Translation Quarterly

No. 55 2010

香港翻譯學會出版

Published by
The Hong Kong Translation Society

翻譯季刊

二〇一〇年
第五十五期

《翻譯季刊》 *Translation Quarterly*
二〇一〇年三月 第五十五期 No. 55, March 2010

ISSN 1027-8559-55

The Hong Kong Translation Society has entered into an electronic licensing relationship with EBSCO Publishing, the world's most prolific aggregator of full text journals, magazines and other sources. The full text of the *Translation Quarterly* can be found on EBSCO Publishing's databases.

翻譯季刊
Translation Quarterly

香港翻譯學會

The Hong Kong Translation Society

Translation Quarterly **No. 55, March 2010**

目錄 CONTENTS

Chief Editor's Note

Coming after two combined issues last year, each of them commemorating a special occasion, the present issue presents once again the regular fare of articles submitted to the editorial office and reviewed by experts in the field. The journal has been able to accommodate only two this time. Yasmin H. Hannouna's "Is Arabic Machine Translation a Dream or a Reality? A Quality Assessment of Three Arabic Systems" is in a different category from what we have been used to publishing, giving as it does an extended discussion of problems in machine translation. It is impressive not just in the treatment it gives to the subject matter in question, but also in its scope. In 63 pages Hannouna details the investigation she has carried out into the merits and drawbacks of three currently available English-to-Arabic MT systems. Taking issue with predecessors in the field, she adopts a comparative and adequacy-oriented approach, and boldly proposes ways of tackling the deficiencies she has discovered. Li Bo's "Heteroglossia, Dialect and Literary Translation: A Case Study of Wang Chen-ho's *Rose Rose I Love You* and Its English Translation" engages with a topic about which too little has been said—that of the relevance of Mikhail Bakhtin's theories on heteroglossia to translation studies. His re-reading of Howard Goldblatt's translation of the Taiwanese writer Wang Chen-ho's best-loved and most controversial novel forces us to rethink the

possibilities that can be opened when borrowed insights from other fields are applied to translated texts.

Members of the Hong Kong Translation Society are once again reminded that if they wish to secure hard copies of the journal, they will have to approach our Secretary and pay a modest administrative fee (see the Society's *Bulletin*, no. 48). For those who are interested in viewing the online version, it will be necessary to check at the libraries. EBSCO Publishing databases are available in library systems around the world. There is no charge for readers of EBSCO Publishing content. Those who use a library which has subscribed to the "Communication & Mass Media Complete" can explore the contents of the *Translation Quarterly* whenever they wish.

Leo Chan

March 2010

Is Arabic Machine Translation a Dream or a Reality?
A Quality Assessment of Three Arabic Systems

Yasmin H. Hannouna

Abstract

The present study investigates the overall quality of three currently available English-into-Arabic machine translation (MT) systems. The evaluation deals with selected quality characteristics and various text types. The quality requirements (both functional and non-functional) have been expressed in terms of characteristics and sub-characteristics. Much attention has been paid to the functional criteria in this study as they determine the quality of an MT output. The theoretical construct adopted in this study is based on the Framework for Evaluation of MT in ISLE (FEMTI). The criterion under evaluation, appropriate metrics, scoring scheme and methods regarding the overall judgement of the product are identified and described. The proposed model for the functional criteria is of the black-box type, and is a comparative and adequacy-oriented evaluation. As for the non-functional criteria, the evaluation model is of the comparative performance and adequacy-oriented type. The sample consists of a total of 268

English sentences taken from twelve different special-domain texts. Despite the good translation quality of some text types, the overall comparison of the three systems confirms that English-into-Arabic MT systems suffer from many drawbacks, especially relating to the grammar and meanings of the translated sentence. Their output reflects certain deficiencies in translation and they all need to be improved upon.

1. Background

1.1 Evaluating MT Systems

Evaluation is an implicit aspect of all human activity. With respect to MT, it remains an open fundamental issue and one of the most important stages in the life cycle of an MT system. It is the measure of the quality of a system, in a given context, as stated by the definition of quality as "the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs" (ISO/IEC-9126 1991: 2). It is the process of system appraisal which leads to global, overall, quantification of performance.

Thus, with the rapid growth in sales of MT software and the increasing availability of MT services over networks, there is an urgent need for MT researchers, developers, and vendors to agree and implement objective, reliable, and publicly acceptable benchmarks, standards and evaluation metrics.

1.2 Evaluation of Existing Arabic MT Systems

It is only since about 1992 (ATA Software Technology 1997: 7)

that some Arabic MT systems have achieved a certain degree of marked acceptance. There seems to be much confusion concerning the level of capability possessed by these products.

In order to adequately answer the question of whether Arabic MT of texts and documents is practical, it is necessary to rigorously evaluate the currently existing English-to-Arabic MT systems at the local market. These are: Golden Al-Wafi 1.00 (2002) (henceforth GW); Al-Mutarjim Al Arabey 3.00 (2002) (henceforth MA) and An-Nakel Al-Arabi 2.00 (2000) (henceforth NA).

It is worth mentioning here that web-based systems (e.g., Google, Babylon, Tarjim and Sakhar) are excluded from the present evaluation task because:

(1) The major objective in this study is to test the Arabic MT systems available in our local market that have been used by a large number of students and people representing different social positions and who cannot have access to the Internet always for one reason and another. These people badly rely on MT systems to serve their various needs.

(2) The access to the online systems like Sakhr and Tarjim is not free. Even Sakhr is a very expensive desktop MT system that not everybody can buy. As for Tarjim, it has the same engine as Al-Mutarjim Al Araby produced by ATA Company in England. They operate the same engine online at the Al Misbar website which is not free and many users do not find it easy to have access to and use. The researcher tested Tarjim with the same texts used in evaluating the desktop MT systems in the present study. Its output was found similar to that of MA systems, although the desktop system has more flexibility in designing dictionaries that can make the MT output of a better quality. The online systems lack this potentiality. The aim in this study is to test MT systems where there is a possibility to interact with them and that have

the potentiality to perform other processes.

Further, there is much emphasis in the literature of MT and MTE that MT systems must share common features and criteria for the adequacy of the comparative study and the reliability of its results. Therefore, it is inappropriate to compare desktop MT systems to online ones (where it is not possible to evaluate the computational criteria of these systems and to compare them to these of desktop ones). Consequently, this will lead to ineffective, incomplete and unreliable quality assessment of systems.

The present study is intended to provide the end-user of the available systems with adequate information, based on an objective evaluative study. This concerns the extent to which each system can serve the needs of the end-user and which system is the best in performing its task.

The evaluation of MT systems is a complex task indeed. This is not only because many different factors are involved, but because measuring the overall translation performance of an MT system is itself difficult (Arnold 1994: 270).

1.3 Significance

With the increasing number of MT systems available on the market, the demand for valid ways of assessing their translation capabilities from the users' as well as from the developers' point of view becomes even greater. Evaluation is one of the most important stages in the life of any software. This is not only for its potential users and buyers, but also for researchers and developers. Evaluation is also crucially important to scrutinize any MT system, analyzing the quality of its output, classifying the errors it makes and improving it. The evaluation can thus be concerned with the technical quality of the system. It can also target the computational and linguistic limitations of the

system, the software engineering aspects and/or the cost and benefits.

In fact, MT users need to know which system is appropriate for their specific requirements. This cannot, of course, be decided by the layman by simply saying this or that system is good or bad. Evaluators, who are, in most cases, translators, system designers or researchers can perform various tests based on objective and scientific evaluation standards to indicate the quality of such systems. So, the present research is significant for it concerns itself, through the application of special metrics and methods, with the evaluation of different Arabic MT systems available in the local market, taking into account a variety of quality criteria and text types. Thence, it points out: (a) is an MT system good? (b) which of two or three systems is better for a certain text type? and (c) what do "good" and "better" mean in this context?

Thus, a study like the present one is of value for people who want to compare several MT systems. They can browse and select the characteristics that best reflect their circumstances by finding associated evaluation measures and tests. It is also useful for people who want to design a new MT system or upgrade an old one as they can learn about the needs of users and find niche applications for their system.

Above all, to the best of the researcher's knowledge, this study is the first of its kind. It aims at detecting the evaluation of English-to-Arabic MT systems and comparing recently released versions of these systems within the framework of the formal international standards of evaluation. The aims are to:

(1) make sure to what extent the claims of the manufacturers of these systems are true as they are meant for advanced and professional levels in the translation of a variety of scientific and humanities texts;

(2) investigate the strengths and weaknesses and types of errors for subsequent improvements by designers and developers, and for designers of MT systems in general to avoid present problems;

(3) indicate the economic value of such systems as they are very expensive with respect to the quality of translation they perform;

(4) be in line with the international advancements and research projects in the field of MT and MTE; and

(5) determine where and how they could be utilized by the end-user.

This whole question of evaluation of MT is, in fact, a hot topic especially in recent years, and it is hoped that the present study is a good contribution in this field.

2. The Evaluation Task of the Study

2.1 The Purpose of Evaluation

Regarding the present study, the purpose of evaluation involves:

(1) Evaluating the external quality characteristics of each of the MT systems under testing with respect to selected attributes (e.g., functional and non-functional). The types of evaluation required here are black-box methods of declarative and performance evaluation, where global properties of the system will be evaluated, such as the semantic fidelity of the translation or the intelligibility of the TL output.

(2) Testing the translation external quality of each MT system in this study with respect to texts representing various domains that each of these systems is expected to translate.

(3) Comparing the results obtained from the evaluation of the three competitive MT systems under testing (comparative evaluation).

(4) Assessing both positive and negative effects of each MT system under evaluation and its adequacy in translation with respect to various text types.

2.2 The Framework Model of MTE

This section outlines the model of MTE based on the *International Standards of Language Engineering* (ISLE) taxonomy 3 "*A Framework for the Evaluation of Machine Translation in ISLE* (FEMTI)" (*FEMTI 2003*) with all the requirements, criteria, metrics, evaluation types and quality characteristics relevant to the present evaluation, taking into consideration the EAGLES 7-step recipe (See *EAGLES Report* 1999) and the stages in the evaluation process (*ibid.*).

The first step in any evaluation procedure should be to analyze user requirements, from which relevant quality characteristics are deduced and broken down in terms of reportable attributes. With reference to the definition of the user, the objects (e.g., MT systems) of evaluation are also defined in terms of attributes. Relevant metrics are then developed. Having defined attributes and metrics for the particular evaluation environment, the researcher decides what methods should be applied in order to perform the measurement and to obtain values for the attributes (EAGLES 1996: 32).

It should be noticed here that the identification and explanation of all the aspects of evaluation mentioned above is mainly based on Van Slype's critical study of MTE (1979), the EAGLES Report (1996), the EAGLES Report (1999) and the ISLE's Evaluation Framework (2003) which is comprehensive, up-to-date and accessible (FEMTI 2003).

Thus, the three MT systems will be evaluated to fulfil the needs of the end-user regarding the purpose of evaluation. These Arabic MT systems are evaluated and compared with respect to the "external quality". The selected functional and non-functional criteria listed below are investigated in this study (*ibid.*). The evaluation aims to compare these systems and to judge the quality of their output in the light of the *FEMTI* taxonomy (which provides and proposes metrics and methods for the evaluation of the characteristics of MT systems) and the

EAGLES (1999) 7- major steps of MTE. As for the methods of evaluation, these were partly selected from the literature and partly developed by the researcher (for more details, see, EAGLES 1999: 6-17; Hovy and King 1999: 6-10; Koh et al. 2001; Hannouna 2004: Chaps. 2 and 3; Estrella et al. 2005 and Popescu-Belis et al. 2006).

1. System External Characteristics
    - 1.1 Functionality
        - 1.1.1 Suitability
            - 1.1.1.1 Readability (or Fluency, Intelligibility or Clarity)
        - 1.1.2 Accuracy
            - 1.1.2.1 Fidelity
            - 1.1.2.2 Terminology
        - 1.1.3 Wellformedness
            - 1.1.3.1 Grammar/Syntax
            - 1.1.3.2 Morphology
    - 1.2 Non-Functional Criteria
        - 1.2.1 Efficiency
            - 1.2.1.1 Time Behaviour
                - 1.2.1.1.1 Input-to-output Translation Time
        - 1.2.2 Portability
            - 1.2.2.1 Adaptability
        - 1.2.3 Flexibility
        - 1.2.4 Storage

2.3 Automatic Methods of MTE

One of the most difficult things in MT is the evaluation of a proposed system. Evaluating machine translation hypotheses is a very important part of the ongoing research. Because humans are the golden standard for using language, obviously human evaluation is the *holy grail*

for the evaluation of machine translations. The problem, however, is that human evaluation is very time-consuming and expensive and therefore not always an option. Automatic scoring metrics allow a fast evaluation of translations and a quick turnaround for experiments. They are also cheap and re-usable (Zwarts 2009).

According to Callison-Burch et al. (2007), automated metrics of MT evaluation have been receiving significant attention in recent years. The way that automated evaluation metrics work is to compare the output of an MT system against a reference HT and then correlate it with human judgments. Human judgments come in the form of adequacy and fluency quantitative scores (Lavie and Agarwal 2007). Some of the most common automatic metrics are:

(1) BLEU (Papineni et al. 2002) is an IBM-developed metric and is probably the best known and most used in the MT community (Callison-Burch et al. 2007). It is based on calculating an n-gram precision between the system output and the reference translation and a brevity penalty if the output differs from the reference.

(2) TER (Translation Error Rate) measures the amount of editing required to change an output translation into a reference translation. The possible edits in TER include insertion, deletion, and substitution of single words.

(3) METEOR was initially proposed and released in 2004, designed to explicitly address several observed weaknesses in IBM's BLEU metric (Lavie et al. 2004). This automated metric is based on an explicit word-to-word matching between the MT output being evaluated and one or more reference translations (Banerjee and Lavie 2005). METEOR's matching can also match words that are simple morphological variants of each other (i.e., they have an identical stem) and words that are synonyms of each other.

(4) WER (Word Error Rate) is based on the minimum number of

substitutions, deletions, and insertions that have to be performed to convert the generated text into the reference text.

(5) NIST is closely related to BLEU and aims at upgrading BLEU metric (Doddington 2002).

(6) The F-measure was proposed as a comprehensible alternative for MT evaluation, and it can be defined as a simple composite of unigram precision and recall.

While the disadvantages of human metrics are seen in the fact that they are time and labor consuming, expensive and not re-usable, automated metrics have more disadvantages. They require text preparing for testing both the source text to be analyzed and one or more reference texts. All the automated metrics are not necessary reliable and rather subjective since they are based on human evaluation (the source text is compared with already adjusted and evaluated texts). In addition, automated metrics do not produce very reliable sentence-level scores. These metrics do not give any details about the nature of translation errors and it has recently been found that BLUE does not correlate with human judgment well enough and suffers from several other deficiencies (Turian 2003; Elliott et al. 2004). Further, Callison-Burch et al. (2007) maintain that human metrics are ultimately what we are interested in and they allow us to perform high-level analysis of the evaluation process.

Therefore, in this study, human metrics and scoring schemes have carefully been chosen from among the most common and reliable models of MTE like Van Slype, EAGLES, ISLE and FEMTI. Automatic metrics are not used as they do not capture a lot of plausible variations and some are not set up to work for Arabic (e.g., METEOR).

3. Experimental Design

The evaluation aims at investigating which of the three MT systems under consideration shows the best quality of translation and performance with respect to the various text domains it translates.

3.1 Text Sampling

A sample of English STs was selected for English-into-Arabic MT and evaluation. The texts chosen were representative of twelve various special-domains that were most suitable for MT, i.e., scientific, technical and commercial fields. They included the fields of: Biology, Chemistry, Computer Science, Banking and Finance, Legal, Military, Medicine, Petroleum, Physics, Social Sciences and Commerce. In addition, they represented most of the domains that the manufacturing companies claim their systems can best translate. Most of these texts were selected on the basis of being rich in domain-specific terminologies. It should be noted that for each of the sample texts in this study, an effective Arabic human translation (henceforth HT) was found. These texts acted as adequate translations for the sake of comparison. The SL texts and their parallel texts (i.e., MT and HT) were aligned and segmented into equal number of sentences. The unit of translation to be analyzed and evaluated, in the study, was the entire sentence. Extracts of 20-29 sentences constituting "self-contained" texts (i.e., complete paragraphs or sections) were taken from each of the twelve text types under evaluation. For each SL text, in this work, there were three MT texts and one adequate HT in Arabic, i.e., the sample of texts consisted of: (12) SL texts, (36) MT texts and (12) HT texts. So, a total of (268) English SL sentences were used as input data in testing and estimating the cognitive and linguistic properties of sentence level translations of the three Arabic translation software packages.

3.2 A User-Oriented Model of Test Types

In order to evaluate the MT systems' functionality, the task-based evaluation employed standard, rather than randomly-chosen, tests of MT output quality selected from the *FEMTI* (2003) with respect to the methods, metrics and scoring schemes. This is to estimate the values of the functional criteria of each MT system considered in terms of a black-box comparative and adequacy-oriented evaluation.

Further, some non-functional criteria (i.e., speed of translation and storage) were evaluated in terms of special methods of performance. Others, like "adaptability" and "flexibility" were tested by means of benchmark tests using a special checklist designed for this purpose. In this evaluation experiment, the non-functional (i.e., computational) criteria were tested by a comparative performance and adequacy-oriented evaluation. It is important to emphasize that the goal was to test the systems' capabilities, not the hardware used.

3.3 Implementation of Evaluation Methodology

Prior to the process of identifying, classifying and discussing the results of the tests, the metrics and methods directly associated with each quality characteristic under investigation were applied. The measures were carried out in accordance with the standard metrics and methods. Different weightings were assigned by each evaluator to the individual measures, in order to calculate the overall general software quality (henceforth GSQ). Moreover, the researcher used descriptive statistical techniques and scoring schemes whenever required in the detailed evaluation task. To get a certain reliability in the assessment of these aspects, a second opinion is required. Thus, a specialist in MT was approached and she closely analyzed and evaluated the sample texts, using the same metrics and methods. Her judgements were almost identical with the researcher's.

Following the *FEMTI* (2003), the order of criteria to be evaluated in this study matches the order of tests to be performed in the evaluation sequence. The result is a top-down ordering of tests.

3.3.1 Testing the functional criteria

The researcher suggested five questions designed to assess what each MT system under testing can do, in order to fulfil the pre-specified requirements of its users. So, the following are the questions on which the whole analysis of the MT outputs in every test was based, each with a clear discussion of the procedure followed in its application (see Hannouna 2004: Chaps. 3, 4 & 5).

3.3.1.1 Readability rating

To what extent is the output of an MT system understandable, fluent and readable?

The above question motivates inquiry into the ease with which the output text can be read. In other words, it is an investigation of the degree to which some discernible meaning is conveyed by a sentence in the TT without any reference to the ST.

Four metrics were decided upon for purposes of measuring "readability" of the output text (Van Slype 1979b: 67-68; Vanni and Miller 2001: 3; Dabbadie et al. 2002; FEMTI 2003: 42). The two evaluators assigned a "clarity" score to the same set of (268) output sentences representing 12 different domains. The same scoring scheme and procedure of evaluation was repeated with each MT system under investigation. For each TT sentence a score of 3, 2, 1 or 0 was given, depending on whether a sentence was readable or not.

3.3.1.2 Fidelity rating

To what extent is the content of the source language sentence

preserved in the translation?

Fidelity is designated the most important characteristic for the translator's task. It is a measure of the information successfully conveyed from the SL text to the TL output. As the evaluators have good knowledge of the SL (English), this test was performed by judging the "fidelity" of the translation with respect to the original text only. Fidelity scores for each text in this evaluation experiment were computed (Van Slype 1979b: 72-78; Miller et al. 2001; Hovy 2002a; FEMTI 2003: 48), with each sentence assigned a value from a 4-point scale. The scoring was performed with values ranging from 0 to 3, based on the amount of information in the MT output after being compared with the original input text. The numbers of texts corresponded to those in the "readability" test using the same MT systems.

