N° d'ordre :
N° de série :

الجمهورية الديمقراطية الشعبية الجزائرية
وزارة التعليم العالي والبحث العلمي

جامعة الشهيد حمة لخضر - الوادي
كلية العلوم الدقيقة
قسم الإعلام آلي

مذكرة نهاية الدراسة لنيل شهادة

ماستر أكاديمي

الميدان: رياضيات واعلام آلي
الشعبة: اعلام آلي
التخصص: أنظمة موزعة وذكاء اصطناعي

من اعداد وتقديم:

- غريسي نور الدين
- غريسي ياسين

الموضوع:

نمذجة وتصميم نظام ترميز دلالي من أجل الترجمة الآلية (إنجليزية - عربية)

نوقشت بتاريخ 2019/06/20 أمام أعضاء اللجنة:

الأستاذ: بن على عبد الكامل	رئيس ومناقش
الأستاذ: نديوي عبد الحميد	مناقش
الأستاذ: مفتاح محمد شرف الدين	مؤطر

الموسم الجامعي: 2018/ 2019

ملخص

تعد مهمة الترجمة الآلية من أهم المهام التابعة لفرع المعالجة الآلية للغة الطبيعية، لكن النظم الخاصة بها لازالت بعيدة جدا عن تحقيق آداء قريب من الترجمة البشرية المثالية بسبب العديد من العوائق و الصعوبات التي من أهمها : عدم محافظة الترجمة على الدلالة التي يقصدها الكاتب وكذلك مشكلة الغموض بأنواعه المختلفة. في هذا العمل نقوم بتصميم نظام ترميز دلالي يتمثل عمله الأساسي في ارفاق المعنى بالنص أثناء تحريره من طرف الكاتب، و رفع الغموض عنه بهدف انتاج ملف يحتوي على نص جاهز للترجمة لاحقا بواسطة النظم المختصة بذلك. وفي الأخير نقوم باختبار هذا النظام من أجل تقييم آداءه.

الكلمات المفتاحية : لغويات الحاسوب، المعالجة الالية للغة الطبيعية، الترجمة الالية، الترجمة الالية بمساعدة الانسان، الدلالة، الغموض.

Abstract

The task of Machine Translation (MT) is one of the most important tasks related to the branch of Natural Language Processing (NLP), but its systems are still very far from achieving a performance close to the ideal human translation because of many obstacles and difficulties, the most important of which: not to maintain the meaning intended by the writer as well as the problem of ambiguity and its different types. In this work we design a semantic coding system whose main function is to attach the meaning to the text while it is being edited by the writer, and to raise its ambiguity in order to produce a file* containing text ready for translation later by the relevant systems. Finally, we test this system to evaluate its performance.

Key words : Computational Linguistics, Natural Language Processing, Machine Translation, Computer Aided Human Translation, semantics, ambiguity.

Résumé

La traduction automatique (MT) est l'une des tâches les plus importantes de la branche du traitement automatique du langage naturel (NLP), mais ses systèmes sont encore très loin d'atteindre une performance proche de la traduction humaine idéale en raison de nombreux obstacles et difficultés, le plus important étant: ne pas conserver le sens voulu par l'auteur ainsi que le problème de l'ambiguïté et de ses différents types. Dans ce travail, nous concevons un système de codage sémantique dont la fonction principale est d'attacher le sens au texte pendant son édition par le rédacteur et de lever son ambiguïté afin de produire un fichier contenant du texte prêt à être traduit ultérieurement par les systèmes concernés. Enfin, nous testons ce système pour évaluer ses performances.

Mots clés: Linguistique informatique, Traitement automatique du langage naturel, traduction automatique, traduction humaine assistée par ordinateur, sémantique, ambiguité.

الفصل الأول : مدخل و نظرة عامة

نحاول أن نعطي في هذا الفصل، وإن كان بشكل موجز، نظرة عامة عن مجال الترجمة الآلية من خلال النقاط التالية: نبذة تاريخية عن الترجمة الآلية، المقاربات والبنيات المختلفة لنظم الترجمة الالية التي تم تطويرها، تدخل الانسان في عملية الترجمة الالية، طبيعة مستخدم نظام الترجمة الالية، مختلف الصعوبات و التحديات التي تواجه هذه الانظمة، وبدقة نحاول الإجابة على سؤال مهم وهو كيفية تقييم الترجمة الآلية ؟

الفصل الثاني : نظم الترجمة الآلية الدلالية

نتعمق في هذا الفصل اكثر في أنظمة الترجمة التي تعتمد على الدلالة والمعنى في عملها، ثم نقوم باختبار و مقارنة بعض من أشهر نظم الترجمة المعروفة حاليا. هنا، نحن نركز على كيفية استخدام الدلالة في أنظمة MT و نذكر اهم تحدياتها والمشاكل الخاصة بها. بادئ ذي بدء، نتحدث عن "ماذا نعني بالدلالي؟". ثانياً، نناقش التحديات والصعوبات التي تواجه هذه الأنظمة. ثم نذكر بعض الأنظمة الحالية القائمة على الدلالة. وأخيرًا، نقوم بمقارنة هذه الأنظمة، مع توضيح عيوب ومزايا كل نظام.

الفصل الثالث : نظام الترميز الدلالي المقترح

في هذا الفصل نقدم مقاربتنا المتمثلة في نظام ترميز دلالي مساعد على الترجمة الالية، والذي يتكون من ثلاث مكونات رئيسية وهي مكون العرض ، مكون التحكم و مكون النموذج. حيث يحتوي مكون العرض على واجهة تحرير النص التي تلعب دور الوسيط بين الكاتب والنظام، في حين يحتوي مكون التحكم على مختلف وحدات التحليل التي يمر بها النص اثناء معالجته وأخيرا مكون النموذج الذي يحتوي على كافة البيانات التي يحتاجها النظام.

الفصل الرابع : التنفيذ و التجريب

بعد ان شرحنا مقاربتنا في الفصل السابق، نقوم في هذا الفصل الاخير بتقييم النظام المقترح عبر تقييم كل عملية على حدى: فك الغموض المعجمي، فك الغموض الثقافي وفك الغموض التركيبي، حيث اظهر هذا التقييم نتائج مشجعة على المواصلة في هذا الطريق واضافة العديد من التعديلات التي من شأنها تحسين أداء النظام.

شـــــكـــــر

قال رسول الله صلى الله عليه و سلم:

"من لم يشكر الناس لم يشكر الله"

صدق رسول الله صلى الله عليه و سلم

الحمد لله على إحسانه و الشكر له على توفيقه و إمتنانه و نشهد أن لا إله إلا الله وحده لا

شريك له تعظيما لشأنه و نشهد أن سيدنا و نبينا محمد عبده و رسوله الداعي إلى رضوانه

صلى الله عليه و على آله و أصحابه و أتباعه و سلم.

بعد شكر الله سبحانه و تعالى على توفيقه لنا لإتمام هذا البحث المتواضع نتقدم بجزيل الشكر إلى والدينا العزيزين الذين أعانونا و شجعونا على الإستمرار في مسيرة العلم و النجاح، و إكمال الدراسة الجامعية و البحث؛ كما نتوجه بالشكر الجزيل إلى من شرفنا بإشرافه على مذكرة بحثنا الأستاذ الدكتور " مفتاح محمد شرف الدين" الذي لن تكفي حروف هذه المذكرة لإيفائه حقه بصبره الكبير علي، ولتوجيهاته العلمية التي لا تقدر بثمن؛ و التي ساهمت بشكل كبير في إتمام و إستكمال هذا العمل؛ إلى كل أساتذة قسم الإعلام الآلي؛ كما نتوجه بخالص شكرينا و تقديرنا إلى كل من ساعدنا من قريب أو من بعيد على إنجاز و إتمام هذا العمل.

" رب أوزعني أن أشكر نعمتك التي أنعمت علي و على والدي و أن أعمل صالحا ترضاه

و أدخلني برحمتك في عبادك الصالحين"

إهـــــــــــداء

إلى الذي وهبني كل ما يملك حتى أحقق له آماله، إلى من كان يدفعني قدما نحو الأمام لنيل المبتغى، إلى الإنسان الذي إمتلك الإنسانية بكل قوة، إلى الذي سهر على تعليمي بتضحيات جسام مترجمة في تقديسه للعلم، إلى مدرستي الأولى في الحياة،**أبي الغالي** على قلبي أطال الله في عمره؛

إلى التي وهبت فلذة كبدها كل العطاء و الحنان، إلى التي صبرت على كل شيء، التي كانت سندي في الشدائد، و كانت دعواها لي بالتوفيق، تتبعتني خطوة خطوة في عملي، إلى من إرتحت كلما تذكرت إبتسامتها في وجهي نبع الحنان **أمي** أعز ملاك على القلب و العين جزاها الله عني خير الجزاء في الدارين؛

إليهما أهدي هذا العمل المتواضع ليك أُدخل على قلبهما شيئا من السعادة

إلى **إخوتي و أخواتي** الذين تقاسمو ا معي عبء الحياة ؛

كما أهدي ثمرة جهدي لأستاذي الكريم:**مفتاح محمد شرف الدين** الذي كلما تظلمت الطريق أمامي لجأت إليه فأنارها لي و كلما دب اليأس في نفسي زرع فيا الأمل لأسير قدما و كلماسألت عن معرفة زودني بها و كلما طلبت كمية من وقته الثمين وفره لي بالرغم منمسؤولياته المتعددة؛ إلى كل أساتذة قسم العلوم الإقتصادية و علوم التسيير؛

إلى كل هؤلاء أهدي هذا العمل.

الطالب: غريسي نورالدين

إهـــــــــــداء

الى تلك التي لم تدخر نفسا في تربيتي -**أمي** الحبيبة .

الى من تشققت يداه في سبيل رعايتي – **أبي** الحنون .

الى كل من علمني حرفا بدئا بأساتذة الابتدائي ومرورا بالمتوسط والثانوي وأخيرا وليس أخرا اساتذتي في طور الجامعة وعلى رأسهم الأستاذ المؤطر لنا " **مفتاح محمد شرف الدين** " .

الى اخوتي : **كمال** و **الحبيب** و**علي** و**أحمد** .... و **أخواتي** الكريمات .

الى أصدقائي **عبد النور** و **عصام** و **طارق** و **عبد المجيد** و **جبوري** .

الى كل من نصحني لأدرس **اعلام الي** والى كل من مد يد العون في إتمام هذا العمل .

الى **زوجتي** المستقبلية وابنائي ان كتب لنا الله في طول العمر .

وفي الأخير ارجو ان يكون هذا العمل والمشروع خالصا لوجه الله وان تكون فيه منفعة لمن يطلع عنه ويرزقنا الاجر والثواب .

الطالب : غريسي ياسين

جدول المحتويات

ملخص I

شكر III

إهداء IV

قائمة الرسوم التوضيحية VI

قائمة الجداول VII

قائمة الصور VII

1. مقدمة عامة **11**

2. الفصل الاول: مدخل ونظرة عامة **14**

1.2 المقدمة 14

2.2 تعريف 14

3.2 نبذة تاريخية 14

4.2 الصعوبات والتحديات 16

1.4.2 الغموض المعجمي 16

2.4.2 الغموض التركيبي 17

3.4.2 الغموض الثقافي 18

5.2 تدخل الانسان في عملية الترجمة 19

6.2 طبيعة المستخدم 21

1.6.2 الترجمة الالية للمراقب 21

2.6.2 الترجمة الالية للمراجع 22

3.6.2 الترجمة الالية للمترجم 22

4.6.2 الترجمة الالية للمؤلف 22

7.2 مقاربات 22

1.7.2 النظم القائمة على القواعد 22

2.7.2 النظم التجريبية 23

8.2 التقييم 25

1.8.2 **BLEU (BiLingual Evaluation Understudy)** (دراسة تقييمية ثنائية اللغة) 25

2.8.2 **WER (Word Error Rate)** (معدل الخطأ في الكلمات) 25

3.8.2 **PER (Position-independent word Error Rate)** (معدل الخطأ لكلمة الموضع المستقل) 25

4.8.2 **(Translation Error Rate) TER** (معدل خطأ الترجمة) **.....** 25

9.2 الخاتمة 25

3. الفصل الثاني: نظم الترجمة الآلية الدلالية 27

1.3 المقدمة 27

2.3 ما هو علم الدلالة 27

3.3 لماذا ندرس الدلالة 27

4.3 الصعوبات والتحديات 27

1.4.3 التمثيل المحوسب للمعاني **.....** 27

2.4.3 بناء قاعدة المعارف **.....** 28

5.3 مشكلة الغموض المعجمي ومساهمات الاخرين في حلها 28

1.5.3 طرق تعتمد على قواعد المعارف**.....** 28

2.5.3 طرق تعتمد على تعلم الالة **.....** 29

6.3 دراسة بعض نظم الترجمة الالية الموجودة حاليا 31

1.6.3 الأنظمة العربية **.....** 31

2.6.3 الأنظمة الاجنبية**.....** 31*

3.6.3 مقارنة**.....** 32

4.6.3 مناقشة النتائج**.....** 35

5.6.3 خلاصة التجربة **.....** 35

7.3 الخاتمة 36

4. الفصل الثالث: نظام الترميز الدلالي المقترح 38

1.4 المقدمة 38

2.4 ماذا نقصد بالترميز الدلالي 38

3.4 بنية النظام 39

1.3.4 مكون العرض: واجهة تحرير النص **.....** 39

2.3.4 مكون النموذج **.....** 40

3.3.4 مكون التحكم **.....** 47

4.4 التفاعل بين النظام وكاتب النص..... 66

5.4 خاتمة 69
5. الفصل الرابع: التنفيذ و التجريب 71
1.5 المقدمة 71
2.5 أدوات، واجهات البرمجة و بيئات التطوير المستخدمة 71
1.2.5 لغة بايثون **Python 71**
2.2.5 بيئة طوني **Thonny 71**
3.2.5 حزمة **NLTK..... 71**
4.2.5 حزمة **Tkinter.....72**
3.5 الواجهة الرسومية 72
1.3.5 خطوات تحرير النص **.....73**
4.5 التجارب، النتائج و المناقشة 74
1.4.5 تقييم عملية فك الغموض المعجمي **.....75**
2.4.5 تقييم عملية فك الغموض الثقافي **.....76**
3.4.5 تقييم عملية ترجمة الكلمات من خلال السياق المحلي **.....78**
4.4.5 تقييم عملية فك الغموض التركيبي **..... 83**
5.5 الخاتمة..... 86
6. الخاتمة العامة 87
المراجع VIII

قائمة الرسوم التوضيحية

رسم توضيحي 1 : الترجمة البشرية 19
رسم توضيحي 2 : نظم التحرير السابق 20
رسم توضيحي 3 : نظم التحرير اللاحق 20
رسم توضيحي 4 : ترجمة بشرية بمساعدة الآلة 20
رسم توضيحي 5 : الترجمة الآلية بمساعدة الإنسان 21
رسم توضيحي 6 : ترجمة آلية تلقائية بالكامل 21
رسم توضيحي 7: مثلث Vauquois 23
رسم توضيحي 8 : موقع النظام المقترح في عملية الترجمة بأكملها 38
رسم توضيحي 9 : بنية النظام المقترح 39
رسم توضيحي 10 : العلاقات بين الحقول المختلفة للقاموس 42
رسم توضيحي 11 : : مخطط فئات او ما يسمى class diagram يوضح بنية القاموس 42
رسم توضيحي 12 : الطبقة العليا من التسلسل الهرمي لأسماء ووردنت 44
رسم توضيحي 13 : الطبقة العليا من التسلسل الهرمي لمجالات ووردنت 46
رسم توضيحي 14 : بنية بيانات التدريب الخاصة بالسياق 47
رسم توضيحي 15 : بنية ملف xml الخاص بالنص بعد مرحلة التجزيء 48
رسم توضيحي 16 : بنية ملف xml الخاص بالنص بعد مرحلة التحليل الصرفي 48
رسم توضيحي 17 : يوضح خوارزمية التحليل التركيبي 50
رسم توضيحي 18 : شجرة التحليل النحوي المعتمد على علاقات التبعية 51
رسم توضيحي 19: خوارزمية التحليل الدلالي 53
رسم توضيحي 20 : تداخل المجالات في النصوص 54
رسم توضيحي 21 :مثال على توزيع معاملات بعض المجالات على معاني كلمة معينة ضمن مشروع مجالات ووردنت الموسعة XWND . 57
رسم توضيحي 22: خطوة اختزال شجرة التركيب 58
رسم توضيحي 23 : خطوة تحميل المعاني من القاموس 59
رسم توضيحي 24 : خطوة نشر و توزيع العلاقات النحوية بين المعاني 59
رسم توضيحي 25: تقييم العلاقات النحوية بين معاني كلمتين: 60
رسم توضيحي 26 : تقييم العلاقات النحوية بين معاني كل الكلمات 60
رسم توضيحي 27 : تجاهل العلاقات المرفوضة بين المعاني 61
رسم توضيحي 28 : تصفية المعاني 61
رسم توضيحي 29 : رسم توضيحي 29 : حساب نسبتي التشابه المباشر و العكسي 62
رسم توضيحي 30 : استنتاج خصائص سياقية جديدة بفضل علاقات الشمول في ووردنت 66
رسم توضيحي 31 : مختلف المهام التي يمكن للكاتب القيام بها 67
رسم توضيحي 32 : التفاعل بين النظام و الكاتب 69
رسم توضيحي 33 : نتيجة اختبار فك الغموض الثقافي 78
رسم توضيحي 34 : مقارنة دقة الاداء بين الانظمة المختلفة 79
رسم توضيحي 35 : نتيجة اختبار فك الغموض التركيبي 84
رسم توضيحي 36 : شجرتي التحليل التركيبي لجملة غامضة 85

قائمة الجداول

جدول 1 يحتوي على جمل تظهر فيها كلمتي break و sound بمعاني مختلفة 17
جدول 2 يحتوي على جمل ذات غموض تركيبي 17
جدول 3 : تعابير اصطلاحية مباشرة 18
جدول 4 : تعابير اصطلاحية غير مباشرة 19
جدول 5 : بنية القاموس 41
جدول 6 : شرح بعض العلاقات النحوية 52
جدول 7 : مثال يحاكي كيفية توزيع معاملات مجال معين على كل المعاني ضمن مشروع مجالات ووردنت الموسعة XWND 56
جدول 8 : تقييم العلاقات النحوية بين معاني كلمتين 64
جدول 9 : معاني كلمة window 65
جدول 10 : نتيجة اختبار فك الغموض المعجمي 75
جدول 11 : نتيجة اختبار فك الغموض الثقافي 77
جدول 12 : نتيجة اختبار ترجمة كلمة break من خلال السياق المحلي لها 79
جدول13 : التفاصيل الكاملة للنتائج المتحصل عليها 81
جدول 14 : نتيجة اختبار فك الغموض التركيبي 84

قائمة الصور

صورة 1 : قواعد التحليل التركيبي 51
صورة 2 : بنية ملف xml الخاص بالنص بعد مرحلة التحليل التركيبي 52
صورة 3 : بنية ملف xml الخاص بالنص بعد مرحلة التحليل الدلالي 52
صورة 4 : الواجهة الرسومية للنظام 72
صورة 5 : خطوات تحرير النص 74
صورة 6 : لقطة شاشة توضح نتيجة اختبار فك الغموض المعجمي 75
صورة 7 : لقطة شاشة توضح نتيجة اختبار فك الغموض الثقافي 77
صورة 8 : لقطة شاشة تظهر نتيجة اختبار فك الغموض التركيبي 83

1. مقدمة عامة

تعتبر بحوث الذكاء الاصطناعي أهم ميادين تكنولوجيات الاعلام والاتصال التي برزت نتائجها في شكل تطبيقات يستعملها الجميع. ومن أهم مجالات الذكاء الاصطناعي، المعالجة الآلية للغات الطبيعية أو ما يطلق عليه أيضا اللسانيات الحاسوبية والتي أثمرت نتائج تطبيقية لا يستغني عنها أحد مهما كان مجال عمله، كالتدقيق اللغوي الآلي، البحث المعلوماتي، التعرف الآلي على الكلمات، قراءة النصوص، الملخص الآلي، التصنيف الآلي و تأتي في مقدمتها كأقدم و أكبر مجال للبحث في ميدان اللسانيات الحاسوبية: الترجمة الآلية، والتي ظل الباحثون يقاومون و يتحدون تعقيدات اللغة و يقترحون مختلف المقاربات و المنهجيات لجعل الآلة قادرة على ترجمة نصوص من لغة إلى أخرى.

