

Automatic German-to-Arabic Text translation with Translation Post-Editing and translation Revision Systems

Prof. Dr. Reda Houssine Abo Aleze
Prof. Dr. Mohamed fared Zaphlol
Prof. Dr. Ass. Salwa al said hamada
Prof. Dr. Ass. Ahmad Almahdy
Eng. Dahey Gaber anem

Abstract

After we make a translation errors analysis by pen and paper for the first eleven chapters of the first book of the A Song of Ice and Fire saga – called A Game of Thrones – that has been translated from German to Arabic by Google translation system. We developed an automatic translation Post-Editing System (TPES) and translation revision System (TRS) to detect and fix these errors automatically.

We developed TPES in order to solve the fluency and adequacy errors in the target Arabic text by using a Maximum Spanning Tree (MST) in order to construct the correct German and Arabic parse tree.

We developed TRS depend on frequent pattern tree growth (FPT-Growth) algorithm to fix the Meaning Transfer (Accuracy) and Language (Mechanics) errors in the target Arabic text.

The manual and automatic evaluations show that, our TPES and TRS systems are able to improve the quality of the target Arabic text.

Keywords: Translation Studies, Text Analysis and Translation, Arabic Translation Post-editing system, Arabic Translation Revision system, Translation Errors, German text, and Arabic text.

1. Introduction

We make translation errors analysis by using the parameters presented in the translation revision model created by Brian Mossop in 2001 [1, 2]. The parameters used in the present study as categories of a tool for the analysis of translation errors, Table 1 show the translation Post-editing parameters and the translation revision parameters. So, from the 73 chapters of A Game of Thrones (German-Arabic), 11 were analyzed. From these 11 chapters, 129 occurrences of translation errors were found. From these 129, 98 were classified according to 6 of the categories extracted from Mossop's Model – as shown in Table 1 – and 31 occurrences of a similar nature that have no equivalent category in the model. As a consequence, a new category had to be created for them: Consistency (the fluency and adequacy) and all errors were counted as discrete errors. In the errors correction step, the source and target sentences and all the flagged errors are passed to the suggestion generator. The rule based TPES and TRS make the needed corrections automatically by the following seven action categories based on the action required to correct the error: {Edit, Add, Merge, Split, Delete, Move, Other} [3,4].

Parameters categories	Parameters types	Sub-parameters types
The translation Post-editing parameters	Group A (Fluency) parameters	1. Wrong term 2. Syntactic errors 3. Punctuation errors
	Group B (Adequacy) parameters	4. Omission 5. Word-structure 6. Misspelling 7. A miscellaneous errors.
The translation revision parameters	Group A (Transfer)	1. Accuracy 2. Completeness
	Group B (Content)	3. Logic 4. Facts
	Group C (Language)	5. Smoothness 6. Tailoring 7. Sub-languages 8. Idiom 9. Mechanics
	Group D (Presentation)	10. Layout 11. Typography 12. Organization

Table 1: the translation Post-editing parameters and the translation revision parameters

Our translation Post-editing system (TPES) performs two functions
1. Fluency parameters
2. Adequacy parameters

• *Author name is Eng.dahey gaber anem PHD degree program in electric power engineering in University ALAzhar, Country eygpt . E-mail: adahey@yahoo.com*

International Journal of Scientific & Engineering Research, Volume
ISSN 2229-5518

Our Translation Revision System (TRS) performs two functions

1. Meaning Transfer (Accuracy): Does the translation reflect the message of the source text?
2. Language (Mechanics): Have the rules of grammar, spelling, punctuation, house style been observed?

Categories	Number of Occurrences
Idiom	38
Accuracy	35
Smoothness	13
Completeness	7
Logic	3
Mechanics	2
TOTAL:	98

Table 3: Occurrences of Translation Errors according to Mossop's Model

To make the translation errors correction automatically we use Google Translation System (GTS) to translate the source German text to the target Arabic text. The output of GTS is frequently extremely grammatically incorrect due to the absence of linguistic rules for the language pair being applied. Grammatical error not only fails the fluency and adequacy of the translation. In translating German to Arabic by Google translation system, the BLEU scores [5] range only between 0.21 and 0.29, depending on test sets and numbers of reference translations. Table 4 shows the Occurrences of Translation Errors.