3.3.1.3 Domain terminology rating

How precisely does the system translate subject-matter terminologies?

The correct translation of the domain-specific terms is one of the most important tasks an MT system is expected to perform in this test. Untranslated terms, wrongly translated terms, synonym usages, the occurrence of forms irrelevant to the TL and other issues were taken into consideration in the evaluation of this criterion. A human translated text for each SL text acted as a guideline for raters, so that they considered anything other than an exact match as wrong even though the variations encountered might be correct from the language perspective.

The metric is the percentage of correctly translated terms (Miller et al. 2001; Vanni and Miller 2001; FEMTI 2003: 48). For this test, the researcher designed user dictionaries and preference dictionaries for MA and only user dictionaries that require insertion of grammatical

and semantic data for each entry for NA. This helped in increasing the potentiality of these MT systems in handling the exact and precise translation of the domain-specific terms and expressions. Such a facility is not available in GW. Therefore, it merely depended in its translation on the general dictionary. Testing the effect of these dictionaries on the lexical capabilities of the MT systems investigated was one of the aims of this evaluation.

3.3.1.4 Syntax rating

To what extent does the MT output respect the reference grammatical rules of the TL? How syntactically well-formed is the output?

Here, the researcher suggested adopting a procedure of evaluation based on the human rating of MT sentences on a 5-point scale, where five represented a sentence was perfectly grammatical. It is worth mentioning that although HT texts were translated and revised by experts in the subject matter of the sample text and in translation, an expert in Arabic was also consulted to do the necessary corrections, if any, and the final editing and revision (White et al. 1994; Hovy and King 2001: 20-22).

3.3.1.5 Morphology rating

Does the MT system produce a morphologically correct output?

The aim of this test is to investigate the degree to which the MT output respects the inflectional morphological rules of the TL. The inflected words were identified and underlined in the reference HT of each text type. This facilitated the work of the raters in detecting the morphological errors and making the corrections needed in the MT output with respect to number, gender, case, tense, aspect, etc.

In the present evaluation experiment, the raters counted and

underlined the inflectable words in the MT output of each sample text. Then, they started by making all the morphological corrections in the output texts. The morphological score of each sample text was calculated as the ratio of the number of morphological corrections to the inflectable words in the MT output. Further, the final morphological score was calculated as the average value of all the texts under evaluation for each MT system with respect to the correlation coefficient of the scores of the two raters.

Finally, the value of "morphology" for an MT system in terms of its relative importance in the calculation of the overall GSQ was calculated.

3.3.2 Testing the non-functional criteria

Similar to the procedure followed in testing the functional criteria, the researcher formulated four questions upon which the evaluation of the performance of the three MT systems considered was based. The sub-sections below present the details of testing each of the computational criteria involved.

It should be emphasized that testing the non-functional criteria achieved the greatest possible degree of objectivity and the results of the two evaluators seemed completely identical.

3.3.2.1 Testing the input-to-output translation

How fast is the MT system in performing the translation of each text?

The ultimate purpose of the translation task determines the importance of the "speed" factor. In general, the monitoring of a relatively large volume of texts and on-line translation requires high speed.

Using a stop-watch, the two evaluators worked separately in calculating the time which each text in the twelve selected domains

takes in translation. Each measure was repeated three times by each rater for constancy of results. When the translation time of each sample text was expressed as "number of words/time of translation", the evaluators started computing the average value of speed of translation for each of the three systems (JEIDA 1992: 114; FEMTI 2003: 55).

3.3.2.2 Testing adaptability

How is the MT system adaptable to different specified environments? To what extent can an MT system be easily transported to different operational settings?

This question aims at investigating the operational setting of an MT system and the extent to which it is adaptable to different PCs, operating systems and other software it requires (if any). This inspection was based on the evaluators' observations and tests during system operation. In this respect, benchmark testing using values to be assigned on a special checklist was adopted. Here, a rating score covering five features is used for performance rating (see EAGLES 1999: 44; Hovy 2003; Popescu-Belis 2003).

As it is important to obtain percentage values of the results of this test for the estimation of the GSQ of each MT system considered and for comparison between systems, a special rating scheme was used. A rating score covering five features (ranging between 0-4) was designed for this purpose.

3.3.2.3 Testing storage

What is the largest size in Kb of the text or document the MT system can translate and display in its English and Arabic text boxes?

Two MT systems, MA and GW, translate English texts that should be not more than 40 Kb (about 20 pages and 5,000 words), which is considered very good, i.e., there is a 40 Kb limit on displaying text in the

text boxes for English and Arabic. The manufacturing company (Cimos) of NA, on the other hand, does not give any indication of the size of text or document the system can display and store in its text boxes.

To translate documents larger than 40 Kb, both MA and GW have the option of direct or batch translation in the background.

During the benchmark tests, the final percentage value of "storage" for each MT system under evaluation (the largest size of text in Kb it can translate and display in its text boxes which represents the storage value of the system) was divided by the typical storage suggested in this study (i.e., the highest storage in Kb that any of the three Arabic MT systems can achieve).

Finally, the value of this criterion for each MT system in terms of the relative importance of "storage" in the calculation of the overall GSQ is found.

3.3.2.4 Testing flexibility

Is there a possibility of adding new words and dictionaries to the MT system?

The present test aims at investigating the system's extensibility. Regarding the three MT systems under testing, there is a possibility of free Internet updates for the most recent versions of MA and GW. This mainly concerns the systems' dictionaries, i.e., their general dictionary and all the specialized science dictionaries. Such updates often concern the addition of new words, expressions and special-domain terminologies.

As GW lacks the potentiality of adding or making dictionaries, the researcher designed user dictionaries for MA and NA. In addition, a preference dictionary for each text type in the present evaluation was made for certain words and terms automatically and randomly selected by MA. For example, the word "development" has four Arabic meanings listed in the dictionary of the system as follows: (*tnmiyah, tawsee?, nimou,*

tatawer) with the preferred Arabic word *tatawer*. The user can change this to any of the four Arabic nouns according to the subject matter of the text and always use it. This is not possible for the other systems because of their limited flexibility in this respect. In addition, An-Nakel AL-Arabi enables its users to create a TM database after post-editing processes and their own idioms dictionary too.

The two raters based their judgments and scores merely on the extent to which it was possible for the three MT systems to accept the addition of new words, expressions and dictionaries. The systems' capacity in this respect was also taken into consideration. The evaluation of "flexibility" of the MT systems in this study involved the use of a checklist where a rating score covering three features was used for performance rating.

3.3.3 GSQ weighting scheme

Regarding the final calculation of the average value of General Software Quality (GSQ) in the present evaluation, there is a need to combine different scores into the "mother node" as a general score. According to Hovy (2002b), a simple method of combining these scores is to decide, in each domain, how important the sub-characteristics of general criteria are relative to each other and whether they are equally important—or whether one is more important than the other(s). Moreover, Popescu-Belis (2002a) confirms that there is no "fair" distribution of the weights (or percentage values of criteria) here, because these should be fixed according to the destination of the MT systems. This depends on the needs of the users or the context, i.e., it must capture the purpose of the system and how it will be used. In this study, concentration was on the functional criteria, since they represent the main function of the MT system and the major requirements of the current evaluation. They are also the most essential and difficult

problem of any evaluation process. Such a weighting scheme enabled the evaluators to measure the overall value of the GSQ of an MT system after obtaining the results of the calculation of the total average values of each MT system in terms of the functional and non-functional criteria selected (see Figure 1 on p. 21).

According to the EAGLES (1996: 30), measures may be compound. In other words, the value of an attribute (i.e., criterion) may itself be structured. The value taken by an attribute may therefore be a composite, calculated on the basis of putting together an average or even a weighted average of the values assigned to two or more other attributes. Any one of these attributes may in its turn be a composite. In other words, there is no theoretical limit on the depth of the hierarchy of attribute-value pairs. In brief, the values of the sub-attributes combine to a value for a mother node to reflect the relative importance and strength of the components in a specific evaluation. Further, finding out the GSQ or the total performance of an MT system is essential when comparing systems (as it is the case in this study).

In the present evaluation experiment, the functional criteria were critical and of major importance. Thus, they were assigned the highest percentage values with respect to the weighting scheme of the overall value of the GSQ of an MT system. To avoid counter arguments and as the sub-characteristics under "Accuracy", "Suitability" and "Wellformedness" are closely related from the translation viewpoint, they were given equal percentage values in the weighting scheme of the GSQ. The non-functional sub-characteristics were also equalized regarding their percentage values in this weighting scheme for the same reason. The ones chosen for the present evaluation (i.e., speed of translation, adaptability, storage and flexibility) are relevant to the translation process.

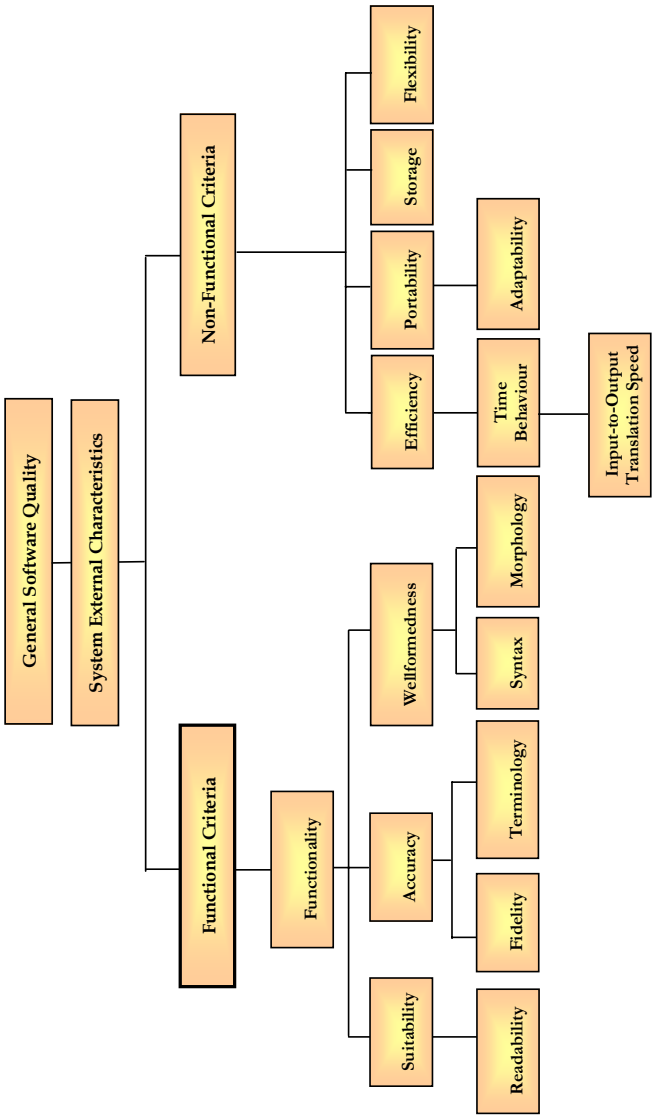

Figure 1: Model for Evaluating the Total Performance (External Quality of Translation) of Arabic MT Systems Based on the FEMTI 2003 Classification

On the basis of this, the distribution of percentage values of the functional and non-functional criteria in the weighting scheme of the overall GSQ is shown below. It is worth noticing that the total performance of an MT system (i.e., the GSQ) gets a sum value of 100%. So, to make the present distribution as fair as possible with respect to the relative importance of criteria (giving primacy to the functional ones), a percentage value of 80 was assigned to the functional criteria, whereas the non-functional criteria were given a percentage value of 20. Thus:

(1) The percentage values of the functional criteria in the weighting scheme of the overall GSQ are as follows: "Functionality"=80% (where the sub-characteristic of "Suitability"=16%; while each of "Accuracy" and "Wellformedness"=32%). Each of the terminal sub-characteristics under these three mother node criteria, in turn, takes the percentage value of 16% for the reasons mentioned above.

(2) The percentage values of the non-functional criteria in the weighting scheme of the overall GSQ are as follows: "Non-functional criteria"=20% (where each of the sub-characteristics under this mother node criteria, i.e., "Efficiency"/"Speed of Translation"; "Portability"/ "Adaptability", "Storage" and "Flexibility", takes a percentage value of 5%).

Hovy (2002a and b) maintains that the distribution of percentage values like the above-mentioned depends on the personal preference for relative weightings that is based on the aims and requirements of a specific evaluation (here, the requirements are related to the end-user). For any study of this type, the most important thing is to select some weighting scheme (such as this) and then also to record all the raw numbers as well. This is also confirmed by Guessoum (2002a and b). In addition, Popescu-Belis (2002b and c) indicates that the important thing is not to have one "absolute" evaluation scheme, but to adapt this

to the user needs. So, there are no "absolute" criteria, but whatever the conventional criteria are, they should be the same for everybody. Further, King, et al. (2003: 3) maintain that the important point here is not the exact definition of any one quality characteristic or of its sub-characteristics; it is rather that MT has a multitude of potential uses in a multitude of different contexts. In any specific context, some characteristics may be important, others not, to the point where a characteristic which is a *sine qua non* in one context may be completely irrelevant in another. And it is precisely because the relative importance of individual quality characteristics is never the same in two different work contexts that the MT evaluator is tempted to feel that he is tackling a problem which has never been tackled before, and therefore, to design his evaluation from scratch.

For calculating the overall GSQ, the following formula was suggested and applied by the researcher according to the distributed criteria weighting scheme:

GSQ%=0.16R+0.16F+0.16T+0.16S+0.16M+0.05P+0.05A+0.05G+0.05X

(where: R=Readability; F=Fidelity; T=Terminology; S=Syntax; M= Morphology; P=Speed of Translation; A=Adaptability; G=Storage; and X=Flexibility)

4. Analysis and Classification of Results

4.1 Evaluation Results of Individual Software Criteria

The following sections and subsections concern themselves with the presentation, classification and discussion of the results obtained by testing the three Arabic MT systems in terms of various criteria types.

The results of each test and the overall quality assessment of the participating systems are presented and classified first. Then, some observations and problematic issues relevant to the evaluation results are discussed together.

4.1.1 Functional evaluation results

As it is stressed in the MTE literature, the criteria of "suitability", "accuracy" and "wellformedness" are the most interesting from the user's point of view in determining the quality of an MT output. In this study, the MT systems under investigation were tested in terms of the classical cognitive criteria such as "readability" and "fidelity", as well as the more objectively measurable criteria like the consistent and correct use of terminology and the linguistic criteria, i.e., morphology and syntax. Here, the tests were designed to show what a system could manage and whether or not it could meet the pre-defined user's tasks. A detailed analysis based on the black-box approach and adequacy/declarative evaluation of twelve various text types for each of the three Arabic MT systems revealed the following results.

These results are classified and presented on the basis of: (a) variation in scores between raters; (b) comparison of systems for text types; and (c) interrelation among functional criteria. For the limits of space, no details of the results calculation are given (see Hannouna 2004).

4.1.1.1 Variation in scores between raters

It should be noted that the total sentences' score values (resulting from the summation of the sentences' ratings of each score on the scale) were basically used in the calculation of the percentage values of the above-mentioned criteria for both raters. These percentages and their average values are demonstrated in Appendices A, B and C, respectively.

To be evaluated using a scoring technique based on subjective

scales, at least two raters should be involved for the reliability of the testing results. This concerns the criteria of "readability", "fidelity" and "syntax". In order to find the level of correspondence in the results between the two raters, it is important to use the widely used statistical measure, i.e., the correlation coefficient value r (see Appendix D). This basically refers to measuring the relationship between two variables. In this study, the two variables represent the scoring values of evaluations 1 and 2 for these criteria. The obtained correlation coefficient values are based on Pearson's linear correlation (Downie and Heath 1983: 103). The Excel results are tabulated in Table I below, where the input data for each value represents twelve values for each evaluator that stand for twelve text types (i.e., twenty-four values for each individual criterion and each MT system).

Criterion	MT System		
	Golden Al-Wafi	Al-Mutarjim Al-Arabey	An-Nakel Al-Arabi
Readability	0.98	0.97	0.96
Fidelity	0.98	0.96	0.97
Syntax	0.93	0.94	0.93

Table I: Correlation Coefficient between the Results of Evaluator 1 and Evaluator 2

According to Downie and Heath (*ibid.*), in every usage an r of 0.8 and above is considered a high coefficient where all the points on the scatter gram tend to fall along a straight line. In addition, an r around 0.5 is considered moderate while that of 0.3 and below indicates a low coefficient. Thus, the results of the tested systems in this study reflect very high correlation coefficient values (most of which are above 0.9)

which means a statistically significant correlation. This points to the fact that the scores of the two evaluators are in close agreement where a strong correlation exists, although for all the three tests not exactly the same sentences were given the same grades by these raters. So, the results of only one evaluator can better be relied on in this respect than the average scores between the raters. In this study, only the researcher's (i.e., eval. 1) results were taken into consideration.

4.1.1.2 Comparison of systems for text types

For each of the functional criteria in this study, the performance of the individual MT systems was investigated in terms of the various text types. This is to show in which domain a particular system was able to produce fluent, accurate and well-formed output for the end-user. For the results of the analysis of (36) sample texts to show the ranking of the MT systems involved with respect to these criteria and the relevant illustrative graphs, see Hannouna (2004: 197-204).

4.1.1.3 Interrelation among functional criteria

The correlation coefficients between the five functional criteria for each of the participating MT systems obtained by using the cross correlation matrix (which is also based on twelve texts' values for each individual criterion) are shown in Tables II, III and IV below. The significance of these values are usually tested by subjecting them to a t-test which represents a measure of a standard error of r and should be computed by the following formula:

$$t = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$$

as t is a function of $t\left[1-\frac{\alpha}{2};\right]$ where α is the level of significance, the most frequently of which is 0.05, 0.01 or 0.001.

	Readability	Fidelity	Syntax	Terminology	Morphology
Readability	1.0	0.72	0.72	0.81	0.67
Fidelity	0.72	1.0	0.61	0.73	0.66
Syntax	0.72	0.61	1.0	0.53	0.8
Terminology	0.81	0.73	0.53	1.0	0.6
Morphology	0.67	0.66	0.8	0.6	1.0

Table II: Correlation Coefficient Matrix of the Functional Criteria of Golden Al-Wafi

	Readability	Fidelity	Syntax	Terminology	Morphology
Readability	1.0	0.74	0.68	- 0.085	0.61
Fidelity	0.74	1.0	0.42	- 0.21	0.27
Syntax	0.68	0.42	1.0	- 0.20	0.61
Terminology	- 0.085	- 0.21	-0.20	1.0	0.35
Morphology	0.61	0.27	0.61	0.35	1.0

Table III: Correlation Coefficient Matrix of the Functional Criteria of Al-Mutarjim Al-Arabey

	Readability	Fidelity	Syntax	Terminology	Morphology
Readability	1.0	0.8	0.63	- 0.134	0.11
Fidelity	0.8	1.0	0.8	- 0.06	0.33
Syntax	0.63	0.8	1.0	- 0.1	0.34
Terminology	- 0.134	- 0.06	- 0.1	1.0	0.55
Morphology	0.11	0.33	0.34	0.55	1.0

Table IV: Correlation Coefficient Matrix of the Functional Criteria of An-Nakel Al-Arabi

In this study, the critical rc value (0.576) is directly taken from a developed table of values (Downie and Heath 1983: 329) depending on 5 percent level of significance and a degree of freedom df=n–2 or 10 in the present work, since n refers to the number of pairs used in computing r. Accordingly, it is unnecessary to go through any calculations.

A look at the correlation coefficient values for each system in the tables above shows that the r values are compared with rc and it is statistically emphasized that each should have a value larger than (0.576) to be significant. In this respect, it is noticeable that for GW 88% of the r values are significant, indicating that the data of the interrelation among the functional criteria fit into Pearson's model (*ibid.*: 103) and are of a linear relation. On the other hand, only 33% of the r values are significant for MA, while 22% of these values are significant for NA. It is worth stressing, here, that there are often situations where the relationship between two sets of variables is not linear, and then the calculated r is an underestimation of the true relationship between the two variables.

