الترجمة الآلية بمقاربة إحصائية : تحديات وآفاق لحوسبة اللغة العربية

وفاء بن تركي، أستاذة بالمعهد العالي العربي للترجمة
د. ناصر الدين سمار، أستاذ زائر بالمعهد العالي العربي للترجمة

كلمات مفتاحية: الترجمة الآلية بمقاربة إحصائية -حوسبة اللغة العربية– المعالجة الآلية للغات الطبيعية. Big Data اللغوية - المعطيات الضخمة - المتون

الملخص

منذ اختراع أول حاسوب، سعى خبراء الذكاء الاصطناعي إلى تكريس القدرات الفائقة للحاسوب لمعالجة اللغات الطبيعية آليا من خلال محاكاة القدرات الإدراكية البشرية لاكتساب ومعالجة وصياغة اللغة الطبيعية.

وتمثل الترجمة الآلية الاحصائية أحد أهم هذه التطبيقات الذكية التي واكبت عصر المعلومات Information Age، الذي أوجد عالما افتراضيا تتسابق إليه الأمم لإيجاد مكان لثقافتها وبالتالي لغتها من خلال اثراء محتوى رقمي يعكس رصيدها الحضاري والمعرفي. ولما كان المحتوى الرقمي باللغة العربية لا يفوق ٣٪ من اجمالي المحتوى الرقمي على الشابكة، بات لزاما علينا ايجاد حلول تواكب هذا الانفجار المعرفي دون إغفال خصوصية اللغة العربية.

وفي إطار هذا البحث سنركز على ما يلي:

١- التحديات التي تواجه المعالجة الآلية للغات الطبيعية بصفة عامة مرورا بالتحديات التي تواجه اللغة العربية بصفة خاصة.

٢- سياق التطوير الذي يتسم بانفجار معرفي تعكسه المعطيات الضخمة Big Data المتدفقة عبر الشابكة من جهة وفجوة معرفية Digital Gap من جهة أخرى وسبل استغلال البيانات الضخمة في مجال الترجمة الآلية.

٣- سنتعرض إلى تقنية الترجمة الآلية بمقاربة احصائية: سنلقي الضوء على بنية هذه النظم ثم نتناول مواطن الضعف والقوة، ووصولا إلى معايير تقييم جودة الترجمة في نظم الترجمة الآلية.

ويخلص البحث إلى خلاصة وتوصيات من شأنها الاسهام في تحسين أداء نظم الترجمة الاحصائية.

- الفروق الموجودة بين اللغات:

- ترتيب الكلمات
- صعوبة إزالة مكامن اللّبس المرتبطة ارتباطا وثيقا بمنظومة الكتابة في أي لغة طبيعية وبالتالي تعقيد الخوارزميات المخول لها فك هذا اللبس.
- مشكلة الاشتراك اللفظي: يمكن لمفردة واحدة أن يكون لها عدة معاني ألم : أَلَمْ - أَلَمْ- أَلَمَّ- ألمّ
- مشكلة اختلاف المعاني (بين اللغة الأصل ولغة الوصل).
- مشكلة ربط الكلمات وتكوين الجمل: يصعب على الحاسوب على غرار المترجم البشري إيجاد نظائر للتراكيب اللغوية بين اللغة الأصل Source Language ولغة الوصل Target Language ، دون إدخال تغيير ولو طفيف في الصيغة الأصلية لأن كل لغة تتميز بأساليبها الخاصة ولا يمكن فرض أسلوب اللغة الأصل على لغة الوصل لأن ذلك قد يؤثر سلبا على مستوى الترجمة.
- مشكلة ربط الجمل وتكوين الفقرات.
- المشكلات المرتبطة أساسا باللغة العربية العبارات المسكوكة Idioms : غياب معاجم مختصة في العبارات المسكوكة (التعابير الاصطلاحية) إضافة إلى غياب تقييس موحد للمصطلحات Terminology.
- استعمال العبارات المجازية: «فتحرير رقبة مؤمنة») سورة النساء الآية ٩٢(إن ورود كلمة تحرير رقبة تعني تحرير المؤمن من الرق أو عتقه كما ورد في تفسير ابن كثير وقد استعملت مجازا كلمة رقبة للتعبير عن النفس المملوكة.
- اللَّبْسُ النَّحْوِيُّ: في غياب حركات التشكيل
- مثال: يمكن أن تشير كلمة يعد إلى كلمة:

يَعِدُ: يلتزم بوعد She /he promises

- يَعُدُّ: يحصي He calculates /Computes
- يُعَدُّ: يُعْتَبَرُ considered

- الإبهام الناتج عن تعقيد البنية التركيبية للنص الأصلي:
- مثال: ألم ألم ألم بدائه ★★★ إن آن آن آن أوانه

أَلَمْ أَلَمَّ أَلَمْ أَلَمَّ ب دَائِه ★★★

إِنْ آنَ آنْ آنَ نَلْهَأَ

في البيتين السابقين للمتنبي نلاحظ إبهاما على مستويين :المستوى الأول هو الإبهام الناتج عن الاشتراك اللفظي)نفس

المتتالية الحرفية أو نفس اللفظ يعبر عن معاني متباينة ، أما المستوى الثاني من الابهام فهو راجع لاستعمال هذه الألفاظ كلها معا في نفس العبارة مما يضفي على النص تعقيدا يصعب المهمة على المترجم البشري ونظام الترجمة الآلية بدرجات متفاوتة من الصعوبة وما زاد الأمر تعقيدا غياب حركات التشكيل.

- اللبس الناتج عن المتلازمات اللفظية التي يفترض أن تترجم « كمقطع واحد » بدل أن ترجمة كل كلمة على حدة.

مثال: عبارة رَغِبَ بنفسه عن الشيء والتي تعني تَرَفَّعَ عن الشيء.

الترجمة البشرية: He rised above

نتيجة مترجم جوجل : Wishes for himself

نتيجة مترجم بينج: Same thing with the desired نلاحظ بان الترجمة الآلية في هذه الحالة بعيدة كل البعد عن الترجمة البشرية والسبب هو غياب متون لغوية ثرية لغويا وتتضمن المتلازمات اللفظية الواردة في اللغة العربية.

- نقل البنية التراكيبية للغات الأجنبية إلى اللغة العربية :

مثال: أكثر مهارة - عوض أمهر – more skilled

أكثر قدسية – عوض أقدس more sacred رغم أن الإيجاز الذي يميز اللغة العربية يجعل المتون اللغوية العربية أقل حجما من غيرها إلا أنه قد يكون سببا من أسباب اللبس التي تواجهها نظم الترجمة الآلية.

- ندرة استعمال علامات الوقف في النصوص العربية تصعب عملية تحديد نهاية الفقرات والجمل مما يؤثر على تحديد دلالة الفقرة/الجملة وبالتالي يؤثر حتما على الترجمة المنتجة.

- لا يوجد سبيل لتحديد أسماء العلم ضمن النصوص العربية، فخلافا للنصوص

العربية، نجد أن النصوص الإنجليزية أو الفرنسية تستعمل تفخيم الحرف الأول Capitalization للتمييز بين اسم العلم وباقي كلمات النص.

٢- الانفجار المعرفي The Knowledge Explosion :

أدخل عصر المعلومات Information Age تغييرا جذريا على كل مجالات الحياة وغير مفاهيم واستحدث أخرى فضلا عن أنه أثرى معاجم اللغة بمصطلحات جديدة مثل العالم الافتراضي virtual world والشبكات الاجتماعية social networks، والمعطيات الضخمة Big Data. وقد عرفت البشرية نهضة علمية لا نظير لها في كل مجالات الحياة بفضل التقانات التي واكبت هذه الثورة المعلوماتية إضافة إلى الكم الهائل من المعلومات المتداولة على الشبكة والذي تتزايد نسبتها يوما عن يوم، وقد ورد في بحث)هيلبيرت ولوبيز٢٠١١- Hilbert & Lopez، ٢٠١١) (أن نسبة النمو السنوي للمعلومات المخزنة بلغت ٢٣٪ ولو عدنا قليلا إلى الوراء لوجدنا أن القدرة العالمية على تخزين المعلومات قد تضاعفت خلال الفترة الممتدة بين ١٩٨٦ إلى ١٩٩٦ بنسبة ٨٠٠٪،

المتتالية الحرفية أو نفس اللفظ يعبر عن معاني متباينة ، أما المستوى الثاني من الابهام فهو راجع لاستعمال هذه الألفاظ كلها معا في نفس العبارة مما يضفي على النص تعقيدا يصعب المهمة على المترجم البشري ونظام الترجمة الآلية بدرجات متفاوتة من الصعوبة وما زاد الأمر تعقيدا غياب حركات التشكيل.