تعتبر الترجمة الآلية أهم ميدان بحث في اللسانيات الحاسوبية هدفها انتاج تطبيقات قادرة على ترجمة نص (مكتوب) أو كلام (منطوق) من لغة الى أخرى. قد يتخيل القارئ أن الترجمة الآلية تتمثل في ارفاق كلمة في اللغة الهدف الى كلمة في اللغة المنطلق بمجرد معاينة قاموس ثنائي اللغة، لكن الواقع أكثر تعقيدا مما يظهر، فهذا النوع من الترجمة يؤدي الى ضياع شكل وهيكل الجملة، والأهم من ذلك كله هو ضياع المعنى. لإيجاد حل فعلي لتعقيدات اللغة وانتاج ترجمة تولد جملا صحيحة في اللغة الهدف وتحافظ على المعنى الأصلي في اللغة المنطلق دون ارتباط ببنية الجملة الأصلية ظهرت عدة مقاربات وحلول لتجاوز هذه العقبات. كالمقاربة القواعدية، التجريبية والهجينة. تعتمد المقاربة القواعدية على مجموعة من قواعد لغوية و دلالية مصممة من طرف خبراء بشريين، هذه القواعد تساعد على تحليل جملة باللغة الأصل صرفيا و تركيبيا ودلاليا من اجل الوصول الى تعبير مكافئ بلغة الهدف. على عكس المقاربة القواعدية لا تعتمد المقاربة التجريبية على القواعد البشرية من اجل اكتساب معرفة الترجمة وانما تستمدها من ذخيرة لغوية على شكل ثنائيات ترجمة متوازية. أما المقاربة الهجينة فهي عبارة عن دمج لمقاربتين او اكثر من المقاربات القواعدية و التجريبية.

سنقدم ضمن هذا المشروع تصميم نظام ترميز دلالي يهدف الى تحرير ملفات نصية قابلة للترجمة وذلك من اجل تسهيل مهمة نظم الترجمة فيما بعد، بعد إزالة اهم العقبات امامها وهي : الغموض المعجمي والثقافي والتركيبي. تعتمد مقاربتنا أساسا على النهج القواعدي يتخلله النهج التجريبي في بعض العمليات، حيث تمر كل جملة عند معالجتها على مراحل عدة وهي : تجزيء النص، التحليل الصرفي، التحليل التركيبي و التحليل الدلالي. في المرحلة الأولى تقسم الجملة الى اصغر وحدة بنيوية فيها وهي الكلمات، وفي مرحلة التحليل الصرفي تُجَرَّد كل كلمة من ميزاتها الصرفية مع ارجاعها الى اصلها وإيجاد نوعها (اسم، فعل، صفة ...)، اما في مرحلة التحليل التركيبي فيتم اكتشاف بنية الجملة وتحليلها الى مكوناتها الجزئية (جملة اسمية، جملة فعلية، جملة جار و مجرور ...) مع إيجاد كل العلاقات

النحوية الرابطة بين الكلمات، واذا تم اكتشاف غموض تركيبي يُطْلَبُ من الكاتب إضافة الاقواس. وأخيرا في مرحلة التحليل الدلالي نظرا لكثرة المعاني المخصصة لكل كلمة فان هذه الأخيرة تمر بتصفية آلية للمعاني عبر عمليتي تصفية : الأولى حسب المجال العام للنص، اما الثانية فتكون حسب السياق المحلي للكلمة الغامضة، بعد هذا يأتي دور الكاتب في تصحيح التحديد الخاطئ للمعنى او تحديده ان بقيت غامضة، مع الإشارة الى انه في حالة عدم بقاء أي معنى مقبول للكلمة فانها تعتبر مجازية (او غامضة ثقافيا) فيطلب من الكاتب صياغتها بتعبير اخر عام.

الفصل الاول

مدخل و نظرة عامة

2.الفصل الاول: مدخل ونظرة عامة

1.2 المقدمة

بعد 65 عامًا، يعد هذا الحقل أحد أقدم تطبيقات أجهزة الكمبيوتر. على مر السنين، كانت الترجمة الآلية محور اهتمام الباحثين اللغويين وعلماء النفس والفلاسفة وعلماء الكمبيوتر والمهندسين. لن يكون من المبالغة القول أن العمل المبكر على الترجمة الالية ساهم بشكل كبير في تطوير مجالات مثل اللغويات الحاسوبية والذكاء الاصطناعي ومعالجة اللغة الطبيعية الموجهة للتطبيق. يمكن تعريف الترجمة الآلية، المعروفة باسم Machine Translation ، بأنها "الترجمة من لغة طبيعية (لغة المصدر) إلى لغة أخرى (اللغة الهدف) باستخدام أنظمة محوسبة، مع أو بدون مساعدة بشرية". نحاول أن نعطي في هذا الفصل، وإن كان بشكل موجز، نظرة عامة عن مجال الترجمة الآلية من خلال النقاط التالية: نبذة تاريخية عن الترجمة الآلية، المقاربات والبنيات المختلفة لنظم الترجمة الالية التي تم تطويرها، تدخل الانسان في عملية الترجمة الالية، طبيعة مستخدم نظام الترجمة الالية، مختلف الصعوبات و التحديات التي تواجه هذه الانظمة، وبدقة نحاول الإجابة على سؤال مهم وهو كيفية تقييم الترجمة الآلية؟

2.2 تعريف

يشير مصطلح الترجمة الآلية (Machine Translation MT) إلى الأنظمة المحوسبة المسؤولة عن إنتاج الترجمات بمساعدة بشرية أو بدونها. عادة ما يتم التمييز بين الترجمة بمساعدة الإنسان (Human Aided Machine Translation) والترجمة البشرية بمساعدة الآلة (MAHT Machine Aided Human Translation). يشتمل الأخير على أدوات ترجمة معتمدة على الكمبيوتر تدعم المترجمين من خلال توفير الوصول إلى القواميس على الإنترنت، قواعد بيانات المصطلحات عن بُعد، ونقل واستقبال النصوص، ومخازن النصوص المترجمة سابقًا ("ذاكرات الترجمة")، والموارد المتكاملة، والتي يشار إليها عادة باسم محطات عمل المترجم. يستخدم مصطلح الترجمة بمساعدة الحاسوب (Computer Assisted Translation) أحيانًا لتغطية جميع أنظمة الترجمة المعتمدة على الكمبيوتر. [1]

3.2 نبذة تاريخية

على الرغم من أنه من الممكن ان نتتبع أصول الترجمة الآلية إلى أفكار القرن السابع عشر(اللغات العالمية والفلسفية والقواميس الميكانيكية)، إلا أنه لم يتم تقديم اقتراحات عملية حتى القرن العشرين. يمكن تقسيم تاريخ الترجمة الآلية إلى خمس (05) فترات: [2]

الفترة الأولى (1948-1960): البداية

• 1949: اقترح وارن ويفر في مذكرته لعام 1949 الأفكار الأولى حول استخدام أجهزة الكمبيوتر في الترجمة، من خلال اعتماد مصطلح الترجمة الحاسوبية.

• 1952: الندوة الأولى للترجمة الآلية، بعنوان مؤتمر الترجمة الآلية، والتي عقدت في يوليو 1952 في معهد ماساتشوستس للتكنولوجيا تحت قيادة يهوشوا بار هيليل.

• 1954: تطوير أول مترجم آلي (أساسي جدًا) من قبل مجموعة من الباحثين من جامعة جورج تاون بالتعاون مع IBM، والذي يترجم إلى أكثر من ستين (60) جمل روسية إلى الإنجليزية. ادعى المؤلفون أنه خلال ثلاث إلى خمس سنوات، لن تكون الترجمة الآلية مشكلة.

• 1954: نشر Victor Yngve أول مجلة على الترجمة الالية، بعنوان "الترجمة الميكانيكية المكرسة لترجمة اللغات بمساعدة الآلات".

الفترة الثانية (1960-1966) التحليل التركيبي وخيبة الأمل

• أوائل الستينيات من القرن العشرين، تم تقديم هذا التحليل التركيبي باعتباره الوسيلة الوحيدة الممكنة للبحث لتعزيز الترجمة الآلية. وبالتالي هناك بالفعل العديد من المحللون المطورين من أنواع مختلفة من القواعد، مثل قواعد اللغة وقواعد التبعية.

• 1961: في فبراير من هذا العام، وُلدت اللغويات الحاسوبية، وذلك بفضل المحاضرات الأسبوعية التي نظمها ديفيد ج. هايز في مؤسسة راند في لوس أنجلوس. سيتم إدراج هذه المؤتمرات كأوراق في المؤتمر الدولي الأول للترجمة الآلية للغات وتحليل اللغة التطبيقية في تدينغتون في سبتمبر 1961 بمشاركة اللغويين وعلماء الكمبيوتر المشاركين في الترجمة مثل: بول جارفين، سيدني م. لامب، كينيث هـ. هاربر وتشارلز هوكيت ومارتن كاي وبرنارد فاكويس.

• 1964: إنشاء لجنة ALPAC (اللجنة الاستشارية لمعالجة اللغة التلقائية) مع الحكومة الأمريكية لدراسة وجهات نظر وفرص الترجمة الآلية

• 1966: نشرت ALPAC تقريرها الشهيرة التي خلصت فيها إلى أن أعمالها في الترجمة الآلية هي مجرد إهدار للوقت والمال ؛ هذا التقرير هو أنه كان له تأثير سلبي على بحثهم في الترجمة الالية لعدة سنوات.

الفترة الثالثة (1966-1980): ولادة جديدة وأمل

• 1970: بدء المشروع REVERSO من قبل مجموعة من الباحثين الروس.

• 1970: تطوير نظام SYSTRAN1 (روسي-إنجليزي) بقلم بيتر توما، الذي كان في ذلك الوقت عضوًا في مجموعة تبحث عن جورج تاون.

• 1976: إنشاء نظام الطقس في مشروع TAUM (الترجمة الآلية في جامعة مونتريال) تحت إشراف Alai Colmerauer للتنبؤات الجوية للترجمة الآلية لعامة الناس، تم إنشاء هذا النظام من قبل مجموعة من الباحثين.

• 1978: إنشاء نظام ATLAS2 من قبل الشركة اليابانية FUJITSU، واستند هذا المترجم على قواعد كما أنه قادر على الترجمة من الكورية إلى اليابانية والعكس بالعكس.

الفترة الرابعة (1980-1990): الاجتياح الياباني

• 1982: تقوم شركة SHARP اليابانية بتسويق مترجمها التلقائي DUET (الإنجليزية - اليابانية)، وقد استند هذا المترجم إلى قواعد وهو نهج لنقل الترجمة

• 1983: بصفتها عملاق الكمبيوتر، تقوم NEC بتطوير نظامها للترجمة استنادًا إلى خوارزمية تسمى "المحور". تم تسويقه تحت اسم Honyaku Adapter II، النسخة العامة لنظام الترجمة من NEC يعتمد أيضًا على طريقة المحور، باستخدام اللغة الوسيطية.

• 1986: تطوير نظام PENSEE بواسطة OKI3، وهو مترجم (ياباني-إنجليزي) استنادًا إلى القواعد.

• 1986: قامت مجموعة Hitachi بتطوير نظام الترجمة الخاص بها استنادًا إلى القواعد (وهو نهج تم اتخاذه عن طريق التحويل)، والمعروف باسم HICATS (نظام الترجمة بمساعدة الكمبيوتر من هيتاشي / اليابانية - الإنجليزية).

الفترة الخامسة (منذ 1990): الويب، الشيء الغامض و الجديد على المترجمين

• 1993: مشروع C-STAR (اتحاد البحوث المتقدمة لترجمة الكلام) هو تعاون دولي. موضوع المشروع هو الترجمة الآلية للإفراج المشروط في مجال السياحة (وكيل سفر عميل الحوار)، عن طريق الفيديو. هذه المشاريع ولدت نظام C-STAR I الذي تعامل مع ثلاث (03) لغات (الإنجليزية والألمانية واليابانية) وقدم أولى المظاهرات عبر الأطلسي بثلاث لغات في يناير 1993

• 1998: تسويق المترجم REVERSO من قبل شركة Softissimo.

• 2000: تطوير نظام ALPH بواسطة المختبر الياباني ATR، يأخذ هذا المترجم (اليابانية-الإنجليزية والصينية - الإنجليزية) مقاربة تستند إلى أمثلة.

• 2005: ظهور أول موقع ويب للترجمة التلقائية، مثل (Google (http://translate.google.fr/.

• 2007: METIS-II هو نظام للترجمة الآلية الهجينة، حيث يتم استخدام رؤى من الترجمة الآلية المستندة الى الاحصاء والى الامثلة و إلى القواعد (SMT و EBMT و RBMT على التوالي).

• 2008: استخدم 23٪ من مستخدمي الإنترنت الترجمة الآلية و 40٪ يفكرون في القيام بذلك

• 2009: 30 ٪ من المهنيين استخدموا الترجمة الآلية و 18 ٪ يقومون بتصحيح التجارب المطبعية.

• 2010: 28٪ من مستخدمي الإنترنت، استخدموا الترجمة الآلية و 50٪ يخططون للقيام بها.

4.2 الصعوبات والتحديات

1.4.2 الغموض المعجمي

تكون الجملة في النص المصدر غامضة من الناحية المعجمية حين تحتمل الكلمة أكثر من معنى. يكون للكلمة المتعددة المعاني معنى أول وهو المعنى الشائع او الجوهري ـأو الاساسي، ككلمة " stage " مثلا، لها معنى أول وشائع وهو "خطوة" ومعنى آخر أقل شيوعا ويتمثل في " خشبة مسرح " وقد يكون للكلمة معنى حقيقي ومعنى آخر مجازي، ولكن لا تتسع مساحة البحث إلى التطرق إليه. [3]

مثال: يعرض هذا الجدول معاني مختلفة لكلمات "break" و "sound" اعتمادًا على السياق الذي تظهر فيه:

	الجملة	المعنى
1	The boy broke the window	كسر الولد النافذة
2	The medicine will break you of smoking	سوف يخلّصك هذا الدواء من التدخين
3	Some people break the law	بعض الناس يخترقون القانون
4	Please, do not break your promise	من فضلك لا تخلف وعدك
5	The dawn breaks at 5 O'clock today	يبزغ الفجر عند الساعة 5 اليوم
6	I like the sounds of birds	أحب أصوات الطيور
7	Your suggestion sounds reasonable	يبدو اقتراحك معقولا
8	It is a sound basis	إنه أساس صلب
9	That sound is quite narrow	ذاك البرزخ ضيق جدا

جدول 1 يحتوي على جمل تظهر فيها كلمتي break *و* sound *بمعاني مختلفة* [3]

2.4.2 الغموض التركيبي

هو غموض يظهر على مستوى البنية في النص المصدر، وتكون الجملة غامضة إذا كان لها أكثر من معنى، ويزول الغموض إذا كانت الجمل منطوقة، ألن التنغيم والوقف يساعدان على فهم المعنى المقصود. [3]

مثال: يعرض الجدول التالي جمل مختلفة ذات غموض تركيبي :

	الجملة	المعنى 1	المعنى 2
1	The **old men and women** stayed at home	The old **(men and women)** stayed at home	The **(old men)** and women stayed at home
2	They fed her rat poison	They fed her (rat poison)	They fed (her rat) poison
3	He saw a man eating fish	He saw a man (eating fish)	He saw a (man eating) fish
4	Students hate annoying Professors	Students hate annoying (professors)	Students hate (annoying professors)
5	I said I would see you on Tuesday	I said I would see you on Tuesday	*Last Tuesday* I said I would see you

جدول 2 يحتوي على جمل ذات غموض تركيبي [3]

3.4.2 الغموض الثقافي

نقول ان هناك غموض ثقافي عندما يكون للجملة معنى اخر(غير المعنى الحقيقي) شائع ومصطلح عليه في المجتمع المتحدث باللغة الأصل. وإذا ما ترجمت هذه الجمل ترجمة مباشرة فستجعل النص الهدف يبدو غريبا، وهو ما سماه شلايماخر بتغريب النص حين ميّز ما بين الترجمة الحرفية والترجمة الحرة. ويظهر هذا الغموض خاصة في ترجمة الامثال والحكم والتعابير الإصطالحية، ولا تخلو دراسات الترجمة من نظريات ترجمية تعنى بالمشاكل الثقافية والحلول المناسبة لها .[3]

وفيما يلي أمثلة عن التعابير الاصطلاحية وكيفية ترجمتها في الترجمة الالية :

التعابير الاصطلاحية: هي عبارات جاهزة تحمل المعنى نفسه ولا يتغيّر بتغيّر المتكلم او السياق.