However, from the linguistic viewpoint, the five evaluated functional criteria are, in fact, highly interrelated. In other words, any deterioration in one affects the quality of the others. For instance, any syntactic, morphological or lexical defect in a sentence may lead to a total or partial distortion of its form and meaning, and, in turn, result in low fidelity and readability. Although such relations are very close, the resulting r approaches zero.

Further, as an example, a deviation from linearity is apparent on a scatter gram, made for the two variables (readability and terminology) of MA (see Figure 2). Here, it is found that the line is not apparent and the points are spreading evenly in all directions giving an indication of no relationship. Nevertheless, it is possible that a very high, but non-linear relationship will appear to be very low on the basis of Pearson's model (*ibid.*).

Thus, according to the findings of this study, the bivariate relationship is considered curvilinear and the eta coefficient (i.e., correlation ratio) can be used, which reflects the variance accounted for by the best-fitting line, be it curved or straight. This is out of the scope of the present work and further research on this particular aspect may be desirable.

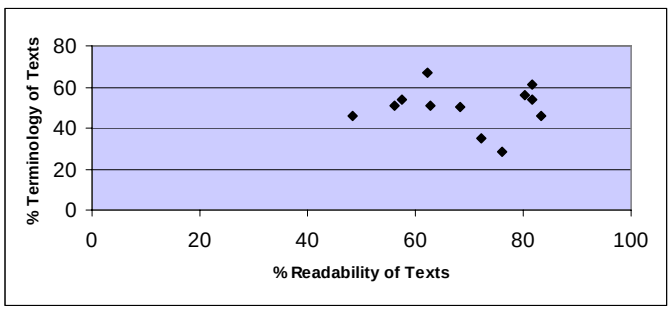

Figure 2: Scatter Plot of the Readability and Terminology Results of the MA System

4.1.2 Ranking of MT systems for non-functional criteria

As mentioned earlier in this study, the top level tasks the participating systems are required to address are the functional (linguistic and cognitive) requirements. In addition, four non-functional (computational) quality characteristics were investigated for the adequacy of a system's performance in fulfilling certain tasks relevant to the translation process. Therefore, benchmark tests were carried out to show the central properties for product choice and to check the possibilities the individual systems offer for the end-user (adequacy/ performance evaluation). Besides, the tests intend to demonstrate the differences among these systems and the systems' ranking with respect to the stated criteria (comparative evaluation).

In general, the average values of MA are the highest with top results of 100% except for the speed of translation where its value is lower than that of GW. Nevertheless, this does not affect ranking the system with the best non-functional performance. Although its average value for speed is higher than that of MA and its average values for adaptability and storage are equal to these of MA, GW comes second in rank after MA regarding the overall non-functional performance. This result is mainly due to its fair flexibility. Further, the results of NA are the worst among these systems as the average values it obtained rank the system in the lowest position for non-functional performance, except with flexibility as its value is the same as that of GW.

It is worth noticing that the results of the two evaluators (i.e., the researcher and a computational expert) are remarkably identical. These are tabulated and presented below without details for reasons of space. For more information of these results, see Hannouna (2004: 207-215).

4.1.3 Evaluation results of GSQ

After obtaining the average values of the functional and non-functional criteria for each of the participating MT systems, the overall GSQ was calculated for purposes of final assessment and comparison in terms of two levels: (a) criteria type and (b) text type. As mentioned earlier, no direct comparison between the two evaluators is given here, since their results are almost identical and only the researcher's values are taken into consideration in this respect.

4.1.3.1 Quality assessment and comparison for criteria

The overall average value of software quality for each of the evaluated MT systems is estimated. This is obtained from the average values of the functional and non-functional criteria in this study. In fact, the MT systems were judged and compared for their total

performance and the evaluation aimed at investigating which system's components work adequately. Thus, the results are presented below with respect to their average values and values of relative importance:

(1) According to the results graphically represented in Figure 3 below, MA has the best system's total performance in comparison with GW and NA. It obtained the highest percentage value of 58% for GSQ. This indicates that with respect to all the functional and non-functional (except for speed of translation) quality characteristics in this study, it gave the best results. However, both GW and NA gained identical results, which are, in fact, not very encouraging and below average (i.e., below 50% in general quality performance). Their total percentage values in the final evaluation of GSQ are 44% and 45%, respectively. Thus, it is believed that there is an elemental need to make more effort to develop and improve the major components of these systems.

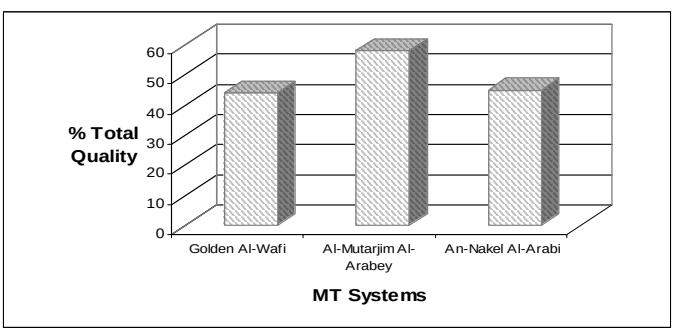

Figure 3: The Percentages of the Total Quality for the Three Arabic MT Systems

(2) In the final stage of classification and presentation of the evaluation results for GSQ, the Arabic systems are compared for their best functional and non-functional total performance. The functional criteria are favored for their direct effect on the translation quality of

the output. In general, the values for both functional and non-functional total performance of MA are the best among these systems. It obtained a value of 40.15% for the former and 17.85% for the latter (see Figures 4 and 5 and Table V below). This shows the fair quality of the MT output and the high computational capabilities of the system. On the other hand, although the values for GSQ of GW and NA are almost identical, NA shows a better functional performance than GW, i.e., 34.35% in comparison with 28.5%, while GW is better in the computational capabilities than NA. The values of these results are 15.5% and 10.65%, respectively. Thus, the quality of the MT output produced by NA is better than that of GW which is faster in performing the translation and capable of translating larger texts and documents.

MT System	Functional Total Performance 80%	Non-Functional Total Performance 20%	GSQ 100%
GW	28.5	15.5	44
MA	40.15	17.85	58
NA	34.35	10.65	45

Table V: GSQ Based on Functional and Non-Functional Total Performance

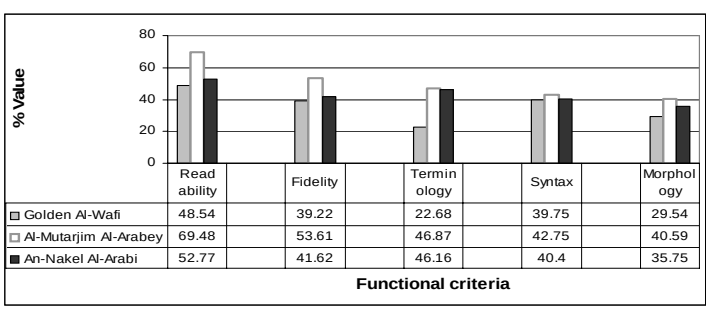

Figure 4: Average Values of Functional Criteria for Arabic MT Systems

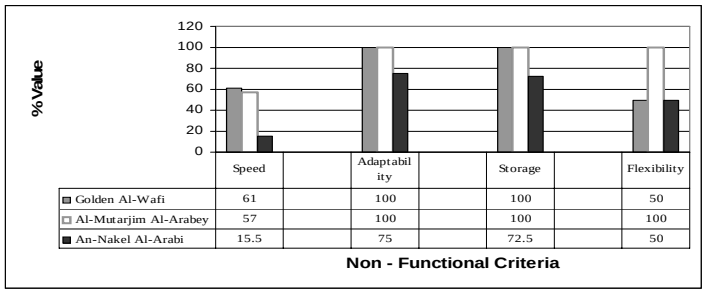

Figure 5: Average Values of Non-Functional Criteria Showing the Range of Arabic MT Systems

4.1.3.2 Quality assessment and comparison for text types

The GSQ of each of the evaluated systems is also calculated in terms of text types, i.e., a total quality obtained from the average values of all the text types translated by an MT system. So, the final evaluation of the participating systems gives the following results (for the limits of space, only the general results will be discussed and the others could be obtained from Figure 6):

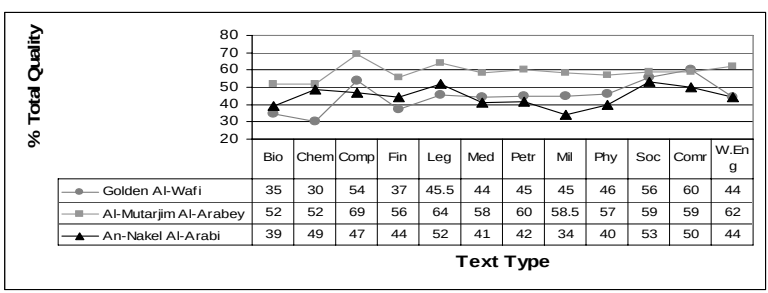

Figure 6: The Variation of Total Quality for the Three MT Systems Based on Text Types

(1) With respect to GW, the total quality it obtained in terms of

text types was 44%, which was below average. Although ATA Company recommended GW for best performance in the scientific fields, it was proved that the system was only better in translating humanities and computational texts. Thus, showing certain defects in the grammatical, lexical and semantic components, GW can produce average MT output (i.e., with a total quality of about 50%) with fair fluency and accuracy for computational, social and commercial texts. Otherwise, the quality of the MT output of this system is poor.

(2) The results of MA are more encouraging than those of GW, with the best total quality performance of 58% in comparison with the other systems. The system showed the highest performance with the scientific texts, although the average values of all the texts reflect no significant differences. In brief, MA can provide fast, readable and accurate MT output of average quality in most cases.

(3) The system that comes second in rank after MA with an almost identical total quality performance with GW (i.e., 45%) is NA. This result also reflects a total quality that is below average. Like GW, NA is better in translating humanities texts. It has a low performance with respect to most of the text types recommended by its manufacturers and even with its best translations, the result is slightly above average in most cases.

4.2 Problems of MT Quality

The preceding sections give the results of the experimental analysis and evaluation of the functional and non-functional criteria, as well as the systems' quality and total performance for criteria types and text types as employed in the data. The next few sections will discuss these results, outlining the types of MT problems, their extent and causes. Because of the limits of space, illustrative examples of Arabic MT reflecting each type of problems below will not be given. These are

available in Hannouna (2004).

4.2.1 Cognitive problems

Various types of problems are identified in the data. These can be considered as direct causes of unintelligible and inaccurate MT output. They virtually result in complete distortion of meaning and unfaithful conveyance of ST content. Indeed, linguistic phenomena are particularly troublesome. The main factors which give rise to incomprehensibility in the Arabic MT output and affect fidelity to the ST are illustrated below.

4.2.1.1 Grotesque word order and choice

In relation to the fluency and understandability of TL sentences, syntactic correctness, organization and coherence, as well as stylistic appropriateness are crucial. As already noted, a considerable number of sentences translated by the Arabic MT systems dramatically reflect alteration and distortion of meaning. Word choice, syntactic arrangements and alternative expressions are generally bizarre. Besides, there are critical words untranslated. These sentences are, in fact, unintelligible and they tend to read like nonsense. It appears that no amount of study and reflection can reveal the meaning of the sentence.

The analysis of the Arabic MT texts shows a combination of grammatical and word usage problems in the MT output from the three Arabic systems. Needless to say, the translations done by MA and NA are better than these of GW, mainly due to the effect of the user dictionary. The translation style of GW seems almost like that of MA since the two systems are produced by the same company. They use the same MT modules, but GW has a less extensive lexicon. In general, what affects the clarity of the output of these systems can be attributed to incorrect grammatical forms and problems in the relationships between the sentence elements. A common feature concerns preservation of

the SL word order in the TL text. In other words, the texts are translated literally, word-for-word following the word order of the SL; therefore, no dependencies between words are created. In addition, untranslated words, poor style and unhappy word choice are also noticeable.

4.2.1.2 Understandable but clumsy translation

Here, the sentence seems intelligible, but actually it is more unintelligible than intelligible. The general idea is somewhat there, but the sentence is hard to read due to the awkward translation. In other words, problems in grammar, word usage and/or style and similar phenomena are present, but constitute mainly "noise" through which the main idea is still perceptible. For the three Arabic MT systems, the meaning can only be guessed after careful study, if at all. To clarify the meaning of the sentences within the various texts, revisions are needed.

4.2.1.3 Unknown and mistranslated critical words

Sometimes the sentence could be generally clear and intelligible, although style and syntactical arrangements are somewhat poor. However, what badly affects readability and clarity of meaning is the untranslation or mistranslation of critical words. Post editing could help create this in nearly acceptable form. It is observed that the translation of GW is poorer than that of the other systems. Neither MA nor NA produces good translations, although the gist can be extracted. The positive effect of the user dictionary of these two systems is apparent in their output too. In spite of certain grammatical and stylistic infelicities in these sentences, the untranslation and mistranslation of some critical words are the direct cause of the problems of readability here.

4.2.1.4 Partially correct translation

Another recognizable phenomenon is that some sentences

translated by the Arabic MT systems are partially intelligible, as only certain parts are correctly rendered. This has a negative effect on the overall readability and clarity.

4.2.1.5 Adherence to SL structures

One of the most distinctive features of texts produced by MT systems is their unnatural literalness. The principal reason for the absence of fidelity in the output considered is due to the complexities of structural transfer and the generation of Arabic grammatical constructions. In fact, structural changes are so common when translating from one language into another that the "low-level" ordering of basic elements like nouns, verbs, adverbs and adjectives should be expected from any MT system. In Arabic output, for instance, the syntactic dependencies and relationships between the sentence elements should be correct Arabic—e.g., *alasaleeb alakhraa li?nwanat althakira* instead of *thakira u'khraa -yu?nwin asaleeban*, which is a word-for-word translation done by NA of the English construction (other memory-addressing modes). Not only is order of significance here, but all the other factors (like word forms and inflections as well as the grammatical categories and arrangements) should be considered with reference to the TL grammatical rules in addition to their faithfulness to the ST (e.g., passive or plural forms, etc.). These must be regarded as the minimum requirements, and any MT system which fails in this respect must be suspected to have deficiencies elsewhere, probably of a grave nature.

4.2.1.6 Loss and distortion of information

In addition to problems of syntactic fidelity, there exists a kind of lack of semantic faithfulness in a number of sentences in the data. This indicates that the content of the input sentence is not conveyed in the MT output sentence at all. In other words, the output is not a proper

sentence. This is mainly due to: (a) loss of information (e.g., where words or even whole sentences are not translated or clauses or phrases are missing); (b) interference (i.e., noise) where words or expressions are added by the system; and (c) a distortion of form (a combination of loss and interference when some words are badly translated or when they do not have an equivalent in the TL).

A look at the Arabic MT output demonstrates serious deficiency in maintaining precision and correctness. In most of these sentences, infidelity to the input SL sentence is due to untranslated words (i.e., either left out or kept in their SL form), bad rendition, inappropriate word choice and inadequate word-for-word translation.

4.2.2 Linguistic problems

The linguistic factors are generally most significant; they are immediate clues to the fact that a system produces the text. Actually, the complexity of the linguistic structure of the Arabic language seems to have posed somewhat of a handicap for MT systems. The result is often unreadable and intangible translation both from English into Arabic and vice versa. The most typical errors which occur in English-into-Arabic and Arabic-into-English translation tend to be structural and lexical; also, several semantic and morphological errors are evident in an MT analysis. In this respect, the data reveal a number of linguistic problems peculiar to the grammatical structure of words and sentences and domain-specific terminology. It is worth noticing that these phenomena constitute further causes for the lack of fluency and accuracy in the MT output. Main translation problems encountered in the present evaluation involve the following:

4.2.2.1 Inadequate terminology interpretation

Lexicalization is concerned with choosing the right word or

expression. This can go wrong in a number of ways. The most obvious and trivial case is where words are incorrectly and inappropriately associated with underlying concepts. MT systems with such a fault may produce something completely nonsensical. Indeed, any MT system is determined above all by the quality and range of its dictionary information. In other words, no MT system will produce a good translation if its dictionaries are inadequate.

Despite the poor performance of the three Arabic MT systems in the terminology test, the results show that in some cases MA is more successful than NA. But, both are clearly better than GW for they have the potential of creating domain-specific/user dictionaries which contribute to much better translation quality. It should be confirmed here that because the guidelines of this test were precise, raters were strict in their implementing them and only restricted themselves to exact comparisons with HT. This may be one of the reasons of poor system performance in this test. However, various types of problems could be identified in the data of these systems as major causes of inadequate rendition of terminology.

(1) Untranslated terminology: Untranslated words are generally a great source of semantic errors and wrong syntactic categorization. This eventually results in unintelligible and unfaithful translation. Despite the use of domain-specific dictionaries, many instances of untranslated words (i.e., either missing or left in their English forms) appear in the MT output of both MA and NA. On the other hand, this phenomenon is more apparent in the translations produced by GW as it does not have the option of making user dictionaries. So, this can be considered one of the reasons of its poor MT quality. It should be stressed here that it is evident in the data of MA that this system translates a given term via the user dictionary, but when it occurs in other sentences in the same text, it is left untranslated (i.e., it reappears in its SL form).

(2) Forms irrelevant to the TL: Some ST terminology occurs as deformed words which are irrelevant to the TL in the MT output of the Arabic systems. Especially when they are key words and expressions, they cause changes and distortion of meaning. This is mainly observed in the data of MA and NA as a result of duplicated letters or additional letters to these words. Besides, sometimes strange words appear due to a literal rendition.

(3) Synonym usage: One of the major reasons for the inexact rendition of words and terminologies in the data of the Arabic MT systems is the heavy reliance on synonyms (i.e., not optimally used words and terms). This is a marked feature especially in the output of GW. In fact, this phenomenon negatively affected the evaluation results of the terminology test.

(4) Inappropriate word choice: As meaning is heavily dependent on context, contextual information and a broad coverage of word senses need to be encoded in a system's dictionary. As the data of the three Arabic MT systems reveal, they are unable to use context to increase the accuracy of word choice. The MT system must use more information from the context to determine which translation is the best. This is a fundamental problem which the programmers of these systems must overcome in order to improve the quality of their products.

(5) Mistranslations: It is noticeable in the data of the Arabic systems considered that some bad translations of critical words and terms are attributed to word-for-word interpretation and transliteration. The result is mostly inappropriate word usage such as, for instance, translating the SL word into two or three words. This spoils the accuracy and clarity of meaning.

(6) Wrong ordering of phrasal constituents of terms: There are other lexical errors in the Arabic MT sentences of the three systems. Although most of the necessary domain-specific terminologies were

inserted into the user dictionaries of both MA and NA, their translations, in most cases, were not precise or correct, due to word-order and other grammatical problems. Without such problems, the accuracy in the terminology rendition of these systems could have been improved.

(7) Errors of category and word class: Other grammatical phenomena affecting the precise and appropriate interpretation of words and terminologies involve errors of categories of nouns and verbs (e.g., number, gender, tense, etc.) and wrong parts of speech (e.g., noun, verb, adjective, pronoun, preposition, etc.).

4.2.2.2 Incorrect rendition of SL grammatical structures

A detailed analysis of the evaluation results revealed that none of the Arabic MT systems involved in this study is good at rendering English syntax correctly. It seems clear that more complex restructuring may well be beyond the capacity of the Arabic systems (e.g., the treatment of the plural). The English-into-Arabic MT system must be capable of recognizing the equivalent structure in Arabic. Thus, it is essential that Arabic MT output follows Arabic grammatical rules and at the sometime preserves its fidelity to the SL structure and meaning. In most cases, this requires the application of four transformational syntactic processes: rearrangement, addition, deletion and replacement.

