

Comparative Translation Assessment

Comparative Translation Assessment:

Quantifying Quality

By

Ljuba Tarvi

**Cambridge
Scholars
Publishing**



Comparative Translation Assessment: Quantifying Quality

By Ljuba Tarvi

This book first published 2026

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Copyright © 2026 by Ljuba Tarvi

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN: 978-1-0364-6474-5

ISBN (Ebook): 978-1-0364-6475-2

For Yákov & Véra,
my dear parents

Denken Sie grob!
Think crudely!
—Berthold Brecht

TABLE OF CONTENTS

Preface	x
Introduction	xiv
Chapter 1	1
Comparative Translation Assessment: the Twilight Zone	
1.1 Statement of Problem.....	2
1.2 Terms and Definitions	2
1.3 Position in Translation Studies	4
1.4 Critics' Dilemma.....	6
1.5 Major Approaches.....	10
1.6 Key Problems.....	12
1.7 Statement of Goal	15
Chapter 2	18
Comparative Translation Assessment: Yesterday and Today	
2.1 Researchers' Dilemma	18
2.2 Computer-Based Systemic Models	21
2.3 Human-Based Systemic Models	23
2.3.1 Source Text ↔ Target Text.....	24
2.3.2 Source Text ↔ <i>Tertium Comparationis</i> ↔ Target Text	27
2.4 Semi-Systemic Models	29
2.4.1 Prose Samples	30
2.4.2 Poetry Samples	32
2.5 Quasi-Systemic Models	36
2.6 Merits and Demerits of the Models.....	38
Chapter 3	40
Alexander Pushkin's <i>Eugene Onegin</i> in Russian and English	
3.1. Alexander Pushkin's Life and Literary Heritage	40
3.2 <i>Eugene Onegin</i> in Russian (Source Text)	42
3.2.1 General Description.....	42
3.2.2 Content	43
3.2.3 Form (the Onegin stanza).....	45

3.3 <i>Eugene Onegin</i> in English (19 Target Texts).....	46
Chapter 4	68
Token Equivalence Method: Introduction	
4.1 Verbal Plane.....	70
4.2 Token Frame	72
4.2.1 Criteria of Token Isomorphism	75
4.2.2 Types of Labelling	77
4.2.3 Examples	81
4.2.4 Inventorying Texts	90
4.3 Poetic Plane.....	93
4.4 Rhyme Type.....	95
4.4.1 Masculine Rhyme	95
4.4.2 Feminine Rhyme	97
4.4.3 Criteria of Poetic Isomorphism	99
4.4.4 Inventorying Texts	100
4.5 Joint Plane.....	104
4.5.1 CTA and DNA	105
4.5.2 Two-Frame Module.....	106
4.5.3 Inventorying Texts	106
4.6 Selection of Samples.....	109
4.7 Modifications of the Token Equivalence Method.....	112
Chapter 5	116
Token Equivalence Method: Results	
5.1 Verbal (Content) Isomorphism	117
5.1.1 Sample I: One-Frame (TF) Module.....	117
5.1.2 Sample Two: One-Frame (TF) Module.....	120
5.1.3 Sample III: One-Frame (TF) Module	122
5.1.4 Verbal Plane: Cumulative results	124
5.2 Poetic (Form) Isomorphism	128
5.2.1 Sample I: One-Frame (RF) Module	128
5.2.2 Sample II: One-Frame (RF) Module	130
5.2.3 Sample III: One-Frame (RF) Module.....	131
5.2.4 Poetic Plane: Cumulative Results.....	131
5.3 Joint (Content and Form) Isomorphism	134
5.3.1 Sample I: Two Frame (TF+RT) Joint Module	135
5.3.2 Sample II: Two Frame (TF+RT) Joint Module.....	136
5.3.3 Sample III: Two Frame (TF+RT) Joint Module.....	136

5.3.4 Joint Plane: Cumulative Results	137
5.4 Qualitative Implications of Qualitative Results	140
Chapter 6	143
Token Equivalence Method: Comparison	
6.1 Other Comparative Translation Methods: Results	143
6.1.1 EO translators: Self-Assessment	144
6.1.2 Professional Critics: <i>EO</i> Assessment	153
6.1.3 University Lecturers' Survey	163
6.2 Token Equivalence Method: Application to Other Texts	164
6.2.1 Prose Samples	164
6.2.2 Poetry Samples	168
Chapter 7	184
Nabokov's Riddles	
7.1 Pushkin Revisited.....	184
7.1.1 The Great EO Debate in the 1960s.....	186
7.1.2 Nabokov's Metric EO Clones	188
7.1.3 Nabokov's EO Commentary	194
7.1.4 Nabokov and Pushkin: Kinship.....	195
7.2 Pushkin Beyond Russia.....	200
Chapter 8	211
Comparative Translation Assessment: Tomorrow	
Appendix A	215
<i>Eugen Onegin</i> in English:(the Corpus)	
Appendix B.....	217
Token Equivalence Method (Abbreviations)	
Appendix C.....	218
List of Tables	
Bibliography	220

PREFACE

The Internet has been called a lot of bad names, but it undoubtedly has more benefits than drawbacks. For instance, despite being full of fakes, it can be brutally honest, the quality of its answers is directly proportional to that of the questions asked; and it seems to be the only driving force behind bringing to light publications that have not been marketed. Being multiply mentioned on the net is a proof of texts' pragmatic or aesthetic value beyond PR campaigns – it just happens.

During the twenty years after I defended this thesis, I published regularly in scholarly journals, mostly on Vladimir Nabokov, but only two of my papers have passed the "Internet Test:" "Translation as a Conceptual Metaphor» (*Forum*, 2013) and "Chronotope and Metaphor as Ways of Time-Space Contextual Blending: Principle of Relativity in Literature" (*Bakhtiniana*, 2015). The third winner was the book you are reading, my PhD thesis (2004) published by my university in a hundred copies, so the net references to my PhD are nothing short of a miracle. A self-managing gestalt of global haphazard information and worldwide feedback-equipped readership, the Internet is an embodiment of one of the basic laws of quantum physics – Heisenberg's uncertainty principle.