وتصنّف إلى مجموعتين :

- المجموعة الاولى: وهي التعابير الاصطلاحية المباشرة، و يمكن أن تترجم ترجمة مباشرة ولكنها تفهم بطريقة غير مباشرة، أي تترجم ترجمة حرفية ولكن ليس ذلك هو المعنى المقصود، مثل his name : was in the tip of my tongueإذ يمكن القول: كان اسمه على رأس لساني. ولكن المقصود هو " كنت على وشك ذكر اسمه".
- المجموعة الثانية: وهي التعابير الاصطلاحية غير المباشرة، فهي لا تترجم ترجمة مباشرة بل ترجمة حرة مثل My car is second hand إذ لا يمكننا القول "سيارتي يد ثانية" بل سيارتي مستعملة.

وفيما يلي أمثلة خاصة بالنوعين للنظر في كيفية ترجمتها ترجمة آلية :

المجموعة الاولى:

	الجملة	الترجمة الحرفية	الترجمة الحرة
1	His name was on the tip of my tongue	كان اسمه على رأس لساني	كنت على وشك ذكر اسمه
2	Some people put the car before the horse	يضع بعض الناس العربة أمام الحصان	يعكس بعض الناس الأمور
3	They apply the law of the jungle	يطبّقون شريعة الغاب	يطبّقون قانون القوي يأكل الضعيف
4	Let us turn a new page	لنفتح صفحة جديدة	لنبدأ من جديد

جدول 3 : تعابير اصطلاحية مباشرة [3]

المجموعة الثانية:

	الجملة	الترجمة الحرفية	الترجمة الحرة
5	Let us talk shop	لنتحدّث عن الحانوت	لنتحدّث عن العمل
6	My car is second hand	سيارتي يد ثانية	سيارتي مستعملة
7	He received a golden handshake	تلقى مصافحة ذهبية	تلقى مكافأة ضخمة (تكريما عظيما)
8	My brother does not stand a dog's chance	ليس لأخي فرصة كلب	ليس لأخي أية فرصة
9	There is far too much monkey business going on around here	هناك كثير من العمل القردي يجري حولنا	هناك كثيرمن الأشياء المريبة تجري من حولنا

جدول 4 : تعابير اصطلاحية غير مباشرة [3]

5.2 تدخل الانسان في عملية الترجمة

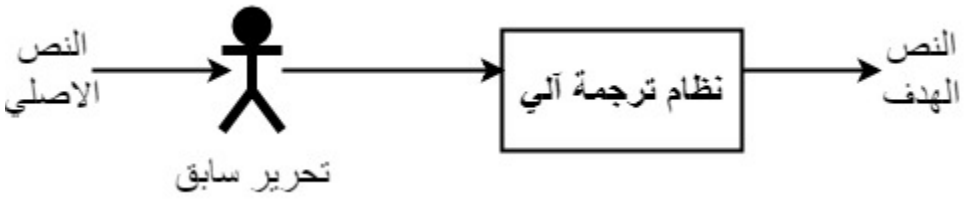
تنقسم أنظمة الترجمة إلى 6 أنواع وفقا لمدى تدخل الانسان في عملها: [4] [5]

- **الترجمة البشرية** : يقوم المترجم البشري بتنفيذ جميع الخطوات في عملية الترجمة، باستخدام الكمبيوتر فقط كمعالج نصوص، و ربما لا يستعمله اصلا.

رسم توضيحي 1 : الترجمة البشرية

- نظم التحرير السابق :

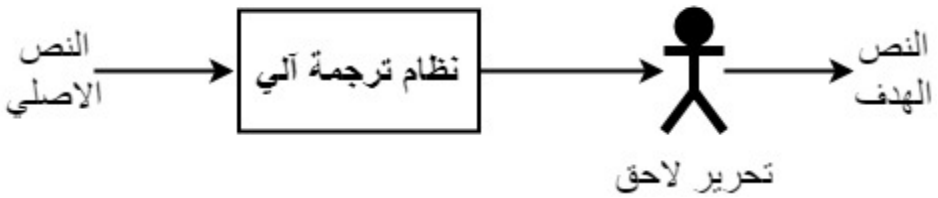
التحرير السابق هو معالجة النصوص قبل الترجمة الآلية. يتضمن عادةً تصحيح الأخطاء في النص المصدر (بشكل أساسي قواعد اللغة وعلامات الترقيم والإملاء)، وإزالة الغموض وتبسيط الهياكل. بالنسبة إلى الترجمة الالية الإحصائية، قد يتضمن أيضًا تكييف النص بطريقة تجعل نص الإدخال أقرب إلى النصوص التي تم تدريب المحرك عليها، مما يمكن أن يساعد محرك الترجمة الالية على الأداء بشكل أفضل.



رسم توضيحي 2 : نظم التحرير السابق

- نظم التحرير اللاحق

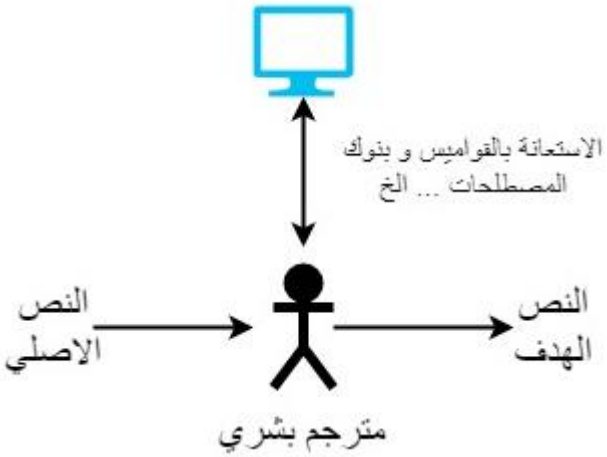
عملية التحرير اللاحق هي العملية التي يقوم بها المترجمون أو اللغويون المدربون تدريباً مهنياً بمراجعة وتصحيح إخراج الترجمة الالية لإزالة كل من الأخطاء الدلالية واللغوية. يمكن تقسيمها إلى فئتين: تحرير لاحق سريع و تحرير لاحق كامل. يركز الأول بشكل أساسي على نقل المعنى الصحيح، مع تجاهل أي مشاكل أسلوبية. في سيناريو التحرير اللاحق الكامل، على النقيض من الاول، يجب أن يصل النص إلى مستوى عالي الجودة يمكن مقارنته بمستوى الترجمة البشرية ويجب معالجة جميع المشكلات.



رسم توضيحي 3 : نظم التحرير اللاحق

- ترجمة بشرية بمساعدة الآلة

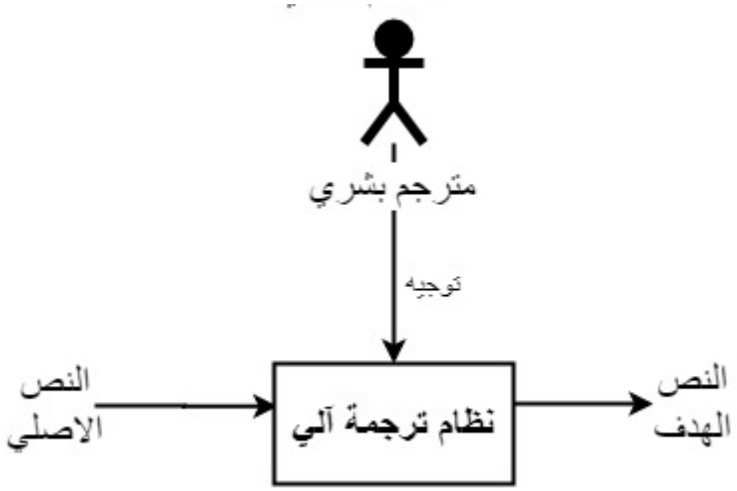
يتم إجراء الترجمة بواسطة مترجم بشري، لكنه يستخدم الكمبيوتر كأداة لتحسين أو تسريع عملية الترجمة. يُطلق عليها اسم الترجمة بمساعدة الحاسوب (CAT) من قِبل الأشخاص في مجال الترجمة.



رسم توضيحي 4 : ترجمة بشرية بمساعدة الآلة

- الترجمة الآلية بمساعدة الإنسان

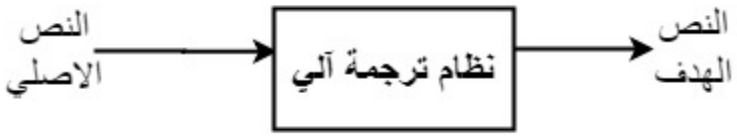
يتدخل الانسان في عملية الترجمة إما قبل أو أثناء أو بعد ترجمته بواسطة الالة.



رسم توضيحي 5 : الترجمة الآلية بمساعدة الإنسان

- ترجمة آلية تلقائية بالكامل

يتم إدخال نص اللغة الاصل في الكمبيوتر كملف، وينتج الكمبيوتر ترجمة تلقائيًا دون أي تدخل بشري. يشار إلى هذا أحيانًا باسم "batch mode" .



رسم توضيحي 6 : ترجمة آلية تلقائية بالكامل

6.2 طبيعة المستخدم :

تنقسم نظم الترجمة الآلية حسب طبيعة مستخدمها الى الانواع التالية : [2]

1.6.2 الترجمة الآلية للمراقب

هذا مخصص للقراء الذين يريدون الوصول إلى بعض المعلومات المكتوبة بلغة أجنبية والذين هم أيضًا على استعداد لقبول الترجمة السيئة المحتملة بدلاً من لا شيء. كان هذا النوع المتصور من قبل الرواد. جاء ذلك مع الحاجة إلى ترجمة الوثائق التكنولوجية العسكرية. كان هذا تقريبا الترجمة المعتمدة على القاموس بعيدا عن الترجمة الآلية القائمة على اللغة.

2.6.2 الترجمة الآلية للمُراجِع

يهدف هذا النوع إلى إنتاج ترجمة أولية تلقائيًا بجودة مماثلة لجودة المسودات الأولى التي ينتجها الإنسان. لا يمكن اعتبار مخرجات الترجمة إلا كمسودة تحتاج الى مراجعة، بحيث يمكن ترقية المترجم المحترف الذي تم تحريره من هذه المهمة المملّة جدًا والمستهلكة للوقت إلى مراجع.

3.6.2 الترجمة الآلية للمترجم

يهدف هذا إلى مساعدة المترجمين البشريين على القيام بعملهم من خلال توفير قواميس على الإنترنت وموسوعة المرادفات وذاكرة الترجمة. عادة ما يتم دمج هذا النوع من نظام الترجمة الآلية في أدوات الترجمة المعتمدة على الكمبيوتر. "أدوات للمترجمين الفرديين كانت متاحة منذ بداية اتمتة المكاتب" وتلك الأنظمة التي تعمل على منصات عادية ومتكاملة مع العديد من معالجات النصوص هي التي حققت النجاح العملي والتجاري.

4.6.2 الترجمة الآلية للمؤلف

هذا النظام موجه إلى المؤلف الراغب في ترجمة نصوصه إلى لغة واحدة أو عدة لغات قابلا الكتابة تحت سيطرة النظام و في نفس الوقت يساعده على إزالة الغموض التام في الكلام بحيث يمكن الحصول على ترجمة مرضية دون أي مراجعة. وفي هذا يختلف عنه نظام الترجمة الالية التفاعلية، في ان التفاعل أثناء التحليل و النقل في هذا الاخير يتم من قبل المتخصصين في النظام واللغة، وليس من قبل المؤلفين. باختصار، لم تكن هناك نجاحات عملية حتى الآن في هذا النوع من النظم، لكن التصاميم أصبحت موجهة نحو المستخدم بشكل متزايد وتوجه نحو النوع الصحيح من المستخدمين المحتملين، الأشخاص الذين يحتاجون إلى إنتاج ترجمات، ويفضل أن يكون ذلك بعدة لغات.

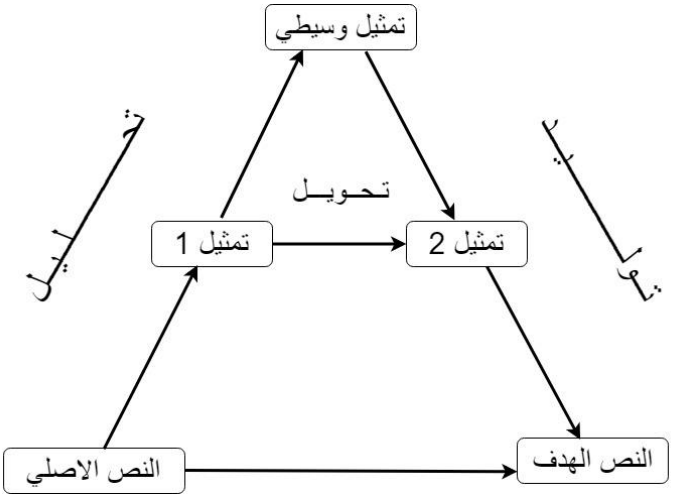
7.2 مقاربات

يقوم نظام الترجمة الآلية (MT) أولاً بتحليل إدخال اللغة الاصل وإنشاء تمثيل داخلي. يتم معالجة هذا التمثيل ونقله إلى نموذج مناسب للغة الهدف. ثم في الاخير يتم إنشاء الإخراج في اللغة الهدف. يمكن تصنيف أنظمة الترجمة الآلية وفقًا لمنهجيتها الأساسية. تحت هذا التصنيف، يمكن العثور على نموذجين رئيسيين: النهج القائم على القواعد والنهج القائم على أساس الذخيرة. في النهج القائم على القواعد، يحدد الخبراء البشريون مجموعة من القواعد لوصف عملية الترجمة، بحيث يتطلب الأمر قدرا هائلا من المدخلات من الخبراء البشريين. من ناحية أخرى، في ظل النهج القائم على أساس الذخيرة، يتم استخراج المعرفة تلقائيًا من خلال تحليل أمثلة الترجمة من مجموعة متوازية صممها خبراء بشريون. أدى الجمع بين ميزات التصنيفين الرئيسيين لأنظمة الترجمة الآلية إلى إنشاء نهج الترجمة الآلية الهجينة. [6] [7] [8]

1.7.2 النظم القائمة على القواعد

تعتبر الترجمة الآلية المستندة إلى القواعد (Rule Based Machine Translation RBMT)، والمعروفة أيضًا باسم الترجمة الآلية المعتمدة على المعرفة والنهج الكلاسيكي لـ MT، مصطلحًا عامًا يشير إلى أنظمة الترجمة الآلية المستندة إلى المعلومات اللغوية حول اللغات الاصل واللغات الهدف التي يتم الحصول عليها أساسًا من القواميس والقواعد (ثنائية اللغة). تغطي الانتظام الدلالي الرئيسي، المورفولوجي، والنحوي لكل لغة على التوالي. بعد ادخال جملة (في لغة اصل معينة)، يقوم نظام RBMT بتوليدها لإخراج جمل (في بعض اللغات الهدف) على أساس التحليل الصرفي والنحوي والدلالي لكل من الاصل واللغات الهدف المشاركة في مهمة ترجمة ملموسة.

هناك ثلاثة أبنية مختلفة لأنظمة الترجمة الآلية المستندة إلى القواعد. إنها الترجمة الآلية المباشرة والقائمة على النقل و الوسيطية على التوالي. وهي تختلف في عمق تحليل اللغة المصدر ومدى محاولة الوصول إلى تمثيل مستقل عن اللغة للمعنى أو النية بين المصدر واللغات المستهدفة. يمكن ملاحظة اختلافاتهم بوضوح من خلال مثلث Vauquois الذي يوضح مستويات التحليل هذه كما هو موضح أدناه.



رسم توضيحي 7: مثلث Vauquois

الترجمة الآلية المباشرة

تسمى الترجمة المباشرة أو الترجمة المعتمدة على الكلمات أو الترجمة القائمة على القاموس. تتمثل طريقة الترجمة المباشرة في ترجمة الكلمات كما يفعل القاموس كلمة بكلمة، عادةً دون أي ارتباط كبير بالمعنى بينهما.

الترجمة الآلية التحويلية

مع تحسين طريقة الترجمة الحرفية، ظهرت الطريقة التحويلية. وهي تتمثل في تحليل بنية الجملة، وعلى أساس الترجمة من كلمة إلى كلمة، يتم توليد نص اللغة المستهدفة وفقًا للقواعد اللغوية المختلفة للغات المختلفة. هناك ثلاثة قواميس متوفرة: قاموس اللغة المصدر، وقاموس ثنائية اللغة المستهدفة والمصدر، وقاموس اللغة المستهدفة.

الترجمة الآلية الوسيطية

جنبا إلى جنب مع تطوير طريقة الترجمة المباشرة والطريقة التحويلية، ظهرت طريقة قائمة على اللغة الوسيطية. أولاً، يتم تحليل اللغة المصدر، وتحويلها إلى لغة الوسيطية، وهو تمثيل مجرد مستقل عن اللغة وينطبق على جميع اللغات. وعلى هذا الأساس، يتم تحويل اللغة المصدر إلى اللغة الهدف.

2.7.2 النظم التجريبية

او الترجمة الآلية القائمة على الذخيرة (يشار إليها أيضًا باسم الترجمة الآلية المعتمدة على البيانات) هي طريقة بديلة للترجمة الآلية للتغلب على مشكلة اكتساب المعرفة المتمثلة في الترجمة الآلية القائمة على القواعد. تستخدم الترجمة

الآلية المستندة إلى الذخيرة(CBMT Corpus Based Machine Translation) ، كما يشير اسمها، مجموعة متوازية ثنائية اللغة للحصول على المعرفة للترجمة الواردة الجديدة. يستخدم هذا النهج كمية كبيرة من البيانات الخام في شكل هيئة متوازية. تحتوي هذه البيانات الأولية على نصوص وترجماتها. يتم استخدام هذه الذخيرة للحصول على معرفة الترجمة. يتم تصنيف النهج القائم على الذخيرة بشكل إضافي الى صنفين: الترجمة الآلية الإحصائية و الترجمة الآلية المستندة إلى الامثلة.