A further and more detailed discussion of Arabic MT syntactic problems is presented in the following:

(1) Bizarre syntactic arrangement: In spite of the serious lexical problems of untranslated words and the systems' inability to handle polysemy, the major handicap in Arabic MT (leading to confused structure, poor style and awkward translation) is that of word order. The free word order of Arabic makes the parsing of sentences very difficult. In fact, the three participating Arabic systems fail to produce sentences that exhibit TL syntactic rules and arrangements. In most

cases, sentences start with the noun and place the verb towards the end position, preserving the SL word order which is the reverse order of the TL. This problem is observed at all grammatical levels of the Arabic MT sentences in the data (i.e., clausal and phrasal). Even when words and terms are appropriately selected on the basis of context, the form and meaning of the whole sentence are violated. Thus, sound post-editing processes in the rearrangement of syntactic structures of sentences are badly needed to improve the quality of Arabic MT output.

(2) Lack of grammatical agreement: Another phenomenon leading to incorrect rendition of MT sentences concerns syntactic agreement or concord between sentence elements. Differences in gender, number or case are another cause of relatively frequent errors. Arabic is a language with strong requirements with respect to gender, number and case agreement—not only between subject and verb (gender and number agreement) but also in several other types of constructions. In the data examined, it is found that no such agreement exists between the sentence elements in most cases. The three systems reveal considerable shortcomings in this respect.

There is an error in relation to gender in the inappropriate choice of the singular demonstrative pronoun *hada* instead of *hadihi* by NA. This is due to word-for-word translation irrespective of context and TL (here, Arabic) syntactic rules, for Arabic demonstratives have gender distinctions not found in English.

(3) Differences in cases: In general, Arabic has three cases (nominative, accusative and genitive) which are expressed explicitly in nouns, pronouns and adjectives. English, on the other hand, has two cases for nouns (common and genitive), three for pronouns (nominative, accusative and genitive) and none for adjectives (Aziz 1989: 132). These cases are distinguished by diacritics, morphological inflections and prepositional constructions in Arabic. For example, the Arabic equivalents

for "an engineer" and "engineers" (the singular and plural forms for all cases in English) are: nominative *muhandis*, *muhandisaan*, *muhandisoon* and accusative/genitive *muhandisa*, *muhandisi*, *muhandisayn*, *muhandiseen*. Such differences in cases, of course, result in translation errors, because the main bilingual dictionaries of the Arabic MT systems do not contain valency and sub-categorization information.

(4) Incorrect pronoun translation: This is another known area of deficiency in Arabic MT systems. Again, the data of the three participating systems manifest certain incorrect pronoun translations mainly due to differences in gender and number, homographs and other problems in addition to the complexity of the Arabic pronoun system. The English personal pronoun "you" for instance, is rendered into Arabic in no less than five ways: *Anti/Anta/Antumaa/Antum/Antuna*.

(5) Wrong part of speech categorization: An obvious syntactic error in the data of the three Arabic MT systems that results in mistranslation and alteration of the structure and meaning of the SL sentence occurs when the wrong part of speech is selected in the TL output.

(6) Inconsistent preposition handling: The problem of the Arabic MT systems' inability to handle prepositions correctly and appropriately is quite serious. Often, the wrong preposition appears in the MT sentence, or two successive prepositions are used, or there is a missing preposition, or there is a need to delete one.

(7) Failure to treat the definite article: Errors regarding the rendition of the Arabic definite article are evident in the MT output of the three systems. This is an indication of the systems' incapability to treat this syntactic aspect appropriately due to the missing definite article in the Arabic text or vice versa (i.e., it is arbitrarily added by the system where it does not fit). Moreover, it is even interpreted wrongly in some instances where it is replaced by the preposition "L".

(8) Problems of tense, aspect and voice: For both the passive and perfective tenses, the Arabic MT systems often fail to treat the English verbal combination as one unit. Rather, they translate the auxiliary and the lexical verbs separately into their equivalents in Arabic as a matter of word-for-word rendition.

In sum, there might be other syntactic problems, but the ones discussed above are the most obvious and the major causes of inefficient and inadequate MT output. This is mainly attributed to the systems' failure to deal with the Arabic grammatical rules and to certain deficiencies in the translation process itself. The overall effect is that of limited and poor syntactic quality.

4.2.2.3 Erroneously recognized TL morphology

Another serious and very complex linguistic problem that results in ill-formed MT sentences concerns the incorrect TL morphological structure. An important consideration in the development of Arabic MT systems is Arabic morphology. Most of the researches in Arabic NLP systems mainly concentrate on the field of morphological analysis (Ditters 2001; Othman et al. 2003). In the present study, it is found that all the Arabic MT systems exhibit uniformly low morphological representations. This is because they are incapable of recognizing the morphological differences and the equivalent structures between the two languages for certain defects in their morphological analyzers and rules, as well as in the grammatical information contained in the lexicon of each system.

Three types of morphological problems were found to be recurrent in the data:

(1) Inappropriate choice of prefixes: One major factor that often impedes correct morphological forms in the Arabic MT output is that of definiteness, i.e., the Arabic definite article "al" is found absent where

it should be used or vice versa. Other problems relevant to the attachment of prefixes to lexical items involve errors in word class categorization, gender, tense and doubling of prefixes.

(2) Errors of word infixes: Due to certain syntactic processes involving the derivation of lexical items (e.g., deriving the plural form "sawareekh" from the singular noun *sarookh*, "missile"), some morphological changes in the internal structure of words occur. Very few errors of infixes (i.e., the morphological inflections that occupy the medial position of words) were observed in the data of the participating MT systems. These mainly concern errors in the derivation of plural noun forms and incorrect choice of the part of speech.

(3) Inability to maintain correct suffixes: The morphological structure and formation of Arabic words is governed to a great deal by the syntactic relationships among the sentence elements in terms of agreement in number, gender, case, tense, etc. Suffixes determine the grammatical significance of lexical items which is realized by different morphological inflections and is expressed differently in each language. The noun "engineer", for instance, can be changed into the plural form by attaching the inflectional suffix "-s" to the end, as in "engineers", while in Arabic the plural form of *muhandis* is realized by the inflection "-oon" for regular plural masculine nouns.

4.2.3 Operational problems

In addition to the cognitive and linguistic problems that directly contribute to poor MT quality, other types of deficiencies regarding the operation of the Arabic MT systems and the translation process itself were also recognized. Two types of such problems were well identified. The first is related to the measurement of the translation speed, while the second concerns the design of the user dictionaries.

4.2.3.1 Impediments in measuring speech

In measuring the speed of translation some errors hindered the process in the case of NA only. This is mainly attributable to the fact that the system currently available in the local market is a copy and not the original version. Therefore, during the process of translation, the system did not work adequately and suffered from constant pauses.

Further, in other cases, during the translation of a text some suggestions regarding the correct rendition of certain words appeared in a message box requiring confirmation on the part of the user. This is another reason for increasing the time the system needed for the translation process, which has a negative effect on the computational quality and the overall quality of the system. However, this takes only a few seconds and the speed of translation of NA is considered good in general. Such a pause has significance, in particular, for evaluation purposes especially these involving the comparison of systems.

From the evaluation results based on this criterion, it is proved that NA has the lowest performance, while GW has the highest. This is partly due to the reasons mentioned above. Besides, what may slow down the translation speed of NA is the extra time it requires in searching for the correct and most appropriate equivalent to the SL meaning and structure making use of its four dictionaries and TM (i.e., Translation Memory), in addition to the user dictionary. GW, on the other hand, performed a translation faster than the other Arabic systems as it has only one dictionary involved (i.e., the basic bilingual dictionary).

4.2.3.2 Limitations of user dictionaries

The researcher designed user dictionaries for the twelve special-domains involved for MA and NA. Although these dictionaries enhanced terminology rendition of the two systems (as GW lacks such a characteristic), their supply of translations for the specific terminology

while the systems were running was not free of errors. The researcher even encountered certain difficulties in this respect especially with NA. It is worth stressing, here, that even though most domain-specific terms and certain expressions were entered into the user dictionaries of these systems, their influence over the total quality of the translated output was proved to be ineffective.

(1) MA System: The following are the limitations of the user dictionary and the preference dictionary of MA:

(a) Even though some terminologies and expressions were entered and encoded into the user dictionaries of the system, it selected other equivalents in the TL that do not fit the context for unknown reasons.

(b) Another defect of MA's user dictionary concerns the word forms. It was realized that if the term or expression entered is in the plural form, the system did not translate it when it recurred in its singular form and vice versa. Thus, it is concluded that any change in subsequent occurrences of a word form will result in its mistranslation or untranslation.

(c) Terminology and expressions should be entered and encoded into the user dictionary of MA according to their sequence in the sentence or text. Any change in this respect leads to translation errors in word positions. The researcher took this into account when designing the user dictionaries in the present evaluation.

(d) Another dictionary intended to enhance translation accuracy is the preference dictionary. It can be designed by the user after the first translation of the text. It is worth stressing that all the words in this dictionary were randomly selected by the system. The user's role here is only to highlight the most appropriate word meaning that fits the context. This takes

place after the system's selection of the first meaning in the first attempt which is inappropriate in most cases. In the biological text, for instance, the word "culture" has the meaning *ziraa?ah*. In the preference dictionary, the system chose the first meaning in the list of alternative meanings—*thakaafah* where it did not fit. Thus, the researcher highlighted the word *ziraa?ah* among *ziraa?ah*, *mustanbet*, *turaath*, *thakaafah*; it is the most appropriate to the context and suggested in the HT. Accordingly, this meaning correctly appeared in the second translation of the text. Despite the positive effect of this dictionary on translation quality, there are certain disadvantages: (i) In many cases, all the suggested words given in the list of alternative meanings do not fit the context, or the optimal meaning is not mentioned. Here, the user has no alternative but to let the system select the first meaning in the list, as it often does. As a result, the SL word will be mistranslated or deformed in the MT output. (ii) Some functional words (e.g., both, in, through, etc.) that do not greatly affect the accuracy of the overall meaning of the sentence or text are randomly selected by the preference dictionary, which assigns them the exact meaning in the TL. Such an option, in fact, should instead be restricted to the important key words, i.e., content words and terminology. But, it was found that although such words directly provide the exact meaning and effectively convey information, they were mostly not selected.

(2) NA System: As mentioned earlier, since the system under evaluation is not the original version of NA, some difficulties were encountered:

(a) There are few buttons in the user dictionary dialogue box (e.g., "open", "close", "add", "delete", "modify", "ok" and

"cancel") which are essential for its adequate operation and performance. These commands were not shown on the buttons at all and it was really so difficult to proceed without knowing which button was being pressed and for what purpose. With the help of the manual of the system and after many attempts, the researcher could at last realize how the commands were represented. This is the same for the idiom and specific dictionaries.

(b) In many cases, either when a new word was entered into the user dictionary or when a specific button was pressed, a message box appeared indicating that "This program has performed an illegal operation and will be shut down" while another one pointed out that "If the problem persists, contact the program vender". Consequently, the system closed and had to be restarted again.

(c) In general, the design of NA's user dictionary is more complex than it is with MA. While such an operation requires solely the listing of terms and expressions needed with MA, it was essential to get more information regarding the semantic features (e.g., human, animal, concrete, etc.), the main grammatical categories of the content words (verb, noun, adjective and adverb) or terms such as number (singular/plural); gender (masculine/feminine); type (expression, proper noun, title, etc.) and verb (without object, one object, two objects, etc.). Needless to say, such a process was time-consuming and required special effort when the commands were not shown on the buttons. Above all, the designer of such dictionaries must be highly competent in the linguistic systems of the two languages involved.

(d) When any word class was typed in the dialogue box, it did

not appear even though it was added to the list of words in NA's user dictionary. This was confirmed by a second attempt to enter the same word, when a list of items alphabetically arranged was displayed, showing the word highlighted.

(e) If the word entered into the user dictionary was capitalized, NA did not translate it when it recurred with a small letter in a particular text.

(f) On certain occasions, NA did not translate—or mistranslated—the words or terms for unknown reasons. It did not take the correct and appropriate translation of these words and expressions from the user dictionary.

(g) Another serious defect in the user dictionary of NA concerns the selection of the optimal and most appropriate meaning for a word or expression. Here, it was necessary to highlight the intended meaning in the long list. On certain occasions, none the meanings in the list fit the context. Thus, the researcher entered and highlighted a new meaning. As the system mostly selected the first meaning because of some functional defects, all the alternative meanings had to be deleted. In addition, it was noticed that when the intended meaning was highlighted in the list, it did not appear in the second translation of the text. Again, this required the deletion of all the alternative meanings, leaving only the one desired so that the system could select this in the second translation.

(h) In other cases, the alternative meanings of verbs or other lexical items did not appear in the meaning box, but when the cursor was placed on different positions in the box only the highlights showed up. As it was impossible to see and know the hidden meanings, the researcher had to delete all the highlights and entered the word intended instead.

Because of all these deficiencies, NA does not operate adequately and a lot of time and effort are needed to perform certain functions to improve the quality of its MT output.

Nevertheless, with respect to the grammatical structure of words and expressions, NA was found to be better than MA. In other words, the items entered into its user dictionary changed according to the grammatical categories and contextual requirements. Further, when a word was added to NA's user dictionary, it appeared in the same form, or in different forms, wherever it occurred in the same text.

In general, the effect of the user dictionary on the accuracy of translation and the improvement of MT quality is apparent in the translations of MA and NA systems with some instances of good quality (for examples, see Hannouna 2004).

5. Conclusions and Recomendations

5.1 General Conclusions

The present evaluation confirms the following drawbacks with these systems:

A. Overall systems quality performance

(1) In general, all the three Arabic MT systems evaluated produce average or below average quality, i.e., with a total quality performance of 58% for MA, 45% for NA and 44% for GW. In fact, their outputs reflect many deficiencies in translating various text types. In short, they all need serious improvements.

(2) MA is better than the other systems in performing scientific translations and the three systems can only manage commercial and computational texts with a "quite satisfactory" translation.

(3) The end-user can use the Arabic MT systems to grasp the general idea of the ST, or translate short, simple texts.

B. Scoring schemes

(1) For the evaluation of "readability", "fidelity" and "syntax", the usual methods of rating the quality of MT output have relied on human judges assigning scores on a graded scale (such as 0-3 and 1-5). Although this scoring technique is rather subjective, automatic MTE measures are even less reliable and are still very far from being able to replace human judgement. Nonetheless, the present study shows that MT can be evaluated using well-known evaluation measures and the metrics and methods adopted proved significantly reliable and are the most commonly used in the MTE literature (based on the ISLE's most recent Evaluation Framework, i.e., EFMTI of 2003, and the researcher's suggested methods). The approach presented is, indeed, the right way to move towards a trustworthy systematic evaluation of MT quality. In general, human judgement still seems to be the most popular way of assessing quality characteristics.

(2) To obtain reliable evaluation results, the correlation coefficient values between the scorings of the two evaluators with regard to the criteria of "readability", "fidelity" and "syntax" are above 0.9, which indicate a statistically significant correlation.

(3) The correlation coefficients between the five functional criteria for each Arabic MT system are tested using the cross correlation matrix. This is to statistically determine the degree of interrelation among these criteria since they are linguistically highly interrelated. The results are tabulated showing that for GW, 88% of the r values are significant, indicating that the data of the interrelation among the functional criteria fit into Person's model with a linear relation. However, for both MA and NA, this test points out to a relationship between two sets of variables,

but it is not linear. The aim of these tests is only to show the possibility of carrying out statistical studies on the basis of the different evaluation results.

C. Major systems' problems: functional criteria

(1) Three major types of problems are identified in the data of the Arabic MT systems showing the main causes of poor quality. These are cognitive, linguistic and operational problems. These could be seen in the previous sections and sub-sections.

(2) Cognitive problems reflecting unclear and inaccurate MT output and unfaithfulness to the ST result from a variety of phenomena, most of which are linguistic ones.

(3) The linguistic problems are peculiar to the grammatical structure of words and sentences and domain-specific terminology. Such deficiencies in the Arabic MT output are due to: (a) inadequate terminology interpretation; (b) incorrect rendition of SL grammatical structures; and (c) erroneously recognized TL morphology.

(4) Various types of lexical and semantic problems have been identified in the data of these systems. These are direct causes of inadequate rendition of terminology.

(5) There are major syntactic and morphological deficiencies in the data of these systems that significantly alter the structure and meaning, badly affecting the fidelity of the MT output to the original ST and the understandability of the TT.

D. Major systems' problems: non-functional criteria

(1) Two types of operational problems were recognized during the running of the Arabic MT systems and the translation process. These concern certain impediments in measuring the translation speed and some limitations in the design of the user dictionaries. Building the

user dictionaries of NA was not an easy task at all. In addition to some operational deficiencies, the process required high linguistic competence in the two languages involved, i.e., English and Arabic. Thus, the system can be better used by translators or knowledgeable users (see Hannouna 2004).

(2) MA and GW can translate large texts with typical system storage of 40 Kb in the English and Arabic text boxes. They can also translate large documents of whatever size in the background (e.g., 5.84 Mb/ 1.165.823 words in 55 minutes).

(3) MA achieved the best computational performance among the systems studied except for the speed of translation, as GW was faster, though the difference was not big here. MA had the adaptability and flexibility of 100% as it ran under any version of the operating system and could easily be installed on PCs of all specifications. As it has the highest security, it can be registered on one computer only. This system showed great flexibility with respect to user and preference dictionaries building and updating of its general and special-domain dictionaries.

(4) NA suffered from many pauses during the translation process and it could not translate long and grammatically complex sentences. The largest size it can translate should not exceed 29 Kb, and the system is unable to translate large documents in the background. It is not adaptable to the latest versions of the operating system (i.e., Windows 2000 and XP), but it can easily be installed and run on PCs of all specifications, since it is only a copy of the original system. Further, its flexibility is restricted to building user dictionaries.

(5) GW was the fastest system, since no extra time was required to search for domain-specific terminologies or the most appropriate alternative meanings that went with the theme of the text, as it was the case with the other MT systems.

5.2 Some Sources of Arabic MT and MTE Problems

Some possible causes of problems regarding the poor performance of Arabic MT systems are:

(1) Variations between SL and TL linguistic structures.

(2) Deficiencies of systems' dictionaries and components.

(3) The frail nature of MT.

(4) Lack of awareness of the strategic nature of Arabic MT.

(5) Lack of awareness of the significance of MTE.

5.3 Recommendations

For the purpose of taking Arabic MT and MTE to the stage where MT for other languages has reached, the following recommendations are suggested:

(1) There should be increasing awareness of the strategic nature of Arabic MT.

(2) MT staff development schemes are required.

(3) Teaching advanced and standard MTE techniques is required.

(4) Cooperation among Arabic MT companies is essential to working on problems of common interest and directing effort and expertise towards one goal.

(5) Governments and Pan Arab organizations should enrich Arabic research in NLP and Arabic MT theory and practice.

(6) Evaluation of MT should concentrate on the evaluation and improvement of the lexicons of such systems. Domain-specific dictionaries should be run during the translation process, and a special code should be made for each domain and text type to ensure the precise selection of terms.

(7) Contextual information, knowledge of the world and a broad coverage of word senses are essential, and they should be encoded in a system's dictionary. Other grammatical, pragmatic and semantic aspects

should also be taken into account.

(8) It is necessary to design and improve systems that are restricted to special text types and domains so that there is high quality MT output with a minimum number of errors and maximum precision and speed.

(9) There is a need to concentrate on the production of interactive MT systems where a human translator can interact with the computer and make the necessary changes and corrections during the translation process for output of better quality.

(10) Emphasizing the advantages of pre-editing and post-editing processes is a necessity.

References

Arnold, Doug (1994). *Machine Translation: An Introductory Guide*. London: Blackwell.

ATA Software Technology (1997). *Al-Mutarjim Al-Arabey User Manual.* <Http://www.atasoft.com/almutarjim. htm/>. Cited 2 July 2009.

