ΜΕΤΑ

The Application of Argumentation Theory to Translation Quality Assessment

Malcolm Williams

See table of contents

Explore this journal

Cite this article

Article abstract

Translation quality assessment (TQA) models may be divided into two main types: (1) models with a quantitative dimension, such as SEPT (1979) and Sical (1986), and (2) non-quantitative, textological models, such as Nord (1991) and House (1997). Because it tends to focus on microtextual (sampling, subsentence) analysis and error counts, Type 1 suffers from some major shortcomings. First, because of time constraints, it cannot assess, except on the basis of statistical probabilities, the acceptability of the content of the translation as a whole. Second, the microtextual analysis inevitably hinders any serious assessment of the content macrostructure of the translation. Third, the establishment of an acceptability threshold based on a specific number of errors is vulnerable to criticism both theoretically and in the marketplace. Type 2 cannot offer a cogent acceptability threshold either, precisely because it does not propose error weighting and quantification for individual texts. What is needed is an approach that combines the quantitative and textological dimensions, along the lines proposed by Bensoussan and Rosenhouse (1990) and Larose (1987, 1998). This article outlines a project aimed at making further progress in this direction through the application of argumentation theory to instrumental translations.

The Application of Argumentation Theory to Translation Quality Assessment

MALCOLM WILLIAMS
University of Ottawa, Ottawa, Canada

RÉSUMÉ

Les modèles d'appréciation de la qualité des traductions peuvent se diviser en deux groupes principaux : (1) les modèles quantitatifs, tels que le SEPT (1979) et le Sical (1986), et (2) les modèles non quantitatifs et textologiques proposés par Nord (1991) et House (1997), entre autres. D'une part, le premier type accuse plusieurs lacunes importantes, qui découlent de l'approche microtextuelle (échantillonnage et analyse intra-phrastique) et de la quantification des fautes, dimensions inhérentes au modèle. En effet, (1) en raison des contraintes de temps, il ne peut servir à porter un jugement sur le texte entier que sur la base de probabilités statistiques, (2) l'analyse microtextuelle entrave forcément l'évaluation du contenu global de la traduction, et (3) en établissant un seuil d'acceptabilité fondé sur un nombre précis de fautes, on prête le flanc à la critique sur le plan théorique et sur le marché de la traduction. D'autre part, le deuxième type ne présente pas de seuil convaincant non plus, parce qu'il n'intègre pas la pondération ou la quantification des fautes à l'analyse d'une traduction donnée. Il faut donc viser un modèle qui s'inspire, à l'instar de ceux proposés par Bensoussan et Rosenhouse (1990) et Larose (1987,1998), des deux dimensions quantitative et textologique. Le présent article résume l'un des principaux axes d'un projet en cours qui proposera quelques pistes de solution grâce à l'application de la théorie de l'argumentation aux textes pragmatiques.

ABSTRACT

Translation quality assessment (TQA) models may be divided into two main types: (1) models with a quantitative dimension, such as SEPT (1979) and Sical (1986), and (2) non-quantitative, textological models, such as Nord (1991) and House (1997). Because it tends to focus on microtextual (sampling, subsentence) analysis and error counts, Type 1 suffers from some major shortcomings. First, because of time constraints, it cannot assess, except on the basis of statistical probabilities, the acceptability of the content of the translation as a whole. Second, the microtextual analysis inevitably hinders any serious assessment of the content macrostructure of the translation. Third, the establishment of an acceptability threshold based on a specific number of errors is vulnerable to criticism both theoretically and in the marketplace. Type 2 cannot offer a cogent acceptability threshold either, precisely because it does not propose error weighting and quantification for individual texts. What is needed is an approach that combines the quantitative and textological dimensions, along the lines proposed by Bensoussan and Rosenhouse (1990) and Larose (1987, 1998). This article outlines a project aimed at making further progress in this direction through the application of argumentation theory to instrumental translations.

MOTS-CLÉS/KEYWORDS

quality assessment, quantitative models, non-quantitative models, argumentation theory, instrumental translations

> The assessment of translator performance is an activity which, despite being widespread, is under-researched and under-discussed.
>
> Hᴀᴛɪᴍ ᴀɴᴅ Mᴀsᴏɴ 1997: 199

1. Introduction

Translation quality assessment (TQA) is not a new field of inquiry. Moreover, it has the distinction of being one that interests a broad range of practitioners, researchers and organizations, whether their focus is literary or instrumental (pragmatic) translation. Concern for excellence in literary translation or translation of the Scriptures dates back centuries. Quality in instrumental translation as a subject of discussion is a more recent phenomenon, but as far back as 1959, at an FIT international symposium on quality in Paris, E. Cary and others were already debating the requirements of a good translation. More recently still, with the advent of globalization, the coming of age of translation as part of the language industries, and the concomitant emphasis on "total quality" and ISO certification in private industry at large, special issues of *Circuit* (1994) and *Language International* (1998) have been devoted to quality assurance processes, professional standards, and accreditation, and German and Italian standardization organizations have issued national translation standards.

The reasons for people's interest in quality and TQA have, of course, evolved: where they were once primarily aesthetic, religious and political, they are now primarily professional and administrative (e.g., evaluation of students) and economic and legal (e.g., predelivery quality control/assurance; postdelivery assessment to ensure that terms of contract have been met by supplier).

In short, the relevance of, and justification for, TQA is stronger than ever. Yet whereas there is general agreement about 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. National translation standards may exist but, as the organizers of a 1999 conference on translation quality in Leipzig noted:

> there are no generally accepted objective criteria for evaluating the quality both of translations and of interpreting performance. **Even the latest national and international standards in this area—DIN 2345 and the ISO-9000 series—do not regulate the evaluation of translation quality in a particular context. […]**
> **The result is assessment chaos.**
> (Institut für Angewandte Linguistik und Translatologie: 1999, my emphasis)

This article covers part of a larger study in which I will propose a full-text, argumentation-centred approach to TQA as a means of resolving this dilemma. Here, I will focus on one component of the study: analysis and comparison of source-text (ST) and target-text (TT) argument macrostructures as an efficient means of determining translation quality and as a tool that could usefully complement other approaches.