طريقة الترجمة الآلية الإحصائية

يعامل النهج الإحصائي الترجمة الآلية كعملية لنقل المعلومات، باستخدام نموذج قناة لتفسير الترجمة الآلية. هذه الفكرة هي أن أي جملة في لغة ما ستكون ترجمة جملة في لغة أخرى، ولكن الاحتمال يختلف. مهمة الترجمة الآلية هي العثور على الجملة ذات الاحتمال الاقصى. لذلك، تتكون الترجمة الآلية القائمة على الإحصاء من الجوانب التالية: مشاكل النماذج، مشاكل التدريب ومشكلة فك التشفير. تتمثل مشكلة النموذج في إنشاء نموذج احتمال للترجمة الآلية، أي لتحديد طريقة الحساب لاحتمال ترجمة الجملة من اللغة الاصل إلى اللغة الهدف. مشكلة التدريب هي الاستفادة من مجموعة الذخيرة للحصول على جميع معالم هذا النموذج. تتمثل مشكلة فك التشفير في إيجاد الحد الأقصى لترجمة الاحتمالات لأي جملة من لغة الاصل على أساس النماذج المعروفة.

الترجمة الآلية المعتمدة على الامثلة

بنفس المقاربة الإحصائية، المقاربة المعتمدة على الامثلة او الحالات تقوم أولاً، بتقسيم نص اللغة الاصل بشكل صحيح إلى جمل، ثم تقسيم الجمل إلى عبارات، وترجمة هذه العبارات إلى عبارات اللغة الهدف عن طريق القياس، وأخيراً، دمج العبارات في جمل. بمعنى آخر، عملية الترجمة لهذه الطريقة هي كما يلي. أولاً، سيتم إنشاء قاعدة محاذاة ثنائية اللغة. إدخال جملة اللغة الاصل، سيقوم النظام بالبحث عن الجملة الأكثر تشابهًا من قاعدة اللغة الاصل، ثم وفقًا لترجمة اللغة الهدف، سيتم تغيير بعض الكلمات أو العبارات في اللغة الهدف وسيتم إجراء الترجمة.

الترجمة الآلية العصبية

تتمثل الترجمة الآلية العصبية في استخدام نماذج الشبكة العصبية لتعلّم نموذج إحصائي للترجمة الآلية. وتتمثل الفائدة الرئيسية لهذا النهج في أنه يمكن تدريب نظام واحد مباشرة على النص المصدر والهدف، حيث لم يعد يتطلب هذا مجموع عمليات النظم المتخصصة المستخدمة في التعلم الآلي الإحصائي.

فعلى عكس نظام الترجمة التقليدي القائم على العبارات، الذي يتكون من العديد من المكونات الفرعية الصغيرة التي يتم ضبطها بشكل منفصل، تحاول الترجمة العصبية الآلية بناء وتدريب شبكة عصبية واحدة كبيرة تقوم بقراءة الجملة وانتاج ترجمة صحيحة.

وعلى هذا النحو، يمكن القول بأن أنظمة الترجمة الآلية العصبية هي أنظمة من طرف إلى طرف حيث لا يلزم سوى نموذج واحد للترجمة. و تكمن قوة أنظمة الترجمة الآلية العصبية في قدرتها على التعلّم مباشرة، بطريقة من طرف إلى طرف، بالتمثيل من النص المدخل إلى نص الناتج المرتبطة به.

3.7.2 النظم الهجينة

من خلال الاستفادة من كل من منهجيات الترجمة الإحصائية والقواعدية، تم تطوير طريقة جديدة، تسمى الطريقة الهجينة، والتي أثبتت كفاءة أفضل في مجال أنظمة الترجمة الآلية. في الوقت الحالي، تستخدم العديد من قطاعات

MT الحكومية والخاصة هذا النهج المختلط لتطوير الترجمة من لغة المصدر إلى اللغة المستهدفة، والتي تستند إلى كل من القواعد والإحصائيات. يمكن استخدام النهج المختلط بعدة طرق مختلفة. في بعض الحالات، يتم إجراء الترجمات في المرحلة الأولى باستخدام نهج قائم على القواعد يتبعه ضبط أو تصحيح الناتج باستخدام المعلومات الإحصائية. بطريقة أخرى، يتم استخدام القواعد لإجراء معالجة مسبقة لبيانات الإدخال وكذلك بعد الخرج الإحصائي لنظام الترجمة القائم على الإحصاء. هذه التقنية أفضل من السابقة ولديها المزيد من القوة والمرونة والتحكم في الترجمة. [6]

8.2 التقييم

تقييم نظام الترجمة الآلية مهم ليس فقط للمستخدمين والمشترين المحتملين، وكذلك للباحثين والمطورين. تم تطوير أنواع مختلفة من التقييم، مثل: [2]

1.8.2 BLEU (BiLingual Evaluation Understudy) (دراسة تقييمية ثنائية اللغة)

كان مقياس BLEU، الذي اقترحه Papineni في عام 2001، أول مقياس تلقائي يتم قبوله كمرجع لتقييم الترجمات. مبدأ هذه الطريقة هو حساب درجة التشابه بين ترجمة المرشح (الآلة) وترجمة مرجعية واحدة أو أكثر بناءً على دقة n-gram معينة.

2.8.2 WER (Word Error Rate) (معدل الخطأ في الكلمات)

يقيس مقياس WER، الذي اقترحه Popovic و Ney في عام 2007. يستخدم أصلاً في التعرف التلقائي على الكلام، تشير فرضية الجملة إلى جملة تستند إلى مسافة Levenshtein. يستخدم أيضًا في الترجمة الآلية لتقييم جودة فرضية الترجمة بالنسبة للترجمة المرجعية. لذلك، تتمثل الفكرة في حساب الحد الأدنى لعدد التعديلات (الإدراج أو الحذف أو استبدال الكلمة) المطلوب إجراؤها على ترجمة الفرضية لجعلها مماثلة للترجمة المرجعية.

3.8.2 PER (Position-independent word Error Rate) (معدل الخطأ لكلمة الموضع المستقل)

مقياس PER، الذي اقترحه تيلمان في عام 1997. قارن بين كلمات الترجمة الآلية وكلمات المرجع بغض النظر عن تسلسلها في الجملة.

4.8.2 TER (Translation Error Rate) (معدل خطأ الترجمة)

مقياس TER، الذي اقترحه Snover في عام 2006. يتم تعريفه على أنه الحد الأدنى لعدد التعديلات اللازمة لتغيير فرضية بحيث تتطابق تمامًا مع أحد المراجع. تتضمن التعديلات المحتملة في TER إدراج كلمات مفردة وحذفها واستبدالها، وتعديلًا ينقل تسلسل الكلمات المتجاورة. تطبيع بواسطة متوسط طول المراجع. نظرًا لأننا مهتمون بالحد الأدنى لعدد التعديلات اللازمة لتعديل الفرضية، فإننا نقيس فقط عدد التعديلات إلى أقرب مرجع.

9.2 الخاتمة

قدمنا في هذا الفصل مقدمة ونظرة عامة عن الترجمة الالية عموما، حيث بدأنا بنبذة تاريخية عن تطورها ثم ذكرنا اهم الصعوبات التي تواجهها ووضحنا كيف يتدخل الانسان في عملية الترجمة و طبيعة مستخدم أنظمة الترجمة الالية، ثم تطرقنا الى مقاربات وبنيات هذه الأنظمة، وأخيرا ذكرنا اهم المعايير التي يتم على أساسها تقييم هذه الأنظمة.

سنركز في الفصل القادم اكثر على أنظمة الترجمة التي تعتمد على الدلالة والمعنى في عملها، ثم نقوم باختبار و مقارنة بعض من أشهر نظم الترجمة المعروفة حاليا.

الفصل الثاني

نظم الترجمة الآلية الدلالية

3. الفصل الثاني: نظم الترجمة الآلية الدلالية

1.3 مقدمة

لقد كانت الدلالة ولا تزال مسألة مهمة في الترجمة الآلية منذ بداية البحث في هذا المجال. ومن النقاط المهمة في هذه المسألة هي تلك المتعلقة بمركزية الدلالة في نظم الترجمة الالية وشكل ومدى المكونات الدلالية.

هنا، نحن نركز على كيفية استخدام الدلالة في أنظمة MT. بادئ ذي بدء، نتحدث عن "ماذا نعني بالدلالي؟". ثانياً، نناقش التحديات والصعوبات التي تواجه هذه الأنظمة. ثم نذكر بعض الأنظمة الحالية القائمة على الدلالة. وأخيرًا، نقوم بمقارنة هذه الأنظمة، مع توضيح عيوب ومزايا كل نظام.

2.3 ما هو علم الدلالة ؟

علم الدلالة هو الذي يهتمُ بدراسة دلالة الوحدات المُعجميّة، ودراسة المعنى المقصود منها، ويُعرّفُ أيضاً بالعلم الذي يهتمُ بدراسة الشروط والقواعد التي يجبُ توافرها في الكلمة، وبالرغم من أنه علم قائم بذاته، إلاّ أننا لا يمكن أن نفصله عن باقي العلوم، فهي مرتبطة مع بعضها، ويُعتبر علم الدلالة من العلوم القديمة والتي عرفتها لغات متعددة، إلاّ أنه أخذ وجهته الكاملة في العصر الحديث.

و يعرف علم الدلالة أيضا بأنه دراسة المعنى اللغوي، حيث، في حالة الجملة التصريحية، يتم اعتبار معنى الجملة عادة قضية منطقية تحتمل الصحة او الخطأ تصف الحالة المحتملة. بشكل عام، يهتم علماء الدلالة بنمذجة الطريقة التي تساهم بها معاني العناصر المعجمية "الكلمات" في تكوين معاني العبارات والجمل التي تظهر بها. [1]

3.3 لماذا ندرس الدلالة ؟

من الأمور التي لا جدال فيها قولنا بأن الدلالة ستعمل بشكل عام على تحسين أي نظام للترجمة الآلية. هذا القول، بالطبع، يفترض أن الدلالة تقدم نوعًا من المساهمة في الفعالية الكلية للنظام دون تقليل كفاءته. أنا هنا أتخذ موقفًا مفاده أن الكفاءة من حيث السرعة و / أو التكلفة هي عوامل مهمة في أي نظام للترجمة الآلية. المشكلة، كما أراها، ليست ما إذا كانت الدلالة ستعمل على تحسين نظام الترجمة الالية، بل هي ما إذا كانت ضرورية له، حيث أن الإجماع سيكون كذلك.

4.3 التحديات و الصعوبات

1.4.3 التمثيل المحوسب للمعاني

- *التمثيل المنطقي عالي المستوى* Higher-order logic

يتمثل هذا النهج المعتمد على بناء دلاليات حسابية للغة (بشكل أساسي) في تحديد إجراء ترجمي لوصل كل مكون في شجرة التحليل النحوي إلى تعبير مطابق له في نظام منطقي عالي المستوى. و هذا يشمل تنفيذ برنامج Montague (1974) للترجمة المنهجية للبنيات غير الغامضة تركيبيا من اللغة الطبيعية إلى تعبيرات IL Intensional Logic (منطق مكثف). في إطار Montague، يتم تفسير تعبيرات IL نموذجًا نظريًا. في دلاليات حسابية مطبقة، من الممكن تفسير الصيغ والمصطلحات الخاصة بـ IL بالنسبة لقاعدة البيانات، والتي تعمل كنموذج مقصود للغة الطبيعية. [1]

- *بنية خصائص محددة النوع* Typed feature structure/s

توفر القواعد لبنية الخصائص محددة النوع طريقة بديلة للتعامل مع واجهة دلاليات بناء الجملة. تستخدم الرسوم البيانية الموجهة لتمثيل جميع المعلومات اللغوية كتسلسل للخصائص محددة النوع التي تأخذ قيمًا من أنواع محددة، حيث قد تكون هذه القيم هي نفسها بنية خصائص. [1]

2.4.3 بناء قاعدة المعارف

- *قاعدة المعارف اللغوية:*

هي مجموعة القواعد الهيكلية التي تحكم تكوين الجمل والعبارات والكلمات في أي لغة طبيعية معينة.

- *قاعدة المعارف الخاصة بالعالم الواقعي*:

هي تمثيل منهجي لحقائق العالم الواقعي في شكل مجموعة من القواعد، وهذا يشمل الأنطولوجيا التي هي "مواصفات منهجية للمفاهيم المشتركة" (Gruber 1993).

5.3 مشكلة الغموض المعجمي ومساهمات الاخرين في حلها

الغموض المعجمي Lexical ambiguity هو ما ينتج عن تعدد الاستخدامات المعجمية للفظ الواحد وتعدد الدلالة المعجمية له، مثل تعدد استخدامات ومعاني كلمة (عين). ومثل ذلك في اللغة الإنجليزية استخدام كلمة (Play) والتي تورد المعاجم الأجنبية معان عديدة لها ترجمة بعضها: مسرحية، لعب، يلعب، يؤدي، يشغل، تشغيل، وخدمة، يضارب، يتلاعب، يلهي، يعزف، يفسح، ينافق، لهو، مداعبة، لعبة، مباراة ومعان أخرى عديدة، لا نتبينها إلا من خلال السياق، وقد يكون ذلك صعباً إذا لم يقم السياق بدوره في توضيح المعنى.

ولحل هذا المشكل ظهرت عدة طرق تعرف في الإنجليزية باسم Word Sense Disambiguation approaches (WSD) يمكن ان نقسمها الى قسمين :

- طرق تعتمد على قواعد المعارف
- طرق تعتمد على تعلم الالة

1.5.3 طرق تعتمد على قواعد المعارف :

لقد ظهر النهج القائم على المعرفة خلال السبعينيات والثمانينيات من القرن الماضي، ويستخدم عدة أنواع من الموارد مثل القاموس أو المرادفات، وورد نت، و SemCor، ويكيبيديا، إلخ، لتوفير المعنى المناسب للكلمة في السياق.

يستخدم النهج القائم على المعرفة إما قواعد مكتوبة يدويًا أو قواعد نحوية لإزالة الغموض عن الكلمات وأيضًا يستخدم المعنى الذي يشبه في تعريفه الى حد كبير سياق الكلمة الغامضة.

الطرق المختلفة التي تنطوي ضمن هذا النوع هي:

التفضيلات الانتقائية : هي واحدة من الآليات المهمة المستخدمة لفك غموض الكلمات و التي تعتمد على العلاقة بين كلمة وكلمة ولكن لديها نجاح محدود بسبب عدم اعتمادها على الوظيفة الدلالية للكلمات.

تداخل تعاريف المعاني : وتسمى هذه الطريقة أيضًا باسم gloss overlap.بتميز تداخل المعنى بنوعين من مجموعات البيانات: مجموعة المعنى و مجموعة السياق. تتطلب هذه الطريقة وجود قواميس محوسبة MRDs

وعملها الأساسي هو اكتشاف التداخل بين خصائص تعريف المعاني المختلفة للكلمة الغامضة و خصائص السياق الذي ظهرت فيه.

الطريقة التركيبية : تستخدم هذه الطريقة لتمثيل معلومات تركيبية (تشبه الشجرة) لكل نمط. و تستخدم أيضا للتعرف على الأنماط وقياسها، في السياقين المحلي و العام وهدفها هو تصنيف البيانات (على وجه التحديد، المعاني) استنادًا إلى العلاقات التركيبية للخصائص.

2.5.3 طرق تعتمد على تعلم الالة

النهج الخاضع للإشراف : تستخدم الطريقة الخاضعة للإشراف المصنفات لتحديد المعنى الصحيح لكل كلمة. يحتوي هذا النهج على مرحلتين: مرحلة التدريب ومرحلة الاختبار. في مرحلة التدريب، تستخدم المصنفات مجموعة من المواد التدريبية المعلمة لاستخراج الملامح الدلالية والنحوية. في مرحلة الاختبار، تحاول المصنفات معرفة المعنى الأنسب لكل كلمة بناءً على الكلمات المحيطة في الجملة.

الخوارزميات الأكثر استخدامًا في الطريقة الخاضعة للإشراف هي:

قوائم القرار : تم تقديم مفهوم قوائم القرار من قبل Rivest في عام 1987 و، هو مجموعة من القواعد مرتبة في شكل قائمة تستند إلى مجموعة من قواعد (if-then-else) المزودة بمعاملات weights. يتم استخدامه لاستخراج الخصائص (الوظيفية، الدلالية، النحوية الخ) من مجموعة ذخيرة التدريب.

طريقة Naive Bayes : تعتمد هذه الطريقة على النهج الاحتمالي الذي يعين أقصى الاحتمالات لفئة صحيحة للحصول على نتائج دقيقة. في هذا النموذج، يتم تمثيل النص في شكل حقيبة (متعدد المجموعات) من الكلمات دون استخدام قواعد النحو، وترتيب الكلمات يعتمد على التعدد.

أشجار القرار : قدم كوينلان مفهوم شجرة القرار في عام 1986، بناءً على نموذج التنبؤ. يتم استخدام ذخيرة معلمة المعاني كمصدر يتم من خلاله التدريب. يتم تطبيق قاعدة التصنيف في شكل صيغة "yes-no"، وبمساعدة هذه القواعد، يتم تقسيم مجموعة بيانات التدريب بشكل متكرر. الخوارزميات الأكثر شيوعًا المستخدمة في إنشاء الشجرة هي: خوارزمية ID3، ID4، ID5 أو ID5R و C4.5.

الشبكات العصبية : تم اقتراح نموذج الشبكة العصبية بواسطة McCulloch و Pitts في عام 1943 [33] وهي مجموعة مترابطة من الخلايا العصبية الاصطناعية. الغرض من هذا النهج هو تقسيم سياقات التدريب إلى مجموعات غير متداخلة للاستجابات المطلوبة باستخدام ميزات الإدخال. المشاكل الرئيسية التي تحدث في الشبكات العصبية هي: صعوبات في تفسير النتائج، والحاجة إلى كمية كبيرة من بيانات التدريب، وضبط المتغيرات مثل العتبات.