The input German text	**The Google Arabic text translation**	**Human Arabic text translation**	**Errors category**	**Errors parameter**
Der winter kommt	الشتاء قادم	الشتاء قادم	Post-editing	Consistency
Gared gehörte seit vierzig Jahren der Nachtwache an, als Mann und schon als Junge, und er war es nicht gewohnt, dass man sich über ihn lustig machte.	كان جارد حارساً ليلياً لمدة أربعين عاماً ، كرجل وكصبي ، ولم يكن معتاداً على السخرية منه.	لقد قضي الشيخ أربعين عاما كاملة مع حرس الليل،منذ إلتحق بهم وهو صبي.ولم يكن يروق له أن يستخف به الأخرين	language and style	idiom
Vier Jahre war er auf der Mauer. Als man ihn zum ersten Mal auf die andere Seite geschickt hatte, waren ihm all die alten Geschichten wieder eingefallen, und fast war ihm das Herz in die Hose gerutscht.	كان على الحائط لمدة أربع سنوات. في المرة الأولى التي تم إرساله فيها إلى الجانب الآخر ، كانت جميع القصص القديمة قد عادت إليه ، وكان قلبه ينزلق في سرواله.	ويوم أرسلوه وراء الجدارللمرة الأولي وجد الحكايات القديمة تتدفق من ذاكرته وشعر بأمعائه تتقلص	meaning transfer	completeness
Was gibt es da?	ما هو هناك؟	من هناك ؟	language and style	smoothness
Will trat an den Baum, einen gewölbten, graugrünen Wachbaum, und begann zu klettern.	صعد إلى الشجرة ، شجرة حراسة خضراء رمادية مقببة ، وبدأ في الصعود.	إتجه إلي شجرة الحارس الضخمة ذات الأفرع المقنطرة واللون الأخضروالـ رمادي وبدأ يتسلق	meaning transfer	accuracy
Der Andere zögerte. Will sah seine Augen, dunkler und blauer, als Menschenaugen jemals sein konnten, ein Blau, das brannte wie Eis. Sie richteten sich auf das Langschwert, das dort oben bebte, betrachteten das Mondlicht, das kalt über das Metall lief. Einen Herzschlag	الآخر تردد. سوف يرى عينيه ، أكثر قتامة وأكثر زرقة من العين البشرية يمكن أن تكون ، زرقاء تحترق مثل الثلج. كانوا يستهدفون الفستان الطويل الذي يرتعد هناك ، يراقبون ضوء القمر البارد فوق المعدن. لدقات قلب	توقف الأخر ،ورأي ويل عينيه .كانتا ذات لون أزرق شديد العمق،يحرق كالجليد،أعمق وأكثر زرقة من أي عين بشرية ،وقد ثبتت نظراتهما علي السيف الطويل الذي يرتفع مرتجفا في يد صاحبه،ورا قبتا نور القمر البارد يجري علي المعدن،وللحظة جرؤ ويل علي	content	logic

International Journal of Scientific & Engineering Research, Volume Ǌ(ÚÚÜÏ,Ï)ÜÕÌıĺƔƕ
ISSN 2229-5518

lang wagte er zu hoffen.	تجرأ على الأمل.	الأمل		
Er nannte sie »kleine Prinzessin« und manchmal »Mylady«, und seine Hände waren weich wie altes Leder.	دعاها "الأميرة" الصغيرة" وأحيانا "سيدتي" ، وكانت يديه ناعمة مثل الجلود القديمة.	وقال أمرا "قفي": مكانك "، ثم "إستديري". نعم عظيم. تبدين "... بهية" "كملكة" وكان يمرر يده علي صفحة الماء	language and style	mechanics

Table 4: The Occurrences of Translation Errors.

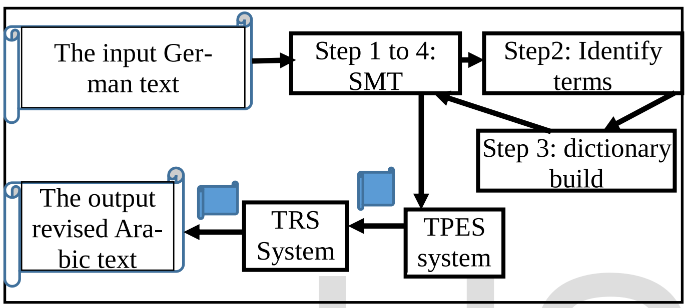

Figure 1: The translation system with translation post-editing and translation revision systems.

The rest of the paper is organized as follows. In Section 2, we present the related works. This is followed by The TPES and TRS Processes (Section 3), Section 4 present The TPES and TRS architecture, Section 5 presents The TPES and TRS Evaluation, Section 6 present Conclusions and Future Works.

2. Related Works

All of post-editing systems known to us perform statistical post-editing (SPE), typically by training an SMT system in a monolingual setting. The approach of RBPE of SMT can be seen as being complementary to SPE of RBMT as performed by [6] and thus can similarly benefit from the difference in issues of the rule-based and statistical approach, combining them to outperform both of them. An SMT system coupled with a RBPE system could also be classified as a hybrid MT system [7], such as Tecto MT [8]. An independent implementation for English-to-Persian exists, called Grafix [9]. We are first researchers to make Arabic text revision system.