I presume it was a purely statistical approach used in my comparative method that might, as a rarity in research, be responsible for my passing the "Internet Test." Indeed, Saussure's structure of a linguistic sign, Russian Formalists' dissection of texts, Chomsky's Context-Free Grammar, Feynman's diagrams, Viney-Darbelnet's audit of translated tokens that I borrowed for my method – all these utterly reductionist and highly formalized approaches made a splash only to be drowned in criticism. Appealing and stimulating, such outlooks are severely criticized but remain examples of innateness amidst the conventionality of the mainstream trends.

The chance to edit my first book was an eye-opener for me since I realized that both the project I am currently working on and the book that I published in 2023 are direct descendants of my PhD. I discovered that I had been consistent in correlating polysemic verbal signs with their certain "solid" quantitative counterparts in an attempt to co-tune the metaphysical facet of textual quality with relevant types of textual quantity. After all, our Universe functions on competing polarities – some obvious, some less

so – like particle-antiparticle, energy-matter, attraction-repulsion, fission-fusion, etc., that ensure its dynamics.

There is little doubt that verbal art, with its distinct reading and writing facets as the dynamic poles, is closely associated with the dual magic of energy/matter conversions governed in language by both the “soft” problems of perception, attention, interpretation, etc. and the “hard” problem of arising consciousness. Language and quantum mechanics, however, have to be correlated with caution since in both the “hard” problem is only hypothesized: both a quantum of energy and a token of information may or may not carry a metaphysical charge of qualia that is a subjective aspect of conscious experience.

The “hard” problem of energy transformation into matter seems obvious but since antiquity people have been speculating what sets it in action: God, Logos, Nature’s self-regulation, the Will, the collective unconscious, la Divina Assoluta del Mondo, etc. In order to compensate for the fact that we still see a two-way highway as a one-way street, metaphorized in *Genesis* by God’s breathing into Adam’s nostrils, we have to invent theoretical “patch” terms to cover up the lacunae, like “virtual” (undetectable but functional) particles in quantum mechanics, the “undecidable propositions” in the Gödel theorem, or the “wave function collapse,” which in common parlance means “We do not know how we come from a quantum probability to what we have classically measured.” We have also learnt to designate the presence of this mysterious “deep state” force with abstract symbols, like the bar in Saussure’s structure of a linguistic sign, the copula in Lakoff’s formula of metaphor, c^2 in the Einstein formula, the internal lines in Feynman’s diagrams, etc.

There have been numerous propositions to solve the “hard” problem by assigning its power to various hypothetical forces, but they still remain, in terms of Karl Popper, falsifiable but not verifiable. One of the popular trends since Hegel’s “self-evolving absolute idea” has been the concept of global consciousness expressed by Carl Sagan: “We are a way for the cosmos to know itself.” Thirty years after Sagan, his idea was appropriated by Federico Faggin in his book-length “discovery” that the hard problem is about global consciousness striving to cognize itself. “Consciousness” is a beautiful concept, but we still do not even know if it is fundamental (panpsychism) or emergent (the Diósi–Penrose model).

My PhD thesis was, as I came to understand, my first try at correlating the soft and hard problems as regards verbal signs by presuming that when the inexactness of a quantitative approach is mapped on that of a conventional emotive one, they might prove synergetic. The premise was simple: quantity is not about numbers but rather about the logic behind

them. My theoretical “patch” was correlating the surface words (quantity) with their deep layer stratum (quality) by a direct comparison of the obtained results, i.e., *a token as a quantum/quale*. I must confess that I was surprised that quantitative and qualitative results proved to be not only comparable but actually congruent.

There was a nearly fifteen-year period after the defense, when I involved myself in the topic that enriched me intellectually but produced a zero result, a clear case of the “Law of the Planning Fallacy.” The goal I set to solve was to compare the stylistic patterns in Vladimir Nabokov’s nine novels written in Russian and nine novels written in English. I did manage to outline a number of fascinating patterns, but, alas, remaining within the firm boundaries of stylistics, since the concept of style requires co-considering an enumerable number of surface expressions. Style, on a par with gravitation or consciousness, has no explicit unit, which precludes its systematic study. My project taught me a lot but it was doomed to fail since neither human nor artificial intelligence is yet capable to cope with the mind-bogglingly complex nature of style.

In my second attempt to correlate surface text features with their implied roots, I returned to a token or rather to a manageable set of tokens in *Modern Fantasies on Love versus Classical Romances* (2023) focused on *a token as information/energy*. To compare models of romantic relations in the seventeen most popular novels of the last two centuries, a limited set of recurrent plot-related textual markers was selected, and the formal analysis revealed three major patterns of novelistic intergender relations – Adaptation (19th century), Alienation (20th century), and Imagination (1990-2011).

My current project in progress that concludes the trilogy, *In the Human Pursuit of Meaning, Magic is the Only Real Thing*, is definitely my most daring approach to the soft and hard problems correlation based on co-considering linguistics, quantum physics and the Bible, i.e., *a token as a quale/quantum*. The first three chapters of *Genesis* are a potted history of the Universe creation via materialization of energy, i.e., it is directly related to the hard problem. The Bible vests the power of shaping the world into the hands of God who creates matter, living matter included, using a word as an energy quantum. In the Bible, the mechanism of creating mass as manifest energy is obviously not revealed, but the produced matter is described in a detail that allows its rough correlation with relevant physical laws and forces, which might, hopefully, provide a fresh perspective on the old unsolvable problem.

My efforts at coordinating the hard and soft problems via correlating recurrent textual markers with their implied deep layer in translation

studies, literature and linguistics remain within the mainstream of modern science – expanding rather than delving into depth. Our inability to cope with a growing volume of information has resulted in fragmentation: like Lawrence Krauss’ virtual particles, objects of scientific attention spring out of nowhere and disappear to be replaced with brighter targets that only increase entropy. Science has stopped being a perception manager, and I wonder what term Thomas Kuhn would coin to designate the present paradigm of scientific evolution. Metaphorically, being unable to blow into God’s nostrils, thereby turning our relationship with him into a conversation, we continue playing with a kaleidoscope that produces new combinations of old bits of glass, fused as ever from the perceived and the implied. The hard problem remains a well-guarded secret, and we seem to be doomed to continue watching shadows on the wall of Plato’s cave.

