# QCRI Machine Translation Systems for IWSLT16

أجري هذا البحث (Durrani, et al. 2017)من قبل فريق Qatar Computing Research Institute معهد قطر لبحوث الحوسبة للمشاركة في مسابقة IWSLT 2016 evaluation campaign وتم بناء هذا النظام بالاعتماد على الترجمة الإحصائية حيث تم بناء نموذجين للترجم من العربية إلى الإنكليزية و من الإنكليزية للعربية , حيث تم بناء نماذج الترجمة بالاعتماد على نظام moses للترجمة الآلية الإحصائية.

* تم تدريب النماج باستخدام UN corpus, Ted Talks وOpen subtitles corpu و قد اتبع الباحثون الخطولت التالية في بناء النموذج
* فلترة بيانات التدريب و ترشيحها بهدف التخلص من البيانات الضارة
* تدريب ألية ترتيب الكلمات و أصناف الكلمات بشكل مستقل
* تم تدريب النموذج أولا على مدونات الأمم المتحدة ثم تحسين النتائج على بيانات Ted Talks
* تم استعمال أطوال الجمل حتى 80 كلمة
* تم الاعتماد على مقاطع لغوية حتى 5-gram

تم استعمال محلل صرفي للغة العربية MADAMIRA لمعالجة الكلمات العربية MADAMIRA

تم بناء نظام الترجمة الآلية بالاعتماد على نظام moses للترجمة الآلية الإحصائية.

تم تدريب عدة أزواج من النماذج أحدهما من بيانات التدريبBaseline تم تدريبه على (مدونات الأمم المتحدة و Open Subtitles Corpus (QED Corpusوأخر كبير نسبيا +bigLM و ا الذي هو مجموع أفضل ثمانية نماذج كما تم أستعمال بيانات تدريب أخرى من مسابقات سابقة و كانت النتائج حسب معيار BLEU كما يلي

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| المعدل | العينة الرابعة | العينة الثالثة | العينة الثانية | العينة الأولى | النموذج |
| 30.4 | 28.6 | 32.3 | 32.4 | 28.2 | Baseline |
| 30.9 | 29.2 | 33.2 | 32.8 | 28.3 | +bigLM |

# English to Arabic Machine Translation Based on Reordring Algorithm

أجري هذا البحث (Mohammed and Aziz 2011)في كلية هندسة المعلومات في جامعة كيبانجسان ماليزيا، هدف البحث هوو بناء نظام للترجمة الآلية من الانكليزية إلى العربية.

يعتمد هذا النموج في عمله على بناء نموذج قواعدي Context Free Grammar (CFG)، بالإضافة إلى استعمال محدد أقسام كلام Part Of Speech لكل كلمة , إعادة ترتيب الجملة حتى يتناسب ترتيب الجملة الإنكليزية مع ترتيب الجملة العربية, تم استعمال قاموس إنكليزي عربي من أجل ترجمة كل كلمة على حدا مع الاستفاد من خرج محدد أقسام الكلام لتحديد صنف الكلمة للحصول على ترجمة وحيدة لهذه الكلمة .

تم بناء قواعدCFG إعادة تالترتيب من اللغة الإنكليزية إلى اللغة العربية بناء لكل نوع جملة أو تعبيير في اللغة العربية

يتم عمل النموذج على مرحلتين

المرحلة الألولى تقسيم جملة الدخل وصولا إلى مستوى الكلمات مع تحديد صنف كل كلمة مع معالجة الجملة وفق قواعد CFG للتناسب كل كلمة مع تصنيفها.

المرحلة الثانية توليد جملى الخرج بالعتماد على القاموس لترجمة كل كلمة و إعادة ترتيب القواعد للتتناسب الكلمة مع تصنيفها.

تم اختبار النظام على 20 نص مختلف حيث عدد الجمل الكلية لهذه النصوص 95 جملة، تم الاعتماد على الحبرة البشرية في تحديد جودة الترجمة و لم يتم الاعماد على أية معيار أخر، وكانت نتائج دقة الترجمة بحسب البحث 81.55% للترجمة من الإنكليزي إلى العربي.

## An English-Arabic Bi-Directional Machine Translation Tool in the Agriculture Domain

A Rule-based transfer approach for translating expert systems))

أجري هذا البحث (Khaled, Ashraf and Ahmed 2010) في الجامعة البريطانية في دبي وكان مخصص في مجال ترجمة نصوص المتعلقة بالزراعة، حيث تم بناء نظام للترجمة من العرببة إلى الإنكليزية و بالعكس و اعتمد هذا النظام على مبدأ transfer-based MT و الذي يتضمن ثلاث مراحل

التحليل: ويتم فيها تحليل الجملة قواعديا و لغوية لتوليد تمثيل مجرد للجملة الأصل.

النقل: يتم فيها توليد تمثيل مجرد للجلة الهدف بناء على تمثيل جملة الأصل.

التوليد: يتم فيها توليد الجملة الهدف بناء على التمثيل التي تم توليده في مرحلة النقل.

تم اختبار النظام على معيار BLEU وكانت نتائج اختبار النظام موضحة بالجدول التالي

|  |  |  |
| --- | --- | --- |
| التقييم بحسب معيار BLEU | التقييم باستخدام ترجمة مرجعية واحدة | التقييم باستخدام ترجمتين مرجعيتين |
| Arabic to English | 0.4581 | 0.8122 |
| English to Arabic | 0.4504 | 0.6427 |

## First Result on Arabic Neural Machine Translation

أجري هذا البحث (Almahairi, et al. 2016) لمجموعة من الباحثين من جامعة نيويورك و جامعة مونتريال حيث تم بناء نظام للترجمة الآلية باستخدام الشبكات العصبية من العربية إلى الإنكليزية و بالعكس أيضاً.