شعاع الدعم الآلي (SVM Support Vector Machine) : تم اقتراح هذه الطريقة بواسطة بوسر عام 1992 والتي تلعب دورًا مهمًا في مشاكل التصنيف. SVMs هي مصنف خطي ينتج عنه hyperplane لفصل أمثلة التدريب الإيجابية والسلبية (تسمى شعاع الدعم) بهامش أكبر. كما أنها تستخدم لحل مشاكل التعرف على الأنماط، والتعرف على الصور والمشكلات المتعلقة بـ NLP مثل التقطيع و التحليل التركيبي.

النهج الغير خاضع للإشراف : يتطلب النهج الخاضع للإشراف بيانات تدريب تم إنشاؤها يدويًا، وهي مكلفة جدًا. تعرف هذه المشكلة باسم مشكلة اكتساب المعرفة. النهج غير الخاضع للاشراف يحل هذه المشكلة عن طريق إدخال مفهوم ان معنى كلمة يعتمد على الكلمات المجاورة لها.

يقسم النهج غير الخاضع للإشراف ظهور كلمة معينة إلى عدد من الفئات من أجل تقرير ما إذا كان ظهور الكلمة له نفس المعنى أم لا. الامر المهم في هذا النهج هو تحديد تجميعات المعاني. الأنواع المختلفة من الطرق المستخدمة في النهج الغير خاضع للإشراف هي: تجميع السياق، وتجميع الكلمات والرسم البياني المشترك.

ولكن لهذا النهج حدود و هي: انه ليست مناسب للحالات واسعة النطاق، والحالات في بيانات التدريب قد لا تعين بالمعنى الصحيح، وتشكيل مجموعات غير متجانسة وعدد التجميعات قد يختلف عن عدد معاني الكلمة المستهدفة.

1) تجميع السياقات

في طريقة تجميع السياقات، يتم تمثيل كل تواجد لكلمة مستهدفة كشعاع سياق في الذخيرة. يتم تجميع هذه الاشعة في مجموعات لتحديد معنى الكلمة الهدف.

إيجابية هذا النهج: كمية كبيرة من بيانات التدريب المشروحة يدويا ليست مطلوبة.

عيب هذا النهج هو: بيانات التدريب مطلوبة لكل كلمة تحتاج إلى أن تفك غموضها.

2) تجميع الكلمات

في هذه الطريقة، يتم تجميع الكلمات المتشابهة لغويًا لتشكيل معنى محدد. تعتمد العلاقة بين كلمة التعريف والكلمة المستهدفة على محتوى معلومات الخصائص الفردية، والتي تتمثل في التبعيات النحوية في الذخيرة (على سبيل المثال، فعل - فاعل، فعل – مفعول به، اسم - صفة، وما إلى ذلك).

3) مخططات الظهور المشترك Co-occurrence Graphs

في هذه الطريقة، يتم إنشاء المخطط على أساس العلاقات النحوية بين الكلمات، والتي أدخلها Widdows و Dorow في عام 2002. في المخطط، تسمى كل كلمة في النص vertex وعلاقة نحوية تسمى edge.

في هذه الطريقة، يتم تعيين معامل لكل edge بالاعتماد على العلاقة. يتم تطبيق خوارزمية تكرارية على المخطط، للعثور على الكلمة (عقدة) التي تملك أعلى درجة وتستخدم في نهاية المطاف شجرة minimum spanning tree لفك غموض الكلمة الهدف.

النهج شبه خاضع للإشراف

يسمى النهج شبه خاضع للإشراف أيضًا بالنهج الخاضع للإشراف بشكل ضعيف أو بمحدودية، حيث يتم استخدام البيانات المصنفة والبيانات الغير مصنفة.

كانت خوارزمية Bootstrapping أول خوارزمية لهذا النهج المستخدم لتحسين أداء WSD. يعالج هذا النهج المشكلة التي يمكن من خلالها استخدام البيانات المصنفة والبيانات الغير مصنفة، لبناء مصنفات أفضل.

فك الغموض المعجمي القائم على النهج شبه خاضع للإشراف هو غيرمضيع للوقت وغير مكلفة بالمقارنة مع الطرق الأخرى.

6.3 دراسة بعض نظم الترجمة الالية الموجودة حاليا :

1.6.3 الأنظمة العربية

نظام "المترجم العربي" الذي طورته شركة أي تي أي ATA التي أسست في لندن عام 1992 حيث اطلق هذا البرنامج عام 1995.وقد طورت الشركة المذكورة برنامجا مصغرا أسمته "الوافي"، يقترح هذان البرنامجان حلول ترجمة من الإنجليزية إلى العربية أو العكس

نظام "الناقل العربي" الذي طورته شركة سيموس CEMOS العربية في باريس، و لدى الشركة المذكورة أربعة برامج للترجمة بين الإنجليزية والعربية وبين الفرنسية 53والعربية - برنامج لكل اتجاه

نظام "عجيب" الذي طورته شركة صخر التي تأسست عام 1982 و الكائن مقرها بمصر. هذا البرنامج يترجم النصوص و مواقع الإنترنيت من الإنجليزية إلى العربية أو العكس

2.6.3 الأنظمة الأجنبية

نظام متيو-تاوم (الترجمة الآلية لجامعة مونتريال): بدأ هذا المشروع عام 1965 و هو برنامج كان في السابق موجها للترجمة العامة و أصبح اليوم مخصصا لترجمة كل التنبؤات الجوية الموجهة للجمهور العام التي تبثّها مصلحة الأرصاد الجوية بكندا وهو يترجم من الإنجليزية إلى الفرنسية (منذ عام 1977 (و من الفرنسية إلى الإنجليزية (منذ عام 1989).و يشمل هذا النظام آلية الترجمة الآلية كليا و الترجمة الآلية جزئيا إضافة إلى محرر نصوص للترجمة البشرية و المراجعة. كما يعتمد البرنامج على مفردات محدودة، تقارب ألفي كلمة و تعبير و يترجم بمعدل ثلاثين ألف كلمة في اليوم أي ما يقارب خمسة و عشرين مليون كلمة في السنة (أو ما يعادل جهد ثمانية و أربعين مترجم بشري). كما يعتمد على نسبة مراجعة أقل من 5 ،% و لأن النظام شديد التخصص فهو يحتوي على مصطلحات محصورة و تراكيب ثابتة

نظام ريفارسو (reverso): يترجم مختلف أنواع النصوص (وثائق، مواقع الويب، رسائل إلكترونية...) على الإنترنيت من الفرنسية إلى اللغات الأوروبية الأخرى و بالعكس (الإنجليزية، الألمانية، الإسبانية و الإيطالية) و يقترح للمحترفين حلول في عدة لغات: الإنجليزية، الألمانية، الإسبانية، الإيطالية، الهولندية، الصينية، الروسية، اليابانية، العربية...إلخ، و يمكنه أن يعمل في الوورد، الاكسيل، الباوربوينت. كما يحتوي على معاجم متخصصة في البنك و الطاقة و الصناعة و الطب و الصيدلة...إلخ.

نظام اوروترا: تسميته هي اختزال لعبارة الترجمة وأوروبا، و هو منتوج التجمع الاقتصادي الأوروبي، بدأ التفكير فيه عام 1977، و يترجم من و إلى اللغات الأوروبية كالفرنسية و الإنجليزية و الإيطالية و الإسبانية...إلخ. وميزة البرنامج أنه بني انطلاقا من نظرية متميزة تعتمد على [تمثيل] تصوير وسطي ودلالي مجرد بين اللّغة الأصل ولغة الهدف ونمط التحويل. عكس البرامج المطورة حول محور والتي تتطلب لغة ثالثة شديدة الهيكلة.

نظام google :

هي شركة أمريكية مقرها في كاليفورنيا، تأسست على يد الري بايج وسيرجي، يعمل بها حوالي عشرين ألف موظّف، تقّدم خدمات عديدة مثل أخبار البورصة، والتس وق والخرائط والبحث الرقمي، ومن أشهر خدماتها ترجمة جوجل عبر fr.google.translate.www منذ سنة 2006 ،وهو يترجم إلى 69 لغة، يترجم الكلمات والجمل وصفحات

الويب. يقوم بالبحث عن أنماط في الماليين من النصوص التي تم ترجمتها من قبل المترجمين البشر إلعطاء أفضل ترجمة، وتسمى عملية البحث هذه عن أنماط في النصوص "ترجمة آلية إحصائية "، وهي تقّدم للمستخدم خيارات عديدة

دخل هذا النظام حيز الاشتغال منذ أكتوبر سنة 2007 وهو نظام إحصائي يعتمد على النصوص و المدونات النصية اعتمادا كليا، دون الحاجة إلى مختصين لغويين إلنشاء القواعد و القواميس كما جرت العادة في الانظمة الالية القواعدية مثل سيستران و يعتمد نظام قوقل على مدونة نصية تشتمل على أكثر من 200 مليار كلمة أخذت من أرشيف الامم المتحدة .كما يعتمد على تقنيات التعلم لبناء منواله الاحصائي و حساب الاحتمالات الممكنة الختيار ترجمة دون الأخرى و هي طريقة ال تتطلب عمال لغويا معمقا.

نظام systran :

يعتمد هذا النظام على القواعد اللسانية و يشتمل على 52 زوجا من اللغات التي يترجم منها و إليها و يعتبر الاضخم حاليا في هذا المجال كما يشتمل على أضخم المصادر المعجمية على الاطالق فقواميسه تضم أكثر من 300000 مدخل معجمي لكل لغة إضافة إلى اثني عشر قاموسا مختصا يحتوى كل واحد منها على أكثر من 100000 مدخل معجمي و انتاجات سيستران تستغلها عديد البوابات على الانترنات Babelfish, yahoo, New stran, linguaNet, lexilogos, Free, altavista. :

إلى جانب برامج الترجمة هناك أيضا مواقع على الإنترنيت تعرض خدمات الترجمة نذكر من بينها

www.freetranslation.com

www.babelfish.altavista.co

www.google.com

http://fr.babelfish.yahoo.com

http://tr.voila.fr

3.6.3 مقارنة

من الجدير بالذكر ان لكل نظام من النظم سالفة الذكر له طريقته و نمطه الخاص في أداء عملية الترجمة الالية، والتي يمكن ان تنطوي تحت احد الأنواع الرئيسية : نظم معتمدة على القواعد, على الإحصاء او هجينة، و بما ان عدد هذه الأنظمة كبير جدا سنختار بعض النظم فقط لدراستها :

- نظام reveso
- نظام google
- نظام systran

لابد للترجمة من نقل المعنى بدقة وأمانة أي نقل المعنى الذي يقصده الكاتب الاصلي، وعلى المترجم الالي او البشري بصفته المتلقي الاول لهذا النص فك الغموض الوارد فيه قبل إعادة صياغته في لغة أخرى، وذلك بالرجوع إلى السياق اللغوي والثقافي للنص، فهل يمكن لهذه النظم ان تتجاوز هذه المشكلة ؟

للإجابة على هذا السؤال اخضعنا هذه النظم لتجربة بسيطة، حيث حضرنا 3 مجموعات من الجمل :

مجموعة تحتوي على غموض معجمي

مجموعة تحتوي على غموض تركيبي

مجموعة تحتوي على غموض ثقافي

ثم ترجمنا نفس الجمل كل مرة بنظام مختلف، يعرض هذا الجدول نتائج الاختبار :

نوع الغموض	الجملة بالانجليزية	ترجمتها بالعربية	ترجمة Reverso	ترجمة Google	ترجمة Systran
غموض ثقافي	My brother does not stand a dog's chance	ليس لأخي أية فرصة	اخى لا تقف فرصة الكلب	أخي لا يحمل فرصة كلب	لا يقف أخي كلب فرصة
	There is far too much monkey business going on around here	هناك كثير من الاشياء المريبة تجري من حولنا	فهناك عمل كثير جدا مونكى بزنس حول هنا	هناك الكثير من الأعمال القرد يحدث هنا	هناك بعيد كثيرا جدا قرد عمل يذهب فوق حوالي هنا
	Let's talk shop	لنتحدث عن العمل	دعنا للحديث	دعنا نتحدث متجر	تركتنا تحدّثت متجر
غموض تركيبي	The old men and women stayed at home	ترجم حسب قصد الكاتب : 1- بقي الشيوخ و النساء في المنزل 2- بقي الشيوخ و العجائز في المنزل	ان كبار السن من الرجال والنساء فى منازلهم	بقي الرجال والنساء المسنين في المنزل	القديم رجل وإمرأة يبقى في البيت
	They fed her rat poison	ترجم حسب قصد الكاتب : 1- أطعموها سم الفئران 2- اطعمو فأرها السم	غذوا لها الفئران	أطعموها سم الفئران	هم غذّواه فأر سم
	Students hate annoying Professors	ترجم حسب قصد الكاتب : 1- يكره الطلاب ازعاج الأساتذة 2- يكره الطلاب الأساتذة المزعجين	الطلبة والاساتذة مزعج الكراهية	الطلاب يكرهون الأساتذة مزعج	طالب يكرهون أستاذ مزعج
غموض معجمي	The boy broke the window	كسر الولد النافذة	الصبى بكسر زجاج	كسر الولد النافذة	broke الفتى النافذة
	The medicine will break you of smoking	سوف يخلصك هذا الدواء من التدخين	سوف تكسر الادوية فى التدخين	الدواء سوف يحرمك من التدخين	سيكسر الالطبّ أنت من يدخّن
	Some people break the law	بعض الناس يخترقون القانون	البعض مخالفة القانون	بعض الناس يخالفون القانون	يكسر بعض الناس القانون

تعليق	كل الأنظمة نجحت في ترجمة معنى break ولو ان نظام Systran ذكر الترجمة الأقل شيوعا للفعل في هذا السياق	كل الأنظمة لم تنجح في ترجمة معنى break،ولكن جوجل اقترب قليلا.	كل الأنظمة نجحت في ترجمة معنى break ما عدا نظام Systran	بغض النظر عن الأخطاء اللغوية اختارت جميع الأنظمة التركيب رقم 2 أي صفة مزعج بدلا من اسم الفعل ازعاج	فقط جوجل اختار التركيب رقم 1 اما ريفيرسو اخطأ في ترجمتها وبالنسبة لسيستران كانت ترجمته غير سليمة نحويا	جميع الأنظمة لم تختر أي من التركيبين ولكنهم نقلوا الجملة بغموضها بغض النظر عن الأخطاء اللغوية التي وقع فيها سيستران	لم تفلح جميع الأنظمة في اكتشاف المعنى المجازي لكلمة shop في هذا السياق	نفس الشيء، لم تنجح جميع الأنظمة في اكتشاف معنى العبارة الاصطلاحية monkey business في هذا السياق	جوجل فقط نجح في اكتشاف معنى العبارة الاصطلاحية dog's chance في هذا السياق

جدول 5 : نتائج الاختبار

4.6.3 مناقشة النتائج :

بالنسبة للغموض المعجمي نلاحظ ان معظم النظم تفشل في تحديد معنى الكلمة اذا استعملت هذه الأخيرة في غير معناها الشائع والمعروف، وبالتالي نستنتج ان الكلمات التي تتصف بهذا النوع من الغموض لا يمكن تحديد معناها الا بالرجوع الى **السياق المحلي للكلمة الغامضة** الذي وردت فيه واعتباره عاملا من عوامل فك الغموض المعجمي.

بالنسبة للغموض التركيبي نلاحظ ان جميع الأنظمة اما ان تفشل في إيجاد الترجمة الصحيحة او تختار احد التراكيب الممكنة لجمل تحتمل عدة تراكيب كتراكيب صحيحة لها حسب **قصد الكاتب** او القائل، وفي حالة ما اذا قصد الكاتب التركيب الاخر للجملة فان هذه النظم ستخطأ حتما في إيجاد الترجمة الصحيحة، وبالتالي لا يمكننا الاستغناء عن **رأي الكاتب** في فك هذا النوع من الغموض، فالكاتب هو اكثر من يعي و يفهم الجملة بطريقة دقيقة.

في المقابل هناك كلمات لها معاني متعددة في مجالات مختلفة مثل كلمة *bank* التي تعني البنك في مجال الاقتصاد والمال وتعني ضفة (النهر) في مجال علم الأرض والجيولوجيا. فمثل هذه الكلمات يكون **عامل المجال العام للنص** فعّالا وكافيا في تحديد معناها.

بالنسبة للغموض الثقافي فكما لاحظنا ان معظم الأنظمة فشلت فشلا ذريعا في اكتشاف المعني المجازي و الخفي وراء العبارات الاصطلاحية و المجازية التي ترجمتها، وفي المقابل ترجمتها ترجمة مباشرة، في حين انه كان من المفروض ان تجد التعبير الحقيقي المقابل لهذه التعابير المجازية و تترجمه ترجمة مباشرة. ولكن التحدي المطروح هو : كيف نكتشف ما اذا كانت العبارة مستعملة في معناها الحقيقي او المجازي ؟ واذا كانت موظفة في تعبيرها المجازي،كيف نجد صياغة حقيقية أخرى مكافئة لها ؟ وهذا ما سنحاول الإجابة عليه في الفصل التالي.

5.6.3 خلاصة التجربة :

كما لاحظنا هناك حالات من الغموض المعجمي يمكن فكها باعتبار مجال النص الذي وردت فيه، ولكن هناك في المقابل حالات من الغموض التركيبي لا يمكن فكها الا بالرجوع الى كاتب النص باعتباره الوحيد الذي يعرف بدقة معنى النص.

وبالتالي نستنتج ان ادخال عامل مجال النص و الرجوع الى كاتب النص في حالات الغموض المستعصية سيرفع لا محالة من أداء نظام الترجمة. وهذا ما سنحاول تحقيقه في الفصل التالي الذي يتحدث عن نظامنا المقترح.

7.3 الخاتمة

تم في هذا الفصل التركيز على أنظمة الترجمة الالية المعتمدة على الدلالة حيث تعرفنا على اهم تحدياتها والمشاكل الخاصة بها، ثم قمنا بتجربة بسيطة لمقارنة أداء اشهر أنظمة الترجمة الالية الموجودة حاليا حيث استنتجنا عدة عوامل لو وجدت لكانت دقة الترجمة الالية احسن بكثير.

سيتم في الفصل القادم تقديم تصميم نظام الترميز الدلالي المقترح من طرفنا من التصميم العام الى التفاصيل الدقيقة.