Upon rereading, I find my thesis a solid piece of research, well supported with earlier findings, substantiated with sound argument and moderated with self-criticism. The introduced changes pertain to streamlining terminology and reducing the tabulated data. The Bibliography is left untouched because this text was written at the golden period when the young science of translation studies was at the peak of its development, when generalizations inspired by Holmes as far back as the 1970s brought about many popular models and approaches.

My dream of luring computer assistance into translation classes is yet to be fulfilled albeit I did my best to describe the suggested algorithm step by step (Chapter Four). Twenty years on, my method is much easier to implement since technology has improved exponentially and the all-pervading Wi-Fi made computer-equipped classes redundant.

Helsinki
December 22, 2025

INTRODUCTION

*Elusive Pushkin! ...
This is my task - a poet's patience
And scholiastic passion blent:
Dove-droppings on your monument.*
— Vladimir Nabokov

In human pursuit of knowledge, the so-called Menon paradox, or the learning paradox, is taken seriously:

... either you know what you are searching for or you do not. If you do know, you already have it, whence inquiry is pointless. And if you do not know, you would not recognize it even if you stumbled on it accidentally; hence, again, inquiry is impossible, pointless. (Nickles 1980:89)

Inquiry, I would argue, is not always pointless, since searching does not always imply finding, as is proved by my failed attempt to study Nabokov's style. From the onset of my inquiry devoted to comparative translation assessment (CTA), I knew what I was searching for. In terms of practice, I was planning to develop a user-friendly and relatively simple method to reliably approximate assessments and self-assessments of translated texts. In terms of theory, I was planning to see if rare formal quantitative CTA results were congruent with predominant emotive assessments based on impression and intuition. To this end, the intention was not only to assess the nineteen English translations of Alexander Pushkin's novel in verse *Eugene Onegin* (hereafter EO) with the suggested method but also to ensure a control set of EO conventional assessments by other authors in order to compare the results. Moreover, the results obtained by other authors on texts other than EO used in the discussion were planned to be counterchecked by applying the suggested formal method to them.

My attraction to the formal stance is prompted by several reasons. First, because, as I will argue below, of the two major approaches to assessment, formal and emotive, the former implies seemingly less subjective estimates because it is based, although regretfully one-sidedly, on purely linguistic-textual analysis. Second, because I am convinced that before embracing macro-level approaches one still has a lot to find at the micro-level of tokens. Third, formal methods allow verification. Fourth, the results

obtained with formal methods imply quantitative expression that avoid the use of “judgmental” assessment words. And, finally, because personally I tend toward certain “solid” patterns that at least partially allow one to imitate stability in the wonderfully unstable textual chaos.

My general approach ties well with the type of pluralism suggested by Richard Rorty in his *Contingency, Irony, and Solidarity* (1989), “a utopia of liberal ironies,” where “irony” implies not prioritizing one’s personal position, “liberal” means no cruelty in treating the views of the opponents, and “utopia” is self-evident. Another dictum behind the project is Samuel Johnston’s maxim, “The only end of writing is to enable readers better to enjoy life or better to endure it.”

The statistical model suggested in this thesis can be defined as descriptive rather than prescriptive, objective rather than subjective, and pragmatic rather than theoretical. I hope that my case study could be a contribution, small-scale and inconclusive as it may be, to both the rapidly developing field of comparative translation assessment and the slowly but surely evolving English-language Pushkin scholarships.

I will start discussing the problems of comparative translation assessment (CTA) by outlining in Chapter One the place it occupies in Translation Studies (TS) and the basic problems this TS branch has been (sadly, still is) facing. Chapter Two is concerned with the CTA past and present from the point of view of the methods and approaches used in translation assessment. Chapter Three describes the corpus of the texts under study by presenting Pushkin’s novel and its nineteen English translations, each exemplified with a translation of the same stanza. Chapter Four offers a fresh look at CTA by developing a formal user-friendly quantitative method of comparing translations, the Token Equivalence Method (TEM). This chapter is thoroughly detailed and profusely exemplified as regards the principles of labeling tokens for their inventorying and of calculating the results. Chapter Five presents the quantitative results obtained with the suggested CTA method by arranging them into matrices, tables, mini-tables, and profiles in the process of their gradual downsizing to one final figure, Translation Quotient (TQ). Chapter Six compares the TEM results with those on the same material obtained by conventional emotive methods by re-representing their emotive estimates in the TEM figures for comparison. Chapter Seven is an attempt to elucidate the role Nabokov’s translation played in the history of EO translation into English. The concluding Chapter Eight is a plea for conducting systematic comparative studies with the aim of developing a unified system of CTA.

My genuine thanks to all who supported my six-year doctoral research. First of all, to my supervisor Professor Ritva Leppihalme of the Department of English Philology at the University of Helsinki for choosing me to be her doctoral student and for her boundless patience in restraining my wild imagination. I am grateful to Professor Andrew Chesterman whose lectures inspired my interest in translation studies. I am deeply indebted to my husband Eljas Tarvainen for supporting me morally and financially. I also thank my fellow-students, post-docs and the faculty of the University of Helsinki whose creative criticism helped me shape my raw suggestions into functional ideas.

My deepest gratitude to my fellow Nabokovians who were a big part of my conference life at the time of writing this thesis and after: to Professor Brian Boyd, the spiritual guru at the helm of the Nabokov studies ever since he published his fundamental biography of Nabokov (1996), and to Professor Charles Nicol for arranging a short-term grant for research in the Library of Congress, Washington D.C. (2000) at the Kennan Institute for Advanced Russian Studies of the Woodrow Wilson International Center for Scholars that allowed me to ensure the control block of this project. I am proud to have been a part of the Nabokov “fan band,” a tight group of devoted academics who used to “jam together” in Europe, America, Japan, New Zealand, etc., to discuss the intricacies of Nabokov’s exquisite verbal art. My special thanks to my Internet fans who surprised me by using my method in their research. Even more astonishing is the fact that they are mostly from the countries where hieroglyphics and the Arabic alphabet are used.

