Machine Translation

Tikrit University

College of Arts

جامعة تكريت

Journal of Al-Frahedis Arts

Academic Journal by

College of Arts University of Tikrit

issue : Special issue of 7th college

ISSN : 2074 - 9554

Tikrit University
College of Arts

Journal of Al- Farahidi's Arts

A Quartly Academic Journal
of
The College of Arts - Tikrit

ISSN: 2074- 9554

THE PROBLEM OF MACHINE TRANSLATION IN "GOLDEN AL-WAFI" SOFTWARE SENTENCE ANALYSIS

Lecturer Ibrahem Muhammid Ali Mustafa

THE PROBLEM OF MACHINE TRANSLATION IN "GOLDEN AL-WAFI" SOFTWARE SENTENCE ANALYSIS

By
Lecturer
Ibrahem Muhammid Ali Mustafa

Abstract

Machine translation (MT) has advanced since it was first conceived in 1949. It's one of the first non-numerical applications for computers. MT research programs pop up in Japan and Russia (1955). Xerox uses Systran to translate technical manuals (1978). Recently, Google announced that Google Translate translates roughly enough text to fill 1 million books in one day (2012). This means that the need for MT, though with many defects, increases day by day. In this paper, we try to pinpoint some defects of the software "Golden Al-Wafi" machine translator through testing a few sentences. Then, an analysis of the resulted translation is done stating the defects. After that, small parts of those sentences are translated by the software the changes in the translation, then some solutions are suggested. The direction of the translation is English to Arabic as it is the function of the software.

1. The Beginnings of Machine Translation

The invention of the digital computer in the 1940s inspired scientists to think of using the unprecedented speed of the computer to translate texts from one language to another. So inspired, scientists started to take practical steps to realize the dream and vision of Descartes who wrote in 1629 about a mechanical process to convert one human language to another. In 1949, Warren Weaver, the pioneer of machine translation, wrote a memorandum to his colleagues making four proposals for machine translation systems that go beyond word for word translation. Warren realized that many words in language were ambiguous and he proposed in his memorandum to solve this problem by examining the immediate context of the ambiguous word (Hutchins et al, 2000:43).

He also drew attention to the analogy between the structure of the human brain and the "logical machine". He concluded that the machine translation problem is solvable. He also suggested using the cryptic methods that linguists used in the Second World War for deciphering the German secret code. These cryptographic methods relied heavily on frequencies of letters, combination of letters and letter patterns. He also believed that underlying the statistical regularities of languages, there is a logical and universal foundation which could represent an alternative to translate from one language to another. At the same time with the beginning with the cold war in the 1940s, there was an urgent need for crude machine translation because the United States decided it was essential to scan and interpret every Russian communication coming out of the Soviet Union. However, there weren't enough translators to keep up with the huge volume of Russian books and papers published in the

Soviet Bloc at that time. The urgent need to translate Russian into English coincided with the invention of computers. It was not surprising then, that developing Russian to English machine translation systems would be one of the first tasks these "miracle" machines were set to perform. The first demonstration of the feasibility of fully automated machine translation took place in New York on January 7th, 1954. On that day, Georgetown University and IBM demonstrated the first non-numerical applications and capabilities of the "new" electronic brain by demonstrating a fully automated Russian English machine translation system. The system embraced the commonly held view that a language consisted of a lexicon and a finite set of rules that could generate an infinite set of sentences. Surprisingly, the first Russian to English machine translation system had only 250 words and 6 syntactic rules. This experiment raised high expectations that probably within five years machine translation systems would be readily available. The promise was to develop a system that does not require pre-editing of the input while produces a reliable translation of the input text in the target language that is clear, intelligible requiring only stylistic modifications. No details were given about the actual linguistic processing in the system. For example, no information about dictionary content and lookup procedures were given. No account of how the syntactic analysis of the Russian sentences was performed and how the target English structure was selected. However, there were some references to reversing the order of pairs of sentences by assigning rules to the lexical items involved. Later on a more detailed description of the system is presented in (Garvin,1967: 46-56). Garvin (ibid) gives more detailed description of the dictionary. For