The problems standing in the way of consensus and coherence in TQA are legion, ranging from the debate over whether and how to factor in conditions of production and difficulty of source text to the degree of importance placed on target-language defects. However, my study bears specifically on the following TQA problems.

(A) Sampling versus full-text analysis

TQA has traditionally been based on intensive error detection and analysis and has therefore required a considerable investment in human resources. It takes time. One means of obviating the problem has been sampling—the analysis of samples of translations rather than of whole texts. Yet this approach has shortcomings. First, the evaluator may not take into account any "compensatory" efforts that the translator may have made in unsampled parts of the text. Second, the evaluator cannot benefit from the co-text in order to grasp the meaning of the text as a whole. Third, not only may the evaluator do an injustice to the translator and the translation because he has not absorbed the whole text, but he may also overrate the translation. This, in Daniel Gouadec's opinion, is what makes the validity of sampling for TQA purposes debatable: "[…] il reste toujours un risque que les erreurs les plus lourdes échappent à l'échantillonnage. Ceci est particulièrement vrai pour les traducteurs confirmés qui demeurent susceptibles de dérapages mal contrôlés mais fulgurants" (1989: 56).

(B) Quantification of quality

Microtextual analysis of samples has been used extensively not only because it saves time but also because it provides error counts as a justification for a negative assessment. Translation services and teachers of translation alike have developed TQA grids with several quality levels, or grades, based on the number of errors in a short text. It is felt that quantification lends objectivity to the assessment. The problem lies with the borderline cases. Assuming that, in order to be user-friendly, such a grid does not allow for well-defined levels of seriousness of error, it is quite possible for a translation containing one more error than the maximum allowed to be as good, if not better than, another translation that contains exactly the maximum number of errors allowed.

(C) Levels of seriousness of error

One way to circumvent the drawbacks of quantification is to grade errors by seriousness: critical/major, minor, weakness, etc. The problem then is to seek a consensus on what constitutes a major, as opposed to a minor, error. For example, an error in translating numerals may be considered critical by some, particularly in financial, scientific or technical material, yet others will claim that the client or end user will recognize the slip-up and automatically correct it in the process of reading.

Furthermore, considerable inconsistency is apparent in the assessment of level of accuracy. Some evaluators will ignore minor shifts in meaning if the core message is preserved in the translation, while others will insist on total "fidelity," even if an omission of a concept at one point is offset by its inclusion elsewhere in the text.

(D) Multiple levels of assessment

Darbelnet (1977: 16) identifies no fewer than nine levels, or parameters, at which the quality of a translation should be assessed: accuracy of individual translation units; accuracy of translation as a whole; idiomaticity; correctness of target language; tone; cultural differences; literary and other artistic allusions; implicit intentions of author; adaptation to end user. Other models provide for an assessment for accuracy,

target language quality and format (appearance of text). The problem is this: assuming you can make a fair assessment of *each* parameter, how do you then generate an *overall* quality rating for the translation?

(E) TQA purpose/function

The required characteristics of a TQA tool built for formative evaluation in a university context may differ significantly from one developed for predelivery quality control by a translation supplier. According to Hatim and Mason, "Even within what has been published on the subject of evaluation, one must distinguish between the activities of assessing the quality of translations […], translation criticism and translation quality control on the one hand and those of assessing performance on the other" (1997: 199).

It is these issues that have provoked the most intense debate, even outcry, over the validity, reliability and usefulness of TQA and have engendered accusations of subjectivity on a regular basis. Clearly, "the devil is in the detail." It is not surprising that it has proved impossible to establish a quality standard that meets all requirements and can be used to assess specific translations. Hence, as noted above, DIN 2345 follows in the footsteps of the ISO 9000 series, erring on the side of caution and proposing guidelines for quality control. What is to be standardized is not the level of quality of a translation but a set of procedures for achieving that level. In *Language International,* Sturz explains the limits of the new German standard (DIN 2345):

> The important issue of measuring the quality of translations by rating them […] cannot be solved by a standard. However, a standard can provide specific rules for the evaluation process. Such measures […] include completeness, terminological correctness, grammar and style, as well as adherence to a style guide agreed to between the buyer and the translator. […] DIN 2345 is not a certification standard. (1998: 19, 41)

In fact, what the new standard offers is a set of "normative statements" about the various parameters of translation, and as such it echoes the concepts of translation norms and laws developed by functional theorists such as Toury.

At this juncture, examination of the specifics of actual TQA approaches will serve to highlight what progress has been made in resolving TQA issues and what areas still require improvement.

2. Existing TQA approaches

Existing TQA models, whether they have actually been put into practice or have merely been proposed, all have one feature in common: categorization of errors lies at the heart of each approach. That being said, their concept of categorization differs, according to whether or not they incorporate quantitative measurement, and they can be divided into two schools on that basis.

2.1. *Models with a quantitative dimension*

2.1.1. The **Canadian Language Quality Measurement System** (**Sical**), the TQA model developed by the Canadian government's Translation Bureau, is the best known one, at least on the Canadian scene. It was developed both as an examination

tool and to help the Bureau to assess the quality of the 300 million words of instru-
mental translation it delivered yearly. Initially based on a very detailed categorization
of errors—over one hundred types were identified and could be assessed by evalua-
tors—Sical had evolved by 1986 into a scheme based on a twofold distinction
between (1) transfer and language errors and (2) major and minor errors and on the
quantification of errors. In this third-generation Sical, texts were judged to be of
superior, fully acceptable, revisable, or unacceptable quality, depending on the num-
ber of major and minor errors in a 400-word passage.

In actual fact, it was the major error that was the determining factor in grading
translations. Accordingly, Sical III included a new, more precise definition of such an
error:

> *Translation*: Complete failure to render the meaning of a word or passage that contains
> an essential element of the message; also, mistranslation resulting in a contradiction of
> or significant departure from the meaning of an essential element of the message.
> *Language*: Incomprehensible, grossly incorrect language or rudimentary error in an es-
> sential element of the message. (Williams 1989: 26)

The "essential element of the message" was not defined but was to be related to the
potential consequences of the error for the client. Where an essential element had
been translated erroneously, the translation in its entirety was deemed undeliverable
without revision.