CHAPTER ONE

COMPARATIVE TRANSLATION ASSESSMENT: THE TWILIGHT ZONE

The interplay of quantities behind
the twilight of qualities.
— Velemir Khlebnikov

The title of the book and the epigraph of this chapter interrelate the concepts of quantity and quality considered here from the viewpoint of a complex field of human verbal activity, comparative assessment of translated texts. Qualitative methods of assessment are common practice but they still lack a systemic approach; quantitative methods of computer linguistics are systemic but too cumbersome for practical purposes. Hence, in practice, use is made of eclectic models combining certain features of both approaches, and the task of reaching a consensus probably utopic.

The idea behind this project is not to confront quantized estimates and qualified assessments but rather to compare the quantitative and qualitative results on the same text, which presupposes that the chosen text should be known well enough to ensure both a number of its translations and a published stock of conventional critical essays on its translation. The aim is not to choose between these approaches but rather, if the results prove to be congruent, suggest a quantitative method as an initial step of qualitative analysis. In order to obtain a set of quantitative data, a statistic comparative method devoid of any kind of impressionist estimates is to be devised and tested. Therefore, the project is aimed at checking if the two groups of obtained results correlate by the one and only chosen parameter – the cline of verbal loyalty to the original – or, in terms of Pushkin, by testing “the laws of harmony with algebra.”

The Russian poet Velimir Khlebnikov (1885-1922), believed that “qualities,” intuited rather than observed in the “twilight” of our perception, are lurking behind the “quantities” revealed by their overt interplay. In quantitative approaches to assessment, quantities are in the forefront to be gauged, while qualities remain beyond our immediate perception, which implies that by co-considering both approaches at the level of the

obtained results, we might be capable of interpreting overtly and quantitatively what we perceive covertly and qualitatively. After all, it is quality that gives meaning to quantity.

1.1 Statement of Problem

This project was initiated by a telephone call. A Russian friend of mine, who by that time had been living in the USA for about a decade, called me to complain that her sons, who had left Russia at the pre-school age, were no longer willing either to speak or read Russian. A doting mother, my friend decided to introduce them to Russian literature through English translations. The question she asked seemed to be simple: “Which of the translations of Pushkin’s *Eugene Onegin* is, both in content and form, the closest to the original?” My research is an attempt to answer that “simple” question, which would, hopefully, bring me closer to solving a much broader problem of why Pushkin, the Prince of Russian bards, has still not occupied a worthy position in the Western canon of Russian literary prodigies.

At that time, I knew that Pushkin’s eight-chapter novel in verse *Eugene Onegin* (1837) had several English translations, and the library search (in the pre-Google era of 1998) revealed a corpus of eighteen full English translations (the nineteenth translation was added at a later stage). Working in the New York Public Library, the British Library in London, and in the Congress Library in Washington, I had to xerox the translations I could not buy, and even to order a photocopy of one of them. Besides, there I collected the crucially important block of conventional CTA articles of professional critics with “impressionist” estimates of EO translations, the control group for comparing with my quantitative results.

From the onset, it was clear that both the scope of the project and its goal – locating the compared translations on a cline of their closeness to the original in terms of content and form – necessitated clearly outlined relevant parameters.

1.2 Terms and Definitions

Comparative translation assessment is a regular human activity, and hundreds of critical remarks about translations are made every day, often in an authoritarian way, many even get published. Very few of these critical, often flippant, remarks are based on much more than a supposed knowledge of the source text language; very often there is neither systemic nor informed discussion nor clear criteria, mostly just an occasional exchange of opinions. Assessment in general, and textual assessment in

particular are automatic human activities deeply rooted in certain in-depth and germane processes studied by a number of sciences dealing with “soft problems,” such as philosophy, psychology, physiology, psycholinguistics, etc., which hampers structuring the act and/or process of assessment either conceptually or phenomenologically.

The twilight zone of comparative translation assessment is normally referred to as Translation Quality Assessment (TQA) since what is finally assessed and graded, whatever method is used, is indeed quality, and it is commonly done by supplying the assessed material with an adjective on the cline between ‘good’ and ‘bad’, since quality is “the degree of excellence which a thing possesses” and “the degree of conformance to a standard” (Webster’s New World Dictionary). At the same time, the integration of the notion of “quality” into the name of Juliane House’s comparative model (1977) was considered to be a probable reason of its low popularity and relatively little use in the North America (Roberts 1992:15).

In this study, I intend to use a milder albeit equally misleading like all definitions, term Comparative Translation Assessment (CTA) because any comparison involves assessment, and any assessment is based on comparison. Besides, this term contains the sub-term “comparative,” which implies the use of the source text as a benchmark for comparison. The problem of a precise definition is germane to all sciences, since little has in fact changed since Spinoza famously said, “*Omnis determinatio est negatio*” [Any definition is negation]. As regards the concept of “quality,” it seems to belong to the most polysemic and hence least definable terms. For instance, the following general definition is offered by the Oxford Advanced Learners’ Dictionary: quality is a “degree of goodness or worth” and “the property of things that can be measured,” which entails a contradiction since quality is a quale that cannot be measured.

Definitions of the term “translation” range between purely pragmatic, like “any target language text which is presented and regarded as such within the target system itself, on whatever grounds” (Toury 1982:27) to highly metamorphic, like “... to borrow an image from general semantics, translations are maps, the territories are the originals” (Holmes 1988:58). The truth is, as usual, buried somewhere in between, often without a headstone: “... the nature and definition of the category of “translation” have been notoriously contentious within the discourses of translation theory and practice. ... there is no simple core identity for the cluster of objects identified as translations by many societies of human time and space” (Tymoczko 1998:654).

Like its synonyms “judgment” and “appraisal,” “assessment” is, broadly speaking, a careful consideration (analysis, argumentative criticism) of

something by someone, followed by a judgment (estimate, evaluation, score) about it. This definition underscores the two-sidedness of the notion which implies both the process and the result of assessment. In the present text, the term “assessment” will be used to denote the process of forming a judgment, while such terms as “estimate,” “evaluation,” and “score” will be used for the quantitative side of the matter which is prioritized in this study.