حيث يناء نظام للترجمة العصبية باستخدام نموذج للترجمة العصبية باستخدام ِAttention Mechanism (Bahdanau، Cho و Bengio 2015)

تمت معالجة الكلمات في اللغة العربية بتقسم الكلمة إلى ملحقات للكلمة و جذر الكلمة حيث تم استخدام ) MADAMIRA (Pasha, et al. 2014)

أما بالنسبة للهيكلية الطبقة العصبية فكانت بالنسبة للـencoder فكانت عن طبقة RNN ثنائية الاتجاه مع طبقتين من 512 وحدة GRU، و بالنسبة للـdecoder فكان عباة عن شبكة RNN وحيدة الاتجاه من 512 وحدة

تم تدريب كل نموذج للترجمة لسبعة أيام تقريبا حتى توقف تحسن تابع الخطا على مجموعة التحقق، كما تم استخدام Dropout (Srivastava, et al. 2014)خلال تدريب الشبكة العصبية.

تم تدريب النظام باستخدام مدونات LDC2004T18, LDC2004T17 , LDC2007T08 و بلغ عدد الجمل 1.2 مليون زوج من الكلمات و 33 مليون كلمة عربية، وتمت مقارنة النتائج مع الترجمة الآلية الإحصائية باستخدام برنامج Moses للترجمة الآلية الإحصائية .

و تمت مقارنة النتائج باستخدام معيار BLEU وكانت النتائج كما يلي

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| AR->EN\* | AR->EN | EN->AR \* | EN->AR |  |
| 48.59 | 51.19 | 31.52 | 35.98 | Statical MT |
| 47.12 | 49.7 | 28.64 | 33.62 | Neural MT |

حيث (\*) لم يتم معالجة الكلمات

بين هذا البحث على أن المعالجة الجيدة للكلمة في اللغة العربية يؤثر على نتائج الترجمة بشكل كبير و خاصة في الترجمة من الإنكليزية إلى العربية.

## On the Properties of Neural Machine Translation: Encoder–Decoder Approaches

أجري هذا البحث (Cho, et al. 2014) في جامعة مونتريال و استخدم لبناء نموذج للترجمة الآلية باستخدام الشبكات العصبية باسخدام نموذج encoder-decoder (Cho, Van Merriënboer and Bahdanau, et al. 2014)

تم بناء للترجمة من اللغة الإنكليزية إلى الفرنسية حيث استعمل في التدريب مدونة مؤلفة من 348 مليون جملة متقابلة، و وتم اعتبار أول 30000 كلمة أكثر تكرارً أم بقية الكلمات فاعتبرت مجهولة، وو بالنسبة للشبكة العصبية فتم بناءها 1000 عصبون في الطبقة الخفية

تم تدريب النموذج في حالتين مرة بدون كلمات مجهولة و مرة باستعمال كلمات مجهولة و تمت مقارنة النتائج مع نظام Moses للترجمة الإحصائية و كانت النتائج حسب معيارBLEU

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| الجمل بطول 10-20 كلمة | | جميع الجمل | |  | |
| Test | Development | Test | Development |
| 20.99 | 19.12 | 13.92 | 13.15 | Neural MT | كل الكلمات |
| 32.00 | 28.92 | 33.30 | 30.64 | Moses |
| 27.03 | 24.73 | 23.45 | 21.01 | Neural MT | لا يوجد كلمات مجهولة |
| 35.40 | 32.20 | 35.63 | 32.77 | Moses |

## NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATE

أجري هذا البحث (Bahdanau، Cho و Bengio 2015) كتطوير على نموذج Encoder-Decoder للترجمة العصبية و يقوم هذا النموذج على مبدأ آلية الانتباه التي تم نفصيلها سابقا، حيث أصبحت هذه البنية عي الطريقة الأساسية في الترجمة الآلية العصبية، حيث أعطت هذا لبنية للشبكة العصبية جودة أفضل للترجمة و خاصة في الجملة الطويلة نسبياً، و مما يميز هذه الطريقة أنها لا تحاول أن تضغط الجملة في شعاع واخدج ذات طول ثابت، وعند ترجمة كل كلمة يولد النموذج يبحث في الجملة الهدف عن المواضع التي تتركز فيها المعلومات.

سبق أن بين بينة هذا النموذج الرياضية و طبقات العصبونات في الشبكة العصبية.

أجريت هذه التجارب عل بناء نموذج للترجمة من اللغة الإنكليزية إلى الفرنسية,و حيث استعمل في تدريب النموذج مدونات من عدة مصادر Europarl (61M words), news commentary (5.5M), UN (421M)

تم تدريب نوعين من النماذج النوذج الأول من هيكلية Encoder-Decoder (RNNencdec)و النموذج الثاني يعتمد على آلية الانتباه (RNNsearch)و تم الاختبار مرة على الجمل حتى طول 30 كلمة و مرة حتى طول 50 كلمة كانت النتائج كما هي موضخة بالجدول التالي كما تم الاختبار في حالتين مرة على جميع الكلمات،و في الخالة الثانية تم الاعتبار لـ 30000 كلمة تكرارا و باقية الكلمات تم اعتبارها unknown، وكانت النتائج موضخة بالجدول التالي

|  |  |  |
| --- | --- | --- |
| No UNK | All | Model |
| 24.19 | 13.93 | RNNencdec-30 |
| 31.44 | 21.50 | RNNsearch-30 |
| 26.71 | 17.82 | RNNencdec-50 |
| 34.16? | 26.75 | RNNsearch-50 |
| 36.15 | 28.45 | RNNsearch-50(More training Time) |
| 35.63 | 33.30 | Moses |