- اللبس الناتج عن المتلازمات اللفظية التي يفترض أن تترجم « كمقطع واحد » بدل أن ترجمة كل كلمة على حدة.

مثال: عبارة رَغِبَ بنفسه عن الشيء والتي تعني تَرَفَّعَ عن الشيء.

الترجمة البشرية: He rised above

نتيجة مترجم جوجل : Wishes for himself

نتيجة مترجم بينج: Same thing with the desired نلاحظ بان الترجمة الآلية في هذه الحالة بعيدة كل البعد عن الترجمة البشرية والسبب هو غياب متون لغوية ثرية لغويا وتتضمن المتلازمات اللفظية الواردة في اللغة العربية.

- نقل البنية التراكيبية للغات الأجنبية إلى اللغة العربية :

مثال: أكثر مهارة - عوض أمهر – more skilled

أكثر قدسية – عوض أقدس more sacred رغم أن الإيجاز الذي يميز اللغة العربية يجعل المتون اللغوية العربية أقل حجما من غيرها إلا أنه قد يكون سببا من أسباب اللبس التي تواجهها نظم الترجمة الآلية.

- ندرة استعمال علامات الوقف في النصوص العربية تصعب عملية تحديد نهاية الفقرات والجمل مما يؤثر على تحديد دلالة الفقرة/الجملة وبالتالي يؤثر حتما على الترجمة المنتجة.

- لا يوجد سبيل لتحديد أسماء العلم ضمن النصوص العربية، فخلافا للنصوص

العربية، نجد أن النصوص الإنجليزية أو الفرنسية تستعمل تفخيم الحرف الأول Capitalization للتمييز بين اسم العلم وباقي كلمات النص.

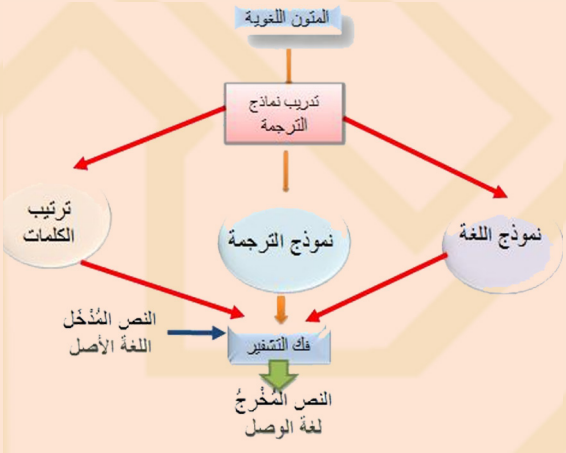
٢- الانفجار المعرفي The Knowledge Explosion :

أدخل عصر المعلومات Information Age تغييرا جذريا على كل مجالات الحياة وغير مفاهيم واستحدث أخرى فضلا عن أنه أثرى معاجم اللغة بمصطلحات جديدة مثل العالم الافتراضي virtual world والشبكات الاجتماعية social networks، والمعطيات الضخمة Big Data. وقد عرفت البشرية نهضة علمية لا نظير لها في كل مجالات الحياة بفضل التقانات التي واكبت هذه الثورة المعلوماتية إضافة إلى الكم الهائل من المعلومات المتداولة على الشبكة والذي تتزايد نسبتها يوما عن يوم، وقد ورد في بحث)هيلبيرت ولوبيز٢٠١١- Hilbert & Lopez، ٢٠١١) (أن نسبة النمو السنوي للمعلومات المخزنة بلغت ٢٣٪ ولو عدنا قليلا إلى الوراء لوجدنا أن القدرة العالمية على تخزين المعلومات قد تضاعفت خلال الفترة الممتدة بين ١٩٨٦ إلى ١٩٩٦ بنسبة ٨٠٠٪،