As regards the object and goal of assessment, Gerard McAlister (1999) attempted to introduce the word “assessment” as an umbrella term to be combined with three additional sub-terms: if the literary qualities of a target text (TT) are reviewed, then we deal with “translation evaluation,” if a target text is analyzed from the functional standpoint, it is “translation criticism,” while comparative linguistic analysis aimed at the explication of the source text - target text (ST-TT) isomorphism is referred to as “translational analysis proper.” It goes without saying that such compartmentalization is not water-tight, since in reality we often have to do with non-systematic review-type translation assessments, which include fragments of “evaluation,” “criticism,” and “analysis.” For instance, within McAlister’s classification, the method to be proposed in Chapter Four can be referred to as “translation evaluation” and “translation analysis,” since it both explicates the ST-TT relationship and takes into considerations the ST’s qualities.

From the same standpoint of the object and goal of its assessment, I would propose to roughly classify the domain of comparative translation assessment (CTA) into the following three fields that partially overlap with Holmes’s four basic divisions of “Applied Translation Studies (1988:77-78): “commercial” CTA, “research/instruction” CTA, and “literary” CTA. From this standpoint, the Token Equivalence Method (TEM) to be suggested in Chapter Four seems to deal with all the outlined fields: it can work with both literary and non-literary texts, it can be used in research/instruction and commercial CTA, and, what is most important, it can be a handy tool of computer-equipped classroom translation instruction.

Let us see what position Comparative Translation Assessment in general and literary Comparative Translation Assessment in particular occupy in the general picture of Descriptive Translation Studies.

1.3 Position in Translation Studies

In his plenary address at the Eighth World Congress of the International Federation of Translators held in Montreal in May 1977, James S. Holmes presented the following thesis:

If and when we ever do arrive at a comprehensive translation theory which meets the rigorous requirements of a true scientific theory explaining the nature of all the phenomena to be considered, that theory – for all science’s desire for the elegance of simplicity – will be much more complex than any that has been presented up to now, since it will have to account for a great many variables which usually either are assumed to be constant ... or are ignored entirely” (1988:101).

Now, half a century later, it is obvious that the field of translation studies still remains patched and seems to be lost in its own attempts of self-cognition. There are both subjective and objective reasons for that since “... the act of translation is subjective; an interpretation of ambiguity. ... A basic component cannot be identified by denotation only; all semantic and semiotic connotations are equally valid elements” (Raguet-Bouvard 1995:134).

In the hierarchy of the much respected and not less criticized map of translation studies, suggested by Holmes in “The Name and Nature of Translation Studies” as far back as 1972, Comparative Translation Assessment, termed by Holmes “translation criticism,” occupies a remote position under the “applied” TS branch. Holmes was certain that translation evaluation “will always elude the grasp of objective analysis to some extent, and so it continues to reflect the intuitive, impressionist attitudes and stances of the critic” (1988:78). Having compartmentalized CTA into objective analysis and impressionist attitudes, Holmes, however, expressed his hope that “closer contacts between translation scholars and translation critics could do a great deal to reduce the intuitive element to a more acceptable level” (ibid.).

Of interest is Holmes’ emphasis on the “intuitive element” in CTA, and his assertion that this element cannot be completely eliminated but can be reduced to “a more acceptable level.” I would like to emphasize that the idea behind my research is to demonstrate that while the intuitive, impressionist attitudes and stances of the critics are practically impossible to attenuate, some complementary CTA methods of a more objective nature could be used alongside to countercheck the validity of either approach.

Holmes does not specify what he implies by “intuitive elements.” Descartes’ dictum, *Cogito ergo sum*, is in fact the expression of individualization, i.e., of one’s own role in the act of perception/assessment. The Webster’s New World Dictionary (WNWD) defines “intuition” as “the ability to perceive or know things without conscious reasoning; immediate apprehension or understanding.” The suggested row of synonyms includes, among others, “instinct,” “perception,” “impression,” and “effect.” Therefore, Holmes’ “intuitive elements” could be viewed as a stumbling block on the way to objective criticism, embracing the whole gamut of the pro-

cesses of comprehending and/or assessing written texts by individuals, starting from perception (WNWD): “mental grasp of objects, qualities, etc. by means of the senses”) guided by intuition to a resulting impression (WNWD: “an effect produced, as on the mind or senses, by some force or influence”).

Human thinking is much richer than its inductive form. There is a gap between a guess or a “leap of faith” achieved with certain rings of the logical chain omitted, and knowledge (understanding), which follows the logical chain ring by ring. Granted the scope and goal of this research, my approach to CTA can be only deductive, logical and systemic. As a reader and a “regular” assessor, I have my priorities based on instinct and immediate impression, but my aim is not to voice or justify my opinions but rather to find a relevant way of quantifying the translations’ closeness to the original in terms of their content and form. As characters of the famous movie *The Matrix* (1999) in all its sequels, we all are attempting to understand the world around that has no sense.

1.4 Critics’ Dilemma

As in translations studies, in literary criticism there are, according to Wimsatt & Beardsley, two types of critics who produce vastly different sorts of criticism: “the critic whose formulations lean to the emotive and the critic whose formulations lean to the cognitive” (1972:354). “Emotive” is defined (WNWD) as “characterized by, expressing, or producing emotions,” with the row of synonyms including such adjectives as “controversial” and “delicate.” “Cognitive” is defined by the Collins Cobuild Lexicon as “relating to the mental process involved in knowing, learning, and understanding things.” Attempts to discover and describe the “impressions” or “effects” produced by verbal art are as old as the history of literature itself; the basic feature of such descriptions being inevitable subjectivity of judgment and heavy metaphoric labeling of the phenomena discussed:

It would be convenient if the passwords of the intentional school, ‘sincerity’, ‘fidelity’, ‘spontaneity’, ‘authenticity’, ‘genuineness’, ‘originality’, could be equated with terms such as ‘integrity’, ‘relevance’, ‘unity’, ‘function’, ‘maturity’, ‘subtlety’, ‘adequacy’, and other more precise terms of evaluation – in short, if ‘expression’ always meant aesthetic achievement. But this is not so. (Wimsatt & Beardsley 1972:338)

