים יותר בפורגראודינג כושל משאר המקטעים. עוד נמצא כי בהערות מחבר מידת העיבוד הרדוד היתה הנמוכה ביותר משאר המקטעים. מדפוס התוצאות הזה נראה כי לקוראים היה קל לזהות שיש להערות המחבר חשיבות ספרותית, ולכן הם ניסו להתעמק ועברו לשלב הפורגראודינג השני, אך סיום תהליך הפורגראודינג היה מאתגר יחסית והקוראים התקשו להגיע לסופו. בנוסף, נמצא כי מקטעי טקסט

בהם היתה קיימת **סטייה לשונית** היו גבוהים בעיבוד רדוד יותר משאר המקטעים. נראה כי האמצעי הסגנוני הזה הקשה על רוב הקוראים להיכנס לתהליך פורגראונדינג, ורק במקרים מעטים הצליחו הקוראים להגיע לפורגראודינג מלא. זה היה האמצעי הסגנוני הכי פחות אפקטיבי מבחינת השלמת הפורגראונדינג, והשפעת הניסיון בקריאה בו היתה מצומצמת.

מתוך מודל הפורגראונדינג הכושל אפשר לחשוב על חלוקה עקרונית לשני סוגים של אפקטיביות. א. האפקטיביות בהכנסת הקורא לתהליך פורגראונדינג. ב. האפקטיביות בהבאת הקורא לסיום מוצלח של הפורגראודינג. ניתוח פרופילי הפורגראונדינג לאמצעים הסגנוניים השונים מוביל להשערה כי יתכן שלמישור שבו קיים הקושי הראשוני ישנו תפקיד חשוב באפקטיביות מהסוג הראשון. שכן כאשר הקושי היה במישור הלשוני רוב הקוראים נטו לפתור את הקושי במישור הלשוני עצמו ומיעטו לעסוק בפרשנות ספרותית. כלומר, כאשר הם נתקלו בבעיה לשונית ניסו רוב הקוראים להבין "מה" נאמר, ולא "מדוע" זה נאמר.

מודל הפורגראונדינג הכושל מאפשר גם לבחון את שכיחותה של **עמדת האסתטיקן רדיקאלי** בקרב קוראים ממשיים. **עמדת האסתטיקן רדיקאלי** זו עמדתם של קוראים שעבורם החוויה הספרותית נובעת ישירות מהקושי, מחוסר היכולת להבין, מעצם המאבק בטקסט סרבן, ומהתדפקות חוזרת ונשנית כנגד דלתו הסגורה של הבלתי מתפענח. לפי עמדה הזו צפוי לפורגראונדינג הכושל והחלקי תפקיד מרכזי בהערכה האסתטית החיובית, או לכל הפחות צפוי שהשכיחות שלהם תלך ותגדל ככול שהניסיון של הקורא גדל. אף אחת מהתחזיות הללו לא התגשמה. גם החיפוש אחר קוראים שנוקטים באסטרטגיות קריאה בהן יש לשני הרכיבים הללו חלק מרכזי לא היה פורה. נמצאו רק ארבעה קוראים שנקטו באסטרטגיית קריאה בה פורגראונדינג כושל הוא רכיב מרכזי, ובדיקה מדוקדקת של הראיונות איתם לא תמכה ברעיון שהם חוו חוויה ספרותית אסתטית רדיקאלית, אלא, שהם נתקלים בעיקר ברעשים סמנטיים שונים, ושכתוצאה מכך הם מתוסכלים, מבולבלים וחסרי מנוחה. לא התקבל הרושם שהם "מתענגים" על המאבק עם הטקסט, אלא דווקא שהם מתוסכלים ממנו עד מאוד. כך שלפחות במקרה שנבחן בניסוי הזה - קריאה ראשונה של טקסט קצר של בורחס - מקרים של רעש סמנטי רב לא קושרו עם הערכה אסתטית-ספרותית אלא דווקא פגמו בה. אם זאת, ישנו במודל רכיב הכולל ערבוב של רעש סמנטי ושל הערכה אסתטית חיובית והוא הפורגראונדינג החלקי. נדמה שישנה מידה מסויימת של קרבה בין

