TITLE	:	Development and Evaluation of German to Arabic Text Translation System with Post Editing and Revision System
SUBMITTED BY	:	Dahey Gaber Abd EL-Fatah Mohamed Anem
SUPERVISED BY	:	Prof. Dr. Mohamed Farid Zaghlool Prof. Dr. Reda Hussein Abo Eez Ass. Prof. Dr. Ahmed Mohamed Nour ALMahdy
PLACE OF EXAMINATION	:	Al-Azhar University
YEAR OF EXAMINATION	:	2021
LANGUAGE OF THESIS	:	English

Ali Mohamed

ABSTRACT

A translation errors analysis was made by pen and paper for the first book of the A Song of Ice and Fire saga - called A Game of Thrones – that which was translated from German to Arabic. A classification for translation errors was used this as a classification regarding data set for training and testing the three systems to detect and fix these errors automatically. ATS, APES and ARS were developed by two methods: Classical method-

A- The ATS was developed for solving three important errors in the target text: word segmentation on terms, unknown word transliteration, and long-distance reordering. ATS achieved the BLEU scores by 34.77%.

B- An automatic post-editing system (APES) based on grammatical rules, was developed The evaluation of APES showed that the output text improves meaningfully, that is, by at least 25%, resulting in high-quality translation output.

C- An ARS for revising and correcting the APES output text errors was developed ARS depend on lexical correction, syntactical correction, and frequent pattern tree growth (FPT- Growth). The manual and automatic evaluations showed that, ARS are able to improve the quality of the output Arabic text.

Deep neural network (DNN) method

A- A deep neural automatic translation system (DNATS) was developed which contained models to translate from large amounts of parallel sentences from German to Arabic texts DNATS was

tested on German to Arabic translation, and obtained significant improvements in the output translation text.

B- A deep neural network was developed based on automatic post-editing system (DNAPES) to improve raw deep neural automatic translation system (DNATS) output. The DNAPES output Arabic text showed statistically significant improvements of 3.96 and 2.68 BLEU points absolute over the original DNATS and phrase-based APES outputs, respectively. Also, human evaluation showed that the DNAPES generated post-editing Arabic text was much better than the original ATS Arabic translated output.

C- DNARS was developed to improve the quality of the Arabic text which resulted from the DNAPES. DNARS evaluation results proved that by interactions through either one of the translation quality could be significantly improved. Greater gains could be the actions, achieved by iteratively performing both actions.

KEY WORDS : ATS ; APES ; ARS ; DNATS ; DNAPES ; DNARS ; German Language ; Arabic Language.

عنوان الرسالة	:	بناء وتقييم نظام ترجمة من نص الماني إلى نص عربى مزودا بنظام تصحيح ونظام تدقيق
اسم مقدم الرسـالة	:	ضاحي جابر عبد الفتاح محمد غانم
لجنــة الإشـراف	:	أ. د. محمد فريد زغلول أ. د. رضا حسين أبو العز أ. م. د. أحمد محمد نور المهدى
جهــة المنـــح	:	كلية الهندسة – جامعة الأزهر
سنــة المنـــح	:	٢٠٢١
لغــة الرســالة	:	اللغة الإنجليزية

ملخص البحث

في هذه الأطروحة بعد إجراء تحليل أخطاء الترجمة بالقلم والورقة للكتاب الأول من ملحمة أغنية من الجليد والنار تسمى لعبة العروش التي تمت ترجمتها من الألمانية إلى العربية أجرينا تصنيف لأخطاء الترجمة ونستخدم تصنيف أخطاء الترجمة هذا كمجموعة بيانات (جسم لغوى ثنائي اللغة) للتدريب واختبار الأنظمة الثلاثة لاكتشاف هذه الأخطاء وإصلاحها تلقائيا قمنا ببناء المترجم الآلي والمصحح الألى بطريقتين :

١. الطريقة الكلاسيكية

- تم بناء المترجم الآلي لحل ثلاثة أخطاء مهمة في النص الهدف تجزئة الكلمات حسب المصطلحات الترجمة الرمزية الغير المعروفة وإعادة ترتيب كلمات الجمل الطويلة. حقق المترجم الآلي قيمة دقة ٣٤.٧٧ نسبة مئوية على مقياس BLEU
- قمنا ببناء نظام مصحح ألي مبنى على القواعد استنادا إلى القواعد النحوية كان من نتائج المصحح الآلي أن النص الناتج يتحسن بشكل ملموس وكانت الدقة تتحسن بنسبة ٢٥ نسبة مئوية على الأقل مما ينتج عنه إخراج ترجمة عالية الجودة.
- قمنا ببناء مدقق ألي لتدقيق وتصحيح أخطاء النص الخارج من المصحح الألى يعتمد المدقق الألى على التدقيق المعجمي والتدقيق النحوي واستخدام خوارزمية شجرة الأنماط المتكررة (FPT-Growth) لتدقيق الأخطاء الدلالية والمعاني في النص العربي بعد التصحيح، تظهر التقييمات اليدوية والآلية أن المدقق الآلي قادر على تحسين جودة النص العربي الناتج.

٢. الطريقة الذكية باستخدام الشبكات العصبية العميقة DNN

- قمنا بتطبيق نظام ترجمة مدرك للمرجعية يساعد على إزالة الغموض بين التشابه بين النص المصدر والنص الهدف وحصلنا على تحسينات كبيرة فى ترجمة الضمائر الضخمة مقارنة بالأسماء الأساسية التي يعود عليها الضمير.
- قمنا ببناء نظام مصحح ألي ذكي باستخدام الشبكات العصبية ذات التعلم العميق (DNAPES) لتحسين ناتج نظام الترجمة الألى. (DNAPES) يظهر النص العربي كمخرج وفيه تحسينات ذات دلالة إحصائية قيمة دقة قدرها ٢.٦٨ على مقياس BLEU على النص العربي الخارج من المترجم الذكي وقيمة دقة على مقياس BLEU قدرها ٣.٩٦ أيضا يظهر التقييم البشرى أن المصحح الذكي الذي تم بناؤه للنص العربي ينتج نصا أفضل بكثير من النص العربي الأصلي الناتج من المترجم الذكي.

- قمنا ببناء مدقق ألي ذكي باستخدام الشبكات العصبية ذات التعلم العميق (DNARS)لتحسين جودة النص العربي الناتج عن المصحح الألى الذكي (DNAPES) تثبيت نتائج تقييم DNARS أنه من خلال التفاعلات باستخدام الإجرائين السابقين في نظام المدقق الألى الذكي، يمكن تحسين جودة الترجمة بشكل كبير ويمكن تحقيق مكاسب أكبر من خلال تنفيذ الإجرائين السابقين بشكل متكرر (أكثر من مرة).

الكلمات المفتاحية : نظم الترجمة الآلية ؛ نظام التصحيح الآلي ؛ نظام التدقيق الآلي ؛ اللغة الألمانية ؛ اللغة العربية.