3. The TPES and TRS Processes

Figure 2 represent the cognitive Processes of post-editing and Revision. As shown, translated text post-editing and revision calls on a range of hierarchically organized sub processes:

1. Translation text—characterizing the translated text's goals, the context (social, organizational, historical, cultural) in which the translated text is being post-editing and revised, the constraints under which the post-editing and the revision is taking place.
2. Detecting-seeing or noticing errors.
3. Diagnosing—characterizing or describing the translated text's errors.
4. Selecting strategies—choosing among optional methods for solving identified errors (revising or editing).
5. Fixing errors (correct)—taking action to remove the errors. The research from which this model was developed exposed dramatic differences in the abilities of experienced and inexperienced post-editors and revisers to involve in and carry out these processes.

Within each of these sub processes, the TPES and TRS have a variety of options. The ability to recognize available options and to make changes that actually improve text was found to distinguish experienced from inexperienced post-editors and revisers.
In fact, in some cases inexperienced post-editors and revisers introduce new errors and make the translated text worse instead of better [10]. Figure 3 represent The Processes of post-editing and Revision planner stages.

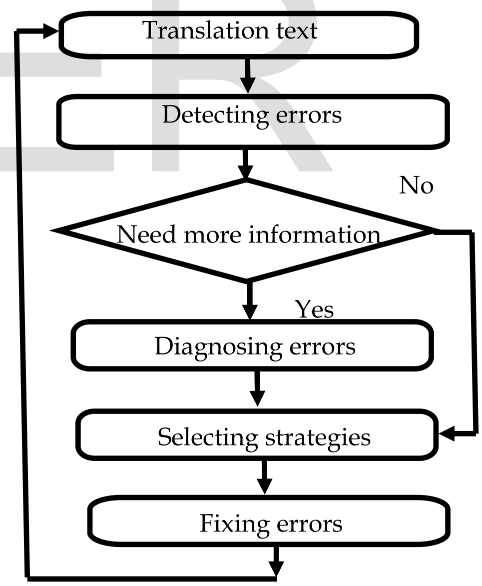

Figure 2: The cognitive Processes of post-editing and Revision

International Journal of Scientific & Engineering Research, Volume ʆǸ(ÚÚÜ Ì ł Ǹ)ÜÕÌ ı l ɣ
ISSN 2229-5518

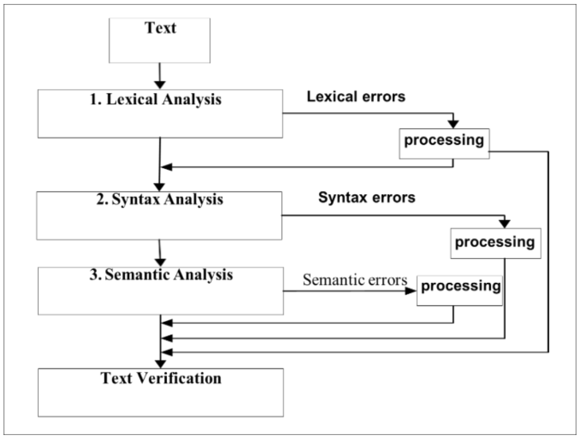

Figure 3: The Processes of post-editing and Revision planner stages.

4. The TPES and TRS architecture

4.1 The TPES components

We develop TPES system have three modules: lexical post-editing, shallow post-editing and deep post-editing. Figure 4 show The ATPE system architecture.

4.1.1 Lexical post-editing (PE)

1. Out Of Vocabulary (OOV) remover is a substitution rule. This rule replaces a German word in the MT output with the correct translation in Arabic.
2. A transliteration also is used to complement the operation of OOV remover, to implement the transliteration component we used Google search engine had improved its search algorithm for Arabic query by using some rules on letters which can be mistaken [11].

We use the NLTK toolkit (http://www.nltk.org/) to pre-process each sentence to obtain a more accurate representation of the information as POS tagger [12], which uses the most popular Penn Treebank POS tag set. The Arabic source text was analysed and tokenized using MADA+TOKAN [13]. Each MT system used a different tokenization scheme, so the source sentences were processed in two separate pipelines. Separate named entity recognizers (NER) were built for each pipeline using the Stanford NER toolkit [14]. Each translated German text was cascaded using Google and then analysed using the Stanford Core NLP pipeline to get part-of-speech (POS) tags [15] and NER [16].