In theory, then, a fully acceptable translation of 400 words could contain as
many as 12 errors of transfer, provided no major error was detected. However, the
designers of Sical III predicated the lowering of the tolerance level on the statistical
probability that a translation with 12 such errors would also contain at least one
major error.

The typology of errors established in this context—a typology modelled on that
of Horguelin (1978)—is indicative of the fact that the quality system by and large
focussed on the word and the sentence, not on the text as a whole. Larose sums up
the approach as follows:

> La grille Sical porte principalement sur les aspects syntaxique et sémantique des textes.
> Elle n'est pas axée principalement sur leur dimension discursive, c'est-à-dire ce qui est
> *au-delà* de la proposition et *entre* les propositions. (Larose 1998: 175)

Issues of coherence and cohesion are not apparent in the Sical guidelines, even
though the primacy of consequence of error and the relationship of major error to
text message would seem to open the door for a discursive, or textological, approach.

Notwithstanding the good judgement and competence of the evaluators and
quality controllers involved and the greater flexibility and precision afforded by Sical
III, the application of a quantified standard continued to spark debate and dissatis-
faction among translators inside and outside the Bureau. Working conditions, dead-
lines, level of difficulty of the source text and the "overassessment" of target language
errors were regularly cited by opponents of the Sical-based quality system

The year 1994 signalled a major shift in the Translation Bureau's approach to
TQA. Implicit in the application of Sical and the quantification of errors was recog-
nition of the fact that translations deemed deliverable contained errors—officially as
many as 12. Since the Bureau was to enter into direct competition with the private
sector in 1995, management concluded that a "total quality" approach was necessary.

Thenceforth "zero defects" was the order of the day; the Bureau was committed to delivering error-free translations to its clients. There was no longer any question of a tolerance threshold and of determining whether that threshold had been crossed in one or more samples. The quality of contractors work is still vetted by sampling, but a quantified tolerance level is no more and translation should, in principle, be examined in their entirety before delivery. Sical continues to be used for examinations as well as for predelivery and performance evaluation purposes.

In general, other translation organizations in Canada have adopted Sical or adapted it to their specific circumstances and requirements (e.g., Ontario government translation services, Bell Canada).

2.1.2. The **Council of Translators and Interpreters of Canada** (CTIC) uses a comparable model for its translator certification examinations, except that "no single error may be considered sufficient to fail a candidate" (CTIC 1994: 3). Each type of error is given a quantitative value (e.g., -10, -5) and the total of the values of errors in the candidate's paper is subtracted from 100: the candidate with 75% or higher passes. Unlike Sical, the definition of major and minor error does not relate error to an essential part of the message:

> *Translation*: Major mistakes, e.g., nonsense, serious mistranslation denoting a definite lack in comprehension of the source language.
> *Language*: Major mistakes, e.g., syntax, grammar, gallicism.

2.1.3. Using works by van Dijk (1980), Widdowson (1979), Halliday and Hasan (1976), and Searle (1969) for the theoretical underpinnings of their model, **Bensoussan and Rosenhouse** (1990) propose a TQA scheme for **evaluating student translations by discourse analysis**. The tool would serve to assess students' comprehension of English in a TEFL context. This pedagogical tool is based on the premise that translation operates on three levels of understanding: "surface equivalence, semantic equivalence (propositional content, ideational and interpersonal elements), and pragmatic equivalence (communicative function, illocutionary effect) […]. Thus a truly equivalent translation […] would reveal the translator/student's understanding on all three levels" (1990: 65).

A student's translation is to be graded according to its fidelity on linguistic, functional and cultural levels. In this regard the authors make a distinction between errors based on lack of comprehension and those resulting from other shortcomings or problems. Comprehension is assumed to happen simultaneously on the macro level and the micro level. Accordingly, they divide errors into (1) misinterpretations of macro-level structures (frame, schema) and (2) micro-level mistranslations (of propositional content, word-level structures including morphology, syntax and cohesion devices).

To demonstrate the model, the authors subdivide a chosen (literary) text of approximately 300 words into units ranging from one to three sentences in length and proceed to identify and characterize errors at the macro and micro levels, giving points for correct translations of each unit. They then generate frequency tables for each category of error.

Note that translations are not graded against a defined standard; a single mark is

given for the number of utterances correctly rendered. The model is therefore criterion-referenced: Did the trainee achieve a specific translation objective?

They conclude, among other things, that mistranslations at the word level do not automatically lead to misinterpretations of the frame or schema. *In other words, the overall message may be preserved in translation, notwithstanding microtext error.* This prompts them to present an interesting hypothesis:

> In this study, discourse analysis was applied to student translations. One inconvenience of evaluating translation is its cumbersomeness. Evaluating according to misinterpretations may solve this problem. Research with different kinds of texts and students of different linguistic backgrounds may contribute to the development of a technique of evaluation by translation.
>
> (Bensoussan and Rosenhouse 1990: 80)

Would it be possible to modify the hypothesis and use macro-level TQA for assessing the translations of professional practitioners as well as students?

2.1.4. **Larose** (1987) proposes a multilevel grid for **textological** TQA, covering microstructure, macrostructure (isotopes, theme/rheme, logemes—in short the overall semantic structure), superstructure (narrative and argumentative structures) and "peritextual" or extratextual factors including the conditions of production, intentions, sociocultural background, etc. Furthermore, the higher the level of the translation error is (microstructure being the lowest), the more serious it will be.

Larose cuts a wide swath through the whole gamut of contemporary literary and linguistic theories, so his application of any one of them to TQA is necessarily cursory: for example, his treatment of argumentative structures is limited to the syllogism. Note also that the model is not demonstrated.

In a more recent article (1994), Larose proposes a more explicit grid for a multicriteria analysis in which translations are evaluated against each quality criterion separately and the value of each criterion is weighted according to its importance for the contract. Larose illustrates the grid with translations of literature, each rendering of lines from Aristophanes' *Lysistrata* being rated against seven criteria: referential meaning, poetic character, humorous imitation of Spartan speech, expression of contrast between Athenian and Spartan speech, rimes and concision. The criteria may be far removed from those of instrumental translation, but it would certainly be possible to devise a relevant set of criteria for instrumental TQA. For training purposes, the evaluator could make a comparative assessment of several translations of the same source text, with each trainee's individual solutions being given a score—for example, on a scale of 1 to 5—which would then be multiplied by the weighting factor. The best translation is the one with the highest cumulative score (such a model could be used for competitions). Larose points out that the number of criteria must be limited if the model is to be workable.