To illustrate Wimsatt & Beardsley’s opinion and to demonstrate the scope of the descriptive terms used in the evaluative or judgmental CTA, some

typical examples of “emotive” criticism of translated texts will be cited below. It is an assorted selection of quotations from other authors’ reviews on *Eugene Onegin* (EO) translations to be considered in Chapter Six where they will be properly labeled for reference. It should be borne in mind that what is presented in this section is not criticism but rather an idea of what I wish to avoid in my analysis. “Judgmental criticism” always finds translations falling short of the ideal of exactly reproducing their originals at any level of comparative analysis chosen. Such criticism mostly lacks the key condition of a systemic approach, i.e., contextualization of individual phenomena on the basis of a theoretically sound and argued set of procedures. I have divided the quotations below into the four fuzzy-boundary groups and italicized the key adjectives and adverbs.

Let us start with evaluative adjectives and adverbs used when comparing various English EO versions: X’s obvious determination to be *as literal as possible, clearly* placed the main emphasis upon form, a *notable* attempt at a compromise, a *careful* rendering of Pushkin’s meaning. Or, the following characteristics are encountered: Y’s translation is *unforgivably trite, leaden, and unmusical*; his text is *admirably straightforward and lucid and direct*, but the style is there and is *vigorously maintained*; some of the lines, of course, are *clumsy*; on the whole an *accurate* and very *talented* translation. The basic feature of all these strong evaluative epithets is their unsubstantiated subjectivity, which describes the impression this or that version produced on the reviewer. Therefore, Rule One: no impressionistic epithets are allowed in my analysis.

Another group of citations is intended to define the “degree” of this or that quality of a translation, and starts therefore with a question: *How good* is the new translation of *Onegin*? Among possible answers, one can find the following: there is *a lot of padding and verbal contortion*, inevitable in a rhymed version, *but less than in other such versions*; Z simply modifies the measure *slightly*, or it *represents a tremendous improvement over X’s* version. Or, for instance, such evasive estimations as: *Pushkin’s tone of voice has been caught very nicely* indeed: *light, civilized, slightly bantering; slightly tongue-in-cheek, slightly ironic, more than slightly Byronic; not enough flat to spoil the general effect*; Z’s lines *sound a little flabby*. In my analysis, I would like to answer the question ‘*How good* is the new translation of *Onegin*?’ as regards one parameter of the analysis and in terms of figures. If I say that “there is a lot of padding,” then I should be able to substantiate this claim in quantitative terms. If I say that in this or that respect the translation “is rather weak,” I should be able to specify the degree of weakness quantitatively.

Next comes a group of judgments aimed at comparison proper, i.e., these evaluations involve various grades of qualities, like *the best translation now available*, it remains *the most intelligent and accurate rhymed version that we have*, *superseding all previous English translations of Eugene Onegin*, among English translations of *Eugene Onegin* into verse this is *the most faithful and the most readable*, X is *more modern and more direct*, Y's more "humble" version, *[m]ore seriously at fault is the line*, Z's work provides an aesthetic experience *something like the original*, and it *suffers no more from "inaccuracies" than* does X's translation, etc. If in my analysis, having compared several versions, I come to the conclusion that one of the translations is 'the closest to the original', it should mean that I have proved it quantitatively (within the requirements and conditions of my analysis). If I say that a certain version is "less readable" than some other translation, I should be able to substantiate my conclusion in terms of figures by specifying to what degree it is less. Therefore, the last two groups of citations can be united by Rule Two: no unspecified and unsubstantiated degrees of qualities are allowed in my analysis.

Finally, there are citations that demonstrate their authors' subjectivity sometimes bordering on mysticism: X's translation ... is *too much of a tour de force* and not ultimately readable, *the verbal magic has vanished*, *the result is a diluted verbal soup*, *there is real poetry in this translation*, he also *catches much of Pushkin's poetry*, *they in no way spoil the general impression* and *do not distort the picture as a whole*, *she seems to me to lack Pushkin's masculinity*, *Y has reduced the sparkling choreography of the Onegin stanza to the dull plod of a schoolboy crib*, *while quite usable*, *this translation lacks the felicity of Z's version*. There are also such evasive descriptions as: *apart from obvious accuracy of meaning (being true to the original) there exists an accuracy of a higher order ... accuracy of context, of shades of meaning, accuracy of the orchestration of the verse and of its sonority*, it is *contemptuous to the poetic passion in a way Pushkin never is*, *there is, simply, something objective about clarity, directness, unforced grace, ease of scansion* and so on. Therefore, Rule Three: no subjective or "flowery" metaphors with a metaphysical flavor are allowed in my analysis.

In his book *Case Study Research Methods*, Bill Gillham presents his two-paradigm model of fundamental research as consisting of two complementary branches (2000:3-5): the "natural-sciences style" positivist philosophy that asserts that only observable, and verifiable, phenomena could be the subject matter of science; and the "naturalistic" research concerned with defining and testing evidence. Gillham outlines the following key differences between the former and the latter: experimental vs.

non-experimental methods, deductive theorizing (hypothesis testing) vs. inductive theorizing (hypothesis seeking), preordained vs. emergent research design, objectivity vs. subjectivity, detachment vs. participation, quantitative (determining the significance of results) vs. qualitative (giving meaning to results) data, constructing evidence vs. searching for evidence, etc. (2000:8). Gillham's dichotomies place the above citations from EO criticism directly under one of his poles, "quasi-judicial or naturalistic."

Comparing quantized estimates and qualified assessments requires terminology for the quantitative pole, since it is a very special manual branch of computer linguistics that need its own terms. The idea of calculating the tokens kept in translation by numbering the tokens in the original was suggested by Jean-Paul Vinay and Jean Darbelnet as far back as 1958. The term they used for numbering the tokens was "inventorying," which is convenient grammatically (an inventory, to inventory, inventoried); besides, in modern language of business, it has a required connotation of "assessing the data pertaining to each item in a brand's product line" by tracking it.