Banerjee, S. and A. Lavie (2005). "METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments". *The 43rd 40th Annual Meeting of the ACL.* Ann Arbor, Michigan, 65-72.

Callison-Burch C., et al. (2007). "(Meta-) Evaluation of Machine Translation". *The Second Workshop on Statistical Machine Translation*, Prague, 136-158.

Dabbadie, Marianne, et al. (2002). "A Hands-On Study of the Reliability and Coherence of Evaluation Metrics". *Workshop at the LREC Conference, Machine Translation Evaluation: Human Evaluators Meet Automated Metrics.* Las Palmas, Canary Island, Spain. 27 May. <Http://www.issco.unige.ch/projects/isle/mteval-may02>. Cited 23 March 2009.

Ditters, E. (2001). "A Formal Grammar for the Description of Sentence Structure in Modern Standard Arabic". *Proceedings of Arabic NLP Workshop*

at ACL/EACL. <Http://www.elsnet.org/ac/2001-arabic.htm/>. Cited 27 Aug 2003.

Doddington, G. (2002). "Automatic Evaluation of Machine Translation Quality using N-gram Co-occurrence Statistics". *Second Human Language Technologies Conference* (HLT-02), San Diego, CA, 128-132.

Downie, N. M., and Robert W. Heath (1983). *Basic Statistical Methods*. New York: Harper & Row Publishers.

EAGLES (1996). *EAGLES Evaluation of Natural Language Proceeding Systems*. Final Report. Center for Sprogteknologi, Denmark.

_____ (1999). *EAGLES Evaluation Working Group*. Final Report. <Http://www.unige.ch/projects/eagles/ewg 99/htm>. Cited 18 May 2001.

_____ (1996-2008). *Electronic Resource*. <Http://www.ilc.cnr.it/EAGLES96/home.html>. Cited 21 September 2008.

Elliott, D., et al. (2004). "A Fluency Error Categorization Scheme to Guide Automated Machine Translation Evaluation". *AMTA*: 64-73.

Estrella, P., et al. (2005). "Finding the System that Suits You Best: Towards the Normalization of MT Evaluation". *Proceedings of the 27th International Conference on Translating and the Computer*, ASLIB, 24-25 November, London.

FEMTI (2003). *The FEMTI Classification of Machine Translation Systems*. <Http://www.issco.unige.ch/projects/isle/taxonomy3/>. Cited 18 May 2001.

Guessoum, Ahmed (2002a). "GSQ Weighting Scheme". E-mail to the researcher. 30 September.

_____ (2002b). "Values of Functional and Non-functional Criteria". 7 October.

Hannouna, Yasmin H. (2004). *Evaluation of Machine Translation Systems: The Translation Quality of Three Arabic Systems*. Dissertation, Al-Mustansiriyya University.

Hovy, Eduard (2002a). "Distribution of Scores". E-mail to the researcher. 1

October.

_____ (2002b). "Overall Summed Score". E-mail to the researcher. 3 October.

_____ (2003). "Adaptability". E-mail to the researcher. 14 February 2003.

Hovy, Eduard, and Maghi King (2001). "Customizing Machine Translation Evaluation". *Machine Translation Journal* 19.8: 1-37.

Hutchins, John (2001a). "Machine Translation and Human Translation: in Competition or in Complementation?" *International Journal of Translation* (Special Issue on MT) 13.1-2: 5-20.

ISO/IEC-2126 (1991). *ISO/IEC 9126: 1991 (E)–Information Technology-Software Product Evaluation-Quality Characteristics and Guidelines for Their Use*. International Organizational Electron technical Commission, Geneva.

JEIDA (1992). *JEIDA Methodology and Criteria on Machine Translation Evaluation (JEIDA Report)*. Ed. H. Nomura. Japan Electronic Industry Development Association, 1-17.

King, M., et al. (2003). "FEMTI: Creating and Using a Framework for MT Evaluation". *Proceedings of MT Summit IX*. New Orleans, Louisiana, U.S.A. 23-27 September. <Htttp://www.amtaweb.org/summit/w52/MT_Summit_IX_Papers.htm>. Cited 5 November 2003.

Koh, Sungryong, et al. (2001). "A Test Suite for Evaluation of English-to-Korean Machine Translation Systems". *Proceedings of MT Summit VIII*. Santiago, Spain. 18-22 September. <Http://www.eamt.org./summit VIII/papers.html>. Cited 26 May 2002.

Lavie, A., and A. Agarwal (2007). "Meteor: An Automatic Metric for MT Evaluation with High Levels of Correlation with Human Judgments". *The Second Workshop on Statistical Machine Translation*, Prague, 228-231.

Lavie, A., et al. (2004). "The Significance of Recall in Automatic Metrics for MT Evaluation". *The 6th Conference of the AMTA*, Washington DC.

Miller, Keith, et al. (2001). "Evaluation of Machine Translation Output for an Unknown Source Language: Report of an ISLE-Based Investigation".

Proceedings of the MT Summit VIII. Santiago, Spain. 18-22 September. <Http://www.eamt.org/summit VIII/papers.html>. Cited 26 January 2002.

Othman, Eman, et al. (2003). "A Chart Paper for Analyzing Modern Standard Arabic". *Proceeding of the MT Summit IX*. New Orleans, Louisiana, U.S.A. 23-27 September. <Http://www.amtaweb.org/summit/w52/MT_Summit_IX_papers.htm>. Cited 5 November 2003.

Papineni, K., et al. (2002). "BLEU: A Method for Automatic Evaluation of Machine Translation". *The 40th Annual Meeting of the ACL*, Philadelphia, 311-318.

Popescu-Belis, Andrei (2002a). "Text Types and Distribution of Criteria Scoring". E-mail to the researcher. 7 October.

_____ (2002b). "GSQ Weighting Scheme". E-mail to the researcher. 18 October.

_____ (2002c). "Final Weighting Scheme". E-mail to the researcher. 19 October.

_____ (2003). "Adaptability". E-mail to the researcher. 5 February.

Popescu-Belis, Andrei, et al. (2006). "A Model for Context-based Evaluation of Language Processing Systems and its Application to Machine Translation Evaluation". *LREC 2006 Fourth International Conference on Language Resources and Evaluation*, Genoa, Italy, May.

Turian, J. P. (2003). "Evaluation of Machine Translation and Its Evaluation". *The 9th MT Summit*, New Orleans, 386-393.

Van Slype, Georges (1979). *Critical Study of Methods for Evaluating the Quality of Machine Translation* (Final Report). Prepared for the Commission of the European Communities, Brussels.<Http://www.issco.unige.ch/projects/isle/van-slype.pdf>. Cited 24 January 2002.

Vanni, Michella, and Keith J. Miller (2001). "Scaling the ISLE Framework: Validating Tests of Machine Translation Quality for Multi-Dimensional Measurement". *Proceedings of the MT Summit VIII*. Santiago, Spain. 18-

22 September. <Http://www.eamt.org/summit VIII/papers.html>. Cited 26 January 2009.

White, J., et al. (1992-1994). *(D)ARPA Workshops on Machine Translation: Series of 4 Workshops on Comparative Evaluation*. PRC Inc., Melean, VA.

Zwarts, Simon (2009). *Evaluation of Machine Translation*. <Http://web.science.mq.edu.au/~szwarts/MTEvaluation.pdf>. Cited 15 September 2009.

Appendices

Appendix A: The Experimental Values of Texts % Readability

Text type	Evaluator	System Type		
		Golden Al-Wafi	Al-Muterjim Al-Arabey	An-Nakel Al-Arabi
Biological	1	34.36	56.06	30.30
	2	30.33	51.51	24.24
Chemical	1	30.55	48.44	46.61
	2	32.66	41.77	50.11
Computational	1	68.66	81.66	76.66
	2	64.00	81.66	70.00
Financial	1	36.46	62.82	50.00
	2	40.04	64.10	51.43
Legal	1	50.00	80.36	59.00
	2	50.21	79.33	53.33
Medical	1	40.96	72.27	43.93
	2	42.12	69.57	45.45
Petroleum	1	32.17	76.16	66.21
	2	29.35	80.75	71.35
Military	1	44.51	62.12	27.21
	2	43.12	64.18	22.18
Physical	1	46.06	57.57	39.39
	2	42.51	60.60	41.25
Social	1	64.00	68.33	51.66
	2	60.66	65.04	55.00
Commercial	1	80.12	83.33	76.60
	2	84.47	86.06	80.36
Water Engineering	1	54.66	81.66	65.66
	2	52.33	79.00	72.33
Average Value of Eval. 1		48.54	69.48	52.77

Appendix B: The Experimental Values of Texts % Fidelity

Text type	Evaluator	System Type		
		Golden Al-Wafi	Al-Muterjim Al-Arabey	An-Nakel Al- Arabi
Biological	1	25.75	36.36	21.21
	2	25.30	35.34	18.24
Chemical	1	15.72	37.72	41.66
	2	11.33	36.11	39.25
Computational	1	36.66	58.33	50.66
	2	36.33	60.34	52.48
Financial	1	35.64	52.30	42.30
	2	33.33	58.12	45.25
Legal	1	41.66	60.33	53.33
	2	38.35	63.45	48.86
Medical	1	34.21	63.33	43.93
	2	35.86	66.14	42.12
Petroleum	1	40.03	61.28	48.73
	2	41.22	63.42	52.13
Military	1	43.33	53.33	20.24
	2	48.16	50.06	17.87
Physical	1	43.75	52.42	36.84
	2	42.09	58.11	38.26
Social	1	60.00	58.33	50.00
	2	65.43	62.22	52.75
Commercial	1	60.31	57.96	53.96
	2	65.55	60.33	55.36
Water Engineering	1	33.66	51.66	36.66
	2	35.25	56.48	42.02
Average Value of Eval. 1		39.22	53.61	41.62

Appendix C: The Experimental Values of Texts % Syntax

Text type	Evaluator	System Type		
		Golden Al-Wafi	Al-Muterjim Al-Arabey	An-Nakel Al- Arabi
Biological	1	33.63	34.54	32.72
	2	33.09	36.75	30.17
Chemical	1	30.00	42.50	41.66
	2	32.35	38.85	40.27
Computational	1	53.00	55.00	44.00
	2	50.66	57.32	48.55

Financial	1	38.46	41.53	46.15
	2	40.35	43.22	46.38
Legal	1	44.00	46.00	48.00
	2	42.95	47.32	43.63
Medical	1	38.18	41.81	44.54
	2	34.22	43.55	48.35
Petroleum	1	40.00	49.65	41.37
	2	38.86	53.75	45.10
Military	1	32.72	33.63	20.36
	2	30.43	37.28	18.03
Physical	1	37.27	35.45	36.36
	2	39.88	40.00	32.18
Social	1	50.00	42.00	45.00
	2	53.86	45.38	41.51
Commercial	1	43.80	41.90	40.61
	2	45.73	43.15	38.32
Water Engineering	1	36.00	49.00	44.00
	2	38.63	52.24	45.36
Average Value of Eval. 1		39.75	42.75	40.40

Appendix D: Statistical Aspects

The (Pearsons' Linear Correlation Coefficient) usually has a symbol r and is a pure number without units or dimensions and it always lies in the interval -1 to +1. The following formula is well adapted for use with the pocket calculator or specific statistical software:

$$r = \frac{n\sum XiYi - (\sum Xi)(\sum Yi)}{\sqrt{[n\sum Xi^2 - (\sum Xi)^2][n\sum Yi^2 - (\sum Yi)^2]}}$$

n = number of individuals

Xi = dependent variable

Yi = independent variable

r is a pure number without units or dimensions and it always lies in the interval -1 to +1

- When two variables are inversely related, it means that as one increases, the other decreases.
- The absence of any relationship between variables is denoted by a correlation coefficient of 0.00 or thereabouts.
- When pairs of values are plotted in such a way that they fall along a straight line from the lower left of the scatter gram to the upper right, then the correlation coefficient is equal to +1.00, and when the points fall a long a straight line, which runs from the upper left part of the scatter plot to the lower right, then this indicates a perfect negative relationship, a value of -1.
- The size of *r* reflects the amount of variance that can be accounted for by a straight line, whether the data are essentially linear or not.
- Positive values indicate a tendency for both variables to increase (or decrease) together, while negative values indicate that the high values of one variable are associated with the low values of the other variable.
- The levels of significance most frequently encountered are 0.05, 0.01, and 0.001.
- The level of significance =0.05 means that 5 percent of the values are significant.

About the Author

Yasmin H. Hannouna is an assistant professor of translation at the United Arab Emirates University and a member of the International Association of Translation and Intercultural Communication. Her fields of research interests are: translation quality assessment, machine translation and evaluation of machine translation systems, localization and translation and intercultural studies. E-mail: <yasminhannouna@uaeu.ac.ae>.

雜語、方言與文學翻譯
——以王禎和的《玫瑰玫瑰我愛你》及其英譯為例

李 波

Abstract

Heteroglossia, Dialect and Literary Translation: A Case Study of Wang Chen-ho's *Rose Rose I Love You* and Its English Translation (*by* Li Bo)

By criticizing individualistic subjectivism and abstract objectivism, Bakhtin proposes heteroglossia in "Discourse in the Novel"*, which attaches great importance to the social stratification of language within a national language and beyond. In Bakhtinian sense, the social dialects play a very important role in the novel, and even make the novel a locus of ideological struggle. In this paper, the author tries to elaborate on the relationship between heteroglossia and social dialects and its implication for literary translation. Wang Chen-ho's* Rose Rose I Love You *and its English translation will serve as a case study of this particular literary phenomenon.*

一、方言與雜語

巴赫金（Bakhtin）在〈長篇小說的話語〉中提出了“雜語”（heteroglossia）這一概念，並且詳細分析了長篇小說組織和引入雜語的主要模式。巴赫金在闡釋“雜語”這個概念的時候指出，“在歷史存在的任何時刻，語言從上到下具有雜和性，它表示了現在與過去，過去的不同時代，現在的不同社會意識集團、不同傾向、學派、團體等等之間的社會意識形態矛盾的共存，各有自己的體現形式”（Bakhtin 1981: 291）。巴赫金一直以來非常重視方言在文學中的作用，特別是方言在建構雜語方面所發揮的作用。有統計，巴赫金在其討論長篇小說話語的長文中，總共使用“方言”這個詞彙超過二百多次，可以看出方言在巴赫金眼中的重要性。方言成了構成雜語的主要因素，那麼，方言、標準語與雜語之間又是什麼關係呢？巴赫金在〈長篇小說的話語〉中指出：

> 語言在其形成過程的每一時刻裏，都不僅僅分解為嚴格意義上的語言學裏的方言（根據語言學的形式標誌，主要是語音標誌）；對我們更重要的，是還分解為不同社會意識的語言，即社會集團的語言、“職業語言”、“體裁語言”、幾代人的語言，如此等等。從這個角度看，規範語本身也只是雜語中的一種，而且它自身又可分解為不同的語言（不同體裁、不同思潮的語言等）。（巴赫金，1998: 50）

巴赫金指出了語言學裏的方言對語言發展過程的重要作用，但更

加重視同一時期語言內部的社會分層（social stratification），也即分解為“不同社會意識的語言，即社會集團的語言、‘職業語言’、‘體裁語言’、幾代人的語言”等社會方言。這樣的分層，對於理解小說中的雜語才具有更加重要的意義。也就是說，我們應該怎麼樣來看待雜語與方言這兩者之間的關係？一方面，巴赫金在討論〈長篇小說話語的發端〉時，特別強調方言在歐洲國家語言和文學（特別是小說）發展中所發揮的重要作用（巴赫金，1998: 503）；而同時，在〈長篇小說的話語〉中又特別強調方言之外語言分層的重要意義。我們可以看到，雜語和方言之間既有重合的地方，又各自有自己的特徵。

巴赫金特別重視社會雜語對小說的重要作用，反覆強調語言的內部分野，他指出：

> The novel can be defined as a diversity of social speech types (sometimes even diversity of languages) and a diversity of individual voices, artistically organized. The internal stratification of any single national language into social dialects, characteristic group behavior, professional jargons, generic languages, languages of generations and age groups, tendentious languages, languages of the authorities, of various circles and of passing fashions, languages that serve the specific sociopolitical purposes of the day, even of the hour (each day has its own slogan, its own vocabulary, its own emphases)—this internal stratification present in every language at any given moment of its historical existence is the indispensable prerequisite for the novel as a genre. (Bakhtin 1981: 263)

巴赫金對於社會語言多樣性的分類並不是窮盡式的，而且，隨着語言學對語言研究日益加深，對社會語言的豐富性也多了些瞭解，比如《語言學詞典》（*Encyclopedia of Linguistics*）就這樣界定社會語言（sociolect），“依賴於社會因素的各種形式的語言行為就稱之為‘社會語言’，因為年齡差異、職業差異、性別差異、地方差異等都會產生語言的差異性”（Strazny 2005: 975）。而且，因為國家歷史不同，會出現“高低雙語現象”（diglossia）；[1] 或因國家的語言使用而出現“雙語”（bilingual）乃至“多語”（multilingual）現象；也可能會出現“混合編碼”（code-mixing）或者“語碼轉換”（code-switching）等情形；也可能會因為被殖民經歷而產生洋涇浜語言（pidgin language）、creole 等等，這些都使得社會語言變得複雜和多樣。在此，本文作者並不想、也不可能窮盡所有的社會語言（而且國家與國家、文化與文化的傳統不同，都會有各自特殊的情況），只是想指出，這些豐富的社會語言之間，通過作家藝術地組織到小說中去，會為小說話語帶來“雜語”的效果，而這些雜語又會引起意識的衝突與對話，從而體現巴赫金的語言對話性思想和小說的複調性特徵。所以，托多羅夫（Todorov）指出，“重點不在於其多樣性，而在於它們之間的差異性”（1984: 56）。也就是說，我們觀察小說中的雜語建構，目的並非要窮盡這些雜語的建構模式，而在於發現這些構成雜語的各要素之間處於什麼樣的關係之中。

儘管如此，方言作為雜語進入小說，在不同的寫作語境下發揮着不同的作用，有的時候是影響整個小說的敘事和寫作意圖的重要工具。比如，索米寧（Suominen 2001）就專門討論了芬蘭語

小說 *Tuntematon sotilas*（直譯為《無名的士兵》）如何利用芬蘭語方言來製造雜語效果，而這些方言在英語和希臘語譯本中又是如何處理的。作者指出，“雜語小說內含有雜語聲音。這些聲音被視為意義的載體——社會文化的、個體的、以及互文性的意義，同時還有依附於權力關係的意義。對譯文的分析，主要是看譯文有沒有成功地將作為意義載體的雜語聲音傳達出來”。[2] 作品主要圍繞三個士兵展開：一個說方言的普通士兵、一個處於方言和標準語“中間狀態”（in-betweenness）的士兵、以及一個說標準語的長官。通過描寫他們之間的關係，作者試圖反映不同的社會階層對於發生在二戰期間的“繼續戰爭”（Continuation War）的態度問題。[3] 按照作者的寫作意圖，這是一部表現反戰題材的作品，但是因為出版商的策劃和操作，卻將作品當成一部戰爭題材來處理。這直接影響到譯者對直接體現寫作意圖的方言的處理方式。

結果，在英譯本中，大部分的芬蘭語方言被“標準化”了，同時也導致了士兵和軍官對戰爭態度的差異消失殆盡。索米寧對這種譯法提出了批評，但也表現出一種無奈。因為作者指出，自己曾經徵求過許多芬蘭、希臘語作者的意見，詢問該如何處理方言的問題。得到的答案大多是偏向於利用譯入語中的標準語來進行語言標準化處理，因為這樣既可以增加閱讀的流暢性，也可以避免導致對譯作的批評，因為在譯作中大量使用目的語中的方言，容易使讀者將作品過多地與本地情形聯繫起來。特別是在希臘語譯本中，譯者利用了希臘的地方方言來處理原文中的方言，但卻使人將這個戰爭題材的小說與希臘聯繫起來。所以，作者提出了一個解決的辦法，就是要創造性地利用文學語言，或者說是