His model is thus not only macrotextual but also multicriteria-referenced, in that weighting factors are used to generate numerical values for performance against each criterion. At the same time he distances himself from a Sical-type model based on error count.

Referring in a later article (1998) to the fact that the Translation Bureau has distanced itself from Sical, he notes that there is a fundamental contradiction between sampling for TQA purposes and the contemporary focus on total quality and

zero defects. At the same time, he points out that the objective of zero defects is probably unrealistic—hence the Bureau's return to systematic revision of the whole translation (Larose 1998: 181).

Larose concedes that the creation of a truly comprehensive TQA grid is probably impossible, because of the number of parameters or criteria, the complexity of their relationships, and the time and resources required to implement it:

> Idéalement, il faudrait recenser à la vitesse de la lumière tout le sémantisme et le sémiotisme du texte en fonction de la totalité des normes qui régularisent la traduction dans un milieu précis à un moment précis. Une telle grille n'existe pas. Et elle n'existera sans doute jamais, car non seulement faudrait-il que cette grille intègre l'ensemble des contraintes pragmatiques qui s'exercent sur le tissu linguistique du texte […] mais qu'elle les gradue par ordre d'importance et d'évaluation afin de rendre possible et cohérente chacune des prises de décision en matière de traduction et d'évaluation.
>
> (Larose 1998: 175)

Accordingly, any grid is necessarily reductionist and based on the most relevant parameters and criteria.

2.2. Non-quantitative models

2.2.1. In her **Skopstheorie** model, Christiane Nord (1991, 1991, 1992) elaborates on Reiss's (1981) premise of translation as intentional, interlingual communicative action and proposes an analytical model based on the function and intention (skopos) of the target text in the target culture and applicable to pragmatic as much as to literary documents.

The evaluator must take the TT skopos as the starting point for TQA, assess the TT against the skopos and the translator's explicit strategies and then do an ST/TT comparison for inferred strategies. Nord emphasizes that error analysis is insufficient: "[I]t is the text as a whole whose function(s) and effect(s) must be regarded as the crucial criteria for translation criticism" (1991: 166). This is a key qualification, for on the basis of a selection of relevant ST features, the translator may eliminate ST items, rely more heavily on implicatures, or "compensate" for them in a different part of the text. Indeed, as Van Leuven-Zwart points out in developing an interesting corollary of translation-oriented analysis, the "shifts in meaning" that account for many "unsatisfactory" ratings in professional translation should perhaps not be considered as errors at all, given that equivalence is not feasible (1990: 228-29). In short, microtextual error analysis is insufficient.

However, in the examples of translation-oriented text analysis presented to illustrate the model, Nord's judgements are generally parameter-specific, and when there is one, it is not definitive. Indeed, in one case, she states that there "will be no overall evaluation of the translated texts" (1991: 226). For example,

> In spite of *some* imperfections, the English version *seems* to meet the requirements of the translation scope much better than the German version. (1991: 197, my italics)

> Neither translation gives a true impression of the ironic effect produced by the particular features of theme-rheme structure, sentence structure and relief structure. (1991: 217).

> […] the literal translations [of a specific term] are not an accurate description of the subject matter dealt with in the text. (1991: 22)

She does, however, make a definitive, overall judgement on the sample texts as a whole: "[N]one of [the translations] meet the requirements set by text function and recipient orientation" (1991: 231). But how does she generate an overall assessment from the parameter-specific comparisons, particularly when her judgement is based on the nature of the errors, not their number?

2.2.2. In her update of a model first proposed in 1977, **House** (1997) presents a detailed non-quantitative, **descriptive-explanatory** approach to TQA. Like Bensoussan and Rosenhouse, she uses the functional text features explored by Halliday and Crystal and Davey (1969). House dismisses the idea that TQA is by nature too subjective. At the same time, she does not underestimate the "immense difficulties of empirically establishing what any 'norm of usage' is," especially for the unique situation of an individual text (1997: 18), and of meeting the requirement of knowledge about differences in sociocultural norms (1997: 74). She also concedes that "the relative weighting of individual errors [...] is a problem which varies from individual text to individual text" (1997: 45).

She stops short of making a judgement on the text as a whole, stating that "[i]t is difficult to pass a 'final judgement' on the quality of a translation that fulfils the demands of objectivity" (1997: 119). She ultimately sees her model as descriptive-explanatory, as opposed to a socio-psychologically based value judgement:

> Unlike the scientifically (linguistically) based analysis, the evaluative judgement is ultimately **not** a scientific one, but rather a reflection of a social, political, ethical, moral or personal stance.
>
> (House 1997: 116)

In other words, TQA should not yield a judgement as to whether the translation meets a specific quality standard.

2.3. *Overall assessment of approaches*

This cursory, and selective, review highlights the following strengths and limitations:

(1) **Norm-based models are for the most part microtextual**. They are applied to short passages or even sentences.
(2) **Criterion-referenced models** (Bensoussan/Rosenhouse, Larose, Nord, House) **are based on discourse and full-text analysis and factor in the function and purpose of the text.**
(3) The conditions surrounding production of a translation can be many and varied. A common, uniform standard that could factor in all the different conditions would therefore be a complex one and would be difficult to apply.
Nord's "translation instructions" approach is designed to circumvent the problem of uniform standards by assessing quality against a specific work statement prepared for a specific project. However, the approach assumes that the initiator has the time, interest, and understanding of the translation process and product to produce such instructions for all contracts. In actual fact, translators usually have to contact the client if the requirements are not clear or, if time is limited, make their own assumptions based on their knowledge of the client and the text type. In short, the production of an adequate work statement is not always a realistic option in the translation industry.
(4) Two criterion-referenced and textological models (Bensoussan/Rosenhouse, Larose) combine qualitative and quantitative assessment. However, the first is demonstrated only on a short text, and the second is not illustrated at all.