الفصل الثالث

نظام الترميز الدلالي المقترح

4. الفصل الثالث : نظام الترميز الدلالي المقترح

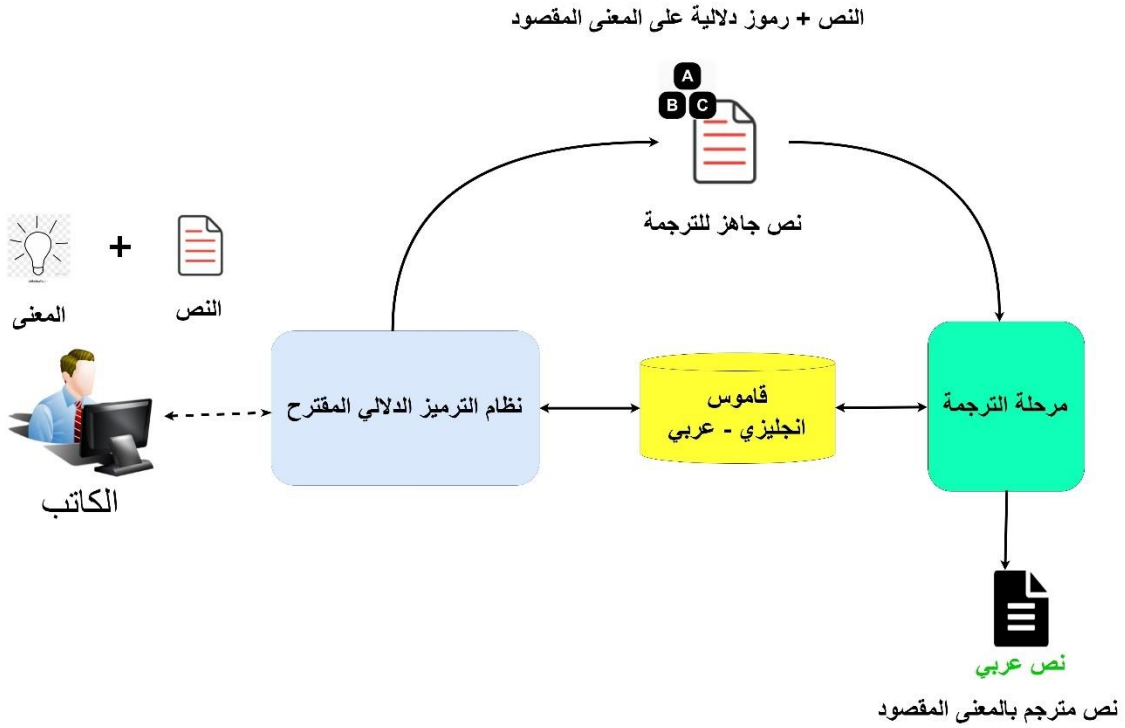
1.4 مقدمة

بعدما خلصنا في الفصل السابق الى ثلاثة عوامل مهمة جدا في تحديد دلالة الجملة المراد ترجمتها وهي : قصد الكاتب، المجال العام للنص و السياق المحلي للكلمة الغامضة. سنحاول في هذا الفصل ان نستخدم هذه العوامل الثلاثة لتصميم نظام الترميز الدلالي الخاص بنا.

2.4 ماذا نقصد بالترميز الدلالي ؟

نقصد بالترميز الدلالي عملية إضافة الرموز والشيفرات الدلالية (حروف، ارقام, اقواس ...الخ) للنص المراد ترجمته اثناء تحرير الكاتب له من اجل إزالة الغموض المعجمي والتركيبي والثقافي عنه، وذلك قصد تسهيل مهمة نظم الترجمة فيما بعد والتي يجب ان تعتمد هي الأخرى مقاربة مشابهة لمقاربتنا وخاصة الرموز الدلالية التي يجب ان يكون متفق عليها من الطرفين.

فالنظام الذي نحن بصدد تصميمه يأتي **قبل** مرحلة الترجمة (في مرحلة **تحرير** النص) ولكنه **يساعد** نظم الترجمة فيما بعد باعتبار انه يزيل كل أنواع الغموض التي ذكرناها في الفصل السابق،

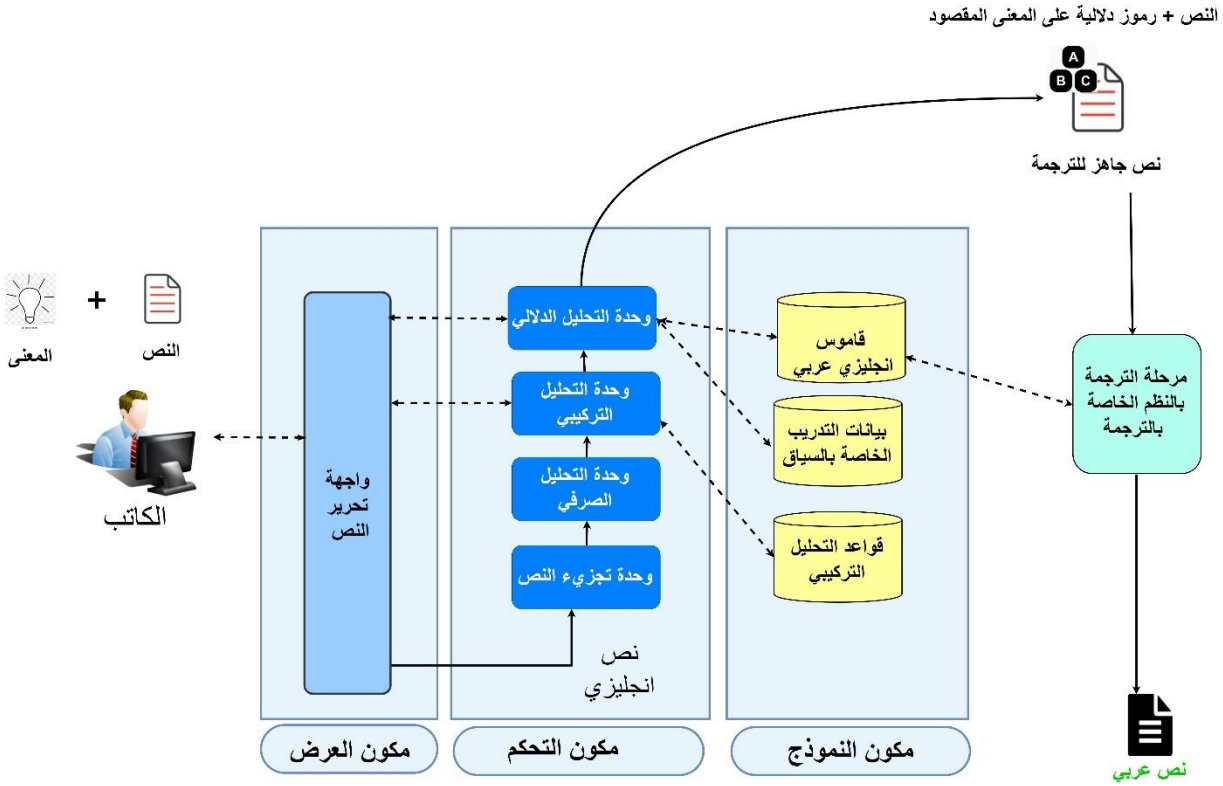


رسم توضيحي 8 : موقع النظام المقترح في عملية الترجمة بأكملها

3.4 بنية النظام

يتكون نظامنا من 3 مكونات أساسية الا وهي :

- مكون العرض : واجهة تحرير النص
- مكون التحكم : جوهر النظام الذي يؤدي مختلف الوظائف المطلوبة
- مكون النموذج : معجم انجليزي عربي



رسم توضيحي 9 : بنية النظام المقترح

1.3.4 مكون العرض : واجهة تحرير النص

تلعب الواجهة دور الوسيط بين الكاتب و النظام حيث من خلالها يستطيع :

- كتابة النص و تعديله (بما في ذلك الإضافة و الحذف .. الخ)
- اختيار مجال النص (علوم, رياضة, سياسة ...الخ)
- تحديد معاني الكلمات التي لم يستطع النظام تحديد معناها
- تصحيح معاني الكلمات التي اخطأ النظام في تحديدها
- .... الخ

2.3.4 مكون النموذج :

أ. قاموس انجليزي عربي

و هو عبارة عن قاعدة بيانات معجمية تتكون من عدة حقول يمكننا ان نقسمها الى مجموعتين : حقول ادخال و حقول اخراج.

حقول الادخال :

- جذر الكلمة الإنجليزية
- نوعها
- المجال المعتبر

حقول الإخراج :

- معاني الكلمة بالنوع المختار وفي المجال المعتبر
- تعريف مختصر لكل معنى
- كلمات عربية معبرة عن كل معنى

بنية القاموس :

يوضح هذا الجدول مختلف الحقول المكونة للقاموس :

ملاحظة : تُعامل أسماء الأعلام بشكل خاص، حيث يخصص لها رمز دلالي خاص بها.

الكلمة بالانجليزية	نوعها	معانيها حسب ووردنت	رمز المعنى	مجالاتها	ترجمتها بالعربية
Break	اسم	وجود الكسر	A	المجال العام	كسر,تحطيم
		هروب من السجن	B	المجال العام	هروب (من السجن)
		الفوز بمجموعة الخصم	C	اللعب، التنس	كسر الارسال
	فعل	يصبح مشهور ومعروف، في الاخبار مثلا	D	المجال العام	شاع ,إنْتشر ,ذاع
		يدمر كيان شيء ما غالبا بالقوة، يقسم الشيء الى اجزاء	E	المجال العام	كسر, حطم
		يتصرف بطريقة مخالفة للقوانين و القواعد	F	القانون، المجال العام	خالف,اخلف
		يقوم بالضربة الافتتاحية التي	G	اللعب، الرياضة	افتتح اللعب

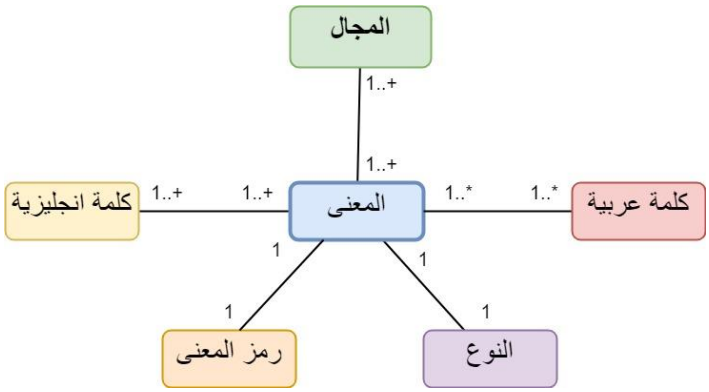
		تبعثر الكرات في لعبة البلياردو			
		يبادل قطع صغيرة من النقود	H	الاقتصاد	صرف (المال)

جدول 5 : بنية القاموس

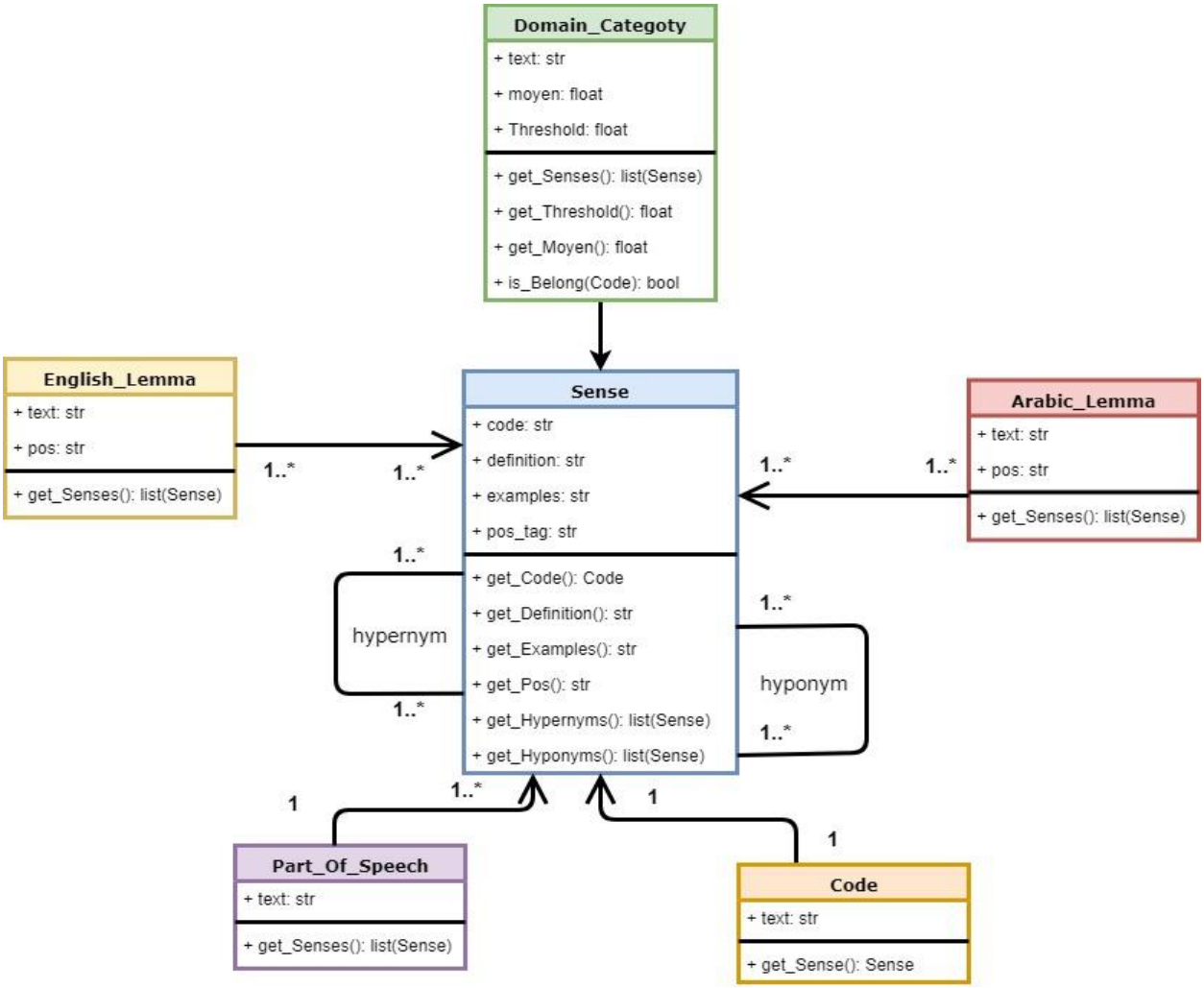
ترتبط حقول القاموس في ما بينها، حيث نجد ان المعنى هو الحقل الرئيسي والذي لا يمكن ان يتكرر اكثر من مرة، لكل معنى :

- كلمة انجليزية او اكثر معبرة عنه
- كلمة عربية او اكثر معبرة عنه
- مجال او اكثر ينتمي اليه هذا المعنى
- نوع وحيد لهذا المعنى

يوضح الرسم التخطيطي التالي العلاقات بين هذه الحقول.



رسم توضيحي 10 : العلاقات بين الحقول المختلفة للقاموس



رسم توضيحي 11 : : مخطط فئات او ما يسمى class diagram *يوضح بنية القاموس*

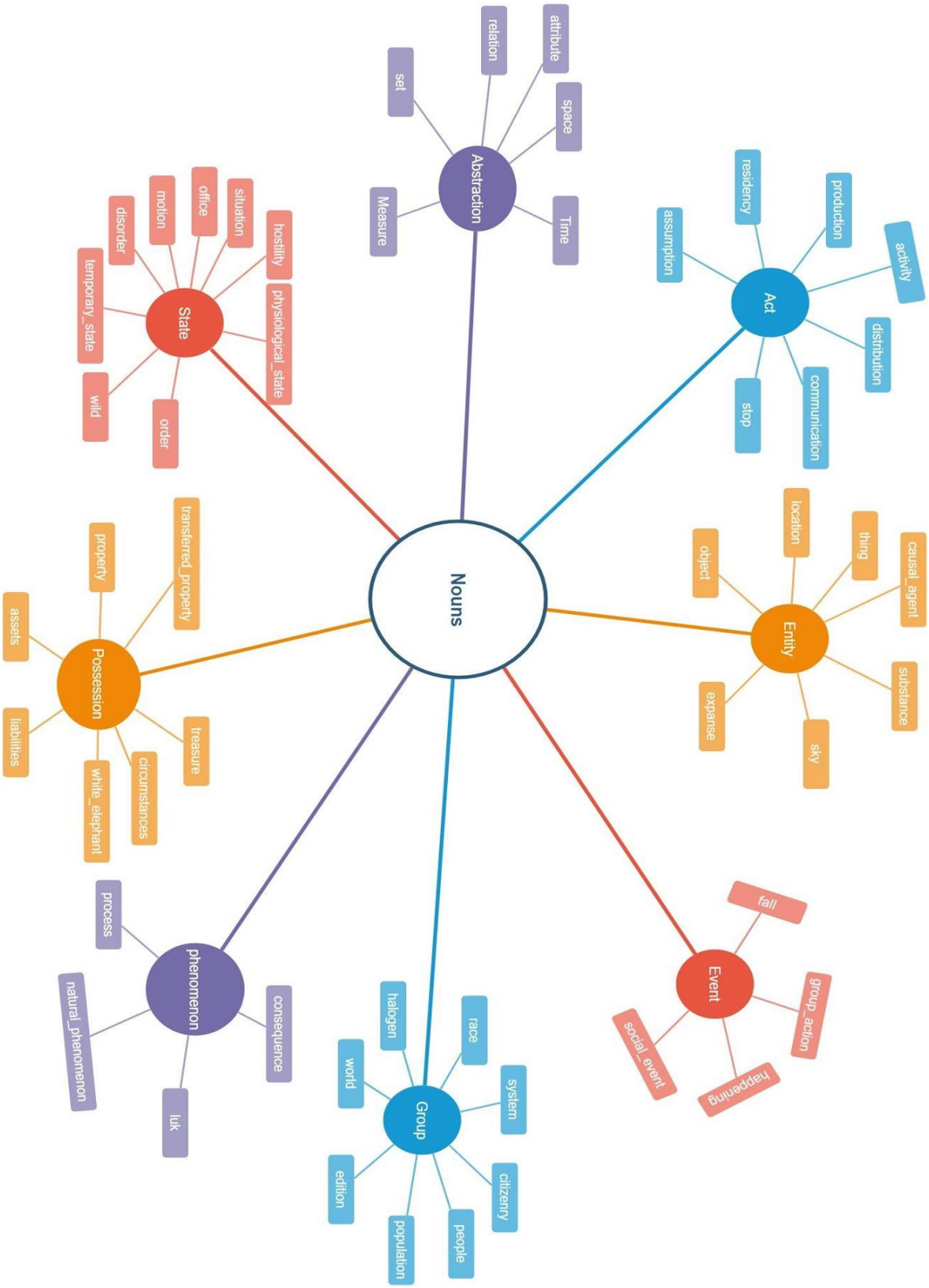
مصادر بناء القاموس

كما ذكرنا من قبل ان القاموس الذي اعتمدناه هو عبارة عن قاعدة بيانات مفرداتية متكونة من عدة حقول : الكلمة بالانجليزية، نوعها، معناها، مجالاتها و مقابلها بالعربية. لكن طبعا بيانات هذه الحقول لم تكن مجموعة من نفس المصدر ولكننا اعتمدنا عدة مصادر في جمعها و هي كالتالي :

- ووردنت WordNet : وهو النواة الأساسية للقاموس، اخذنا منه 3 حقول (الكلمة الانجليزية، نوعها، معناها)
- ووردنت العربي Arabic WordNet : استخدمناه في ملء بيانات حقل الالفاظ العربية
- مجالات ووردنت الموسعة WordNet Domain eXtended : استخدمناه في ملء الحقل الخاص بالمجالات.