إن التزايد المستمر لمخزون العالم المعلوماتي والمتبادل على الشبكة والمتنامي نموا أسيا
يمثل تحديا كبيرا أمام النظم المعلوماتية التي تتحكم في طرق تبادل وتسيير المعلومات غير أنها لا تتحكم كلية في طبيعة المعلومات وذلك ما أدى إلى استحداث برمجيات ذكية لتحليل المعطيات Big Data Analysis، وتكمن أهمية هذا النوع من البرمجيات في استعماله للطرق الاحصائية لتصنيف المعلومات ومن ثم استغلال المعطيات الناتجة للتنبؤ بالتوجهات السائدة في الأسواق العالمية او حتى لتحليل المشاعر ضمن الشبكات الاجتماعيةSentiment Analysis أو لاسترجاع المعلومات Information retrieval

٣- المعطيات الضخمةBig Data واللغة العربية
بالرغم من ضخامة حجم المعلومات على الشابكة إلا أن نسبة المحتوى الرقمي باللغة العربية لا تفوق ٣٪ ، نسبة ضئيلة لا تعكس بأي حال من الأحوال رصيدنا الفكري والحضاري لكنها ليست مسألة حتمية، وبإمكاننا استغلال التقانات المتوفرة وتطويعها لخدمة اللغة العربية فبالنسبة لنظم الترجمة الآلية الاحصائية مثلا والتي تعاني من شح الموارد اللغوية لتدريب نماذج الترجمة على غرار المتون اللغوية، يمكن استغلال الحجم الهائل للمعلومات المخزنة على الشابكة لإنشاء متون لغوية linguistics corpora أحادية اللغة وثنائية اللغة ومتون متخصصة.
ومن ناحية أخرى، يتعين على اللغويين وبمساعدة المختصين استحداث المصطلحات التي
تواكب التقانات المتقدمة وذلك من خلال استغلال الكلمات العربية الصحيحة والأوزان المناسبة.

٤- نظم الترجمة الآلية الاحصائية
تنتهج نظم الترجمة الاحصائية أسلوبا رياضيا في تعاملها مع الترجمة حيث أنها تستعمل متونا لغوية (ذخائر نصية) أحادية اللغة وثنائية اللغة مصفوفة على مستوى الجمل حيث يتم اسناد احتمال لكل ثنائية جمل تمثل ترجمة ممكنة. وتقوم منظومة الترجمة الآلية باختيار الثنائية التي أُسْنِدَ إليها أعلى احتمال من بين كل الثنائيات المتوفرة، لتشكل هذه الجملة في النهاية الترجمة المقترحة ضمن النص المُخْرَج.



بنية نظم الترجمة الآلية بمقاربة إحصائية

-١ مواطن القوة :
- تكمن في قدرتها على معالجة (ترجمة) كم هائل من المعلومات في وقت وجيز.
- نظم قابلة للتحديث: إن بنية النظم الاحصائية الغرضية التوجه object oriented
تسمح بترجمة أي ثنائية لغوية ويكفي توفير متن لغوي أحادي اللغة بالنسبة للغة الوصلtarget language إضافة إلى متن ثنائي اللغة مصفوف على مستوى الجمل.
- يمكن تحسين مستوى الترجمة بإدماج محلل صرفي ضمن نظام الترجمة الاحصائية.
- يمكن تحسين مستوى الترجمة من خلال استعمال متون لغوية موسومة annotated corpora .

٢-٤ مواطن الضعف:
- ترتبط نظم الترجمة الاحصائية ارتباطا وثيقا بالمتون اللغوية، فإذا كانت المتون اللغوية التي تم استعمالها لتدريب نماذج الترجمة محدودة لغويا وذات حجم صغير، فإن ذلك سيؤثر حتما على مستوى ترجمة النص المنتج، إضافة إلى ذلك فإن الكلمات الغير واردة في متن التدريب ووردت في النص المُدْخَل)نص اللغة الأصل Source language (لن تتم ترجمتها.