4.1.2 Shallow post-editing (PE)

1. Incomplete Dependent PE: Relative pronouns such as «هو» in Arabic (German "es") which POS-tagged as SUBR, suggest continuation of the phrase by a dependent clause. If it occurred in a POS sequence without a consequent verb, an incomplete dependent sentence would be identified and a verb should complete the sentence. Currently, in most instances the verb «يكون» (German "ist") is suggested.
2. Incomplete Ended PE: the nouns located at the end of a sentence deemed as incorrect modifiers, so a POS sequence in which a modified. These sequences were removed from the sentence overall since there is no logical translation for given input.
3. Adjective Arrangement PE: In the Arabic language, adjectives usually come after the nouns they describe. For instance (German "schwere Tasche") «حقيبة ثقيلة» is translated literally as "حقيبة ثقيلة". The only exception in this group is superlative adjectives, which are identified by the suffix «أفضل», attached to the adjective. In this special case, the adjective comes before the noun to define it. The appearance of non-superlative adjectives before their described nouns must be corrected by this modifier.

4.1.3 Deep post-editing (PE)

We used an implementation of Dependency Parsing named MST Parser. MST Parser's main algorithm is Maximum Spanning Tree in which the maximum spanning tree should be found in order to find the best parse tree [17].

1. No Subject Sentence PE: Compared to known translation reference sentences, it was seen in some cases that what was parsed as the object in the sentence was actually the subject. Such sentences have a third person verb, no definite subject and an object tagged as POSTP (postposition) in the POS sequence. This transformer is designed to revise the sentence by removing the postposition «و» which is the indicator of a direct object in the sentence. Removal of this postposition changes the sentence to one with a subject [18].
2. Plural Nouns PE: Unlike German, in the Arabic language the word coming after a number is always singular. In SMT output there are instances where plural nouns are located after a number (< PRENUM> POS). This is corrected by removing the plural symbol of Arabic words. The suffix« ون» is the most common plural indicator, which is removed in this rule.

International Journal of Scientific & Engineering Research, Volume
ISSN 2229-5518

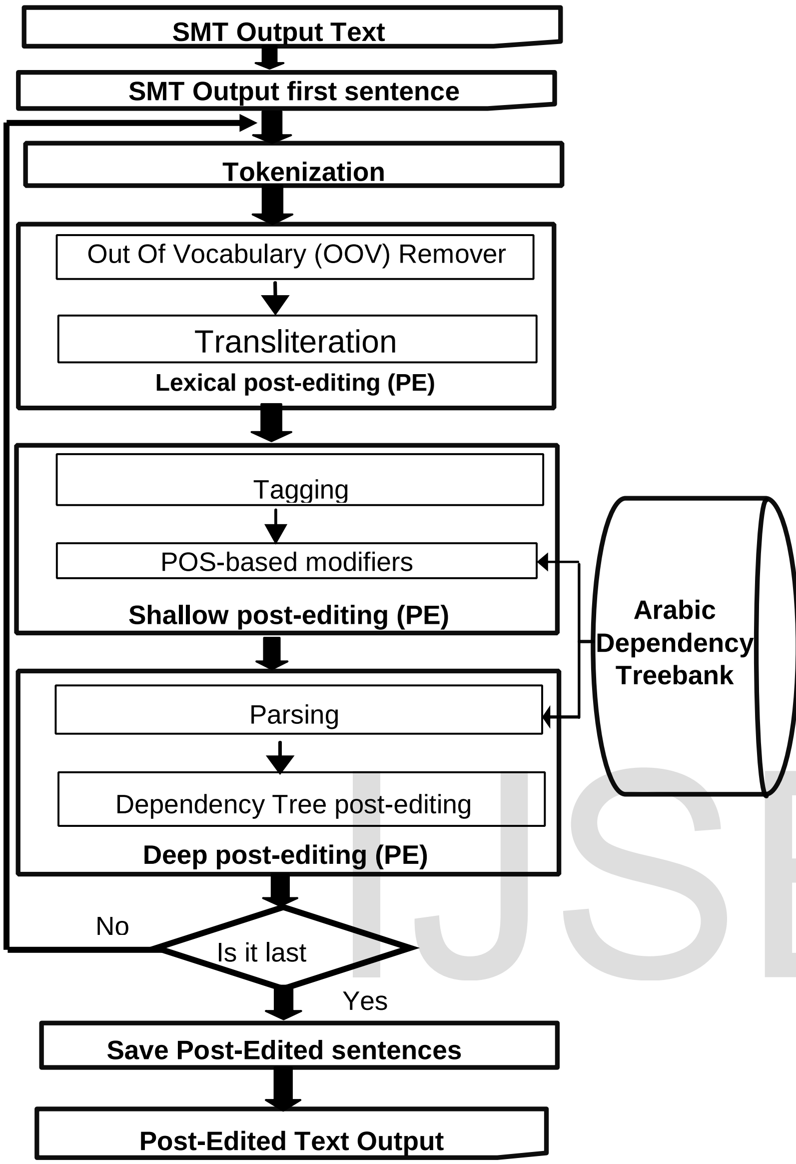

Figure 4: The ATPE system architecture.

