A Model for Translation Accuracy Evaluation and Measurement: A Quantitative Approach

By
Junxiong Huang (Harry)
BA (Beijing Foreign Studies), M Ed (Toronto)

A thesis submitted in fulfillment of the requirement for the degree of
Doctor of Philosophy

Department of Linguistics
Macquarie University
2007
This is to certify that I, JUNXIONG (HARRY) HUANG, being a candidate for the degree of Doctor of Philosophy, am aware of the policy of the University relating to the retention and use of higher degree theses as contained in the University's Doctoral Degree Rules generally, and in particular rule 7(10).

In the light of this policy and the policy of the above Rules, I agree to allow a copy of my thesis to be deposited in the University Library for consultation, loan and photocopying forthwith.

Signature of Candidate

Signature of Witness

Date this 24th day of August 2007

The Academic Senate on 19 February 2008 resolved that Mr. Huang Jun Xiong had satisfied the requirements for admission to the degree of Doctor of Philosophy.

This thesis represents a major part of the prescribed program of study.
Copyright in relation to this Thesis

Under the Copyright Act 1968 (several provisions of which are referred to below), this material must be used only under the normal conditions of scholarly fair dealing for the purposes of research, criticism or review. In particular no results or conclusions should be extracted from it, nor should it be copied or closely paraphrased in whole or in part without the written consent of the author. Proper written acknowledgement should be made for any assistance obtained from this material.

Under Section 35 (2) of the Copyright Act 1968 'the author of a literary, dramatic, musical or artistic work is the owner of any copyright subsisting in the work'. By virtue of Section 32 (1) copyright 'subsists in an original literary, dramatic, musical or artistic work that is unpublished' land of which the author was an Australian citizen, an Australian protected person or a person resident in Australia.

The Act, by Section 36 (1) provides: 'Subject to this Act, the copyright in a literary, dramatic, musical or artistic work is infringed by a person who, not being the owner of the copyright and without the licence of the owner of the copyright, does in Australia, or authorises the doing in Australia of, any act comprised in the copyright'.

Section 31 (1) (a) (i) provides that copyright includes the exclusive right to 'reproduce the work in a material form'. Thus, copyright is infringed by a person who, not being the owner of the copyright, reproduces or authorises the reproduction of a work, or of more than a reasonable part of the work, in a material form, unless the reproduction is a 'fair dealing' with the work 'for the purpose of research or study' as further defined in Sections 40 and 41 of the Act.

Section 51 (2) provides that 'Where a manuscript, or a copy, of material of other similar literary work that has not been published is kept in a library of a university or other similar institution or in an archives, the copyright in the material or other work is not infringed by the making of a copy of the material or other work by or on behalf of the officer in charge of the library or archives if the copy is supplied to a person who satisfies an authorized officer of the library or archives that he requires the copy for the purpose of research or study'.

* Thesis' includes 'treatise', 'dissertation' and other similar productions.
Abstract

Translation quality assessment (TQA) has been part of the translating process since Marcus Tullius Cicero (106-43BCE), and earnest studies on TQA have been conducted for several decades, but there has been no breakthrough in standardized TQA. Though the importance of TQA has been stressed, agreement on specific means of TQA has not been reached. As Chesterman and Wagner summarize, "Central to translation [ ... ]," “[q]uality assessment is so complicated — especially if it is to be objective and reproducible” (2002: 80-81). The approaches to TQA published throughout the past millennia, by and large, are qualitative. “Whereas there is general agreement on the requirement for a translation to be ‘good,’ ‘satisfactory,’ or ‘acceptable,’ the definition of acceptability and of the means of determining it are matters of ongoing debate and there is precious little agreement on specifics” (Williams, 2004: xiv). Most published TQA approaches are neither objective nor reproducible.