example, the dictionary entries were sometimes stems, endings or full words. Each entry is associated with three codes; the first code indicates which of the six syntactic rules would apply, the second code would determine which contextual information are needed to determine the target translation while the third code indicates whether words to be inverted or not. At the technical level, the system represents the first attempt at non-numerical programming which presents developers with many challenges. Developers had to deal with character coding of Russian and how dictionary entries to be stored, what lookup procedure would be followed and how the syntactic rules would be coded and executed.

The Georgetown-IBM experiment and similar other work at the time were significant. First, it was proven that the digital computer could perform non-mathematical tasks such as machine translation. The system took advantage of the speed of the computer compared to human translators. It was also shown that the computer surpassed humans in that it would never forget, could work 24/7 without getting tired, and would never ask for a raise or a vacation. Third, the system demonstrated the need to specify and describe linguistic structures at different levels such the lexical and syntactic levels was demonstrated. Fourth, ambiguity of language was understood to be a problem although it was underestimated. However, as Hutchins (Hutchins, 1995: 431-445) stated correctly, the period from 1956 to 1966 was: "a decade of high expectations and disillusions". The promise to deliver fully automated machine translation systems with no pre-editing and only stylistic post-editing with a 95% accuracy was never achieved. Serious research proved that language structure was much more complex than

previously thought and that translators use huge amounts of linguistic, domain-specific, real world and common sense knowledge that was not considered relevant at the time. The ALPAC report (A report by the Automatic Language Processing Advisory Committee, 1966) concluded that machine translation was not viable given the state of knowledge at the time. Consequently, funding for research in computational linguistics halted in the USA and did not resume until the mid and late seventies.

2. Arabic Machine Translation

Arabic is one of the major languages that have been given attention by machine translation (MT) researchers since the very early days of MT and specifically in the U.S. The language has always been considered "due to its morphological, syntactic, phonetic and phonological properties [to be] one of the most difficult languages for written and spoken language processing. Arabic "differs tremendously in terms of its characters, morphology and diacritization from other languages. Accordingly researchers cannot always import solutions from other languages, and today Arabic machine translation still needs more efforts to be improved, mainly in the area of semantic representation systems, which are essential for achieving high quality translation.

a. Particularistic approaches

Particularistic approaches describe the linguistic features of Arabic and use them for a local processing approach specific to the internal linguistic system of Arabic. They are concerned with the morphological and semantic aspects of Arabic. Sakhr is one of the Arabic speaking groups developing systematically machine processing of Arabic.

b. Universalist approaches

Universalist approaches use the methods and systems proved to be useful in other languages like English or French making some adaptations if necessary. The focus here focus is on the syntactic aspects of the linguistic system in general. This approach is followed by most of the companies producing software applications for Arabic.

(Zughoul,2005:1022–1041)

3. The Transfer Approach to Machine Translation

Problems with the Direct Approach

The direct machine translation approach did not take deep linguistic analysis of the source language into consideration. It involved shallow treatment of the word order of the source sentence to make it appear more related to the order of the target language. For that reason, machine translation scientists quickly realized that the direct method could not deal with the intricacy of normal language. For example, the simple exchanging operation switching the subject and verb from an SVO to a VSO structure turned out to be very complex. So the translation of an English sentence such as '***George likes cake***' into Arabic involves switching the order of the subject ***George*** to come after the verb in Arabic to become '*يحب جورج الكعك*'. However, when the subject of the English sentence becomes complicated as in '***The short man who was wearing glasses and a blue suit and was standing in the middle of the hall with his friends, welcomed us heartily***', identifying the length and borders of the subject requires deep parsing. The direct approach would not yield accurate results for complex sentences like this because it did not incorporate the required syntactic rules for parsing such sentences. There was also a need to develop the technology to efficiently perform deep parsing

and to represent complex disambiguation rules. The transfer approach to machine translation provided considerable contributions on two fronts: the syntactic description of language and a new technology for the representation and processing of deep syntactic parsing. (Farghaly, 2010)

4. Data analysis

In this section , the following sentence is input onto Golden Al-Wafi Translator for translation. Data are analyzed.