(5) **None of the textological models proposes clearly defined <u>overall</u> quality or tolerance levels**. House refuses to pass overall judgements, and Nord's assessments are not related to a measurable scale of values. Further, the models provide for assessment against specific parameters or functions, not for assessment against all parameters or functions combined. This inevitably militates against global assessment unless translations are found wanting in all departments.

(6) The evolution of Sical illustrates above all the problems inherent in a model that is both standard/norm-based and quantitative. The purpose of quantification is to create a more objective, transparent and defensible assessment, but its very transparency opens the door to (a) calls for greater (quantitative) latitude (toleration of more errors) to allow for conditions of production that cannot be factored into a uniform standard and (b) accusations of tolerance of defective products and mediocrity.
Given the above, decisions on borderline cases (acceptable/unacceptable, pass/fail, etc.) should be based on more than error quantity if they are to stand up to scrutiny.

(7) Researchers applying and demonstrating a discourse analysis method use literary, advertising and journalistic texts as examples. **Application to other pragmatic genres has not been demonstrated.**

(8) No evidence has been adduced that models are reliable for a broad range of texts of varying lengths.

(9) **None of the demonstrations cover both professional products and student translations.** However, given the differences in purpose between production-related TQA and training-related TQA, and given the distinction between formative and summative evaluation, development of a comprehensive TQA model would face significant challenges.

(10) No definition of error gravity has been proposed on a scientific, theoretical, textological basis, and evaluators have to rely on ill-defined concepts such as "complete failure to render the meaning" and "essential part of the message." How is the "essential part" to be determined, and can "partial" failure not be just as damaging to an essential part of the message?

In short, the following questions may fairly be asked about the **validity** of the models:

– In textological, parameter-specific TQA, the target of the assessment is multiple—subject matter, composition, cohesion, tenor, mode, etc. How can a valid overall TQA be extracted from the individual assessments?
– In microtextual TQA, how do we prove that the sample is representative of the text in its entirety?

The following questions may fairly be asked about the **reliability** of the models:

– In quantitative TQA, how do we prove that the tolerance level is a reliable measure of acceptability in all cases?
– How do we ensure that the level of significance of the major error is comparable in all cases?

3. Application of argumentation theory

Argumentation may be defined for the purposes of this article as reasoned discourse and embraces the techniques of rhetoric as means of persuading the audience or readership. Research over the last few decades has shown that argumentation and rhetoric are not the preserve of political, legal and political discourse alone. It has been demonstrated that, even in writing in the natural sciences and economics,

where observation, objectivity and accurate measurement supposedly obviate the need for rhetoric, the tools of argumentation are omnipresent. One of the main reasons for the presence of rhetoric in science, it is suggested, is that information, knowledge and ideas are just as argumentative, and arguable, as beliefs and hopes, particularly in today's society of information overload.

Note, too, that the modern proponents of argumentation theory do not present rhetoric in a negative light. Whereas, the study of rhetoric had since the Middle Ages been restricted to aesthetics and the analysis of figures of speech, the "new rhetoric" has rehabilitated the argumentation aspect of rhetoric as an integral part of the creation and communication of knowledge:

> Why should the latest facts not be persuasive? They will not speak for themselves. Why should theories not be articulated? They will not be heard otherwise. Rhetoric, the approach, looks at rhetoric, the language. It asks what the words are doing, […] why these words are chosen to convey the facts and theories. […] Rhetoric is about academic discourse and newspapers, specialist articles and policies, working papers and headlines. The texts are different, and some are scientific; but science also argues and persuades. (Myerson and Rydin 1996: 16)

So even discourse that is strictly informational is arguable; once there is a content to convey, an argument is present, one that transcends and affects all other aspects of the discourse.

3.1. *Features of an argumentation-centred TQA model*

Vignaux (1976: 66-98) divides his analysis of argumentation in discourse into a number of broad components: lexicological elements (choice of words); narration (narrator, mood, type of speech act); ordering operations (order of propositions, but also syntactical ordering devices such as conjunctives and other features of cohesion, thematization and emphasis); logical operations (including specific types of argument based, among other things, on topoï, popular opinion and common values); and rhetorical order (the *dispositio*, or order in which arguments are presented in the text as a whole). Adapting this breakdown for our purposes, I am proposing to develop a TQA model on the basis of the following discourse categories:

1.   Argument macrostructure
2.   Rhetorical topology
     a)   Organizational schemas
     b)   Conjunctives
     c)   Types of argument
     d)   Figures
     e)   Narrative strategy

With specific reference to Item 1, my hypothesis is that argumentation theory and, specifically, argument macrostructure analysis can contribute to progress in TQA by providing the basis for a uniform basic standard of transfer adequacy and uniform definition of major (or critical) error applicable to all translation fields and functions. Put another way, whatever the speciality or purpose, a translation must reproduce the argument structure of ST to meet minimum criteria of adequacy.

Note that the argument macrostructure is related to, but also different from, the conventional *dispositio* of classical rhetoric and propositional macrostructure (see Van Dijk and Kintsch 1983).

3.2. Argument macrostructure

The macrostructure model selected is that of philosopher Stephen Toulmin. He explores arguments in a variety of areas of specialization and draws the conclusion that the phases of an argument are essentially the same in all types of text and that the force of claims and assessments made in texts remain the same as we move from field to field. Objects, ideas and situations under discussion can be labelled "good." "appropriate," "satisfactory," "unsatisfactory," whatever the type, genre, purpose or area of specialization of the text. The key differences are the premises, standards and assumptions by which judgements are made: the criteria against which an assessment or claim concerning the "goodness," "appropriateness," "effectiveness" or "correctness" of an object, idea or situation will vary from field to field. In short, "all the *canons* for the criticism and assessment of arguments […] are field-dependent, while all our terms of assessment are field-invariant" (1964: 38): the generic elements of argument will be invariant while the specific kinds and content of those elements will depend on the field.

In a later work, Toulmin addresses the subject of argument (and reasoning) in terms of universal (field-invariant) and particular (field-variant) rules of procedure, and he proposes a set of elements that are required for an argument in any field—**claims/discoveries**, **grounds**, **warrants/rules**, and **backings**—and two elements that may be required—**qualifiers/modalizers** and **rebuttals/exceptions** (1984: 25). A brief explanation and illustration of each of these terms follows.