My study proposes a model for fuzzy standardized TQA through a quantitative approach, which expresses TQA results in numerical terms in a consistent manner. My model is statistics-based, practice-based and practice-oriented. It has been independently tested by eleven professors from four countries, fifteen senior United Nations translators, and fifty reader evaluators. My contrastive analysis of 23,000 pages of bilingual and multilingual texts has identified the unit of translation — the orthographic sentence in context, which is also verified by the results of an international survey among 66 professional translators, the majority of whom also confirm that they evaluate translations sentence by sentence in context. Halliday and Matthiessen’s functional grammar theory, among others, provides my model for quantitative TQA with its theoretical basis, while the international survey, the necessary data. My model proposes a set of six Fuzzy Functional Translation Grammar terms, a grammar concept general enough to cover all grammar units in the translated orthographic sentence. Each term represents one type of error which contains from one to three sub-categories. Each error is assigned a value — the mean of the professional markers’ deductions for relevant artificial errors and original errors. A marking scheme with sixteen variables under eight attributes is thus created. Ten marks are assigned to each unit of TQA, the sentence. For easy calculation, an arithmetic formula popularly used in statistics ( \( \frac{\sum X}{n} \) ) is adopted. With the assistance of a simple calculator, the evaluator can calculate the grade of a sentence, a
sentence group, and the overall grade for an entire TT, regardless of its length.

Perfect reliability or validity in any form of measurement is unattainable. There will always be some random error or noise in the data (McClendon, 2004: 7). Since it is the first of its type, I do not claim that my model is perfect. Variation has been found in the results of the testing performed by scholars and professional translators, but further testing based on two "easy" (markers' comment) sentences by the 50 reader evaluators respectively achieves 98% and 100% consistency, which indicates that markers' competence may equal constancy or that proper marker training and/or strict marker examination will minimize inconsistency among professional markers. My model, whose formulas withstand testing at the theoretical level and in practice, is not only ready for application, but it has profound implications beyond TQA, such as use in machine translation, and for other subjects like the role of the sentence in translation studies and translating practice.
Acknowledgements

I owe a debt of gratitude to Professor Christian Matthiessen and Dr. Canzhong Wu for giving me the opportunity to develop my idea of quantitative TQA through a formula approach in this thesis. It was their continuous understanding, timely support, and invaluable advice that encouraged me right through to the end of the project. Their emotional support became my spiritual pillar while I was experiencing the numerous challenges and difficulties in my attempt to overcome the impossible.

The over 300 professional translators, writers, editors, university and college professors as well as the approximate 200 university and college students who contributed their translating and marking criteria, time, and wisdom deserve my heartfelt thanks that words fail to convey.

Administrators and support staff of Macquarie University offered me the best support in every way possible. Professor John Hooper, Robyn Guilmette and Lanna Leung, among others, deserve my gratitude.

Finally, special thanks go to my wife and children as well as my mother. My eldest child, Wendy, contributed so much to my work; seven-year-old Alice offered to help me sort out my papers and books in the alphabetical order one winter night; five-year-old Jeffrey has never complained about missing playtime with Daddy. My mother, the most important supporter, encouraged me through the completion of my thesis.

I thank each and every one of the 500 plus people for enabling me to make this little contribution to the field of translation studies and in turn to the human society.

Friendly reminder:

I have done the entire thesis by myself, but I use “we”, “our” and “us” instead of “I”, “my” and “me” in the thesis to acknowledge the important advice offered by Professor Christian Matthiessen and Dr. Canzhong Wu. The reader is kindly advised to interpret the plural pronouns as singular ones for the author, or PhD candidate, accordingly.
I, Junxiong Huang (Harry), declare that this thesis has not been submitted in full or part for a higher degree to any other university or institution. The work is entirely my own, and where sources have been quoted, acknowledgement has been made in full. All research conducted forming part of this thesis has been completed under Macquarie University Human Ethics Committee permit # HE24JUN2005-DO4195.

Signed: [Signature] Date 08/30/2007
Table of Contents

Abstract .......................................................................................................................... ii

Acknowledgements ....................................................................................................... iv

List of Figures ................................................................................................................ xiii

List of Tables .................................................................................................................. xiv

List of Abbreviations ..................................................................................................... xix

Chapter 1  Introduction ................................................................................................. 1
  1.1 The nature of the problem ....................................................................................... 1
  1.2 The significance of the study ................................................................................ 2
  1.3 Scope of evaluation ............................................................................................... 3
  1.4 Methodology ......................................................................................................... 4
    1.4.1 The data ......................................................................................................... 4
    1.4.2 Primary data .................................................................................................. 4
    1.4.3 Participants and the different versions of the survey ..................................... 4
    1.4.4 Calls for participation ................................................................................... 5
    1.4.5 Non-primary data .......................................................................................... 6
    1.4.6 Data collection .............................................................................................. 7
    1.4.7 Data analysis ................................................................................................ 8
  1.5 The conceptual framework for the problem ......................................................... 8