When we input the following sentence onto the program in question, we get results as follows:

1. ***Cars are for getting people from A to B in maximum safety.***

The possible normal/human translations are:

- السيارات هي من اجل نقل الناس من النقطة أ الى النقطة ب باقصى درجة من الامان.
- تستخدم السيارات لنقل الناس من مكان الى اخر بمنتهى الامان.

Golden Al-Wafi machine translator's (henceforth: GW)translation:

- السيارات لتَحْصلُ على الناس من الألف للباء في الأمانِ الأقصى.

We see that

The translation of the individual parts of the sentence by GW are as follows:

Cars: السيارات

Are : no translation

for getting : يُصبحُ

are for getting: يُصبحُ

people from A to B :ناس من الألف للباء

in maximum safety: في الأمانِ الأقصى

If we replace the word ***'getting'*** with '***transporting***' we get the following translation which more appropriate :

Cars are for **transporting** people from A to B in maximum safety.

السيارات لنقل الناس من الألف للباء في الأمانِ الأقصى.

This means that there are insufficient information in the database of the software about, for example, the verb 'get' and its possible combinations with prepositions such as 'get something from -to' and so on.

2. I've come looking for a bit of advice.

- جِئتُ بَحْث عن قليلاً النصيحةِ.

In the above-mentioned sentence, a clear lack of database of idioms such as ***a piece of advice /a bit of advice*** which have to be taught to the machine Translator.

3. I spend a lot on expensive jewelry and clothing...

أصْرفُ الكثير على المجوهراتِ واللباسِ الغالي. . .

In the last sentence, the translation of ***spend*** is not enough another word has to be added to make the sentence sound more natural, i.e., من المالwhich is closely related to the word اصرفthat is to say its collocate.

4. Conclusion

The history of MT has witness gradual development in translation and has been through many difficulties but MT is improving day by day through new and more advanced methods and syntactic and disambiguating rules. In the future, MT is expected to witness tremendous development through teaching the machine more patterns and stretches of languages such as idioms, collocations, phrasal verbs and even readymade sentences. MT could witness a day in which Machine Translators could possibly think translation like human being.

References

ALPAC 1966 Language and machines: computers in translation and linguistics. A report by the Automatic Language Processing Advisory Committee. National Academy of Sciences, Washington, DC.

Farghaly, Ali. 2010 ***Arabic Machine Translation: A Developmental Perspective*** International Journal on Information and Communication Technologies, Vol. 3, No. 3, June 2010

Garvin, Paul. 1967. The Georgetown-IBM Experiment of 1954: An Evaluative Perspective. In Austin William M. (ed) Papers in linguistics in honor of Dostert (The Hague: Mouton, 1967).

Hutchins, John. 1995. MACHINE TRANSLATION: A BRIEF HISTORY in Concise history of the language sciences: from the Sumerians to the cognitivists. Edited by E.F.K.Koerner and R.E.Asher. Oxford: Pergamon Press, 1995.

Hutchins, John. 2000. Early Years in Machine Translation, Ed. John Hutchins, John Benjamin, Amsterdam.

Zughoul, Muhammad; Abu-Alshaar, Awatef (3 August 2005). "English/Arabic/English Machine Translation: A Historical Perspective". Translators' Journal 50 (3).