٥- معايير تقييم الترجمة
وضع خبراء الترجمة الآلية معايير يتم على أساسها تقييم جودة النص المُتَرجَم وذلك من
خلال إسناد درجات لتقدير جودة الترجمة وهي كالتالي:
١- المعيار الأول هو مستوى بلاغة فصاحة النص المُتَرْجَمْ fluency أو ما يُسَمَّى تقنيا بمقاييس «التصويب النحوي»
٢- المعيار الثاني هو «أمانة الترجمة» adequacy ويشير هذا

المصطلح لترجمة أمينة، بلغ مستواها الحد الأمثل المقبول من التناسق مع النص الأصلي.

٣- أمّا المعيار الثالث «الإفادة المعلوماتية» informativeness، فقد قامت وكالة مشاريع البحث المتقدمة ARPA بإضافته، ويتمثل في تقدير كَمّ المعلومات (القيمة المعلوماتية) المُحالَة فِعليًا informativeness من قِبَل النص المُتَرْجَم - Informativeness measure.

٦ – أمثلة عن مقاييس التقييم التلقائي

تم إنشاء هذه البرمجيات لحساب مدى التطابق بين جملة مترجمة آليا وترجمة بشرية تستعمل كمرجع. ونفترض أنه إذا كان الفرق ضئيلا بين الجملتين فإن ذلك يعني أن الترجمة جيدة. وتختلف طرق الحساب من مقياس تقييم آلي لآخر :

فبالنسبة لمقاييس: GTM، NIST ، BLEU ، يتم حساب التطابق من خلال إحصاء عدد الكلمات المشتركة بين الجملة المنتجة من قبل نظام الترجمة الآلية والجملة المرجعية.

ويمثل مقياس TER، فبالإضافة إلى عمليات التنقيح (التعديل) edition والإدراج insertion والاستبدال substitution والحذف deletion تم استحداث عملية الإزاحة shift وتسمح هذه العملية بتغيير مواضع مجموعة من الكلمات المتجاورة contiguous عن طريق إزاحتها إلى اليمين أو إلى اليسار. وتُصَنَّفُ كل عملية إزاحة shift على أنها عملية تنقيح (تعديل) edition واحدة أيًا كان عدد الكلمات المُزَاحَة، ومهما كان امتداد الإزاحةshift ، ويُرْمَزُ لعدد عمليات الإزاحة shift بـ Nshf أمّا صيغة حساب درجة المقياس TER فهي كالتالي :

عند توفر عدة ترجمات مرجعية ، تُحَدَّدُ درجة مقياسTER من جهة، من خلال التعديلات الواجب القيام بها لتعديل الفرق بين الترجمة المرشحة والترجمة المرجعية الأقرب لهذه الجملة، ومن جهة أخرى من خلال متوسط عدد الكلمات الواردة ضمن جمل الترجمة المرجعية.

ملاحظة

بالنسبة لبحثنا استعملنا مقياسBLEU لتقييم نتائج ترجمة Moses لثنائية اللغات (إنجليزي فرنسي) بلغت نسبة ٣١,١٠٪ علما بأن المتن الذي استُعْمِلَ لتدريب نماذج الترجمة يحتوي على ١,٤ مليون جملة إنجليزية مصفوفة مع ترجمتها باللغة الفرنسية وهي مأخوذة من محاضر جلسات البرلمان الأوروبي.

و قد بلغت نتائج تقييم ترجمة نموذجنا ٣١,٣٥٪ بالرغم من استعمالنا لمتن محدود من حيث المضمون ومن حيث الحجم، فإننا نتوقع أن تكون نتائج الترجمة أفضل بكثير إذا توفر لدينا متن لغوي نظير غني ويعكس بحق ثراء ثنائية اللغات (إنجليزية- عربية) ويكون بمستوى وتنوع متن محاضر جلسات البرلمان الأوروبي.