3. Verb Arrangement PE: Arabic language has a preferred word order, with VSO (verb-subject-object) followed by SVO. These two types make up more than 75% of natural languages having a preferred order. One frequently occurring case is sentences in which a main verb as Root does not occur immediately before the period punctuation. The PE treats such cases as follows: the sentence is reordered by moving the root verb and its Non-Verbal Element (NVE) dependents to the end of the sentence.
4. Missing Verb PE: We used the Arabic [verb] valence Lexicon [49] to determine the proper verb for a non-verbal element in the sentence. All required and optional nonverbal elements (main-verb dependents) are listed in this lexicon. For example, the verb «ليستهلك» (German “konsumieren”) is composed of two elements. Searching for «ليستهلك» in this lexicon will return « ل» as the main part of that particular compound.

4.2 The TRS components

The Cognitive processes in revising system derived from observing experienced and inexperienced revisers at work, is intended to capture the thinking processes of post-editors and revisers involved in translated text revision. The revision process has two types; Monolingual revision and full revision and have many parameters listed in [19]. Our TRS system must Achieves the revision principles [19]. Figure 5 shows The Translation Revision System (TRS) components are:

1- Translation Post-editing text: The solving the surface problems includes various kinds of surface changes, such as changes of word order, punctuation, lexical choice, syntactic structure, and so on.
2- Evaluator: The evaluator evaluates the draft, which is an actual text rather than some intermediate representation. The draft is represented as its syntactic structure, which provides helpful information for the evaluation. The evaluator has to detect the structural ambiguities that even humans cannot disambiguate.
3- Revision Planning: The revision planner selects one of the errors detected by the evaluator, and suggests a change to correct it using FP-Growth Algorithm [20].

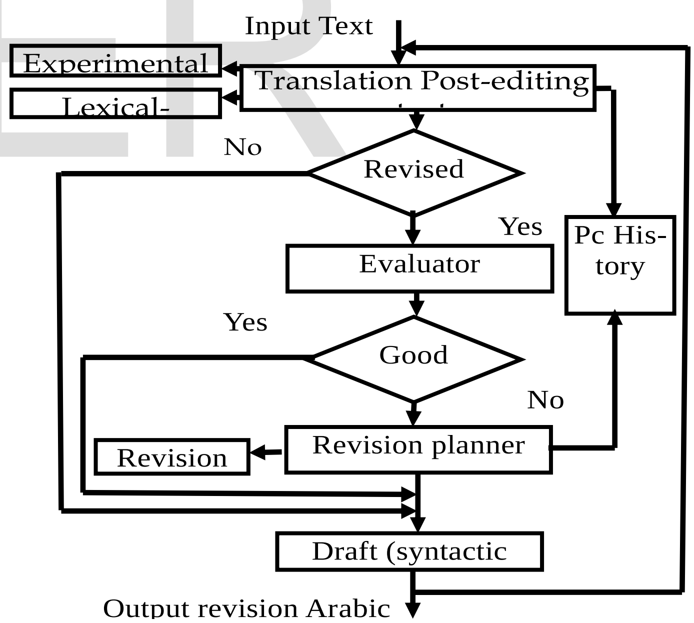

Figure 5: The TRS architecture.

4- The pattern with dual concept: We introduce the PC pair as a resource for the revision planning and the Arabic text change. In the initial Arabic revision generation, the text generator keeps path of the process as the pattern with dual concept. It is represented as a data structure similar to the dependency network used in the truth maintenance system (TMS) [21]. The pattern with dual concept controls both kinds of dependencies. The Infer-

International Journal of Scientific & Engineering Research, Volume
ISSN 2229-5518

ence is used to generate anaphora [22], ellipsis [21], exophora [22], and connectives [23]. They also used to make the pragmatic conclusions [23].

a. suggesting The pattern with dual concept to Change: The revision planner suggests a suitable change and sends the message to the Translation Post-editing Arabic text. The following revision rule will be applied. Generally, the revision rules suggest which feature should be removed from the current draft. Sometimes they even suggest which feature should be added in place of the removed feature.
b. Internal/external Dependencies: The Translation Post-editing Arabic text actually changes the draft with respect to the message. So, there are dependencies among the pattern with dual concept that are applied together to generate a draft. Such dependencies can be found by referring to the pattern with dual concept.
c. Translation Post-editing Arabic text Replacements: Translation Post-editing Arabic text realizes the text changes with respect to the internal and external dependencies.

The pattern with dual concept procedure is:

1. Receive the message from the revision planner,
2. Make the state of the pattern with dual concept in question out, extent the state over the dependency network, and construct the intermediate structure (S) from the features.
3. Request the ordinary Post-editing process,
4. Make the collection of the pattern with dual concept (C) which depend externally on the changed parts,
5. If C is empty, terminate the process, otherwise, choose and remove the pattern with dual concept from C that corresponds to the left most part in the draft,
6. If the pattern with dual concept chosen in 5 still satisfies the inference, go to 5, and otherwise go to 2.