創造文學方言，來彌補因方言遺失而帶來的意義損傷。

在巴赫金看來，小說的雜語歸根到底是源於社會性的雜語現象，而小說則是“用藝術方法組織起來的社會性的雜語現象，偶爾還是多語種現象，又是個人獨特的多聲現象”（Bakhtin 1981: 262）。雖然巴赫金很多時候是集中於討論統一民族語言內部的語言社會分層，但並未將討論局限於此。此處，他也強調了多語種現象對於構建小說雜語的重要作用。艾默生（Emerson）指出：

> In fact he viewed the boundaries between national languages as only one extreme on a continuum, at whose other end translation processes where required for one social group to understand another in the same city, for children to understand parents in the same family, for one day to understand the next. (Emerson 1983: 23)

對巴赫金來說，國家語言之間的邊界，其實就是一個連續軸的一端，而在這個連續軸的另外一端，翻譯也是必不可少的。其隱含的意思就是，在這個連續軸的一端，是國家語言之間的差別，也就是說需要跨語際翻譯來維持溝通與聯繫；而在這個連續軸的另外一端，則是同一國家語言內部的語言分野。在這一端，同樣需要翻譯，才能保證同一城市裏的一個社會群體可以理解另一個群體，一家人裏面孩子才能理解父母，而今天的人們才能理解昨天的人們。而無論是統一民族語言內部的社會分層，還是國家語言之間的共存，都是構成小說雜語的重要元素。

如果打破民族語言的界限，在雙語或多語共存的文化中，語

言——特別是雜語——的問題就變得尤為突出。在這方面，梅萊厄（Meylaerts 2006）的研究發現特別能說明問題。梅萊厄所討論的是一個非常特別，也非常有意思的個案。以二、三十年代的比利時為背景，將語言政治與翻譯結合起來，進行討論，翻譯所使用的語言被認為是意識形態鬥爭的場所（the locus of ideological struggle）。當時比利時的情況是，北部的中下階層說佛蘭德語（方言），南部的中下層說法語的一種方言，而全國上層社會是說標準法語。在北部民族語言意識強烈的群體努力之下，力圖使佛蘭德語成為官方語言，或者使其從“被控制語言”（the dominated language）的桎梏中解脫出來。所以，在將北部佛蘭德語作家的作品翻譯成法語的時候，沒有所謂的原語或者說原語文化，因為在雙語或多語社會中，屬於某個層面上的統一文化體內，不應該被視為普通意義上的跨文化翻譯。比如，原文中有語碼轉換的情形，即說佛蘭德語的父親給小孩唱法語歌。在原文中存在雙語並存的“雜語”情形，但是在譯文中，譯者在譯文後面加注指出“原文為法文”。但作者並不認同這一策略，而在自己的修改版本中，刪掉了這一句注釋。原因是，這一注釋不但強化了文本的翻譯特徵，而且保留了原文中的雙語指涉，並暗示了弱勢文化裏的中層是雙語的社群。這在某種程度上強化了法語在大學和軍隊中的國家語言的身份，這在大多數目標語讀者中引起共鳴，喚醒了他們所認同的社會語言等級制度。然而，這恰恰不是原作者所希望的，於是，作者介入翻譯過程，在最終譯本中刪掉了上面的注釋，打破了目標語讀者的認同機制。

由此可見，巴赫金意義上的方言，是構成小說雜語的重要元素，而這種界定已經遠遠超出語言學上對地理方言的關注，把視

野拓展到統一民族語言內部的地理方言、時間方言、性別方言、階級方言等，同時又將多國族語言納入到雜語的研究視野下。無論是同一語言內的社會分層，還是雙語或多語社會中的語言共存，這種語言分野所帶來的意識形態的衝突，成為雜語小說內部張力的主要構成因素，也必然會引起譯者和翻譯研究者的關注。

二、方言、雜語與翻譯

方言在文學作品中是一種非常普遍的現象，其作為雜語建構的素材，在表現作者創作意圖方面發揮着重要的作用。關於方言文學，世界各地都有不同的討論和研究。[4] 然而，譯者在面對方言的時候，往往會採取不同的翻譯策略，這些翻譯策略會對目的語文本的表現力產生直接的影響。這裏，方言並不局限於單一民族語言內部的地域方言，而是巴赫金意義上的語言分層，包括地理方言、性別方言、年齡方言等，也包括國族語言的雜語建構。在具體文學文本中，這種語言的分野可以有多種形式的體現；但無論以何種方式結合在一起形成雜語建構，我們所關注的是這種雜語建構所發揮的作用。雜語的建構模式多種多樣，而方言是比較重要的構成元素。因此，本文在此以方言的翻譯為例，試圖說明雜語在翻譯過程中的處理模式。

在某種意義上，方言可以說隨着語言的不同有不同的表現形式和特徵，在這一點上，與比喻（metaphor）、雙關語（pun）、文化專有項（culture-specific item）有類似的特徵。針對雙關語的翻譯策略，比利時學者德拉巴斯替塔（Delabastita）分別提出過三

種大同小異的分類法，而香港學者張南峰則將這些策略總結為十種翻譯手段（張南峰，2004: 170-171），並且結合中國語境下的翻譯實例，補充了一種策略，即“雙關語譯為類雙關語”（張南峰，2004: 181）。這些分類模式對我們研究方言的翻譯策略分類，有很大的啟發意義。在翻譯構成雜語建構模式的方言過程中，同樣會有各種各樣的策略，如文本內（外）注釋等；但是，在此，本文簡單地討論以下幾種情形：

（1）方言譯成（真實）方言；

（2）方言譯成（虛擬）方言；

（3）方言譯成非方言；

（4）方言在譯文中消失；

（5）原文中沒有，譯文中創造性使用方言；

（6）（虛擬）方言譯成非方言等。

第一種和第二種情形的區別就在於，第一種情形中，譯者會利用目的語中的真實方言來代替原文中的方言，而第二種情形則是譯者無法在目的語中找到類似的方言結合模式，只好利用目的語資源，創造性的製造方言效果。第一種情形比較多見，比如張谷若用山東方言來代替哈代（Hardy）原文中的英國地方方言。第二種情形比較少見，但是也是一種解決方言翻譯的方法。上面提到索米寧（Suominen 2001）在討論雜語與翻譯的時候，也指出了譯者的這種無奈和創造性主體性的發揮。索米寧指出，“還有一個可能就是創造一種文學方言，而不是使用真實的地理方言，譯者正是使用了這一策略來解決方言翻譯的問題”（Suominen 2001）。

方言譯成非方言是非常普遍的現象了。因為很難在目的語中

重新構建類似的雜語建構模式，很多譯者只好採取標準化的翻譯方法，將原文中的方言用目的語中的標準語來代替，從而在一定程度上造成雜語效果的喪失。當然，方言並不局限在地理方言上，也可以體現在語碼轉換上。比如，在庫許納（Kushner）的小說《美國天使》（*Angels in America*）中有這樣一段兩個男同性戀之間的對話：

> Prior: … You know what happens? When I hear it, I get hard.
>
> Belize: Oh my.
>
> Prior: Comme ça. (He uses his arm to demonstrate.) And you know I am slow to rise.
>
> Belize: My jaw aches at the memory.
>
> Prior: And would you deny me this little solace—betray my concupiscence to Florence Nightingale's stormtroopers?
>
> Belize: Perish the thought, ma bébé.
>
> Prior: They'd change the drug just to spoil the fun.
>
> Belize: You and your boner can depend on me.
>
> Prior: Je t'adore, ma belle Nègre.
>
> Belize: All this girl-talk shit is politically incorrect, you know. We should have dropped it back when we gave up drag.
>
> Prior: I'm sick, I get to be politically incorrect if it makes me feel better. (Kushner 1992: 44)

除了其他的坎普（camp）特徵，語碼轉換也是非常重要的坎普元素。在上面這段對話中，兩個美國同性戀者都採用了語碼轉換，

即在他們的話語中有使用法語的現象。對此，哈威認為可以實現一種“幽默的效果”（Harvey 1998: 300），法語語言和文化滲透到英語世界，透露出一種時尚和都市化品質。同時，哈威也發現，在法國文學裏同性戀人物的語言中，也不時說出英語來，道理同上面的例子如出一轍。這樣的一種雜語建構模式，同樣可以理解為標準語與方言之間的結合。但是，在中文譯文中我們發現，這種效果卻因為標準化的翻譯策略而消失了。譯文如下：

> 普來而：你知道嗎？我一聽到這個聲音，就會硬起來。
>
> 貝里茲：哇噻。
>
> 普來而：像這樣。（他用手臂示範。）你知道，我勃起一向很慢的。
>
> 貝里茲：一想到以前，我的下巴就開始痛起來。
>
> 普來而：你這個南丁格爾該不會背叛我，把這些春夢跟他們講吧？
>
> 貝里茲：我才不會咧，小親親。
>
> 普來而：他們如果把這藥換掉，就不好玩了。
>
> 貝里茲：放心好了，我不會出賣你和你硬起來的那根的啦。
>
> 普來而：我好崇拜你噢，親愛的小黑黑。
>
> 貝里茲：我們講這些有的沒有的閨房悄悄話，態度實在是太差了，你知道嗎。早在我們不再扮女裝那個時候，就不應該再這樣打屁了。
>
> 普來而：我生病了啊，如果這些屁話可以讓我舒服一點，我就要講。你講話口氣怎麼跟路一模一樣。（庫許納，1996: 116-7；底線為本文作者所加）

在譯文中我們可以看到，原文中的語碼轉換被標準化的語言模式取代，也即譯者將方言（這裏指法語）譯成了標準語（譯文中的中文），在一定程度上，失去了哈威指出的雜語建構所帶來的效果。儘管如此，譯者還是做了一定的補償。比如，原文中的 ma bébé 和 ma belle Nègre，就被譯者以“小親親”和“親愛的小黑黑”來代替，這些充滿女性話語特徵的稱謂，用在男同性戀者身上，具有深化其同性戀身份特徵的作用。與語碼轉換相類似的混合編碼也同樣是構成社會雜語的重要方式，在翻譯過程中也會出現與上面提到的語碼轉換類似的處理方式。

與“方言譯成非方言”不同的是，有些方言會因為語言和文化的差異，而在譯文中消失，其中就包括（虛擬）方言的翻譯。在某些文學作品中，作者常以虛擬的方言來塑造人物形象。我們以蕭伯納（G. B. Shaw）的《賣花女》（*Pygmalion*）中的對話為例來說明。

> THE FLOWER GIRL. Ow, eez yə-ooa san, is e? Wal, fewd dan y'də-ooty bawmz a mather should, eed now bettern to spal a pore gel's flahrzn then ran awy athalht pyin. Will yə-oo py me f'them? (Shaw 1957: 9)

原文中，作者蕭伯納創造了一個賣花女的人物形象，而這人物的語言在整個作品中發揮着重要的作用。《賣花女》又譯為《匹克梅梁》或者《皮格馬利翁》。故事大意是講語言學教授息金斯（Higgins），讓出身貧寒的街頭賣花女伊莉莎學習了六個月的標準發音後，改造了她的語言和外表，使她在大使館的晚會上，成

了風度優雅的絕色公主。作者以此說明，“英國資產階級的上層社會，同下層社會的勞苦人民相比，只不過有表面上的差別；上層社會人物只是受過一些資產階級的教育，能操較為文雅的語言，此外並無更多長處”（楊憲益，1982: 1）。因此，作者在塑造賣花女這個人物的時候，在語言上是下了一番功夫的。那麼，在譯文中，應該如何處理這種人物語言的雜語式建構呢？楊憲益的翻譯中，明顯沒有體現這種特徵：

> 賣花女：哦，他是您的孩子嗎？哼，您做媽媽的要是管教管教，他也不能把人家的花給糟蹋完了，就跑開不給錢。您替他給錢吧？（蕭伯納，1982: 9）

作者特意安排的這種人物方言，其實是一種視覺方言（eye dialect），是作者虛擬出來的一種方言。然而，在譯文中這種特徵卻沒有體現出來，因此，無法與人物後來經過語言培訓之後的精緻文雅的語言形成對比和反差。

在《賣花女》開始的部分，伊莉莎的語言充滿了方言的味道，譯者也做了相應的處理。比如，在第一幕伊莉莎看到息金斯在記錄她說的話，以為他會控告自己，於是緊張地說：

> THE FLOWER GIRL. I aint done nothing wrong by speaking to the gentleman. Ive a right to sell flowers if I keep off the kerb. (Shaw 1957: 10)

譯者將這句話譯為“咱跟那位先生說句話不能算是做壞事呀，咱

賣花也不犯法，又沒有在人行道上”（蕭伯納，1982: 13），用“咱”代替“I”，彌補了“aint”的語言效果。在伊莉莎沒有被培訓之前，這種譯法可以突出其出身和社會階層。

當然，原文中沒有方言、譯者創造性地使用譯入語中的方言來塑造人物形象的例子也非常多。但是，無論是地理方言、性別方言、年齡方言、職業方言等，在雜語建構模式中發揮着同樣的功能。在《玫瑰玫瑰我愛你》中，作者使用了大量的方言來製造雜語的效果。而這種效果在英文翻譯中是怎樣處理的呢？

三、《玫瑰玫瑰我愛你》中的方言及其翻譯

（一）《玫瑰玫瑰我愛你》：眾聲喧嘩的世界

臺灣作家王禎和被認為是鄉土文學的代表作家之一，其作品大多反映“書寫小地方小人物的生活，特別是處於文化和經濟巨變時期的矛盾價值觀中的小人物”（Wang 1990: 52），如《人生歌王》、《兩地相思》、《嫁妝一牛車》等等，而《玫瑰玫瑰我愛你》則是其重要代表作品之一。小說描寫了越戰美軍到訪臺灣花蓮度假之前，地方上的錢議員、中學老師董斯文、四大妓院的老闆等積極籌畫，要將來自四大妓院的妓女培訓成全能的吧女，從而服務於來臺度假的美軍士兵。小說集中描寫了吧女速成班開班典禮的幾個小時，並且利用倒敘的手法，以意識流的形式回顧了過去幾天裏的籌備過程，所以小說經常被拿來與《尤利西斯》（*Ulysses*）做比較。同時，小說裏還有中學英語教師董斯文教授準吧女學習英文的滑稽場面，經常讓人聯想到蕭伯納的《賣花

女》。

小說的寫作背景非常獨特，以越戰美軍到臺灣花蓮度假為引線，以中學英語老師董斯文組織籌畫吧女速成班為主要情節，地點又設在花蓮市，使得小說的語言呈現出多樣性；[5] 這種多樣性為雜語的建構提供了語言基礎。在翻譯研究中，雜語經常與語言多樣性聯繫在一起，比如梅萊厄（Meylaerts 2006）、陳德鴻（Chan 2002）、索米寧（Suominen 2001）等等，都是從這個角度來研究雜語與翻譯之間的關係問題。梅萊厄明確指出，“雜語是指在文本中存在外來諺語或社會、地區、歷史等的語言變體，雜語也指文學語言多樣性”（Meylaerts 2006: 86），並追溯雜語這個概念來源於巴赫金的〈長篇小說的話語〉。所以，語言多樣性與雜語之間的關係是非常緊密的，也就是說語言多樣性（無論以何種形式）是雜語構建的語言基礎。

發生在七十年代的臺灣這樣一個歷史和地理背景下，[6] 使得故事在語言上形成了獨特的風格。臺灣歷史上經歷了幾次被殖民的歷史和政權輪替，如 1895-1945 的日據時期，1945 年後的國民黨統治，美國的軍事和經濟影響，甚至於後來的民進黨掌管政權，這一切都導致了臺灣語言和文化的多樣性和混雜性。蕭錦錦指出，“臺灣地區以台語為基礎語言，[7] 日據時代，有了五十年的日語歷史，光復後，全面學習國語，近十年來，英語變成相當時髦的話，要一個大學生說一百句話，中間不夾一點英語，似乎有些困難”（王禎和，1986: 281-2）。這種本土語言與英語的混合編碼，在香港也表現得非常突出。香港近百年的殖民史，使得本地的廣東話在很大程度上與英語形成了混合編碼的語言情形。在香港，混合編碼和編碼轉換是非常普遍的語言現象，而前者尤

為突出。

丘彥明在對王禎和的訪談中提到，王禎和在語言的運用上也做了另一種突破，“中外古今”語彙夾雜的使用，很鮮活的讓時空明確地顯現出來。對此，王禎和說，“我儘量把一個時代的語言寫出來。那時臺灣用的日語、台語、國語和現在一九八三、八四的國語、台語不一樣，我想用對照的方式也許可能使那個時代鮮明的給推演出來”（1986: 274）。王禎和的這種語言觀，在某種程度上，與巴赫金不謀而合。巴赫金所反對或批判的，正是那種獨白式的文學表現模式，所有的文學語言都是官方的標準語，這並不能反映真實的生活和真實的文學語言狀態。但是，這並不是說，只有現實主義的文學作品，才能利用豐富的雜語和真實的語言形態，只能說在現實主義文學作品中，這種雜語的語言模式表現得更加突出，並且發揮着難以替代的作用。同時，巴赫金在〈長篇小說的話語〉中也討論了各種不同的組織和引入雜語的模式，而這些模式並非完全體現在單一文學作品中。

除了台語、國語、日語和英語（源自美國對臺灣在政治、經濟、軍事上的挾制），小說在語言上還有更複雜的表現。蕭錦錦指出：

> 除了這四種正常的語言外，另有臺灣化的日語、英語化的國語、台語，以及最常聽到的“臺灣國語”和客家話，在這麼複雜的語言環境中，王禎和獨到的發現語言的微妙功用——每兩種不同語言的過渡，都含有人生、時代的諷刺，而且是恰到好處的諷刺，一點都不牽強，如“美軍就是美金”，“年歲有澀澀”，“不然就要糟糕一碼事”，“擔，你嗎好啊！”等等。（王禎和，1986: 282）

如果說“四種正常的語言”已經使得小說成為眾聲喧嘩之地，那麼“臺灣化的日語、英語化的國語、台語”這些各具特色的語言變體，使得文本更加豐富、立體。

上面提到的“兩種不同語言的過渡”，其實就是一定意義上巴赫金所說的雜語的建構模式，可以表現為不同人物的不同語言（如董斯文用國語而阿恨則用台語），也可以表現為同一人物在不同語境下使用不同的語言（如錢議員在競選時和當選之後分別使用的台語與國語）；當然，更多的時候是表現為雜合建構的模式。巴赫金強調了文學語言的雜語特徵，其分析的例子也多局限於統一民族語言內部的雜語模式。但是，我們並不一定要將雜語局限在單一民族語言之內，陳德鴻（Chan 2002）的個案研究在這方面就非常有啟發意義。對於臺灣這種語言具有多樣性的社會形態，同樣可以用巴赫金的雜語來加以分析；而且，可以更加立體和多樣地體現巴赫金的雜語觀。另外，巴赫金的雜語常用來分析不同職業、年齡、社會階層、教育背景等的話語主體在言說方式上的差異，但是，我們並不能由此就將巴赫金的雜語觀局限於簡單的語言多樣性的分析，更加應該強調這種“不同語言的過渡”，因為，這種過渡“含有人生、時代的諷刺，而且是恰到好處的諷刺，一點都不牽強”。同時，這些雜語的建構模式，在很大程度上可以體現各種意識形態之間的爭鬥。

小說通過這些語言形式的創造性運用，製造了強烈的諷刺和狂歡效果，以至於王德威將《玫瑰玫瑰我愛你》這部小說定位為“第一部完全版的中國式狂歡喜劇”（Wang 1990: 52）。同時，在談到這部小說的語言的時候，王德威認為小說語言：