The metaphor of monitoring items (tokens) in a warehouse (texts) by creating inventory/stock records (counts, spreadsheets, files) via a step-by-step guide to perform stock-takes remains valid. In modern business, a stock-take is a physical inspection undertaken to determine the quantities of the stock. Stock-takes are performed either manually – whereby employees conduct physical counts of each item in an inventory location – or with digital tools, such as barcode scanners and inventory management software, the former being an exact analog with the suggested method that is still manual but may be computerized. In business, an audit of the existing stock or "stock-taking" or "inventory checking" or "wall-to-wall" is the physical verification of the quantities of items held in an inventory – the essence of the suggested mode of comparison.

In managing the number of translated tokens, their stock-takes, whether manual or digital, are performed in a strict succession of operations. The procedure is manageable for both a document-length texts for translation class, which entails one set of enumeration, and for the stanza-length excerpts of the novel in verse analyzed in this project. The difference is that in the latter case the results obtained for each stanza have to be approximated by using an elementary operation of the arithmetic mean: the results obtained for each stanza in a sample are summed up and divided by the number of the stanzas.

The problem with quality control, as Peter Bush (1998:66) views it, is that "it wants neatness and clarity where there is only complexity in the interaction between words, translation, culture and individual conscious-

ness. Much more is at stake than a tale of two functions.” In a very general way, Gillham’s dichotomy helps resolve the predicament of reconciling, through dichotomizing, the two extremes of human cognitive activity – the discrete logical way of human conscious thinking and the continuous flow of thought. In research, hence, one has to employ not one approach but a combination of at least two. In terms of Velimir Khlebnikov, the metaphor of my research might be studying the “interplay of quantities” hampered by the fact that this interplay is hidden “behind the twilight of qualities” that is crucial in interpreting the meaning of the interplay.

1.5 Major Approaches

This section will be devoted to outlining certain methodological approaches to Comparative Translation Assessment (CTA). None of the models suggested for comparative textual analysis and evaluation can, however, “ultimately enable the evaluator to pass judgments on what is a “good” or a “bad” translation, since any evaluation “depends on a large variety of factors that necessarily enter into a social evaluative judgment” (House 201:155). Having underlined the fact that it is the linguistic analysis that provides grounds for arguing any evaluative judgment, which opinion I share, House claims that the core concept of translation criticism is translation quality (ibid:156). She calls this concept “problematic” because it is often taken to involve individual value judgments alone. “One should be aware,” warns House, “that one will always be forced to move from a macro-analytical focus to a micro-analytical one, from considerations of ideology, function, genre, register to the communicative value of individual linguistic terms” (ibid.). In presenting her way of interrelating the macro- and micro-analytical levels, Maria Tymoczko (2002:9) uses the optical metaphor from the scientific revolution caused by the discoveries in the field of optics in the 1600s: “... translation studies reflect the new shift in the debate between those who assert the preeminence of linguistic approaches to translation and those who advocate primarily cultural studies approaches to translation.”

In order to designate the two poles of comparative criticism, House uses the terms “value judgments” vs. “linguistic analysis,” both terms perfectly relevant for my argument. This “dual, complementary perspective” that a translation critic/assessor has to adopt leads to the problem of the way the two levels could be reciprocally combined. House’s suggestion is tripartite:

The difference between linguistic analysis and value judgment is often ignored when one talks about the quality of a translation. Both components, the

linguistic and the judgmental ones are, of course, implicit in translation evaluation, but they should not be mixed up, nor should the evaluative component be used in isolation from the linguistic component (2001:155-6).

This suggestion could be described as the “Rule of Three Don’ts”: (1) do not ignore the two components, linguistic and judgmental, (2) do not mix them up, and (3) do not use one component in isolation from another. In this set of rules, Rules 1 and 2 are meant to delineate the components and are thus instrumental, while Rule 3 can be viewed as an interface in the sense that it concerns the interrelation of the two components, and is hence the most problematic, e.g., Theo Hermans designates the problem of the “transition from micro-level to macro-level analysis or vice versa” (1999:55) as number one on his list of CTA difficulties.

There are various suggestions as to how Rule 3 could be implemented. For example, Maria Tymoczko (2002:17) believes that one should work “toward the macroscopic from the direction of the microscopic, or vice versa, so that one’s data from the macroscopic level are complemented and confirmed by data from the microscopic level,” while José Lambert and Hendrik van Gorp suggest a spiral way of moving between macro- and micro-levels, probably because their comprehensive model (1985) is multifunctional. Edoardo Crisafulli (2002:26) also argues in favor of an eclectic methodology combining both poles, while insisting on the necessity to allow for translation history.

Within the adopted “judgmental/linguistic” dichotomy, the quotations from the EO translations criticism cited in 1.4 can be undoubtedly called “judgmental” or “emotive.” In this study, the “value” judgment pole of the dichotomy is ensured by the three control groups of collected evidence. The first group presents the results that may be termed “intended function” (6.1.1) and it deals with the self-assessment of the EO translators declared in the introductions to their translations. The second control group of results is the “impressionist” estimates of EO translations by professional critics (6.1.2) exemplified in 1.4. The third group of results (6.1.3) can be described as the only purely functional one because, based on the questionnaire sent to university instructors in the USA, it deals with the “receiving side.”

The linguistic pole of the dichotomy will be covered with the quantitative data obtained by the formal Token Equivalence Method (TEM) suggested in Chapter Four. In this domain of comparative linguistic research, the complex verbal phenomena have to be reduced to their elementary albeit still divisible unit (a token) on a par with atoms in physics. Such a reductionist simplicity is the only mode of assessment that makes possible both direct observation and quantification.

Now that the quality-centered pole of the analysis has been secured with the three control groups of assessment discussed above, and that the quantity-centered pole has been outlined for elaboration, let us return to possible ways of correlating the poles. The culmination of my project is a direct comparison of the two clines of comparative assessment obtained via “emotive” and “cognitive” methods. In other words, House’s Rule 3 is discarded altogether since the obtained results are not combined but are simply compared. If various sets of the obtained results exhibit complementarity and convergence, the suggested hypothesis must be correct – metaphorically speaking, the ouroboros is expected to bite its tail.