ووردنت WordNet

وردنت هو قاعدة بيانات معجمية للغة الإنجليزية. تجمع الكلمات الإنجليزية ذات المعنى الواحد إلى مجموعات من المرادفات تدعى كل مجموعة سينست (بالإنجليزية: synset)،ولكل مجموعة تعريف قصير، عام، كما توضح العلاقات الدلالية المتنوعة بين المجموعات الترادفية مثل علاقة النقيض، علاقة الكل والجزء من، علاقة الشامل والمشمول.



رسم توضيحي 12 : الطبقة العليا من التسلسل الهرمي لأسماء ووردنت

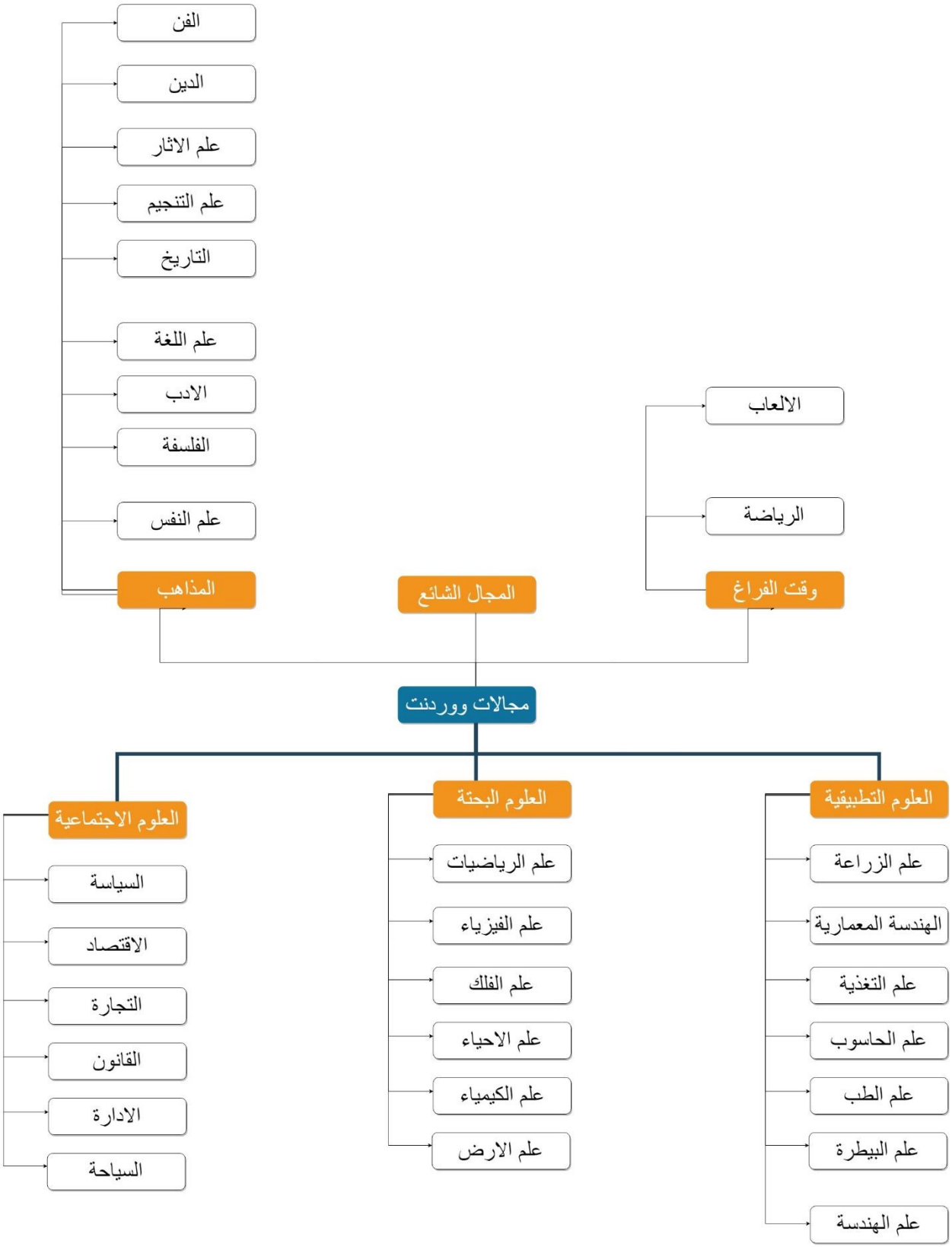
ووردنت العربيَ Arabic WordNet

يتضمن هذا المشروع بناء ووردنت عربي، بعد عملية تطوير ووردنت الاصلي برينستون WordNet و Euro WordNet. وهو يستخدم "الأنطولوجيا المدمجة العليا المقترحة" باعتبارها اللغة البينية لربط ووردنت العربي بووردنت الاصلي المطور مسبقًا.

مجالات ووردنت الموسعة WordNet Domain eXtende

مجالات ووردنت الموسعة (Gonzalez-Agirre et al. 2012a, 2012b) هي عمل مستمر يهدف إلى تحسين مجالات WordNet تلقائيًا. أولاً، تم إسقاط تسميات المجال الأصلية المحاذية لـ WordNet 1.6 إلى WordNet 3.0 باستخدام التعيينات التلقائية عبر إصدارات WordNet. نظرًا لعدم اكتمال التعيين التلقائي (توجد مجموعات جديدة، تغييرات في الهيكل، وما إلى ذلك)، فقد تم ترك العديد من مجموعات المترادفات بدون تسمية. وهكذا، تم القيام بعملية وراثة بسيطة من خلال التسلسلات الهرمية للاسماء و الأفعال. أخيرًا، تم تطبيق تقنية UKB لنشر معلومات المجال من خلال بنية WordNet.

اثبت هذا المشروع نجاعته في مهمة فك الغموض المعجمي Word Sense Disambiguation الشائعة متفوقا بوضوح على المشروع الاصلي بهامش كبير.



رسم توضيحي 13 : الطبقة العليا من التسلسل الهرمي لمجالات ووردنت

ب. قاعدة بيانات التدريب الخاصة بالسياق

تحتوي قاعدة البيانات هذه على جداول (جدول لكل نوع تبعية)، هذه الجداول تتكون من عمودين، عمود خاص بالمعاني التابعة و عمود خاص بالمعاني المتبوعة.

رسم توضيحي 14 : بنية بيانات التدريب الخاصة بالسياق

3.3.4 مكون التحكم :

وهو مكون يحتوي على الوحدات التي تقوم بمختلف الوظائف المهمة المطلوبة من النظام : من تجزئء النص الى تحليله صرفيا مرورا بتحليل تركيبه و أخيرا تحليله دلاليا.

أ. وحدة تجزيء النص :

هنا يتم تقطيع النص الى فقرات والفقرات بدورها الى جمل والجمل بدورها الى كلمات التي هي اصغر وحدة بنيوية للجملة، وهذه بنية الملف الناتج عن هذه المرحلة :

```xml
<text>
    <paragraph>
        <sentence><word>Lucas</word><word>goes</word><word>to</word>...</sentence>
        <sentence><word>He</word><word>has</word><word>many</word>...</sentence>
        <sentence><word>His</word><word>mother</word><word>packs</word>...</sentence>
        ...
    </paragraph>
    <paragraph>
        ...
        ...
    </paragraph>
    ...
    ...
    ...
</text>
```

رسم توضيحي 15 : بنية ملف xml الخاص بالنص بعد مرحلة التجزيء

ب. وحدة التحليل الصرفي :

و هي الوحدة التي تقوم بدراسة بناء الكلمة، حيث تحللها الى عناصرها الأولية و تجردها من صورها الصرفية.

تتكون هذه الوحدة من عمليتين متزامنتين :

- ارجاع الكلمة الى جذرها
- تحديد نوعها (اسم،فعل، صفة) او ما يسمى وسم أجزاء الكلام

الكلمة → (النوع، الجذر)

Word → (root, pos_tag)

Broke → (break, verb)

Children → (child, noun)

بعد هذه المرحلة تكون كل جملة على الشكل التالي بلغة xml :

```xml
<sentence>
    <word root="Lucas"  pos="PROPN" >Lucas</word>
    <word root="go"     pos="VERB"  >goes</word>
    <word root="to"     pos="ADP"   >to</word>
    <word root="school" pos="NOUN"  >school</word>
</sentence>
```

رسم توضيحي 16 : بنية ملف xml الخاص بالنص بعد مرحلة التحليل الصرفي

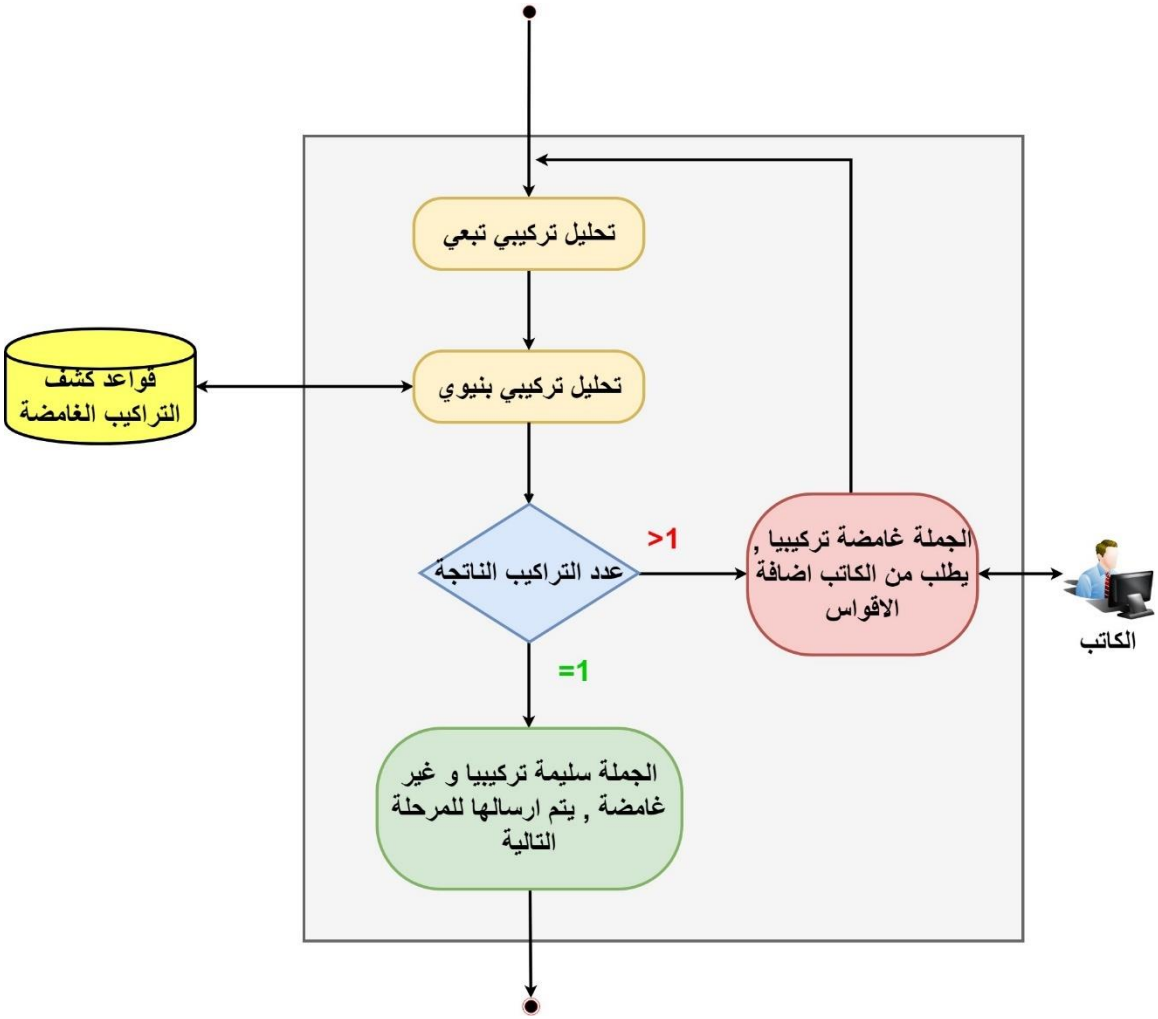
ج. وحدة التحليل التركيبي :

وهذه هي المرحلة الثالثة من مراحل معالجة النص. تمر فيها الجملة كونها الوحدة الأساسية للتركيب بنوعين من التحليل :

- **التحليل البنيوي** : من اجل كشف الغموض التركيبي.
- **التحليل التبعي** : من اجل استخدامها لاحقا في استخراج الخصائص السياقية لكل كلمة.

التحليل البنيوي : الغرض من هذا التحليل هو:كشف التراكيب الغامضة و في حالة وجودها يطلب من الكاتب إضافة الاقواس للجملة التي ظهر فيها هذا الغموض التركيبي. على سبيل المثال، سترفض هذه الوحدة جملة " the old man and woman stayed at home " لان هذه الأخيرة غامضة من الناحية التركيبية، هل تعود صفة old على الاسم man فقط او انها تعود كذلك على الاسم woman، وان لم يضف لها الاقواس ستشكل مشكلة اثناء الترجمة فيما بعد.

في هذه المرحلة تعالج كل جملة من جمل النص على حدى كما هو موضح في خوارزمية التحليل التركيبي، حيث يتم أولا تحليل تركيبها من اجل اكتشاف ما اذا كانت تحمل غموضا تركيبيا ام لا، وذلك بفضل مجموعة من القواعد كتبت يدويا, بعد انتهاء هذه العملية يتم التحقق من عدد التراكيب الناتجة، اذا كانت الجملة تقبل تركيب واحد فإنها غير غامضة تركيبيا و يتم ارسالها مباشرة للمرحلة التالية، واذا كانت تقبل اكثر من تركيب فإنها غامضة تركيبيا فيطلب من الكاتب إضافة الاقواس اليها من اجل إزالة غموضها ثم يعاد ارسالها الى عملية تحليل التركيب الأولى.



رسم توضيحي 17 : يوضح خوارزمية التحليل التركيبي

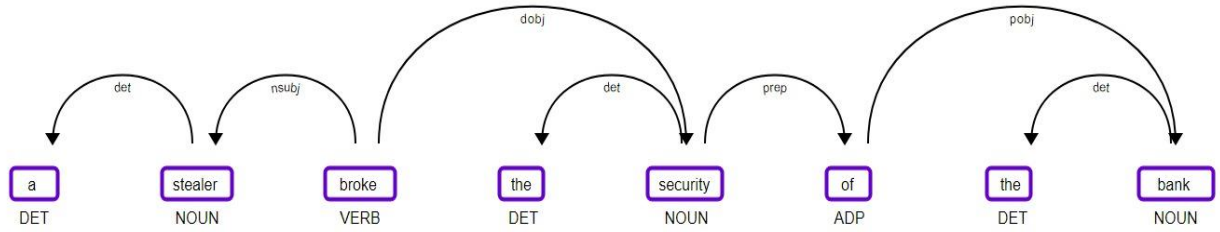
```
S -> ROOT | S PUNCT | S SPACE
ROOT -> CLAUSE | CLAUSE CCONJ CLAUSE | CLAUSE PRER_SCONJ CLAUSE | WRB CLAUSE
CLAUSE -> NP VOP | VOP
VOP ->  VP | VP NP | VP PP | VP NP PP
NP -> NOUN | PRON | ADJ_P NP | DET NP |  NP PP | NP CCONJ NP
ADJ_P -> ADJ | NUM
VP -> VERB | VERB VERB | VERB RB
PP -> PRER_SCONJ NP
NOUN -> 'NN' | 'NNS' | 'NNP' | 'NNPS' | 'EX'
ADJ -> 'JJ' | 'JJS' | 'JJR' | 'PRP$'
VERB -> 'VB' | 'VBD' | 'VBG' | 'VBN' | 'VBP' | 'VBZ' | 'MD' | 'RP'
PRER_SCONJ -> 'IN' | 'TO'
CCONJ -> 'CC' | ','
NUM -> 'CD'
DET -> 'DT'
PRON -> 'PRP'
WRB -> 'WRB'
RB -> 'RB'
PUNCT -> '.'
SPACE -> '\n'
```

صورة 1 : قواعد التحليل التركيبي

التحليل التبعي : حيث تحلل العلاقات النحوية بين الكلمات والتي نسميها "علاقات تبعية" كعلاقة subj التي تعني علاقة الفاعل و obj التي تعني علاقة المفعول به ... الخ.