In our method, because the Post-editing text generates an alternative draft from the same rhetorical structure, the new draft text satisfies not only the grammatical and lexical constraints, but also the pragmatic and textual preferences (the quality of the draft).

5. The TPES and TRS implementation

1- Develop the lexical analysis and the syntax analysis models.
2- Use the prolog language the syntax analysis model.
3- Use Amzi! Prolog application for prolog program.
4- Use Microsoft Visual C# 2010 for lexical and syntactical revision system.
5- Link the database for the lexical analysis and the syntax analysis models with C# language.

Examples of applying PR actions

We additional analyze the performance of our TRS system by examples. Table 6 shows the TRS procedure of improving translation quality for three different sentences.

	Die WHO bestätigte Vietnams sechste tote Vogelgrippefall
	تؤكد منظمة الصحة العالمية وفاة السادس في إنفلونزا الطيور في فيتنام
	(الصحة العالمية) (المنظمة) (تأكيد) (فيتنام) (٦) () (أنفلونزا الطيور) (الموت) (حالة)
Ref	تؤكد منظمة الصحة العالمية حالة الوفاة السادسة من إنفلونزا الطيور في فيتنام
Baseline	أكدت منظمة الصحة العالمية حالات موت أنفلونزا الطيور في فيتنام
PR*1	تؤكد منظمة الصحة العالمية حدوث حالات وفاة لإنفلونزا الطيور1 في فيتنام
PR*2	تؤكد منظمة الصحة العالمية حالة 2 الموت السادسة من إنفلونزا الطيور في فيتنام
	Die nationale Aussöhnung erfordert normalerweise eine gewisse Anzahl von Operationen, und das ist schwierig
	تتطلب المصالحة الوطنية عادة كمية معينة من العملية ، ومن الصعب ذلك
	(الوطنية) (المصالحة) (عادة) (تحتاج) (معينة) () (بالطبع) (،) (لا) (إنجاز في إجراء واحد(.)
Ref	تحتاج المصالحة الوطنية عادة إلى مسار معين ، ولا يمكن تحقيقه في إجراء واحد.
Baseline	المصالحة الوطنية هي عملية صعبة للغاية.
PR*1	تحتاج عملية المصالحة الوطنية عادة إلى معينة صعبة للغاية.
PR*2	المصالحة الوطنية تحتاج عادة إلى مسار معين ، وأنجزت.
PR*3	المصالحة الوطنية عادة ما تحتاج إلى مسار معين ، وأنه لا يمكن أن يتحقق.
	Die zwei Antworten israels können jedoch von den Vereinigten Staaten nicht vollständig gelöst werden.
	(ومع ذلك) (إسرائيل) (الرد) (تفشل) (واضح تماما) (لنا) () (شك) (.)
Ref	ومع ذلك ، فشل رد إسرائيل في إزالة الشكوك بشكل كامل.
Baseline	ومع ذلك ، فإن الرد الإسرائيلي على الإزالة الكاملة للولايات المتحدة.
PR*1	ومع ذلك ، فشلت الاستجابة الإسرائيلية في الشك بوضوح تام. الولايات المتحدة
PR*2	ومع ذلك ، فشل رد إسرائيل في توضيح الشكوك بشكل كامل

Table 6: Examples of applying PR actions multiple times in the German -Arabic translation.

International Journal of Scientific & Engineering Research, Volume ꞁꞂ(ÚÚÜ Ï ꞁꞁ Ꞁ(ꞁ)ÜÕÌ ı l Ƴ ꞁ
ISSN 2229-5518

In the first sentence, two PR cycles lead to a perfect translation. In the first PR cycle (PR*1), revising the translation of "Die sechste" from "٦" to "السادس" improves the neighboring translation. The translation of "Bestätigung" changes from " التأكيد" to "تؤكد ", which a positive effect is. In PR*2, revising the translation of "Fälle" from "حالات" to "حالة" also changes the neighborhood translation, after two PR cycles, the reference translation is obtained. In our current settings, the reference translation could not always be obtained. The maximum achievable BLEU is around 60-70 in general environment.

In the second sentence, "Muss sicher sein" " يجب أن تكون على يقين" is picked in the first PR cycle. Revising the translation from "a" to "need a certain" makes the translation of "normalerweise" "عادي" changing from "يكون" to "دائما". In the next PR cycle, revising the translation of "Prozess" "عملية" from "process" to "course" makes the neighboring translation changing from "," to ", and". Meanwhile, the position of "course" moves to the right place (in front of ","). In the last PR cycle, the translation of "Es ist schwer zu" "من الصعب" is revised from "it" to "it cannot be". After three PR cycles, the translation quality improves significantly. However, the translation is still different from the Reference.