بحوث الإنكليزي

15.	INTERDEPENDENCE OF ALLUSION AND INFERENCE IN ARABIC-ENGLISH TRANSLATION	*Asst. lecturer: Omar Dawood Omar*	1-19
16.	*THE PROBLEM OF MACHINE TRANSLATION IN "GOLDEN AL-WAFI" SOFTWARE SENTENCE ANALYSIS*	Ibrahem Muhammid Ali Mustafa	20-29
17.	RIMBAUD'S LIFE AND CAREER IN FRENCH AND ENGLISH: A TRANSLATION STUDY	Assistant Lecturer Ekhlas Ali Mohsin	30-77

المحتويات

ت	عنوان البحث	اسم الباحث	الصفحة
	بحوث اللغة العربية		
١.	الغريب والمعرب في اللهجة البدوية الشعر البدوي أنموذجا	أ. م. د. احمد عطية علو	١-٢٩
٢.	العربية بين مخاطر الاستغراب ومد العولمة	أ. م. د. نصرة احميد جدوع	٣٠-٥٠
٣.	مهارة الترجمة في العصر العباسي الاول عهد المأمون أنموذجا	أ. م. د. فرحان محمود التميمي	٥١-٩٥
٤.	دعوات التغريب في اللغة والأدب	أ. م. د. حسن سالم هندي أ. م. د. خيري جبير الجميلي	٩٦-١٣٢
٥.	بدر الدين العباسي (٩٦٣هـ) أديبا وناقدا	أ. م. د. حسين لفته حافظ أ. م. د. حسين عبد العال اللعيبي	١٣٣-٢٥٤
٦.	مفهوم التفاعل بين الأدب الورقي والأدب الرقمي	أ. م. د. سلافة صائب خضير العزاوي	٢٥٥-٢٨٩
٧.	العولمة وأثرها على اللغة العربية الاسس والمنطلقات	أ. م. د. اسماء صابر جاسم	٢٩٠-٣٢٥
٨.	التأثر والتأثير بين اللغة العربية واللغة الأسبانية (خرجة الموشح انموذجا)	م. د. سهام صائب خضير	٣٢٦-٣٤٢
٩.	النقد الإسلامي والعولمة: قراءة في خطاب الدكتور محمد إقبال عروي	د. اسماعيل ابراهيم المشهداني	٣٤٣-٣٦٤
١٠.	الألفاظ المستطرفة من الدخيل والمعرّبة في معجمات مجمع اللغة العربية بالقاهرة	د. محمد سالم ياسين الجبوري	٣٦٥-٤٠٣
١١.	صراع الهويات في شعر صدر الاسلام همزية حسان بن ثابت أنموذجاً	د. فنن نديم دحام آل إبليش	٤٠٤-٤٢٣
	بحوث التأريخ والجغرافية والاجتماع		
١٢.	اهمية الترجمة في الكتابة التاريخية تاريخ العراق المعاصر إنموذجاً ١٩١٤ - ١٩٣٤	أ. د. عبد المجيد كامل عبد اللطيف	٤٢٤-٤٣٥
١٣.	رحلة الترجمة العظيمة ودورها في اغناء الفكر الاسلامي بالارث الحضاري القديم	أ. د. باسمة جاسم خنجر	٤٣٦-٤٨٠
١٤.	أبعاد العولمة على المجتمع المحلي (قضاء الحويجة أنموذجا)	م. د. عبد الله صالح علي الجبوري	٤٨١-٥١٧

١٣-يجب أن تكون الخطوط كالآتي:-

أ-اللغة العربية: نوع الخط (Simplified Arabic) حجم الخط (١٤).

ب-اللغة الانكليزية: نوع الخط (Times New Roman) حجم الخط (١٤).

١٤- عمل الهوامش يكون بنظام تلقائي (تعليقات ختامية) في نهاية البحث، ويكون الترقيم مستمراً، مع التدقيق في تسلسل الترقيم.

- **مجالات النشر:**

١- **البحوث العلمية:**

تنشر المجلة البحوث العلمية الأصيلة والمخطوطات المحققة في مجال العلوم الانسانية.