3.2.1. Claim/discovery (C)

The claim (or discovery) is the conclusion of the argument, or the main point toward which all the other elements of the argument converge. The following claims are typical of instrumental texts for translation:

– recommendations in a policy document or discussion paper
– a request for a specific amount of money in a grant application or proposal to a government department
– the announcement of a new health program
– a claim of high energy efficiency of natural gas-heated homes in a survey report
– the judge's decision in an appeal case
– the classification of a newly discovered plant as belonging to a particular order

We can see from this brief list that the areas of specialization are varied. So are the purposes of discourse—to make a recommendation, a request, a public announcement, a claim of superior performance, a legal decision, an announcement of a scientific discovery. In terms of speech act theory, the claim in all cases is the illocutionary point of the text, and justifying the claim through argument and thereby persuading the reader to accept the validity of the claim and act upon it is the perlocutionary point.

3.2.2. Grounds (G)

Claims are not free-standing; they have to be supported by one or more pieces of information, which form the grounds of the argument. These are facts, oral testimony, matters of common knowledge, well-known truisms or commonsense observations,

historical reports, precise statements of legal precedent, and so on, upon which the sender and recipient of the message can agree.

The grounds for announcement of a new health program may be the observation, or report, of overcrowding in emergency departments of hospitals. Note that a claim may be based on more than one ground. For example, the announcement also be prompted by an infusion of new funds into the provincial health budget.

3.2.3. *Warrant (W)*

Warrants are statements indicating how the facts, observations, etc. in the grounds are connected to the claim or conclusion. In our health program example, the logical connection between overcrowding in emergency departments and the new health program is the requirement for rapid response implicit in the emergency department mandate.

However, warrants are not self-validating; they must draw their strength from other considerations, known as backing.

3.2.4. *Backing (B)*

The backing is the overarching principle, value, law or standard governing the issue at hand. In the health program example, the principle of universality enshrined in the *Canada Health Act*, along with human and social values of caring for the sick, would provide support for all the other elements adduced to justify introduction of the new program.

Note that the warrant and the backing may be implicit; they may be presuppositions underlying the communication situation. It is this fact that makes the argument macrostructure different from the "disposition" of the argument. In terms of speech act theory, the argumentative value or force of a text is different from its propositional content.

3.2.5. *Qualifier (Q) / modalizer*

The qualifier or modalizer is a statement or phrase that enhances or mitigates the force of the claim. Thus the new health program may "definitely," "certainly," "probably," or "possibly" be introduced.

Toulmin stresses the importance of qualifying (or modalizing) statements in the argument structure: "Their function is to indicate the kind of *rational strength* to be attributed to C [claim] on the basis of its relationship to G [grounds], W [warrant], and B [backing]" (1984: 86).

Accordingly, the translation evaluator should pay particular attention to the treatment of qualifiers in the target text, and for the purposes of the study I will have to consider how much weight to place on failure to render qualifiers in TT.

3.2.6. *Rebuttal/exception (R)*

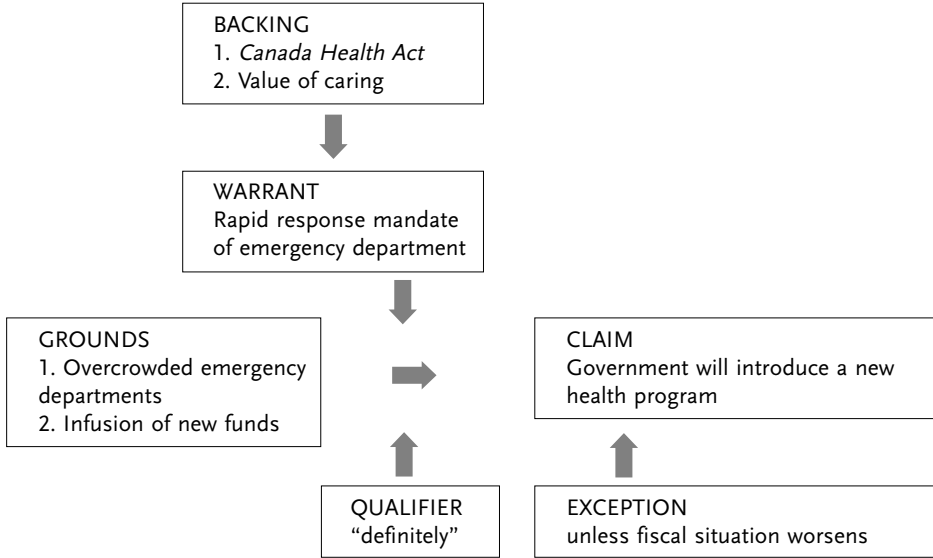This takes the form of a statement of extraordinary or exceptional circumstances that contradicts or may undermine the force of the supporting arguments. It is introduced for the sake of caution or modesty. Thus, in the example, the new health program will be introduced "unless the government's fiscal situation worsens."

The elements of Toulmin's argument macrostructure are very similar to those of

Van Dijk and Kintsch (1983). His model is more refined, however, in that it incorporates the additional features of qualifier and rebuttal and details the logical interrelationship of its main elements.

3.2.7. *Example*


```
BACKING
1. Canada Health Act
2. Value of caring
        |
        v
WARRANT
Rapid response mandate
of emergency department
        |
        v
GROUNDS                              CLAIM
1. Overcrowded emergency    -->      Government will introduce a new
departments                          health program
2. Infusion of new funds
        ^                                 ^
        |                                 |
QUALIFIER                            EXCEPTION
"definitely"                         unless fiscal situation worsens
```

Depending on the complexity of the argument, the claim may be based on several grounds, each of which would require its own B-W-G-C structure. In such an instance, the ground itself becomes a claim that needs to be supported. Furthermore, a long document may contain a number of claims of equal importance, all of which would require support in the interest of sound argument. As a result, the argument structure of the full text will reflect a chain of arguments.