This is because "Auf einen Schlag" should be translated into "دفعة واحدة" instead of "يتحقق". But there is no suitable translation option for it in the current phrase table. So the system cannot generate a perfect translation. The problems will be less significant when real-world human translators are involved. Human translator inputs the correct translation "accomplished in one action", the system will generate the reference translation after constrained decoding (Human).

In the last sentence in Table 9, "Kann nicht" "لا يمكن" is picked as the critical error. Revising the translation from "إلي ال" to "تفشل", leads to an improvement on neighboring phrase (the translation of "Volles Kehren" to "بشكل كامل "), In the second PR cycle, "Israels" "إسرائيل" is picked. Revising the translation from "إسرائيلي" to "اسرائيل"، makes the translation of "Antwort" change from "الإجابة" to " الاستجابة ", which is also a positive effect. However, after two PR cycles, all phrase translations are correct, but the translation is still different from the reference. This is because the language model and lexical reordering model prefer the wrong phrase ordering, which put "الولايات المتحدة" "the US" at the end of the whole sentence. This problem arises from the MT system itself, which may not be solved directly in our TRS.

If more interactions are allowed, for example, performing reordering operations, the above problems could be solved. But the interactions become more complex, and may not be acceptable to human translators. Other solutions include using better statistical models such as neural language models.

5. The TPES and TRS Evaluation

5.1 The TPES and TRS Automatic Evaluation

We used BLEU [5] to evaluate the effect of APE system on translation quality. The results of translation before and after APE are shown in Table 5. The results generally show increases in the metric.

Input	Size Arabic	Size German	BLEU Before ATPE	BLEU After ATPE	BLEU Difference
#1	163	158	0.6523	0.6770	0.0247

Table 5: the Automatic evaluation for ATPE

We conduct experiments on sentences for which the reference could be generated by our current MT system using forced decoding. The decoder forced the unilingual and bilingual Corpora of the book A Game of Thrones in German and Arabic, parallel corpus.

data	BLEU Before TRS	BLEU after TRS
TRS *1	44.59	79.27 (+34.68)
TRS *2	44.59	79.54 (+34.93)

Table 6: Automatic Evaluation for the TRS.

5.2 Manual Evaluation

Both post-editors, who completed the evaluation without discussing or consulting with each other, had very similar judgment of the APE system's output. The results show that the APE system has been successful in improving the quality of the baseline SMT system output by 29.4%. The current developed rules for the APE system are effective for about 37% of the SMT translated sentences.

Both revisers, who completed the evaluation without discussing or consulting with each other, had very similar judgment of the TRS system's output. The results show that the TRS has been successful in improving the quality of the baseline SMT and ATPE systems output. Table 2 show the Manual Evaluation Quality the TRS output.

Subjects	Total Error Points	Revision Quality %	Result
Revisor1	32	93.59	Pass
Revisor1	25	94.80	Pass

Table 7: the Manual Evaluation Quality the TRS output

6. Conclusions and Future Works

6.1 Conclusions

1- We conclude that, additional parallel resources extracted from comparable corpora can improve machine translation for low resource language pairs (e.g., German–Arabic).
2- We conclude that, TPES produces significantly better translations than the Google Translate system which is

a difficult system to beat. This enhancement in translation quality through APE should reduce human PE energy.

3- We conclude that, Brian Mossop's translation revision model into an analytical tool to look at translation errors and also show that this model did not contain all the parameters necessary to look at translation errors.

6.2 Future Work

1- Future work will also investigate a fully integrated second stage MT (APE), based on a neural network approach that will improve translation quality and minimize translation effort and cost by exploring a character/sub-word (using byte pair encoding) based APE system to rectify morphological errors.
2- In the future work, it would also be interesting to explore what would happen to the issue of Consistency by using Translation Memory Systems right in the translation draft.
3- Our two TPES and TRS are by no means limited to German-to-Arabic translation. For many languages, there exist high-performance processing tools, able to provide similar linguistic analyses as the tools we use, but relevant for the respective languages. These could be used to provide ground for developing a similar rule-based and/or statistical post-editing system, which we expect would also be able to improve the quality of machine translation between the respective languages.

References

[1] Robert, Isabelle, Aline Remael and Jimmy J. J. Ureel (2016). "Towards a model of translation revision competence." The Interpreter and Translator Trainer, 10(2). Doi: 10.1080/1750399X.2016.1198183)).

[2] Parton, K., Habash, N., McKeown, K., Iglesias, G., & de Gispert, A. (2012). Can Automatic Post-Editing Make MT More Meaningful? Proceedings of the 16th EAMT Conference.