הפורגורדינג החלקי, לבין חוויה ספרותית חיובית שעיקרה קושי. אפשר לומר שהם נמצאים על אותו הספקטרום, שהפורגוראודינג החלקי הוא מקרה קל של קריאה ספרותית סזיפית אך מענגת. אם זאת, מה שנמצא בנוגע לפורגוראודינג החלקי לא תמך בו כשחקן מרכזי בשדה החוויה הספרותית. תפקידו ביצירת הערכה אסתטית-ספרותית היה זניח, הוא לא אפיין אף אמצעי סגנוני מאלו שבדקנו, והוא לא היווה אסטרטגיה מועדפת על הקוראים בניסוי. כך שהממצאים לא תומכים במאבק בטקסט סרבן כחלק מהותי מהחוויה האסתטית הספרותית.

הפרק האחרון של העבודה מוקדש למתודולוגיה של חשיבה רטרוספקטיבית בקול רם בעזרת על תנועות עיניים. שיטה זו פותרת מספר בעיות מרכזיות הן באיסוף מידע מילולי והן בניחות תנועות עיניים. השיטה מצמצמות את בעיות הריאקטיביות והורטיקליות של המידע המילולי. כלומר, היא לא מפריעה לתהליך הקריאה והחשיבה בזמן אמת, והיא יחסית אמינה מבחינת עקבת הזיכרון, כיוון שתנועות העיניים מזכירות לנבדק את תהליך הקריאה שלו. מסקירת הספרות נמצא כי הסיכוי שיתעוררו בה פבריקציות הוא נמוך, אך הסיכוי להשמטות גבוה. כלומר, לא כל המידע שמופיע בתנועות העיניים מקבל הסבר, אך מה שהנבדק זוכר יחסית אמין. ההשמטות של המידע עוזרות לצמצם את כמות הנתונים, ולפתור את בעיית ה big data שנובעת מעודף נתונים במחקרי תנועות עיניים. הצמצום הזה מתרחש בשני מובנים. ראשית, במקום להתייחס לכל המילים בטקסט, ישנה התייחסות של החוקר רק למקומות בטקסט בהם נמצאה התעכבות בולטת המשותפת לקוראים רבים. שנית, ישנו צמצום נוסף רק למקומות בטקסט אותם הנבדקים זוכרים ויש להם מה לומר אודות ההתעכבות בהם. כך שהשיטה מתמקדת רק בתופעות הגדולות המתרחשות בתהליך הקריאה, באלו המשותפות לקוראים רבים ושמשאירות חותם בזיכרון הקוראים.

השיטה הזו מגשרות בין סוגי תגובות שונות של אותו נבדק, התגובות מילוליות והתגובות פיזיולוגיות. יש בה עושר מילולי רב יחד עם יכולת ללוקליות בטקסט ובזמן. לפי מחקרים קודמים מתחום הממשק המשתמש, וגם לפי לניתוח צירופי מילים שכיחות בניסוי הנוכחי נמצא כי היא מספקת עדויות רבות בעלות רכיב קוגניטיבי. כמו כן נמצא שישנה קורלציה בין ניסיון הנבדק בקריאת ספרות לבין אורך הראיון, ומשמעות הדבר היא שיש לשיטה הזו יתרון יחסי במחקר שעוסק בקוראים בעלי ניסיון בקריאת ספרות, כיוון שהם מצליחים טוב יותר למלל את תהליך

הקריאה שלהם בדיעבד. השיטה לא מוגבלת רק לקוראים בעלי חוויה חיובית של הטקסט, כיוון שגם קוראים בעלי הערכה אסתטית שלילית או שחוו רעשים סמנטיים רבים ספקו אותה כמות של מידע. כך שהשיטה מותאמת למנעד רחב של חוויות קריאה ספרותיות, ובפרט כדי ללמוד גם על הקשיים בהבנה ועל בעיות בתקשורת שמתעוררות אצל הקורא במהלך קריאה ספרותית.