If, as the French philosopher Maurice Blanchot famously remarked, “answers are the diseases of questions,” then “superlatives may be considered infectious when they express an affective state with no backing by critical thought” (Mello 2004:1). My guidelines for devising a formal comparative method forbid the use of metaphoric descriptions or “flowery” metaphors with a metaphysical flavor; instead, they will outline a certain formal “device,” definitely imperfect. However, when the same imperfect “device” is used to measure different items, it does reveal their systemic differences. In the next section, we shall consider some major problems encountered in contemporary CTA models that have to be allowed for in the suggested method.

1.6 Key Problems

Comparative translation assessment is a routine operation carried out on a daily basis by professional critics, teachers, editors, and general readers. Despite an ever-growing international drive to set up criteria for evaluating commercial translations, despite the fact that a number of proposals concerning standards have been suggested in Germany and Italy, this branch of translation studies has not yet produced a unified set of criteria for evaluating translations criteria that are empirically sound, transparent, systemic and intersubjectively reliable – the super complex task on a par with text classification that is doomed to remain unsolved. On the practical level, the known CTA methods have so far failed to produce either CTA techniques for error analysis to be used in instructive CTA, or objective quality evaluation in self-revision to be employed by translators, or assessing the quality of translations and arbitrating between various translation solutions required in editing.

The problems encountered in CTA of literary translations are abundant both at the theoretical and practical levels. In his analysis of CTA models, Theo Hermans describes the following problems as most acute: “the

transition from micro-level to macro-level analysis or vice versa, the definition of the unit of comparison, the place and the role of interpretation in the analysis, the selection of representative fragments ... and the sheer practical applicability of the model in question" (1999:55). One of the least resolved problems is that of the unit of translation/assessment, e.g., Albrecht Neubert claims (1973:52) that "the acceptability of the results reduces with text segmentation going down to the word." Acceptable for what? Acceptable by whom?

These questions bring about the problem of the assessor, or, in terms of Hönig, the "relevant user," who can in theory take on any guise: "... after all," believes Mark Shuttleworth, "each item in ST contains a plexus of possible connotations and allusions, and no one will be able to predict which of these will be perceived as important by the majority of readers, and which will leave the greatest hole if it is not adequately rendered" (1998:81). Hönig's suggestion that research in the cognitive sciences may in the near future provide an indication of the typical reader's reaction seems, in the opinion of many scholars, far too optimistic, since "... judging the reader's reaction involves ... judging the reader's level of present knowledge and even anticipating their future level of knowledge" (Anderman & Rogers 1998:60).

Let us conclude this chapter with a set of criteria that will be applied both to the case studies discussed in Chapter Two and to the presentation of the Token Equivalence Method (TEM). In terms of the object of the study, the method is meant to be product-oriented (Holmes) and text-bound (Reiss), with emphasis on the "intra-textual" (Nord) "formal features" (Wimsatt & Beardsley) of the "linguistic component" (House); in terms of the research paradigm, the method can be described as based on a "micro-level" (Hermans) or "linguistic analysis" (House) conducted within the "positivist paradigm" (Gullham); while in terms of the critical approach, the method may be seen as "objective" (Holmes) and "cognitive" (Wimsatt & Beardsley).

A specific research design must be based on a hypothesis set for testing. In this study, the working hypothesis is as follows: both quantitative and qualitative methods of assessment are expected to produce compatible results. Since the "judgmental" component is provided by the collected critical material of the three control groups, the "linguistic" pole is expected to yield quantitative data on the same text by employing the elaborated comparative method. It should be emphasized that with the goal designated as measuring the ST-TT distance in order to locate several translations on a certain "cline of fidelity" to the source text in terms of preserving its

content and form, translated texts are not assessed as literary texts; what is gauged instead is their formal interrelations with the source text.

It goes without saying that non-quantified research methods are prevailing by a large margin albeit there have been laudable attempts to use quantitative measurements in CTA as well. For instance, Theo Hermans interprets the quantitative findings in Van Leuven-Zwart's CTA method (Chapter Two) as follows: "... the density of shifts [ST distortions] can be expressed in quantitative terms. If roughly one shift, of whatever kind, occurs in each transeme, we have an average of 100%. ... As Van Leuven-Zwart admits himself, however, the percentages as such are not very informative, since the macro-structural impact of a shift depends on its quality rather than on frequency" (1990:88).

Therefore, here we come again to the inevitable "quality" aspect in CTA. To avoid this aspect, in the method to be designed for this project, it is the density of "non-shifts" that will be measured when defining a distance between the target text(s) and the original. Despite the impending criticism of "studying translations in a vacuum," of employing a "stable" unit of comparison and "neat divisions," I intend to test my method. My approach styles itself as descriptive and empirical, context-free, value-free, purely formal, even formalistic and statistical – it is a net with a very coarse mesh. The method, however, allows one to arrange information in various manageable forms like matrices and tables crowned with the Translation Quotient (TQ), a one-figure final quantitative result for each translation.

Once the hypothesis and the goal are formulated, the task is to choose a relevant sample of text for comparison. As regards the sample size, Theo Hermans asserts that there should be "a balance between economy and credibility: the sample should be large enough to be credible in the light of the purpose of the exercise, but small enough to permit appropriate depth" (1990:70). Given the scope of the present project, an eight-chapter novel in verse and its nineteen full translations, "achieving the balance between economy and credibility" is immensely problematic. Various modifications of the Tokens Equivalence Method will be tested (4.4) on various excerpts of the original text, from 35 stanzas (10% of the original text) through six stanzas to one stanza. The idea that large samples are not required in order to obtain reliable quantitative data will be tested.

As was mentioned earlier, the systematic approach inherent in the proposed method will be enhanced with numbering every token in the source text, as was suggested by Jean-Paul Vinay and Jean Darbelnet in their influential "Comparative Stylistics of French and English. A Methodology of Translation:" "... the SL and TL units can be numbered