- الخلاصة و التوصيات

لم تشهد البشرية تدفقا للمعلومات بالحجم الهائل الذي تشهده اليوم ويرجع ذلك لتطور التقانات التي أصبحت اليوم تستوعب كل هذه المعلومات وتتيح تخزينها رغم حجمها الهائل، إضافة إلى تسهيل تنقلها عبر الشابكة بسرعة تتزايد باستمرار، وبناء على ما سبق، فإن التعامل مع هذه الأحجام الهائلة من المعلومات لا يمكن أن يكون إلاّ من خلال الطرق الإحصائية، هذه الطرق التي لا غنى عنها سواء تعلق الأمر بتحليل البيانات الضخمة Big Data ووصولا إلى الترجمة الآلية بمقاربة إحصائية. وقد سعى هذا البحث إلى دراسة آلية الترجمة ضمن نظم الترجمة الآلية الإحصائية مرورا بطرق التقييم الآلية لجودة الترجمة وقد وضحنا نقاط الضعف والقوة ضمن هذ النظم. وقد كانت نتائج البحث مشجعة مقارنة مع الموارد المحدودة التي تم استعمالها فقد بلغت نتيجة تقييم مقياس BLEU نسبة ٣١,٣٥٪ بالرغم من استعمالنا لمتن محدود من حيث المضمون ومن حيث الحجم. وقد توصلنا إلى أن بعض الخطوات البسيطة من شأنها تحسين هذه النسبة:

- بإمكاننا إثراء نتائج هذا البحث من خلال تحسين نتائج التحليل اللغوي وعن طريق إدماج معاجم ثنائية اللغات ومتون لغوية موسومةannotated وباستعمال تقنيات تصفيف المتون النظيرة التي يتبناها الخبراء في نفس الإطار.
- لقد كانت نتائج ترجمة المحرك الإحصائي Moses والذي قمنا بتكييفه لدعم ثنائية اللغات (إنجليزية- عربية) مقبولة إلى حَدٍّ ما. غير أن طموحنا أكبر وقناعتنا أن وجود متون غنية وذات حجم كبير من شأنه إحداث قفزة نوعية على مستوى جودة الترجمة، وتجدر الإشارة إلى أنه قابل للتَّكَيُّفِ مع أي ثنائية للغات ويكفي فقط توفير المتون اللغوية الداعمة للثنائية التي نريد إنشاء النموذج من أجلها
- تواجه اللغة العربية تحديات كبيرة في عالم أضحى يتبنى لغة الأرقام في كل مجالات الحياة، وإذا كانت نسبة تمثيل المحتوى الرقمي العربي على الشابكة لا تعكس ثراءها ورصيدها الحضاري فإنها حتما تعكس هزالة الاستثمارات والمشاريع التي تسعى لتغيير هذا الوضع.
- إن اختيارنا لهذا الموضوع ينبع من حرصنا على اللغة العربية ومن ثم أردنا ان نبين

أن طرق الذكاء الاصطناعي (MOSES) التي تم تطويرها في الأصل لثنائيات لغوية أخرى، يمكن تطويعها وتكييفها لتستوعب اللغة العربية غير أن الكفاءة العلمية لا تكفي في هذه الحالة لرأب الصدع وملء الفجوة العلمية التي تفصلنا عن العالم المتقدم وقد

أدركت الدول المتقدمة خطورة الرهانات المتعلقة باللغة في عصر المعلومات وسارعت إلى تطويرها والاستثمار في البحوث التي من شأنها تأمين بقاء لغاتها على الشابكة إضافة إلى تعزيز مكانتها من خلال استحداث مصطلحات تواكب الانفجار المعرفي الذي يميز هذه الحقبة، ونأمل أن تحذو دولنا حذو هذه الدول .

قائـــمـة المـــراجـــع :

What's new in Statistical MachineTranslation، By Kevin Knight and Philipp Koehn
Information Sciences Institute، University of Southern California (pdf Paper).
٢ Machine translation: a brief history. By W.John Hutchins، From: 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، ١٩٩٥. Pages ٤٤٥-٤٣١]
٣ The Oxford hand book of computational linguistics،By Ruslan Mitkov ،Oxford University Press، ٢٠٠٣ :
- Chapter ٢٧ : pages ٥١١-٥٠٠، Machine Translation :General overview ، By John Hutchins،
- Chapter ٢٨ : pages ٥٢٧-٥١٢، Machine Translation :Latest developements، By Harold Somers.