Chapter 2  Literature review ......................................................................................... 10
  2.1 Untested translation theories ................................................................................ 10
  2.2 Translation theory divorced from practice ............................................................ 10
  2.3 Repeated theses from Cicero to the present-day .................................................. 11
  2.4 The importance of the imperfect translation theory, etc ....................................... 11
  2.5 TQA in the making ............................................................................................... 12
    2.5.1 Translation approaches and TQ criteria in the making .................................. 12
  2.6 Shadow TQA theory since the 1950s .................................................................... 20
    2.6.1 Thought-provoking claims that delimit TQ .................................................. 20
    2.6.2 Development of contemporary TQA theories ............................................. 21
2.7 Definitions of “translation” and “translator” as TQA approaches .................. 22
2.8 TQA criteria divorced from definitions of “translation” .............................. 23
  2.8.1 Nida and Taber’s TQA approach ....................................................... 23
  2.8.2 House’s historic study ........................................................................ 24
  2.8.3 Nord’s bi-directional approach ............................................................ 24
  2.8.4 Schäffner’s active role in TQA research ............................................. 25
  2.8.5 Chesterman’s translation norms .......................................................... 25
  2.8.6 Spyridakis et al’s study on Simplified English ................................. 25
  2.8.7 Campbell’s approach to source difficulty .......................................... 26
  2.8.8 Darwish’s TQA concept ..................................................................... 27
  2.8.9 Al-Qinai’s model for TQA and its limits .......................................... 27
  2.8.10 Liu and G. Huang and SFG in TQA ................................................. 28
  2.8.11 Eric Steiner’s view of co-generation of text ...................................... 28
  2.8.12 Williams’ argumentation-centred approach .................................. 29
  2.8.13 Universally acceptable TQA model yet to come .......................... 30
2.9 TQA in practice ....................................................................................... 30
  2.9.1 The United Nations Chinese Guide (UNCG) .................................... 30
  2.9.2 Two Canadian Models ..................................................................... 32
  2.9.3 NAATI’s model ................................................................................ 33
  2.9.4 A review of the UNCG, the Canadian models and the NAATI model .... 35

Chapter 3 Identification of the unit of translation ................................................. 37
  3.1 Definition of UT adapted ..................................................................... 38
  3.2 Methodology for identifying the UT .................................................... 38
    3.2.1 Sources of contrastive analysis ...................................................... 38
    3.2.2 Sampling for analysis .................................................................... 38
    3.2.3 Surveying for raw data ................................................................. 39
  3.3 Analysis of translation samples .............................................................. 39
    3.3.1 Types of translation analyzed ....................................................... 39
    3.3.2 Analysis of non-literary translation samples .................................. 39
    3.3.3 Analysis of literary translation excluding poetry .......................... 46
    3.3.4 Poetry translation ......................................................................... 53
  3.4 The UT used by professional translators ................................................ 56
    3.4.1 How translators translate .............................................................. 56
5.5 Maximum tolerable inaccuracy at the clause level (MTICL) ........................................... 89
  5.5.1 Responses by language ............................................................................................... 90
  5.5.2 Sub-group means of MTICL ...................................................................................... 90
5.6 MTIRCL Index .................................................................................................................. 92
5.7 Tolerance of inaccurate use of punctuation, capitalization, etc. (TIUPC) ................. 93
5.8 Tolerance of unsuccessfully translated or un-translated paragraphs (TUTUP) .......... 93
5.9 Acceptance of the international mean ............................................................................ 95
5.10 Numerical grades and word/letter grades .................................................................. 95
5.11 Quantifying accuracy and inaccuracy rates ................................................................. 96
5.12 Calculating the minimum acceptability ........................................................................ 97
5.13 ST imperfectness in statistical terms ............................................................................ 98
5.14 TQ assessor qualifications ............................................................................................ 99
  5.14.1 Formal education required for literary TQA ............................................................. 100
  5.14.2 Additional qualifications required of an evaluator ............................................... 101
  5.14.3 Number of translators required as TQ assessors .................................................. 102
  5.14.4 Number of readers required as evaluators ............................................................. 104
5.15 Sampling for TQA .......................................................................................................... 105
5.16 Sampling of sentences .................................................................................................. 105
  5.16.1 Sentence sample sizes ............................................................................................. 106
  5.16.2 Sentence sample sizes in percentages ..................................................................... 107
5.17 Number of mistakes that causes failure .................................................................... 107
  5.17.1 One mistake per ten sentences means failure ......................................................... 107
  5.17.2 Number of sentences to be evaluated ..................................................................... 108
5.18 TT rejection index (TTR Index) ..................................................................................... 108
  5.18.1 Editors’ TT rejection criterion .................................................................................. 108
  5.18.2 Translator editors’ TT rejection criterion ................................................................. 109
  5.18.3 Translators’ TT rejection criterion .......................................................................... 109
  5.18.4 Overall TT rejection criterion .................................................................................. 109
5.19 Mistranslated or un-translated sentences .................................................................... 110
  5.19.1 Unacceptable number of mistranslated or un-translated sentences ....................... 110
  5.19.2 Mistranslated or un-translated sentences rejection index (MUTSR Index) ........ 111
5.20 Cross examination of the evaluation at text level (two sets of poems) .................... 112
  5.20.1 Comparison of group differences .......................................................................... 114
  5.20.2 Testing the validity of the respondents’ grading, etc. ........................................... 115
Appendix D: Complete International Survey in Chinese .................................................. 260
Appendix E: Guide to Mark Deductions ........................................................................ 291
Appendix F: English-Chinese Translation Evaluation ...................................................... 293
Appendix G: Chinese-English Translation Evaluation Set 1 ......................................... 295
Appendix H: Chinese-English Translation Evaluation Set 2 ......................................... 299
Appendix I: Translation Evaluation (2 sentences) by Readers ..................................... 302
References .................................................................................................................... 303
List of Figures