[3] Isabelle, P., Goutte, C., & Simard, M. (2007). Domain adaptation of MT systems through automatic post-editing. MT Summit XI.)).

[4] Papineni, K., S. Roukos, T. Ward, and W. J. Zhu, "BLEU: a method for automatic evaluation of machine translation, "In Proceedings of ACL-2002, 40th Annual Meeting of the Association for Computational Linguistics pp. 311—318, 2002.

[5] K. Oazer and I.D. El-Kahlout. Exploring different representational units in English-to-Turkish statistical machine translation. In Proceedings of the Second Workshop on Statistical Machine Translation, pages 25-32. Association for Computational Linguistics, 2007.

[6] Michel Simard, Cyril Goutte, and Pierre Isabelle. Statistical phrase-based post-editing. In Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference, pages 508{515, Rochester, New York, April 2007. Association for Computational Linguistics.

[7] Andreas Eisele, Christian Federmann, Hans Uszkoreit, Herv_e Saint-Amand, Martin Kay, Michael Jellinghaus, Sabine Hunsicker, Teresa Herrmann, and Yu Chen. Hybrid machine translation architectures within and beyond the euromatrix project. In Proceedings of the 12th annual conference of the European Association for Machine Translation (EAMT 2008), pages 27-34, 2008.

[8] Martin Popel and Zden_ek _Zabokrtsk_y. TectoMT: modular NLP framework. In Proceedings of the 7th international conference on Advances in natural language processing, IceTAL'10, pages 293{304, Berlin, Heidelberg, 2010. Springer-Verlag. ISBN 3-642-14769-0, 978-3-642-14769-2.

[9] Schäffner, Christina (2012). "Translation competence: training for the real world." Séverine Hubscher-Davidson and Michal Borodo (Eds.) (2012). Global trends in translator and interpreter training. Mediation and culture. (pp. 30-44). London: Continuum, 30-44.

[10] McDonald, R., Pereira, F., Ribarov, K., & Hajic, J. (2005). Non-projective dependency parsing using spanning tree algorithms. In Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing (pp. 523-530). Vancouver, B.C., Canada: Association for Computational Linguistics.

[11] Michel Simard, Cyril Goutte, and Pierre Isabelle. Statistical phrase-based post-editing. In Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference, pages 508{515, Rochester, New York, April 2007, Association for Computational Linguistics.

[12] Toutanova, Kristina, Dan Klein, Christopher D. Manning, and Yoram Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network, In NAACL-HLT, pp. 173–180.

[13] Habash, Nizar, Owen Rambow, and Ryan Roth. 2009. MADA+TOKAN: A toolkit for Arabic tokenization, discretization, morphological disambiguation, pos tagging, stemming and lemmatization. Proc. of the 2nd International Conference on Arabic Language Resources and Tools (MEDAR), pp. 242–245.

[14] Finkel, Jenny Rose, Trond Grenager, and Christopher Manning. 2005. Incorporating non-local information into information extraction systems by Gibbs sampling, In ACL, and pp. 363–370.

[15] Toutanova, Kristina, Dan Klein, Christopher D. Manning, and Yoram Singer. 2003. Feature-rich

International Journal of Scientific & Engineering Research, Volume
ISSN 2229-5518

part-of-speech tagging with a cyclic dependency network, In NAACL-HLT, pp. 173–180.

[16] Kübler, S., McDonald, R., & Nivre, J. (2009). Dependency parsing Synthesis Lectures on Human Language Technologies (Vol. 1, pp. 1-127)).

[17] Rasooli, Moloodi, Kouhestani, &Minaei-Bidgoli, valence Lexicon, 2011.

[18] Imran, Ali, “Application of a Mining Algorithm to Finding Frequent Patterns in a Text Corpus: A Case Study of the Arabic, “International Journal of Software Engineering and Its Applications vol. 6, No. 3. 2012.

[19] C. Matthiessen. Lexico (Grammatical) choice in text generation, In Natural Language Generation in Arti cial Intelligence and Computational Linguistics, chapter 10, pages 249-292, Kluwer Academic publishers, 1991.

[20] Parton, K., Habash, N., McKeown, K., Iglesias, G., & de Gispert, A. (2012). Can Automatic Post-Editing Make MT More Meaningful? Proceedings of the 16th EAMT Conference.

[21] D. E. Appelt. Planning natural-language referring expressions. In D. D. McDonald and L. Bolc, editors, Natural Language Generation Systems, chapter 3, pages 69{97. Springer-Verlag, 1988.

[22] M. Elhadad and K. R. McKeown. Generating connectives. In the Proceedings of the International Conference on Computational Linguistics, pages 3:97{101, 1990.

[23] E. H. Hovy. Generating Natural Language under Pragmatic Constraints. Lawrence Erlbaum Associates, 1988.