٤ Translation Engines:Techniques for Machine Translation، By Arturo Trujillo.
Edition: Springer-Verlag London، Limited ١٩٩٩.
٥ Statistical Machine Translation، Philipp Koehn Part ١: Morning Session، University of Edinburgh September، ٢٠٠٧ ،١٠. (pdf Paper)
٦ Multiligual systems (basic strategies) : Machine translation : a brief history . By W.John Hutchins
٧ A Statistical approach to machine translation By Peter F. Brown، John Cocke، Stephen A. Della Pietra، Vincent J. Della Pietra، Fredrick Jelinek، John D. Lafferty، Robert L. Mercer، and Paul S. Roossin. IBM Thomas J. Watson Research Center Yorktown Heights، Computational linguistics revue، volume ١٦ N Hutchins
٨ Moses Statistical Machine Translation System User Manual and Code Guide، By Philipp Koehn، May ،٢٣ ٢٠١١، page ١٤١ to ١٥٧.
٩ A A Statistical approach to machine translation By Peter F. Brown، John Cocke، Stephen A. Della Pietra، Vincent J. Della Pietra، Fredrick Jelinek، John D. Lafferty، Robert L. Mercer، and Paul S. Roossin. IBM Thomas J. Watson Research Center Yorktown Heights، Computational linguistics revue، volume ١٦ N
١٠ Challenges in Statistical Machine Translation By Philipp Koehn Computer Science and Artificial Intelligence Lab -Massachusetts Institute of Technology – ٢٠٠٤
١١ Advancing Best Practices in Machine Translation Evaluation TAUS Labs Report - July ٢٠١٢
١٢ Computational architectures of MT systems By Christian BOITET GETALP-LIG-UJF، Grenoble، France - ٢٠٠٩
١٣ Example–Based Machine Translation: http://en.wikipedia.org/wiki/Example-based_machine_translation
١٤ Design of the Moses Decoder for Statistical Machine Translation، By Hieu Hoang، Philipp Koehn University of Edinburgh ، Software Engineering، Testing، and Quality Assurance for Natural Language Processing، workshop ACL٠٨-: HLT، June ٢٠٠٨ ،٢٠ The Ohio State University Columbus، Ohio، USA. (pdf Paper).
١٥ CGAL library : (Fabi and Al ٢٠٠٠) Paper.
١٦ Moses Statistical Machine Translation system ، Philip Koehn
١٧ http://www.medar.info/index.php.
١٨ Fundamental and New Approaches to Statistical Machine Translation، By Lucia Specia (pdf).
١٩ Human-Computer Interaction and Knowledge Discovery in Complex، Unstructured، Big Data - Third International Workshop، HCI-KDD ٢٠١٣ - SouthCHI ٢٠١٣ Maribor، Slovenia، July ٢٠١٣، Proceedings. Springer Edition.
٢٠ Head First Data Analysis : A learner's guide to Big Numbers، Statistics، and Good decisions– By Michael Milton- O'REILLY Media Edition ٢٠٠٩.

مـراجـع باللـغـة الـعـربـيـة

١. لسان العرب لابن منظور- الاصدار ١,٠٣ pdf book..

٢. المعجم الوجيز، مجمع اللغة العربية، جمهورية مصر العربية الطبعة الأولى، ١٩٨٠.

٣. معجم القاموس المحيط، الطبعة الثانية. مجد الدين محمد بن يعقوب الفيروز آبادي، توثيق: خليل مأمون شيحا. دار المعرفة للطباعة والنشر والتوزيع، بيروت – لبنان، ٢٠٠٧.

٤. معجم إلكتروني: الباحث العربي- قاموس عربي-عربي www.baheth.info الموقع الإلكتروني:

٥. قاموس المورد عربي- إنجليزي لِلدكتور : روحي البعلبكي
٤١st ed، Munir Ba›albki، Dar El-. ٢٠٠٧ ،Al-Mawrid: A Modern English – Arabic Dictionary
Ilm Lil-Malayen، Beirut – Lebanon

٦. قاموس المبرق: للدكتور محمد إبراقن قاموس موسوعي للإعلام والاتصال فرنسي – عربي
طبعة ثانية منقحة، منشورات ثالة- الأبيار، الجزائر، ٢٠٠٧