Figure 3-1: Syntax comparison between Confucius’ ancient Chinese (ST) and English (TT). 44
Figure 3-2: Comparison of sentence structure between English and French of CCRF .......... 45
Figure 4-1: COSB Model ........................................................................................................ 73
Figure 6-1: Clause analysis through the traditional grammar approach .............................. 125
Figure 6-2: Clause analysis through the functional grammar approach .............................. 126
Figure 6-3: Clause analysis through the pedagogical grammar approach ............................ 126
Figure 6-4: FFTG approach to clause analysis in comparison with the traditional approach  128
Figure 6-5: Functional grammar clause analysis (adapted from Halliday & Matthiessen, 2004: 54) ........................................................................................................................................ 131
Figure 6-6: FFTG clause analysis ............................................................................................. 131
List of Tables

Table 3-1: Sentence arrangements in TTs of Articles 10, 11 & 12 of the UDHR ......................... 40
Table 3-2: Sentence arrangements in the sample page of Parallel New Testament ...................... 41
Table 3-3: Mean number of Chinese sentences in TT sample of Parallel New Testament ................. 42
Table 3-4: TT sentence arrangements of Verse 30 of "Surah 30 of The Koran" ......................... 43
Table 3-5: Syntax comparison between ST and TT of the PLPRC ........................................... 46
Table 3-6: Comparison of ST and TT of Soul Mountain through analysis of punctuation .......... 47
Table 3-7: Comparison of Hawkes’ and the Yangs’ TT sentence arrangements ...................... 48
Table 3-8: ST and TT sentence arrangements in Journey to the West .................................... 49
Table 3-9: Some differences between the ST and TT of Pu’s “Lien Shiang” ......................... 50
Table 3-10: Comparative analysis of ST and TT sentence structure in Utopia .................. 51
Table 3-11: Sentence arrangements of S. Zhu’s Chinese translation of Macbeth .................. 51
Table 3-12: Paragraph and sentence arrangements in the TT of The Story Girl .................. 52
Table 3-13: Means of sentences & characters of the 11 TTs of "Of Studies" ...................... 53
Table 3-14: Means of sentences & characters of the 11 TTs in comparison with the ST ... 53
Table 3-15: Features of the Chinese translation of Blake’s “The Tiger” ......................... 53
Table 3-16: Features of the Korean translation of Meyer’s “Pemberton Park” .................. 54
Table 3-17: Features of Xu’s English translation of ancient Chinese poetry .................. 54
Table 3-18: Features of Zheng’s self-translated poems (numbers combined) .................. 55
Table 3-19: Features of Chao’s poems written in English and Chinese ....................... 55
Table 3-20: Summary of translators’ responses to “How do you translate?” .................. 56
Table 3-21: Percentage of translation rendered sentence by sentence by published translators ......................................................... 58
Table 3-22: How translators and editors evaluate and edit translation ....................... 59
Table 3-23: Approaches to evaluation and editing of writing (based on multiple answers) .... 59
Table 4-1: Weight assigned for CP, OS, and SA by professional translators ..................... 66
Table 4-2: Means of CP, OS, and SA in comparison ......................................................... 67
Table 5-1: MOTI index ........................................................................................................ 81
Table 5-2: Translation acceptable to professional translators, writers, editors, etc. .......... 83
Table 5-3: Translation acceptable to the professionals — a breakdown by degree ............ 83
Table 5-4: Translation acceptable to the professionals — a breakdown by profession ........ 84
Table 5-5: Opinions of translators, writers and editors on the possibility of absolute accuracy xiv
Table 5-35: Translator editors’ mean TTR index (per 100 sentences) .................................................. 109
Table 5-36: Translators’ mean TTR index (per 100 sentences) .......................................................... 109