3.2.8. *Generic framework*

Therefore, assuming Toulmin's premise that texts in all fields present essentially the same argumentative structure, we already have a generic working framework for our *TQA model inasmuch as one of the the evaluator's tasks will be to determine whether the basic argument elements (B, W, G, C, Q, R) are accurately rendered in TT if they are present in ST*. The base grid could take the form below:

Element	Present in ST?	Rendered in TT?	Assessment
Claim/Discovery			
Ground			
Warrant			
Backing			
Qualifier/Modalizer			
Rebuttal/Exception			

3.2.9. *Preliminary application of model*

Let us see how the argument macrostructure can be applied in translation. The source text below is the statement of a decision by the Ontario Social Benefits Tribunal regarding an appeal against an earlier, administrative decision to deny disability benefits. The text below is the conclusion of the whole decision document but summarizes the judge's argument.

Source text

Was the test for being disabled met?
Under <u>section 23(10) of the Act</u>, the onus is on the Appellant to satisfy the Tribunal that the decision of the Director is wrong. After reviewing all of the evidence provided by the Appellant, the Tribunal determines that the Appellant has not successfully discharged his onus in this case. The Appellant has not satisfied the onus of showing that he was a person with a disability at the time of the application under review. *The Tribunal found that each part of the <u>criteria in Section 4(1) of the Act</u> were not met. Therefore, the Tribunal finds that the Appellant was not a disabled person within the meaning of the law at the time the Director made the decision.*

The Tribunal took notice of the following information. In the course of his testimony the Appellant testified that he spoke to a female adjudicator at the Disability Adjudication Unit's office, <u>who verbally informed him that his application for ODSP had been approved</u>. He also testified that his family doctor spoke to the same person and was told the same thing. He testified that he tried to re-contact this person many times but she did not return his calls or those of his family doctor. There is a <u>note on file written by hand by the family doctor that related the telephone conversation he had with the Disability Adjudication Unit</u>. The Tribunal does not have the mandate to investigate and therefore can only make a decision based on the **written evidence presented by the Respondent, which included the Health Status Report and the Activity of Daily Living Report as well as the sworn testimonies of the Appellant and his witness.**

<div align="center">

Order
Appeal denied. The Director's decision is affirmed.

</div>

Target text

La personne a-t-elle passé le test pour être handicapée?
Aux termes du <u>paragraphe 23 (10) de la *Loi*,</u> il incombe à l'appelant de convaincre le Tribunal que la décision du directeur est erronée. Apres examen de toutes les preuves fournies par l'*appelant, le Tribunal juge que l'appelant n'*a pas réussi à s'acquitter de son obligation dans ce cas. L'appelant n'a pas réussi à convaincre le Tribunal qu'il était une personne handicapée au moment de la demande en cours d'examen. *Le Tribunal conclut que l'appelant n'a pas satisfait à chaque partie des critères du <u>paragraphe 4 (1) de la Loi</u>. Il conclut donc que l'appelant n'était pas une personne handicapée au sens de la Loi au moment où le directeur a pris sa décision.*

Le Tribunal a pris avis des renseignements suivants. Au cours de son témoignage, l'appelant a déclaré qu'il avait parlé à une experte médicale du bureau de l'Unité de détermination de l'invalidité <u>qui l'avait avisé verbalement que sa demande au Programme ontarien de soutien aux personnes handicapées avait été agréée. L'appelant a également déclaré que son médecin de famille avait parlé à la même personne et qu'elle lui avait dit la même chose.</u> L'appelant a déclaré qu'il avait essayé plusieurs fois de communiquer de nouveau avec cette personne, mais qu'elle n'avait pas rappelé son

médecin ni lui. Le dossier de l'appelant comporte <u>une note manuscrite du médecin de famille qui relatait la conversation téléphonique qu'il avait eue avec l'Unité de détermination de l'invalidité.</u> Le Tribunal n'est pas habilité à enquêter et il ne peut donc prendre sa décision qu'en se fondant sur **la preuve écrite que l'intimé a soumise et qui comporte le Rapport sur l'état de santé et le Rapport sur les activités de la vie quotidienne, ainsi que sur les témoignages sous serment de l'appelant et de son témoin.**

<div align="center">

Ordonnance
*Appel rejeté. La décision du directeur est confirmée.*

</div>

Analysis

The first step is to establish the argument macrostructure of ST. Thus,

claim	=	1. failure to meet criteria 2. denial of appeal (italics)
grounds	=	1. written evidence 2. sworn testimonies (boldface)
warrant	=	1. reliability of evidence and testimony; 2. mandate of Tribunal (shadow)
backing	=	relevant provisions of Act (underlining)
qualifier	=	N/A
rebuttal	=	specific elements of appellant's testimony and written evidence of doctor (double underlining)

Part of the warrant is presupposed. The providers of the evidence and testimony are presumed to be honest and reliable, and the statements are presumed to convey the facts accurately. Note also that, in this example, the rebuttal, which is in fact the argument for benefits advanced by the Appellant, carries little countervailing weight because of the restrictions on the Tribunal's mandate (no authority to investigate).

Furthermore, the claim itself is complex and in fact forms an argument chain in itself. The grounds (the evidence) lead to an initial claim (conclusion that the appellant does not meet the criteria for benefits under the legislation). This initial claim then becomes the grounds for the second claim (denial of appeal). This accumulation of grounds and claims prompts Mendenhall (1990: 211) to subdivide these elements into primary and secondary reasons (grounds) and conclusions (claims). We can present the structure graphically:
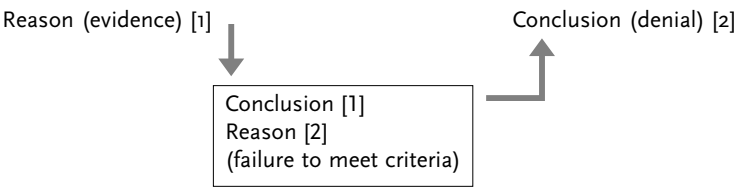
The second step is to establish, through comparative reading, to what extent the argument macrostructure is reflected in TT. We find that it does and we complete the TQA grid as follows:

Element	Present in ST?	Rendered in TT?	Assessment
Claim 1	Yes	Yes	/
Claim 2	Yes	Yes	/
Ground 1	Yes	Yes	/
Ground 2	Yes	Yes	/
Warrant 1	No	N/A	N/A
Warrant 2	Yes	Yes	/
Backing	Yes	Yes	/
Qualifier	No	N/A	N/A
Rebuttal	Yes	Yes	/

We have thus established the correspondence of the argument macrostructure elements in ST and TT. Further testing will be required to determine the impact of non-correspondence of one or more elements on an overall assessment of the macrostructure in translation. Nevertheless, it is reasonable to assume at this stage that Toulmin's six core elements provide a sound theoretical basis for a new definition of major error, in that failure to render accurately one of these elements results in a major/critical defect.