> 是一個混合體，包括臺灣俚語、國語陳詞濫調、洋涇浜英文和不成句的日語。同時，通過互文性手法，將愛國宣教和宗教典籍（patriotic propaganda and religious sciptures）帶入小說中，但卻被滑稽可笑的模仿，帶有了低俗和色情的含義。而所有這些修辭手法營造了一種發音不和諧的話語氛圍（a cacophonous discourse），不但譏諷了臺灣文化的混雜本質，也強調了對正統獨白式小說寫作的不認同和反抗。（Wang 1990: 53）

這麼複雜的語言多樣性，必定會給翻譯帶來巨大的挑戰。這部小說於 1998 年由蔣經國基金出資，由美國漢學家、翻譯家葛浩文（Howard Goldblatt）翻譯，並在美國出版。在英譯本的前言裏，葛浩文專門提到了這部小說的語言特徵，指出“王禎和大量使用方言和低俗的模仿語言（kitsch）來製造強烈的喜劇效果，從而諷刺臺灣的多層文化（multifarious）特徵；同時，也強調了對傳統正統小說獨白模式的反叛”（Goldblatt 1998: viii-ix）。另外，葛浩文也感歎於小說的語言之獨特，認為小說中：

> 對台語、國語（國民黨政權推行的以北京話為基礎的語言）、日語、以及美國英語的有意“濫用”，製造多重的喜劇效果。同時，小說還利用了很多語言形式：如頭音的交換（spoonerism）（常見於漢語中），切割混雜的外國語現象（mangled foreignism, malapropisms），還有大量其他的語言怪體，源源不斷地從王禎和小說人物的嘴裏湧出來。（Goldblatt 1998: viii-ix）

這些語言的形式，對譯者提出了很大的挑戰，以致於譯者葛浩文

認為是“難以克服的挑戰”（formidable challenge），並且需要適量的解釋才能保證讀者的理解。同時，所有這些豐富的語言形態，都使得這部小說成為了眾聲喧嘩之地。其中，議員、中學教師、妓院老闆、妓女、同性戀取向的醫生等等，都在不同的場合發出自己的聲音；並且，作者借助敘述視角的轉換，還特別加強了小說的社會廣度。這些社會語言的形態，通過作家之筆，見諸於書面之上。所以，《玫瑰玫瑰我愛你》不愧為一部體現語言多樣性和社會多面性的雜語小說，是一個完全的雜語世界。面對這些豐富的語言形態，譯者如何處理呢？怎樣才能最好地體現這些語言特徵，並在譯文中保持、改造、或者補償這種雜語性，這些都成為譯者和翻譯研究者應該關注的問題。

（二）《玫瑰玫瑰我愛你》中的方言及其翻譯

本文在前面介紹了王德威、蕭錦錦、丘彥明、以及譯者葛浩文對王禎和這部小說語言的觀察和討論，從中可以看出，方言在王禎和的這部小說裏發揮着重要的作用，以至於這部小說被認為是一部台語寫作的典範。正如索米寧（Suominen 2001）指出的，在《無名的士兵》中，“每個士兵以自己個性化的芬蘭語發聲方式，發出自己的聲音”；[8] 同樣的，在王禎和的《玫瑰玫瑰我愛你》中，作者塑造了董斯文、錢議員、大鼻獅、阿恨、惲醫生等幾個主要人物，而這些人物各有自己的聲音，同樣是有各自獨特的語言方式。小說中，大多數的人物都是以台語作為自己的語言，比如大鼻獅、阿恨，甚至錢議員；而其他的主人公比如董斯文、惲醫生等，則主要以國語為主要用語。在這部小說中，出現了幾個章節的敘述視角的轉換，從開頭的董斯文的敘述視角，逐

漸轉移到大鼻獅、阿恨等人的角度進行敘事，[9] 也同樣引起敘述語言的轉換。這種轉換是作者有意為之的，王禎和指出：

> 這部小說本來是只從董斯文老師的觀點來寫，後來就覺得這個觀點太拘束了，把自己的筆綁得死死的，想想何必如此“自我虐待”呢！於是有好幾個章節就完全不理會那個愛放屁的董斯文的觀點了，改換成龜公大鼻獅、妓女阿恨的對話來發展情節。這樣一放開來編，覺得好像小說比較活潑，而且具多面性。（王禎和，1986: 272-3）

除了敘述角度的轉換導致敘述語言的轉換，也即從董斯文的國語敘述轉換成大鼻獅等人的台語敘述，小說在人物形象塑造和故事情節發展上，也主要是依靠人物語言。而按照巴赫金的說法，“小說引進和組織雜語的另一種形式，也是一切小說無例外全都採用的形式，這便是主人公的語言”（巴赫金，1998: 99）。所以，無論是主人公的直接引語，還是通過敘述視角轉換所帶入的假託敘述者的話語，都以不同的方式引入和組織雜語進入小說，而這種雜語性特徵主要是體現在國語和台語之間的過渡上。我們上面介紹過，臺灣的歷史，特別是在近代歷史上經歷了的殖民史和政權輪換（特別是民進黨打敗國民黨而成為執政黨），這些直接影響到在臺灣的語言政策；而這些語言政策又直接地體現在普通大眾的話語中；同時，也間接地表現在文學作品中。正是在這個意義上，我們來思考和觀察《玫瑰玫瑰我愛你》中的方言問題。

《玫瑰玫瑰我愛你》被認為是鄉土文學的代表，而鄉土文學

與台語緊密地聯繫在一起。由於臺灣的殖民史和政權輪替的歷史，語言在臺灣成為一個敏感的話題。由此，語言成為意識形態鬥爭的場所，而長篇小說則為這種鬥爭提供了廣闊的戰場。巴赫金也充分認識到這一情形，並將此與雜語聯繫在一起。所以，以王禎和的這部小說《玫瑰玫瑰我愛你》來討論臺灣的方言問題，就顯得特別有意義。特別是王禎和將台語與國語、日語、英語等其他語言放在一個空間裏，讓它們相互映照，積極對話與爭鬥，形成眾聲喧嘩的熱鬧局面。可以說，方言在這部小說裏發揮着難以替代的重要作用。

如果說，在大陸語境下我們比較好界定普通話與方言之間的界限的話，那麼，在臺灣語境下，這種關係就比較難以界定。關於台語作為方言與官話之間的關係，林央敏認為：

> 這個觀念無論從什麼觀點來看都錯誤，[10] 因為全世界所有的“國語”、“官話”都是由“方言”而來，只因這些“國語、官話”有經過政治性人為的控制而顯得較有“標準”模式而已。事實上任何方言（語言）只要透過大眾使用和制式改造都可以成為“標準語”。如果說一種語言的語和文都要有某個程度的標準化才能成為“國語”，台語絕對可以當“國語”。（林央敏，1997: 90）

撇開這種論調所表現出來的政治立場不論，僅在語言層面上看，這種觀點還是有道理的，而且巴赫金也認同這種觀點。巴赫金認為“規範語也是雜語的一種”，其隱含的意思就是，在雜語建構這個層面上來說，規範語，或者說標準語（國語），同其他方言一樣是處於政治意識形態之外的中立的實體，也就是說二者

在雜語功能上來說是平等的。所以，沒有所謂高與低的差別，也沒有“雅”與“俗”的分野。[11] 當語言進入意識形態領域，就有了這種種人為的區別。

在臺灣這個語境下，各種語言處於什麼關係之中呢？國語和台語成為黨派和意識形態鬥爭的焦點和工具，甚至出現了將台語與是否愛臺灣聯繫起來的做法。[12] 在臺灣，佔總人口 75% 以上的大多數群體，其母語（或者準確說是族語）是台語（福佬話），然而這個大多數卻沒有能改變台語作為次等方言的命運，這一切被歸咎於內戰後敗退到臺灣的國民黨政權的極端語言政策。1945 年，二戰結束，戰敗國日本撤出臺灣，由當時的國民黨政府接管臺灣。在語言政策上，出於殖民動機，國民黨政府表現出比日本更加急不可待的一刀切，在臺灣大力推行國語，而不考慮臺灣當時大多數族群的語言基礎。在學校裏，學生講台語會被恥笑和受到懲處。[13] 更有甚者，電視臺使用台語竟然與“共匪”陰謀聯繫在一起。[14]

所以，儘管小說也引入了英語、日語和其他形式的語言變體，來製造雜語的效果；但是，小說的主體還是以台語和國語之間的過渡與穿插比重最大。這在一定層面上反映出作者對台語與國語的態度問題，而這種態度在小說語言中有跡可尋。小說的第一主人公董斯文，是受過大學教育的中學教師，所以自然成為國語的代言人，雖然他也會不時說出幾個台語或英語詞彙，但整體上，他是被當成官方語言的代言人來塑造的，而他的國語也就成為他與其他妓院老闆、妓女一類下層人士的分水嶺。因此，董斯文的國語反覆被強調和調侃。比如：

> “的相反。”董斯文又快速地用國語答，臉上有明顯的得意。
>
> “什麼？！”表示聽不懂，矮仔姬猛搖首，大金墜子又大打秋千，兩顆夢露底奶也招搖起來，很招眼。
>
> 其他也有好幾個來賓異口同聲起：“老師，你是在講什麼？”（王禎和，1986: 13）

王禎和刻意用這種反差來製造國語與台語之間的分裂，既是語言上的隔膜，又是語言主體之間的陌生感。在小說的其他部分，有許多地方都是這樣處理的，即將國語與台語置於對立與矛盾之中，同時又表現出對主流意識形態話語，也即國語的狂歡式嘲諷。當然，這種嘲諷被轉嫁到言語主體董斯文身上。比如，董斯文打電話給大鼻獅，交待更換講座地點的事情，並且，要求找的小姐都要 155 公分以上，因為怕被人誤以為是小孩。

> “那個事？那個事是啥咪（什麼）？”話出去後，大鼻獅方頓悟過來，不禁啊啊笑起來。“原來老師是在講那個哦！老師用國語跟我講，有些話，我實在聽莫，甘那（好像）鴨子聽雷！”
>
> “阿尼基，你一定要注意，千萬不要低於一五五的標準——”董斯文仍舊用國語在不斷繼續說着，仿佛不講國語，他就沒辦法說正經的。“你曉得嘛！阿尼基，女人要是個頭太小，你講她——她的那個，也必然很小。你講，是不是如此？——那個要是太小，你講那怎麼能夠和美國人的那個 match（匹配）呢！？Size（尺寸）根本不對攏嘛！你說是不是如此——”
>
> 大鼻獅真的是鴨子聽雷，一句也聽不明白。“老師，你在講啥講啥？我一句都聽莫？是唔是又在講美國話——唔是美國話啊——哦

> －哦－哦－哈－”他笑得鼻翼大展飛向耳際來，嘴上的金煙斗也跟着亂顫亂抖，樂不可支了。（王禎和，1986: 71）

在此，作者還是製造國語與台語之間的距離感，用大鼻獅的話語來強調，“老師用國語跟我講，有些話，我實在聽莫”。大鼻獅的台語，在“聽莫”、“啥咪”、“甘”這些詞彙上體現出來，從而與董斯文的國語區別開來，也才製造了雜語的效果。緊接着，董斯文仍舊用國語說話，“仿佛不講國語，他就沒辦法說正經的”。這是對董斯文為代表的國民黨主流意識形態的諷刺。說些男盜女娼的勾當，竟然也是正經事兒？而且，對董斯文語言的塑造上，又特別添加了一層特徵，那就是混合編碼的運用。董斯文在大學是讀外國文學的，而且現在又是中學教師，但這些設計都是特別安排的。國語可以代表國民黨的語言政策，即將國民黨所推行的以北京話為基礎的國語定為官話，也成為大學、部隊等的官方用語。董斯文恰好可以代表這種官方立場，與說台語的妓院老闆，形成鮮明對比。而混合編碼，則反映出國民黨政權與美國的關係，也即國民黨政權受美國經濟、軍力、以及政治的挾制。

然而，這些雜語所帶來的效果，在英譯本中並未能得到完全的傳達。譯文如下：

> “You-know-what? Just what is you-know-what?” As soon as he said it, Big-nose Lion got the picture, and burst out laughing. “Oh, that’s what you’re talking about! When you speak in Mandarin, teacher, I don’t always get every word, like a duck hearing a clap of thunder!”

> "This is important, Aniki, absolutely no one under five-one—" Dong Siwen continued to make his point in Mandarin, as if that were the only way he could keep the conversation on the proper level. "You know, don't you, Aniki, that if a girl's too small, then her you-know-what will also be too small. I've got that right, haven't I? And if her you-know-what is too small, how's she going to *match* up with an American? The *size* will be all wrong, am I right or aren't I?"
>
> Big-nose Lion was indeed a duck hearing a clap of thunder—for he didn't understand a word the man said. "What's that you're saying, Teacher? I can't understand a word of it. Are you talking American again? You're not? Oh-oh-oh-ah—" By now he was laughing so hard that his nostrils seemed about to fly up into his ears, and his pipe nearly bounced out of his mouth. (Wang 1998: 47)

首先，人物之間在語言上的差異消失了。大鼻獅的台語方言與董斯文的國語，在英文中沒有明顯的區別，並不能達到原文製造的對比的效果，而這種通過主人公語言所構建的雜語效果，是巴赫金所指出的小說進入雜語的主要形式之一。其次，董斯文在講這些"正經事兒"的時候，還是礙於斯文，不便公開討論性器官的大小以及男女的匹配問題。在原文中，通過混合編碼的形式，將這些意思表達出來，凸現了董斯文的虛偽，也諷刺了以董斯文為代表的國語和國民黨政府的虛偽。但是，譯文中卻沒有對此進行補償，只是用簡單的斜體將混合編碼標出，並不能傳達原文所帶來的諷刺效果。最後，混合編碼所表現出來的國民黨和美國之間的曖昧關係，也沒有在譯文中很好地體現出來。

由此，我們可以看到，雜語建構在這部小說中發揮着重要的作用，而作為意識形態鬥爭場域的語言，也就成為體現雜語的主要模式。台語與國語這兩大主要語言，在臺灣的日常生活中扮演着不同的角色，在政治生活中，就更加凸現出兩者之間的意識形態衝突，因此國語和台語在不同的場合發揮着不同的作用。比如，我們來看下面這兩段話，都是錢議員的語言：

（1）“派出所已經把人放出來啦！那好，那好——謝什麼呢！哦，對，我要跟你講：我們已經跟他們談妥了，從下個月十五號開始，公路局就不再跑北埔機場這條線。這條線全讓出來，全讓出來給海星客運——一批預算案等着要通過呢？怎麼敢不肯？——對啦！你要叫阿豬他們趕快做好準備，下個月十五號馬上就到，剩沒多少天咧！——至於花東公路的路權，我跟縣長——”，這時除了經營自己的貨運企業公司也兼任海星客貨運公司名譽董事長的錢議員忽地降低聲音到像在呢喃情話，怕旁人聽到。（王禎和，1986: 34）

（2）“對，褪縛體誰都敢，莫稀罕！如果——”錢銘雄摸着光裸底腹將聲音壓得很低，全場的聽眾都睜眼盯他，連氣都不敢吭似地。“如果是——”聲音又壓低一點，光裸底手向下身一指。“如果是脫褲囊（脫光褲子），是唔是稀罕？是唔是誰都敢？”（王禎和，1986: 45）

我們來看這兩段話有什麼差別。同樣是出自錢銘雄，但卻明顯是兩套話語體系。第一段話是錢銘雄作為議員在電話裏跟人談“公事”，其實是他自己的私事，因為他本人“除了經營自己的

貨運企業公司也兼任海星客貨運公司名譽董事長”。在這樣一個官方場合裏，錢議員自然是用官話，即國語。然而諷刺的是，這分明是一個假公濟私、中飽私囊的貪官污吏，卻堂而皇之地用“官話”在發號施令。再來看第二段話，這是董斯文回憶錢銘雄競選議員時，在拜票場合向民眾講話時的情形。與第一段話中錢銘雄當選後的“官話”／“官腔”相比，競選現場的錢銘雄卻說着地道的台語。而且，是在給一幫起哄的民眾講自己的敢作敢為，說自己既敢褪縛體（即脫上衣），也敢脫褲囊（即脫光褲子），十分滑稽。這一段話的諷刺效果也非常明顯：一是揭露錢銘雄利用台語拉取選票，二來諷刺臺灣民眾在民主政治初期的不成熟。臺灣人口中超過 75% 的是“台語系臺灣人”（林央敏，1997: 15），而台語也成了政治候選人在拜票過程中最多使用的語言，因為這樣可以爭取到佔人口多數的“台語系臺灣人”的選票。我們可以在錢銘雄第二段講話的周圍，聽到許多台語系臺灣人選民的聲音，如：

> 台下一個小女生驚訝得非常地叫：“爸！你看他沒穿衣服！哎唷！羞羞（ㄑㄧㄡ）臉！羞羞臉！”（王禎和，1986: 42）

> 選民看不到好戲，又喧騰起來。
> “當眾換衫，算什麼敢嘛！”
> “是啊！在公眾面前褪衫，什麼稀罕！大家都敢的事情嘛！”
> “也不見得哦！你看查某人敢唔敢？”
> “那裏不敢！跳脫衣舞的查某，連三角褲都脫得離離離！”
> 婦女選民立時有人尖聲叫罵：

“夭壽呢！”

“膨肚短命呢！講這款話。”

“講這款話，唔怕將來生個沒屁股的兒孫！”（王楨和，1986: 43-44）

諸如此類的台語圍繞在錢銘雄話語的前後。所以，在臺灣這樣一個雜語化的社會裏，在政治語言上，要選擇在合適的場合講合適的話。這種情形在今天依然普遍存在。筆者曾經赴臺灣觀察當地的民主選舉，在大大小小的造勢場合，聽到了令選民心潮澎湃的台語演講。最近，又讀到了一篇網上報導，題目就是“馬台語不好不利選情，吳育升：閩南沙文主義”。文中寫道：“馬英九英語比台語好，不利選情？國民黨籍立委吳育升今天表示，語言使用應尊重每個人，如果用閩南語檢驗，是標準的閩南人沙文主義；呼籲民進黨不要因馬英九英文好就吃味，如果以語言角度，臺灣南部不投給馬英九，他覺得不公道，也很悲哀”。[15]

由此我們可以看到，在同樣一個角色的語言裏，也會出現這種差異巨大的現象，這本身就出自不同的社會交際目的和功能。在《玫瑰玫瑰我愛你》這部小說裏，對台語與國語之間的關係，作者自有他的見解。根據上面的語言現象分析，我們不能否認雜語的建構對於作者的寫作意圖是有直接的關係的，而這種寫作意圖，必然與作者的語言、文化身份和政治認同是緊密聯繫在一起的。對於這類政治語言的雜語現象，林央敏總結指出，“選前講台語，選後講國語；選戰用台語，當選用國語”（1997: 83）。錢銘雄當選後，根本就沒有兌現自己的承諾，而且假公濟私地謀取個人私利，這在後面得到回應。在小說最後的第十六章裏，錢

議員被邀請上台給培訓班的吧女們講話，在他開口前，卻被一個山地小姐質問：“錢議員！競選時你說你會叫客運的車走我們豐濱，講得像真正的。嘿！到現在連個車影都沒見到！”（王禎和，1986: 219-220）

如果台語和國語的雜語式結合，對於塑造一個典型人物這麼重要，對於表現作者的寫作意圖這麼明顯，那麼在翻譯的過程中，就應該注意到這一現象，並且尋找合適的方式來再現或補償這種語言模式。然而，在譯文中，我們看不到這樣的痕跡。譯文如下：

> (1) “The police have dropped the charges? Good, good. Why thank me? Oh, by the way, I’ve been meaning to tell you I’ve already worked things out. Starting from the fifteenth of next month, the highway buses will discontinue service to Beipu Airport. Seastar Trucking will have the route all to itself—they still want to pass a budget, don’t they? They wouldn’t dare say no now—right! Tell Ah-zhu and that bunch to get everything ready. The fifteenth of next month will be here before you know it, not many days left—as for the right to operate on the Huadong Highway, the county head and I—” From here on, Councilman Qian, who ran his own trucking company and was honorary chairman of the board of Seastar Trucking, dropped his voice so low it sounded like pillow talk, words not meant for other’s ears. (Wang 1998: 22)