Table 5-37: Overall mean TTR index (per 100 sentences) ................................................................. 109
Table 5-38: Breakdown of those responding to the unacceptable number of mistranslated or un-translated sentences .......................................................... 110
Table 5-39: Unacceptable number of mistranslated or un-translated sentences per 20 paragraphs .......................................................... 110
Table 5-40: Editor MUTSR Index ........................................................................................................ 111
Table 5-41: Editor-translator MUTSR Index ....................................................................................... 111
Table 5-42: Translator MUTSR Index ................................................................................................. 112
Table 5-43: Overall MUTSR Index ..................................................................................................... 112
Table 5-44: Grading of Excerpts 1, 3, 5 and two poems compared by language and profession .............................................................................................................. 113
Table 5-45: The mean grade awarded to each version of the three texts by the average sub-group .............................................................................................................. 114
Table 5-46: Mean grades awarded by professional groups (3 excerpts) ................................................. 114
Table 5-47: Grades awarded for the excerpts by Zhou, Gao & Chao as well as the two poems .............................................................................................................. 115
Table 5-48: Maximum difference allowed between ST readers’ and TT readers’ evaluation 115
Table 6-1: TQ assessors’ highest degrees ............................................................................................ 118
Table 6-2: TQ assessors’ professional titles .......................................................................................... 118
Table 6-3: TQ assessors’ work and institutions (based on all answers available) .................................. 118
Table 6-4: Breakdown of TQ assessors (based on first answers only) .................................................. 119
Table 6-5: Average length of service .................................................................................................... 119
Table 6-6: Mean weight of CP, OS, and SA, and passing line ............................................................... 120
Table 6-7: Summary background information of the international markers ........................................ 132
Table 6-8: Mean deductions for artificial errors of Q1 ........................................................................ 135
Table 6-9: Mean deductions for artificial mistakes created for Q2 ................................................... 136
Table 6-10: Mean mark deduction of “*After came here, I tried to find a job.” .................................. 137
Table 6-11: Mean deductions for “*To drive a car, always need a driver’s license.” .......................... 138
Table 6-12: Mean deductions for “*The computer that is used everywhere to save time and money.” .............................................................................................................. 138
Table 6-13: Mean deduction of missing subjemes as CP ................................................................. 139
Table 6-14: Mean deduction of sentences with missing subjemes .................................................... 139
List of Abbreviations

Following is a list of commonly used abbreviations:

AA  Absolute accuracy
AB  Absolute balance
AccA Acceptable accuracy
AccI Acceptable inaccuracy
AI  Accuracy index
CI  Clause (independent) or independent clause
COSB: Balance between *contenphysique*, *origispirit*, and *stylappearance*, or C.O.S. in Balance
CP  *Contenphysique*
EI  Error index
FFTG Fuzzy Functional Translation Grammar
OS  *Origispirit*
PM  Presence of meaning
RA  Relative/realistic accuracy
RB  Relative balance
SA  *Stylappearance*
ST  Source language text or original text
STQA Standardized translation quality assessment
SVU Subjeme-vheme unit
TQ  Translation quality
TQA Translation quality assessment
TT  Translated text, or target language text
UA  Unit of assessment
UT  Unit of translation
UTQA Unit of translation quality assessment