٢- **المؤتمرات والندوات العلمية:**

تنشر المجلة بحوث المؤتمرات والندوات العلمية المحلية والعربية والعالمية، والتي عقدت حديثاً في مجال العلوم الانسانية.

- **ملاحظات النشر:**

- البحوث المنشورة في المجلة تعبر عن آراء الباحثين ولا تعبر عن رأي المجلة.
- ترتيب البحوث في المجلة يخضع لاعتبارات فنية.
- تستبعد المجلة أي بحث مخالف لقواعد النشر.
- يعطى الباحث نسخة مستلة لبحثه.
- العنوان البريدي: جامعة تكريت عمادة كلية الآداب مجلة آداب الفراهيدي.
- العنوان الالكتروني: magazine_ arts @ yahoo. Com
- العنوان الثاني: arts_college_ t @ yahoo. Com

تصميم وإخراج: محمد عادل جاسم

شروط النشر

١- ان يكون البحث مطبوعاً على الحاسوب، وتزود هيئة التحرير بثلاث نسخ منه مع نسخة على قرص ليزري (CD).

٢- ان لا تزيد عدد صفحات البحث عن (٢٥) صفحة ولا تقل عن (١٥) صفحة من الحجم العادي (A4) ويستثنى من ذلك النصوص المحققة.

٣- ألا يكون البحث جزءاً من رسالة الماجستير أو أطروحة الدكتوراه التي أعدها الباحث، وألا يكون قد سبق نشره على أي نحو كان أو تمَّ إرساله للنشر في مجلة أخرى ويتعهد الباحث بذلك خطياً.

٤- يلتزم الباحث بإجراء تعديلات المحكمين على بحثه وفق التقارير المرسلة إليه وموافاة المجلة بنسخة معدلة في مدة لا تتجاوز (١٥) يوماً.

٥- أن يكون البحث ضمن الاختصاصات الانسانية.

٦- يخطر أصحاب البحوث بالقرار حول صلاحيتها للنشر أو عدمها خلال مدة لا تتجاوز ثلاثة أشهر من تأريخ وصوله لهيئة التحرير.

٧- لا ترد الأبحاث إلى أصحابها سواء نشرت أم لم تنشر.

٨- يلتزم الباحث بدفع أجور النشر المقررة.

٩- في حال قبول البحث للنشر في المجلة لا يسمح للباحث بإعادة نشره في مكان آخر إلا بعد مرور سنة كاملة على تأريخ نشره فيها.

١٠-يطبع البحث ببرنامج (Word)، وتوضع الرسوم أو الاشكال - إن وجدت - في مكانها من البحث على أن تكون صالحة من الناحية الفنية للطباعة.

١١-أن يكون البحث خالياً من الأخطاء اللغوية والنحوية والاملائية.

١٢-يجب اتباع الأصول العلمية والقواعد المرعية في البحث العلمي.

مجلة آداب الفراهيدي

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة تكريت

مجلة آداب الفراهيدي

مجلة علمية محكّمة فصلية تصدر عن كلية الآداب
بجامعة تكريت

الترقيم الدولي: ISSN: ٢٠٧٤ - ٩٥٥٤

رقم الإيداع في دار الكتب والوثائق ببغداد: ١٦٠٢ لسنة: ٢٠١١

العـــدد (١٦) أيلول ٢٠١٣

عدد خاص ببحوث المؤتمر العلمي السابع لكلية الآداب بجامعة
تكريت المنعقد للمدة من ٢٩-٣٠ آذار ٢٠١٣

the Problem of Machine Translation

٤

العدد (١٦) ايلول ٢٠١٣ - ٢٠١٤

جامعة تكريت
كلية الاداب

مجلة آداب الفراهيدي

مجلة علمية محكمة تصدر عن كلية الآداب / جامعة تكريت

العدد: بحوث المؤتمر العلمي السابع

الترقيم الدولي : ISSN: 2074 - 9554