In addition, following up on Bensoussan and Rosenhouse's suggestion that assessment of text-level misinterpretations may avoid the cumbersomeness of other TQA tools, we have established a limited set of six elements on which assessment of overall quality is to be based and which can be applied, in theory, to both students' work and the professional product.

Finally, the translation teacher and the evaluator would expect the trainee and the professional to identify, understand and accurately render the macroelements of a text's argumentation structure. I would suggest that, if the translator meets these requirements, he will have conveyed to the TT readership the core message(s) of the text.

Walton (1989: 276-77) posits "two extreme possibilities of…standards" of precision in making and assessing arguments: the high standard (no chance of error) and the low standard (reasonable assurance of accuracy). Therefore, applying the extremes of Walton's standards continuum to TQA, I would venture to suggest, further, that the core elements of argument also provide the basis for a "low" or minimum standard of translation quality that would be valid and reliable in all situations.

4. Conclusion

The above examination and illustration of one component of argument analysis serves to highlight the potential contribution that an argumentation-centred TQA model can make.

Going back to the brief review of TQA in 2.3., we see that it can offer the following advantages:

1.   It is norm-based, applying a single standard and tolerance level for all lengths and types of text and for all conditions of production (including training) and thereby yielding an assessment of *overall* quality.

2. It provides for full-text analysis, thereby avoiding the main shortcoming of other norm-based models.
3. It combines quantitative and qualitative analysis.
4. It offers a nonempirical definition of major/critical error.
5. It offers an efficient, cost-effective means of conducting full-text TQA.
6. It is potentially **valid** in that the target of assessment is a single parameter and that representativity is not at issue.
7. It is potentially **reliable** in that the same tolerance level (low standard) would be applied in all cases and that the significance of the major error would not vary—it would always be related to one of the core elements of the argument macrostructure.

That being said, we are the low end of Walton's standards continuum, and both trainers and practitioners may require a more detailed assessment of translation quality, leading to measurement against a "high" standard. This will lead us to integrate other aspects of Vignaux's rhetorical topology into our model and also incorporate examination of more traditional aspects of TQA such as target-language quality (the language quality of the French translation in the example above leaves much to be desired, even though the text adequately renders the argument macrostructure).

This is where a multicriteria TQA grid akin to the one proposed by Larose may prove its worth. Depending on the level of assessment detail required, the purpose and end-use of the translation and other variables noted in the introduction, analysis may be extended to other factors of translation quality, with each factor being weighted according to end use or purpose of translation and/or assessment. I hope to demonstrate this multilevel application in the larger study currently underway.

REFERENCES

BENSOUSSAN, M. and J. ROSENHOUSE (1990): "Evaluating Students' translations by Discourse Analysis," *Babel*, pp. 65-84.

CARY, E. and R. W. JUMPELT, eds (1963): *Quality in Translation. Proceedings of the Third Congress of the International Federation of Translators, Bad Godesberg, 1959*, New York, Pergamon Press.

CIRCUIT (1994), numéro spécial «La qualité totale: Engagez-vous, qu'y disaient!» (contrôle de la qualité).

CONSEIL DES TRADUCTEURS ET INTERPRÈTES DU CANADA. Comité de révision de l'examen d'agrément en traduction (1994): *Rapport final*, Hull (Canada).

CRYSTAL, D. and D. DAVY (1969): *Investigating English Styles*, London, Longman.

DARBELNET, J. (1977): «Niveaux de la traduction», *Babel*, 13-1, pp. 6-16.

GOUADEC, D. (1989): *Le traducteur, la traduction et l'entreprise*, Paris, Afnor.

HATIM, B. and I. MASON (1997): *The Translator as Communicator*, London, Routledge.

HORGUELIN, P. (1978): *Pratique de la révision*, Montréal, Linguatech.

HOUSE, J. (1997): *Translation Quality Assessment: A Model Revisited*, Tübingen, Gunter Narr.

LANGUAGE INTERNATIONAL (1998), special issue "Quality: The Struggle over Translation Standards," 10-5.

LAROSE, R. (1987): *Théories contemporaines de la traduction*, Sillery, Presses de l'Université du Québec.

—— (1994): «Qualité et efficacité en traduction: réponse à F. W. Sixel», *Meta*, 34-2, p. 362-373.

—— (1998): «Méthodologie de l'évaluation des traductions», *Meta*, 43-2, p. 163-186.

MENDENHALL, V. (1990): *Une introduction à l'analyse du discours argumentatif*, Ottawa, Presses de l'Université d'Ottawa.

MYERSON, G. and Y. RYDIN (1996): *The Language of Environment: A New Rhetoric*, London, UCL Press.

NORD, C. (1991): *Text Analysis in Translation: Theory, Methodology, and Didactic Application of a Model for Translation-Oriented Text Analysis*, Amsterdam and Atlanta, Rodopi.

TOULMIN, S. (1964): *The Uses of Argument*, Cambridge, Cambridge University Press.

—— *et al.* (1984): *An Introduction to Reasoning*, New York, Macmillan.

VAN DIJK, T. (1980): *Macrostructures. An Interdisciplinary Study of Global Structures in Discourse. Interaction and Cognition*, Hillsdale, Lawrence Erlbaum.

—— and W. KINTSCH (1983): *Strategies of Discourse Comprehension*, Toronto, Academic Press.

VAN LEUVEN-ZWART, K. (1990): "Shifts in Meaning in Translation: Do's and Don't's?," *Translation and Meaning*, 1, Maastricht, Rijkshogeschool, pp. 226-233.

VIGNAUX, G. (1976): *L'Argumentation: essai d'une logique discursive*, Genève, Droz.

WILLIAMS, M. (1989): "Creating Credibility out of Chaos: The Assessment of Translation Quality," *TTR*, 2-2, pp.13-33.