> (2) “Stripping to the waist is no big deal, you got that right. Now if

> …" Qian Mingxiong lowered his voice and rubbed his belly. Everyone held their breath as their eyes snapped open wide. "Now if …" he repeated in a low voice as he pointed to his waistline. "Now if I took off my pants, would that be a big deal? Is that something anyone might do?" (Wang 1998: 29)

從譯文我們可以看出，這兩段話在修辭和詞彙上，都沒有明顯的差異，所以原文中這兩段話語之間的差異被抹去了。而這種雜語帶來的語言差異、文化焦慮和政治暗示，也隨着譯文的不加區別，而被遮蔽了。從這個層面上來說，這個雜語所帶來的效果還是應該引起譯者的關注，並且應該尋找合適的方法來解決這個問題。如果說，在原文語言文化語境下，存在這樣的獨特現象，那麼在很難找到匹配的譯入語文化語境中，應該通過注釋或者其他的方式，將這層意義帶給讀者，從而才能保證小說的充分理解和把握。

所以，在一個語言呈現多樣性的社會裏，各種族群的語言與官方語言之間，很多時候會呈現出非常特殊的關係，而這種關係背後則是非常激烈的意識形態衝突。正是這種語言的多樣性和意識形態的爭鬥，為小說語言提供了豐富的素材，使得作者可以游刃於這些雜語之間，充分發揮這些雜語的建構力量，在文學作品中表現出強烈的社會關注和反思。同樣，在翻譯的過程中，也應該關注這種語言多樣性帶來的挑戰，特別是當語言與語言之間呈現對抗時，要利用各種方法和譯入語資源，來表現、再現、維護、或補償這種對抗的張力。從上面的例子我們可以看出，在小說《玫瑰玫瑰我愛你》中，台語作為方言與作為官方主流話語的

國語之間所體現出來的張力，是非常成功的將社會雜語組織進入小說的個案，在翻譯中，不能不重視這種雜語的效果。

四、結　語

本文借助巴赫金的雜語這一概念，通過語言的社會分層，來凸現小說中的意識形態衝突與鬥爭。在巴赫金意義上，方言不僅僅局限於地理方言，還包括其他社會方言，甚至是雙語或多語共存。但這種語言的社會分層，被作家以藝術的方式組織和引入小說，構成文學意義上的雜語。理解小說的這一特徵，對於我們把握小說內涵和進行文學翻譯有很大的作用。本文從對文學方言的翻譯模式進行了簡單的總結，並結合王禎和的《玫瑰玫瑰我愛你》這個個案，詳細探討由社會方言所建構的文學雜語在小說中所發揮的重要作用。但當我們說社會方言的時候，並不是以僵化的眼光看待它，正如巴赫金指出的：

> 各種局部語進入文學，接近於標準語之後，自然在標準語的土壤上便要失去封閉的社會語言體系的性質；它們會改變形態，實際上已不再是過去的局部語言。然而從另一方面看，這些局部語進入標準語，卻在其中保留着自己作為局部語的彈性，保留着異語的味道，因而也給標準語帶來了變化。標準語也不再是原來的樣子，它也不再是一個封閉的社會語言體系了。（巴赫金，1998: 75）

總之，由社會方言所建構的雜語是小說的一個非常典型的特

徵，這種特徵對小說這一體材來說是非常重要的。而對於譯者來說，要把握和處理這種文學語言現象並不是一項簡單的任務，有時還是難以克服的。本文就這個問題進行了初步的探討，特別是在問題的發現和本質探究上做出了貢獻，但對於具體問題的解決，還需要學界給予更多的關注和研究。

注　釋

[1] 這裏要簡單澄清一下 diglossia 與 bilingual 的區別。前者通常是指一個社會語言區內，同時存在兩種同源卻又有差異的語言現象，通常表現為 H-variety 和 L-variety 之分別，比如中國晚清民初的白話文運動過程中出現的古體漢語與白話文的共時存在；而後者則指統一的社會語言區內，同時存在兩種或以上的民族語言，如香港的兩文三語，兩文就是英語和漢語，屬於 bilingual，而普通話和廣東話就屬於 diglossia 的關係。

[2] 網上電子雜誌，相關內容請參見：<http://www.eng.helsinki.fi/hes/Translation/volume1.htm>。

[3] 繼續戰爭（Continuation War），是發生在二戰期間芬蘭與前蘇聯之間的第二次戰爭，從 1941 年 6 月持續到 1944 年 9 月。

[4] 參見 Snow（1991），Blank（1996），Ch'ien（2004），Lewis（1997），North（1994）等人的論述。

[5] 王禎和本人畢業於臺灣大學外文系，曾任中學英語教師，也曾供職於航空公司和臺灣的電視臺。而小說中的人物董斯文，也有類似的經歷；王禎和是花蓮市人，而且他的小說大多也都以花蓮為背景。

[6] 越南戰爭（1959-1975）提供了一個時間座標；另外，小說中很多地方也有時間提示，表明故事發生的時代背景，比如小說中會出現以括弧的形式標出的詞彙發展嬗變的過程。

[7] 關於台語的名實，參見林央敏（1997: 58-65）。

[8] 同樣見網上連結：<http://www.eng.helsinki.fi/hes/Translation/

volume1.htm>。

[9] 比如第六章整個就是通過大鼻獅的口吻在敘事，而且是隱含了的對話模式。整個敘事的口吻是大鼻獅在跟自己的情婦阿恨講電話，但是卻看不出任何直接引語的痕跡，也沒有阿恨的話語直接出現在小說中。參見王禎和的《玫瑰玫瑰我愛你》。

[10] 這個觀點是指當時社會上有人主張：正式場合用國語（北京話），在家用台語，如此最自然。見林央敏（1997: 90）。

[11] 林央敏（1997: 94-5）。

[12] 林央敏（1997: 102）。

[13] 林央敏（1997: 31）。

[14] 林央敏（1997: 33）。作者曾經舉例說明，在 1972 年的時候，外省人萬年立委穆超質詢電視台選擇方言節目刊登方言廣告，是有日匪或中共匪諜滲透操縱。王禎和曾經供職於電視台，而且這部小說的寫作年代是 1978 年，所以很難否認其寫作動機與這些語言意識形態問題之間的千絲萬縷的聯繫。

[15] 星島環球網 2007 年 6 月 4 日。詳細內容請參看網上資料：<http://gb.chinareviewnews.com/doc/1003/8/1/3/100381317.html?coluid=7&kindid=0&docid=100381317>。

參考文獻

Bakhtin, M. M. (c1981). *The Dialogic Imagination: Four Essays*. Ed. Michael Holquist. Trans. Caryl Emerson and Michael Holquist. Austin: University of Texas Press.

Barbaresi, Lavinia Merlini (2002). "Text Linguistics and Literary Translation". In *Translation Studies: Perspectives on an Emerging Discipline*. Ed. Alessandra Riccardi. Cambridge University Press, pp. 120-132.

Blank, Paula (1996). *Broken English: Dialects and the Politics of Language in Renaissance Writings*. New York: Routledge.

Ch'ien, Evelyn Nien-Ming (2004). *Weird English*. Cambridge: Harvard University

Press.

Chan, Tak-hung Leo (2002). "Translating Bilinguality: Theorizing Translation in the Post-Babelian Era". *The Translator: Studies in Intercultural Communication* 8.1: 49-72.

Emerson, Caryl (1983). "Translating Bakhtin: Does His Theory of Discourse Contain a Theory of Translation?" *University of Ottawa Quarterly* 53.1: 23-33.

Goldblatt, Howard (1998). "Translator's Preface". *Rose Rose I Love You*. By Wang Chen-ho. New York: Columbia University Press.

Hardy, Thomas (1974). *Tess of the d'Urbervilles*. London: Macmillan.

Harvey, Keith (1998). "Translating Camp Talk: Gay Identity and Cultural Transfer". *The Translator: Studies in Intercutural Communication* 4.2: 295-320.

_____ (2000). "Gay Community, Gay Identity and the Translated Text". *TTR* 13: 137-165.

Holquist, Michael (1990/2002). *Dialogism: Bakhtin and His World* (2nd ed). London; New York: Routledge.

Kushner, Tony (1992). *Angels in America: A Gay Fantasia on National Themes* (Part One: Millennium Approaches). London: Royal National Theatre and Nick Hern Books.

Lewis, Richard O. (1997). *Conventional Functions of Black English in American Literature*. San Francisco, CA.: Austin & Winfield.

Meylaerts, Reine (2006). "Literary Heteroglossia in Translation: When the Language of Translation is the Locus of Ideological Struggle". In *Translation Studies at the Interface of Disciplines*. Ed. João Ferreira Duarte, Alexandra Assis Rosa and Teresa Seruya. Amsterdam: John Benjamins, pp. 85-98.

North, Michael (1994). *The Dialect of Modernism: Race, Language, and Twentieth-Century Literature*. New York: Oxford University Press.

Pai Hsien-yung (1990). *Crystal Boys*. Trans. Howard Goldblatt. San Fransisco:

Gay Sunshine Press.

Snow, Donald Bruce (1991). *Written Cantonese and the Culture of Hong Kong: The Growth of a Dialect Literature*. Indiana: Indiana University.

Strazny, Philipp, ed. (2005). *Encyclopedia of Linguistics*. New York: Fitzroy Dearborn.

Suominen, Marja (2001). "Heteroglot Soldiers". *The Electronic Journal of the Department of English at the University of Helsinki*. Vol. 1, 2001. Online link: <http://www.eng.helsinki.fi/hes/Translation/heteroglot_soldiers.htm>.

Tabakowska, Elzbieta (1990). "Linguistic Polyphony as a Problem in Translation". In *Translation, History and Culture*. Ed. Susan Bassnett and André Lefevere. London and New York: Pinter Publishers.

Todorov, Tzvetan (1984). *Mikhail Bakhtin: The Dialogical Principle*. Trans. Wlad Godzich. Minneapolis: University of Minnesota Press.

Wang, Chen-ho (1998). *Rose Rose I Love You*. Trans. Howard Goldblatt. New York: Columbia University Press.

Wang, Der-wei David (1990). "Radical Laughter in Lao She and His Taiwan Successors". In *Worlds Apart: Recent Chinese Writing and Its Audiences*. Ed. Howard Goldblatt. Armonk, N. Y.: M. E. Sharpe.

巴赫金（1998），〈長篇小說的話語〉，《巴赫金全集》（第三卷），石家莊：河北教育出版社。

崔國良（1993），《曹禺早期改譯劇本及創作》，瀋陽：遼寧大學出版社。

哈代（1957），《德伯家的苔絲》，張谷若譯，北京：人民文學出版社。

韓子滿（2002），〈試論方言對譯的局限性——以張谷若先生譯《德伯家的苔絲》為例〉，《解放軍外國語學院學報》4: 86-90。

林央敏（1997），《臺灣文化釘根書》，臺北：前衛出版社。

王禎和（1986），《玫瑰玫瑰我愛你》（三版），臺北：遠景出版事業公司。

楊憲益（1982），〈譯者序〉，載《賣花女》，蕭伯納著，楊憲益譯，北京：中國對外翻譯出版公司。

姚一葦（1986），〈我讀《玫瑰玫瑰我愛你》（代序）〉，載《玫瑰玫瑰我愛你》（三版），王禎和著，臺北：遠景出版事業公司。
張南峰（2004），《中西譯學批評》，北京：清華大學出版社。
朱立立（2002），〈純化與雜化：臺灣現代派小說的語言策略〉，網上資料：<http://guxiang.com/xueshu/others/shijiao/200212/200212260007.htm>。

作者簡介

李波，翻譯哲學博士，現任教於香港嶺南大學翻譯系。研究方向：翻譯理論、文學翻譯、翻譯與性別、翻譯與出版等。電子郵箱：blli@ln.edu.hk。

稿約凡例

《翻譯季刊》為香港翻譯學會之學報，歡迎中、英文來稿及翻譯作品（請附原文及作者簡介）。有關翻譯作品及版權問題，請譯者自行處理。

一、稿件格式

1. 請以電郵傳送來稿之電腦檔案。
2. 來稿請附 200-300 字英文論文摘要一則，並請注明：（1）作者姓名；（2）任職機構；（3）通訊地址／電話／傳真／電子郵件地址。
3. 來稿均交學者審評，作者應盡量避免在正文、注釋、頁眉等處提及個人身份，鳴謝等資料亦宜於刊登時方附上。
4. 來稿每篇以不超過一萬二千字為宜。

二、標點符號

1. 書名及篇名分別用雙尖號（《》）和單尖號（〈 〉），雙尖號或單尖號內之書名或篇名同。
2. “ ”號用作一般引號；‘ ’號用作引號內之引號。

三、子　目

各段落之大小標題，請依各級子目標明，次序如下：

一、／A.／1.／a.／(1)／(a)

四、專有名詞及引文

1. 正文中第一次出現之外文姓名或專有名詞譯名，請附原文全名。
2. 引用原文，連標點計，超出兩行者，請另行抄錄，每行入兩格；凡引原文一段以上者，除每行入兩格外，如第

一段原引文為整段引錄，首行需入四格。

五、注　釋

1. 請用尾注。凡屬出版資料者，請移放文末參考資料部份。號碼一律用阿拉伯數目字，並用（）號括上；正文中之注釋號置於標點符號之後。
2. 參考資料

文末所附之參考資料應包括：（1）作者／編者／譯者；（2）書名、文章題目；（3）出版地；（4）出版社；（5）卷期／出版年月；（6）頁碼等資料，務求詳盡。正文中用括號直接列出作者、年份及頁碼，不另作注。

六、版　權

來稿刊登後，版權歸出版者所有，任何轉載，均須出版者同意。

七、贈閱本

從 2009 年夏天開始，作者可於 EBSCO 資料庫下載已發表的論文。如有需要，亦可向編輯部申領贈閱本。

八、評　審

來稿經本學報編輯委員會審閱後，再以匿名方式送交專家評審，方決定是否採用。

九、來稿請寄：香港屯門嶺南大學翻譯系轉《翻譯季刊》主編陳德鴻教授。電郵地址：chanleo@ln.edu.hk 。

Guidelines for Contributors

1. *Translation Quarterly* is a journal published by the Hong Kong Translation Society. Contributions, in either Chinese or English, should be original, hitherto unpublished, and not being considered for publication elsewhere. Once a submission is accepted, its copyright is transferred to the publisher. Translated articles should be submitted with a copy of the source-text and a brief introduction to the source-text author. It is the translator's responsibility to obtain written permission to translate.

2. Abstracts in English of 200-300 words are required. Please attach one to the manuscript, together with your name, address, telephone and fax numbers and email address where applicable.

3. In addition to original articles and book reviews, review articles related to the evaluation or interpretation of a major substantive or methodological issue may also be submitted.

4. Endnotes should be kept to a minimum and typed single-spaced. Page references should be given in parentheses, with the page number(s) following the author's name and the year of publication. Manuscript styles should be consistent; authors are advised to consult earlier issues for proper formats.

5. Chinese names and book titles in the text should be romanised according to the "modified" Wade-Giles or the pinyin system, and then, where they first appear, followed immediately by the Chinese characters and translations. Translations of Chinese terms obvious to the readers (like *wenxue*), however, are not necessary.

6. There should be a separate reference section containing all the works referred to in the body of the article. Pertinent information should be given on the variety of editions available, as well as the date and place of publication, to facilitate use by the readers.

7. All contributions will be first reviewed by the Editorial Board members and then anonymously by referees for its suitability for publication in *Translation Quarterly*. Care should be taken by authors to avoid identifying themselves. Submissions written in a language which is not the author's mother-tongue should preferably be checked by a native speaker before submission.

8. Electronic files of contributions should be submitted to Professor Leo Tak-hung Chan, c/o Department of Translation, Lingnan University, Tuen Mun, Hong Kong. Email address: chanleo@ln.edu.hk.

9. Given the accessibility, from summer 2009, of the journal via the EBSCO database, authors will no longer receive complimentary copies unless special requests are made to the Editor.

《翻譯季刊》徵求訂戶啟事

香港翻譯學會出版的《翻譯季刊》是探討翻譯理論與實踐的大型國際性學術刊物，學會副會長陳德鴻教授出任主編，學術顧問委員會由多名國際著名翻譯理論家組成。資深學者，如瑞典諾貝爾獎評委馬悅然教授、美國學者奈達博士及英國翻譯家霍克思教授都曾為本刊撰稿。《翻譯季刊》發表中、英文稿件，論文摘要（英文）收入由英國曼徹斯特大學編輯的半年刊《翻譯學摘要》。欲訂購的單位或個人，請與

中文大學出版社聯絡
地　址：中文大學出版社
　　香港 沙田 香港中文大學
電　話：+852 2609 6508
傳　真：+852 2603 6692 / 2603 7355
電　郵：cup@cuhk.edu.hk
網　址：http://www.chineseupress.com

Subscribing to *Translation Quarterly*

Translation Quarterly is published by the Hong Kong Translation Society, and is a major international scholarly publication. Its Chief Editor is the Society's Vice-President, Professor Leo Tak-hung Chan, and its Academic Advisory Board is composed of numerous internationally renowned specialists in the translation studies field. The journal has previously included contributions from such distinguished scholars as the Swedish Nobel Prize committee judge Professor Göran Malmqvist, the American translation theorist Dr. Eugene A. Nida, and the English translator Professor David Hawkes. *Translation Quarterly* publishes contributions in both Chinese and English, and English abstracts of its articles are included in *Translation Studies Abstracts*, edited by UMIST, UK. Institutions or individuals who wish to subscribe to the journal please contact:

The Chinese University Press

Address:	The Chinese University Press
	The Chinese University of Hong Kong
	Sha Tin, Hong Kong
Tel:	+852 2609 6508
Fax:	+852 2603 6692 / 2603 7355
Email:	cup@cuhk.edu.hk
Website:	http://www.chineseupress.com

Translation Quarterly《翻譯季刊》

Subscription and Order Form

To: The Chinese University Press Fax: (852) 2603 7355

Please enter my subscription to *Translation Quarterly*, beginning with No. _____

Subscription (complete volume)	Price
1 year	❑ HK$624 / US$80
2 years*	❑ HK$1,123 / US$144
3 years**	❑ HK$1,498 / US$192
Back issues	❑ HK$180 / US$23 each (Please list issue no.) _____

(please tick your choice)
Prices are at discount rate, delivery charge by surface post included.
* 10% discount.
** 20% discount.

Method of Payment:

❑ Attached is a cheque/bank draft* for HK$ / US$* _____ made payable to **"The Chinese University of Hong Kong"** (* circle where appropriate)

❑ Please debit my credit card account for HK$ _____ . (please convert at US$1 = HK$7.8)

I would like to pay my order(s) by: ❑ AMEX ❑ VISA ❑ MASTERCARD

Card No. _____ Cardholder's Name _____

Cardholder's Signature _____ Expiry Date _____

Please send my journal to:

Name _____

Address _____

_____

Telephone _____ Fax _____ E-mail _____

Subscription Information

- Prepayment is required for all orders.
- Orders may be paid by cheque/bank draft (made payable to "The Chinese University of Hong Kong") in US dollars, or by Visa, MasterCard or American Express in Hong Kong dollars.
- Orders are regarded as firm and payments are not refundable.
- Rates are subject to alteration without notice.

中文大學出版社
THE CHINESE UNIVERSITY PRESS
www.chineseupress.com
The Chinese University Press
The Chinese University of Hong Kong, Sha Tin, N.T., Hong Kong
Tel.: (852) 2609 6508 Fax: (852) 2603 6692 / 2603 7355 E-mail: cup@cuhk.edu.hk