עם זאת, ישנן גם כמה מגבלות לשיטה. היא מותאמת לטקסטים יחסית קצרים: שירים, פרגמנטים, סיפורים קצרים, או פרקים מרומנים. זאת, כיוון שבמידה ומדובר בטקסט ארוך עלולים הנבדקים לשכוח את פרטי תהליך הקריאה שלהם בטרם הראיון. השיטה לא מאפשרת לבדוק תופעות בלתי מודעות לנבדק. השיטה מספקת פחות מידע מנבדקים דלים בקריאת ספרות. כאמור, ישנן סיבות תיאורטיות מדוע יש לצפות בה למינימום בעיות של ורטיקליות וריאקטיביות, וישנם אף מחקרים שתומכים בטענות האלה בתחום ממשק המשתמש, אך בעתיד, במידה והיו מחקרים נוספים שמשתמשים בשיטה הזו במחקר הספרות, יהיה כדאי לגבות את הטענות הללו בעדויות אמפיריות ספציפיות למחקר ספרות.

בסוף העבודה ישנו נספח על רעש סמנטי. זהו פרק תיאורטי בו אני מציג בהרחבה את המונח רעש סמנטי וממקם אותו בהקשרו במחקר האמפירי של הספרות. פרק זה יכול לעמוד גם בפני עצמו, והוא מצורף כנספח כיוון שהוא פורסם כבר בעבר בכתב העת ורסוס (2016). למונח רעש סמנטי יש תפקיד מרכזי בתיאורית הפורגראודינג הכושל, והוא היווה זרז משמעותי לפיתוחה. הפרק בוחן את שאלת הקשר בין קושי העיבוד הקוגניטיבי בקריאה של טקסט ספרותי לבין החוויה האסתטית, ומציג שלוש תיאוריות מתחרות: תיאוריית הפורגראודינג, תיאוריית המיקום האופטימלי, ותיאוריית העיבוד הרדוד, שכל אחת מהן מציעה הסבר שונה במהותו, וכל אחת מהן זכתה לתמיכה מחקרית רחבה. המונח רעש סמנטי משמש כגשר שיכול לעזור לגשר בין הממצאים הסותרים.

לסיכום, שלושת החידושים המרכזיים בעבודה הם אלו:

א. פיתוח מודל הפורגראודינג הכושל, וביסוסו על ידי ניתוח ראיונות עם קוראים.

ב. עצם השימוש בטכניקת חשיבה רטרוספקטיבית בקול רם מודרכת על ידי תנועות עיניים וביסוסה ככלי מתאים למחקר של קריאת ספרות.

ג. שימוש במונח רעש סמנטי לתיאור כשלים בקריאה ופיתוח שיטות למדידתו.

המחקר הנוכחי גילה שפרופיל הפורגראונדינג מושפע מאוד מפרמטרים שמאפיינם את הקורא, ואת הטקסט, כמו הניסיון של הקורא, ההערכה האסתטית, האסטרטגיה והאמצעי הסגנוני. הרגישות הזו של פרופיל הפורגראונדינג לפרמטרים של הקורא והטקסט מעידה על כך שתהליך הפורגראונדינג עצמו הוא "עדין" יותר ממה שמקובל לחשוב - גורמים רבים משפיעים על האפקטיביות שלו. הרגישות הזו של תהליך הפורגראונדינג, יכולה להסביר מדוע ניסויים בתחום התקשו להגיע לתוצאות עקביות ורחבות.

TEL AVIV UNIVERSITY

THE LESTER AND SALLY ENTIN
FACULTY OF HUMANITIES
THE SHIRLEY AND LESLIE PORTER
SCHOOL OF CULTURAL STUDIES



אוניברסיטת תל-אביב

הפקולטה למדעי הרוח
ע"ש לסטר וסאלי אנטין
בית הספר למדעי התרבות
ע"ש שירלי ולסלי פורטר