بعد هذا التحليل يكون الناتج على الشكل التالي :

تحليل جملة : a stealer broke the security of the bank



رسم توضيحي 18 : شجرة التحليل النحوي المعتمد على علاقات التبعية

وبصيغة xml تكون كالتالي :

تحليل جملة : Lucas goes to school

```xml
<sentence>
    <word root="Lucas"  pos="PROPN" dep="nsubj" head="1"  >Lucas</word>
    <word root="go"     pos="VERB"  dep="root"  head="1"  >goes</word>
    <word root="to"     pos="ADP"   dep="prep"  head="1"  >to</word>
    <word root="school" pos="NOUN"  dep="pobj"  head="2"  >school</word>
</sentence>
```

صورة 2 : بنية ملف xml الخاص بالنص بعد مرحلة التحليل التركيبي

هذه بعض العلاقات النحوية ومقابلها بالعربية

رمز العلاقة	اسمها بالانجليزية	مقابلها بالعربية
Nsubj	nominal subject	علاقة الفاعل بين الاسم والفعل
Dobj	direct object	علاقة المفعول به المباشر
Det	Determiner	علاقة التحديد بين الاسم واحدى أدوات التعريف the او التنكير a
Prep	prepositional modifier	العلاقة بين حرف الجر والفعل المتعلق به
Pobj	object of a preposition	العلاقة بين حرف الجر والاسم المجرور
Amod	adjectival modifier	العلاقة بين الصفة و الموصوف

جدول 6 : شرح بعض العلاقات النحوية

د. وحدة التحليل الدلالي (المعجمي والثقافي):

وصلنا الى آخر وحدة في مكون التحكم وهي وحدة التحليل الدلالي semantic analysis وبذلك نكون قد اقتربنا جدا من وصولنا الى الهدف من نظام الترميز الدلالي وهو انشاء ملف جاهز للترجمة.

وحدة التحليل الدلالي تستخدم مخرجات الوحدة السابقة (شجرة تركيب الجملة) لاضافة الرموز الدلالية اليها، وتكمن أهمية مرحلة التحليل الدلالي في أنها مسؤولة بشكل اساسي على تحديد معاني الكلمات و ذلك بإضافة رمز المعنى الى كل كلمة.

من الجدير بالذكر ان هذه الوحدة تتحقق من صحة الجمل من الناحية الدلالية، على سبيل المثال جملة " سيارتي يد ثانية " سترفض على الرغم من انها صحيحة تركيبيا ثم يطلب من الكاتب إعادة صياغتها بالمعنى العام.

بعد هذه المرحلة تكون كل جملة على الشكل التالي بلغة xml :

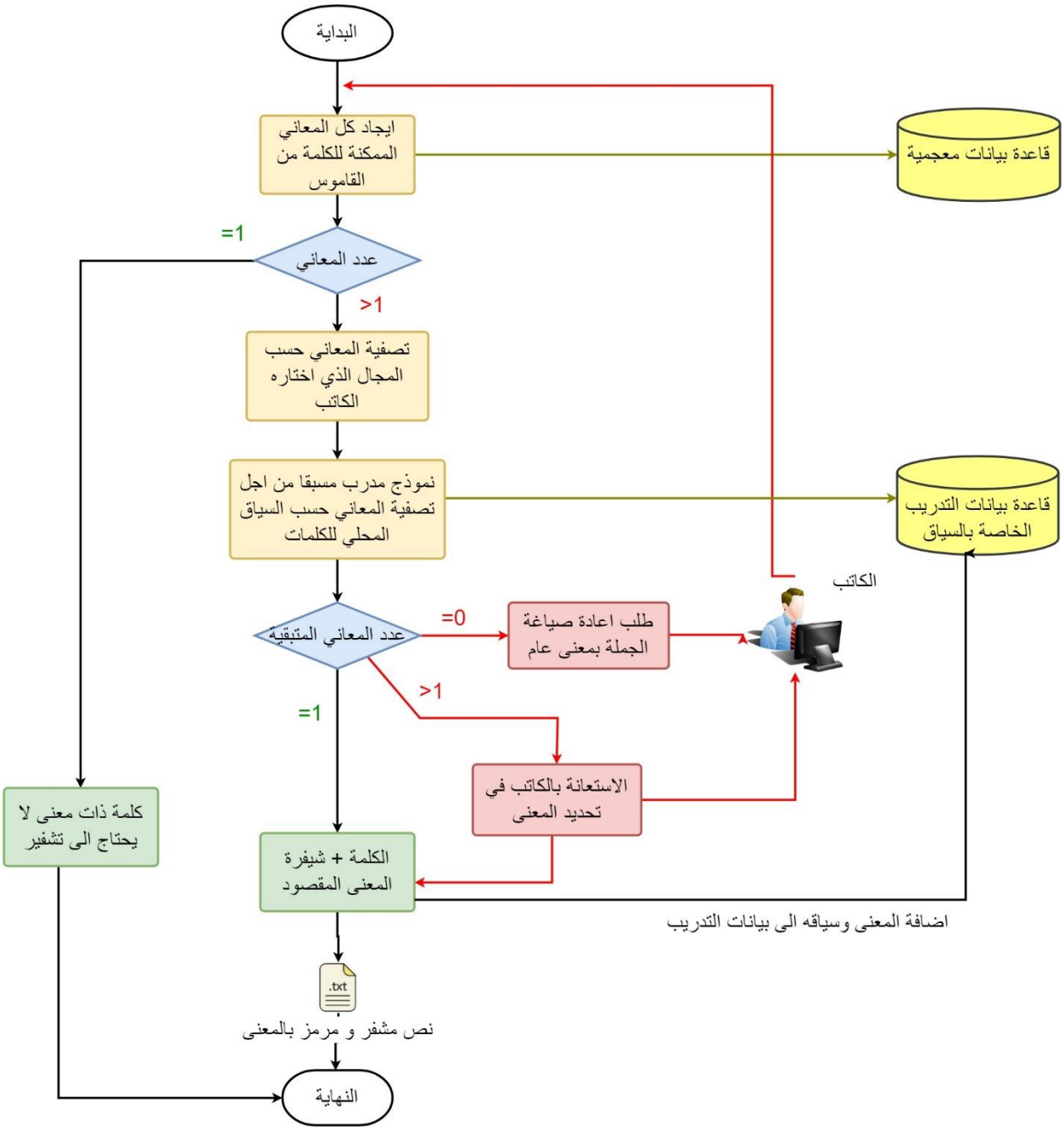
```xml
<sentence>
    <word root="Lucas"  pos="PROPN"  sense_code="B" >Lucas</word>
    <word root="go"     pos="VERB"   sense_code="A" >goes</word>
    <word root="to"     pos="ADP"    sense_code="A" >to</word>
    <word root="school" pos="NOUN"   sense_code="C" >school</word>
</sentence>
```

صورة 3 : بنية ملف xml الخاص بالنص بعد مرحلة التحليل الدلالي

وبالتالي يمكننا القول ان هذه المرحلة تتكون من عمليتين أساسيتين متزامنتين وهما :

- فك الغموض المعجمي (تحديد أي من المعاني المخصصة لكل كلمة هو المعنى المقصود)
- فك الغموض الثقافي (كشف التعابير الاصطلاحية و المجازية و استبدالها)

كما هو موضح في خوارزمية التحليل الدلالي ادناه.

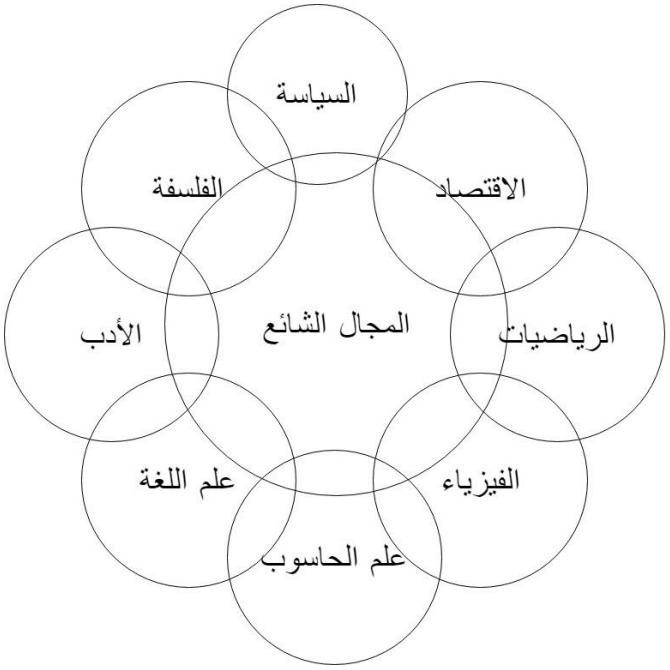


رسم توضيحي 19: خوارزمية التحليل الدلالي

د.1 عملية تصفية المعاني حسب المجال المختار

تُعنَى هذه العملية بقبول المعاني او رفضها بالنظر الى مدى ارتباطها بالمجال الذي اختاره الكاتب، لكننا نعرف ان المجالات متداخلة فيما بينها كثيرا ومتقاطعة حيث ان المجال العام او الشائع factotum متداخل مع كل

المجالات الأخرى بمعنى انه يمكن تظهر معانيه في أي نص خاص بأحد المجالات الأخرى :



رسم توضيحي 20 : تداخل المجالات في النصوص

وبسبب تداخل المجال الشائع مع كل المجالات الأخرى كان لزاما علينا ان نعتبر معنى معين مقبول اذا كان ينتمي الى المجال الشائع او الى المجال الذي اختاره الكاتب.

$$accepted(x, d) \leftrightarrow belong_to(x, Factotum) \; or \; belong_to(x, d)$$

حيث :

- x متغير يمثل المعنى المراد معرفة ما اذا كان مقبول او مرفوض
- $factotum$ ثابت يمثل المجال الشائع
- d متغير يمثل المجال الذي اختاره الكاتب
- $accepted(\;)$ هى الدالة التى تأخذ معنى معين و ترجع قيمة منطقية (مقبول او مرفوض)
- $belong_to(\;)$ هى الدالة التى تأخذ مدخلين معنى و مجال معين وترجع قيمة منطقية تدل على ان هذا المعنى ينتمي الى هذا المجال او لا.

دالة الانتماء الى المجال ($belong_to$) :

بما ان المشروع الذي استخدمناه في استيراد مجالات المعاني الى القاموس لا يعلّم المعاني بالقيمتين 1 و 0 (أي ينتمي او لا ينتمي) وانما حسب كل مجال يعطي لكل معنى قيمة محصورة بين ال1 و 0 سنسميها هنا "معامل المجال" تمثل هذه القيمة مدى ارتباط او قرب هذا المعنى من المجال المعتبر، بما انه كذلك فانه يتوجب علينا تحديد عتبة متعلقة بالمجال فقط من اجل اكتشاف التعابير المجازية نسميها . threshold1 وعتبة أخرى متعلقة بالمجال والكلمة threshold2 من اجل تحديد المعنى.

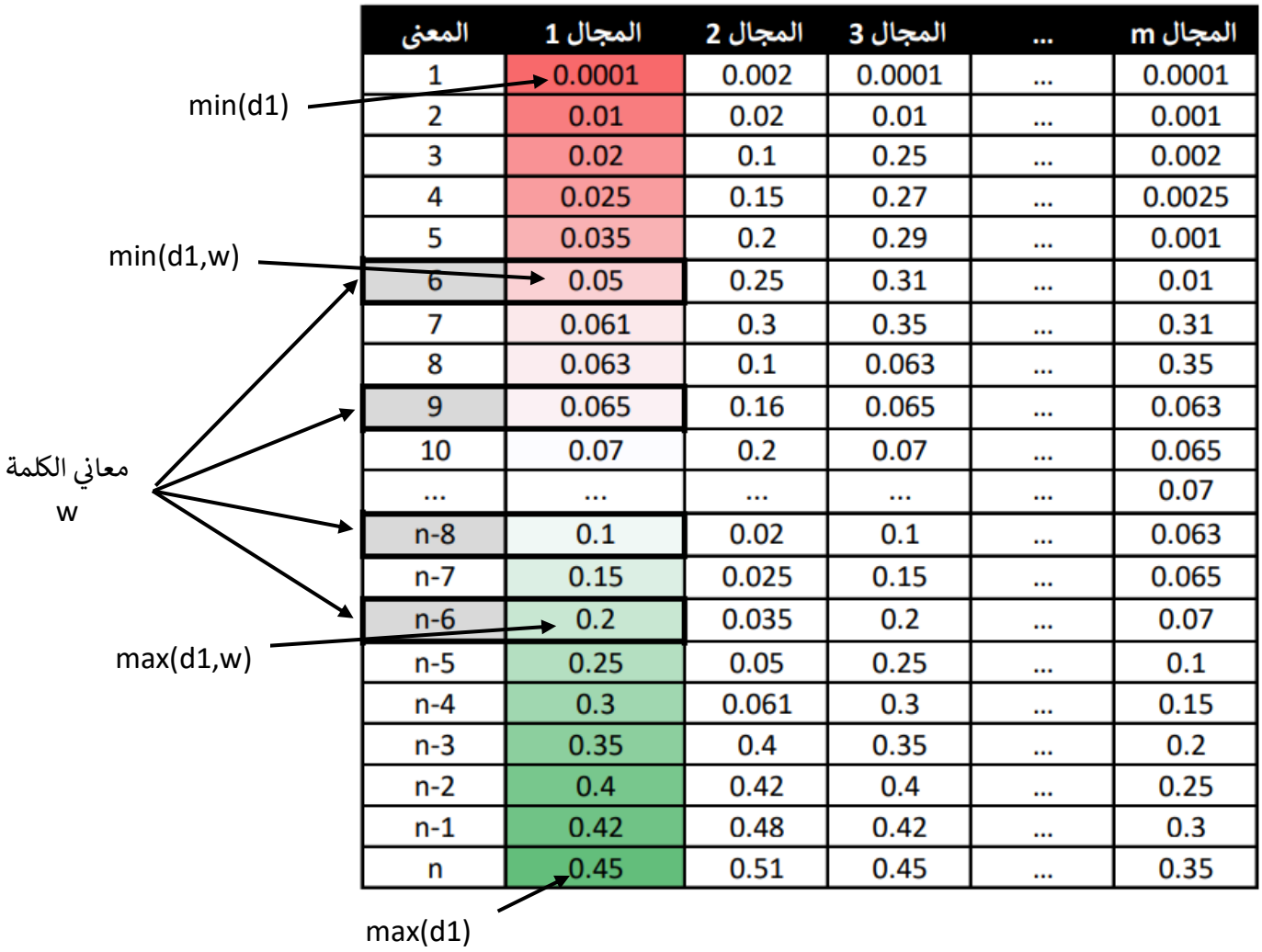
$$belong_to(x,d) \quad \leftrightarrow \quad weight(x,d) \geq threshold_1(d)$$

$$and$$

$$weight(x، d) \geq threshold_2(d، w)$$

حيث :

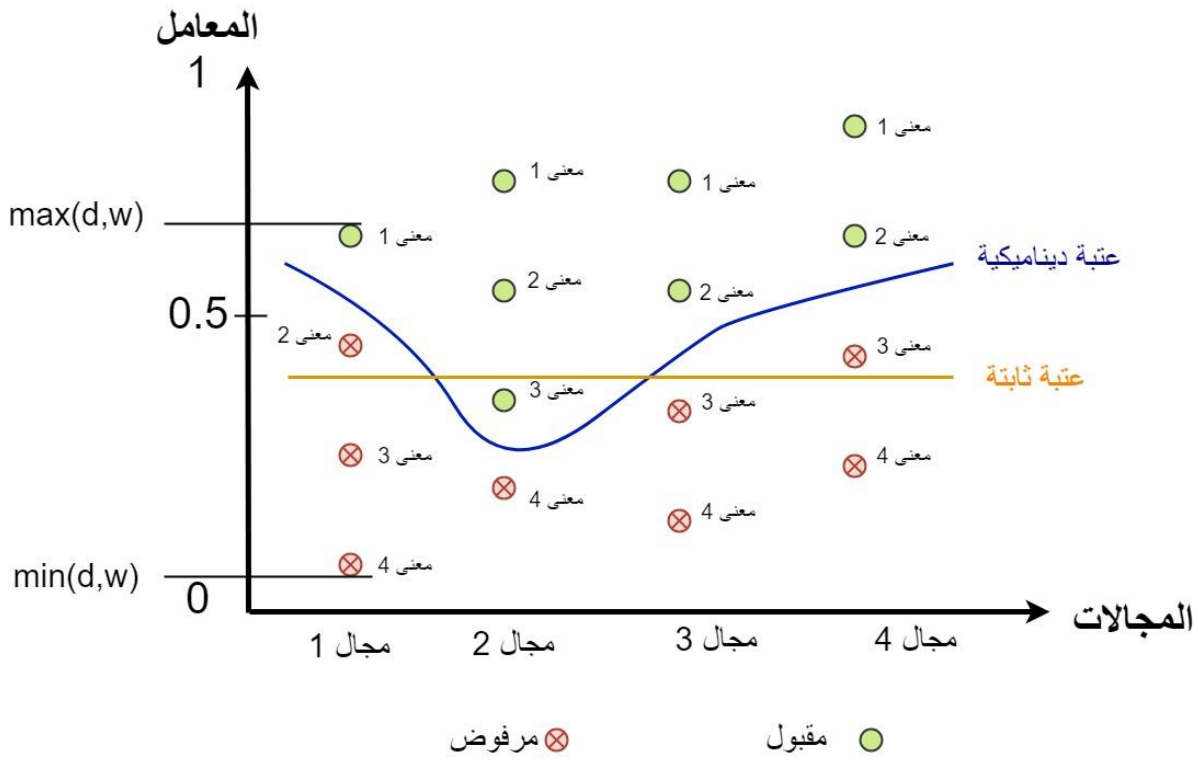
- x متغير يمثل المعنى المراد معرفة ما اذا كان ينتمي الى المجال المعني ام لا
- d متغير يمثل المجال المعطى
- $weight(،)$ دالة تأخذ مدخلين مجال و معنى و ترجع معامل المجال لهذا المعنى.
- $threshold1()$ دالة العتبة الاولى تأخذ مجال معين وترجع قيمة العتبة الخاصة به
- $threshold2()$ دالة العتبة الثانية تأخذ مجال معين وكلمة وترجع قيمة العتبة الخاصة به



المعنى	المجال 1	المجال 2	المجال 3	...	المجال m
1	0.0001	0.002	0.0001	...	0.0001
2	0.01	0.02	0.01	...	0.001
3	0.02	0.1	0.25	...	0.002
4	0.025	0.15	0.27	...	0.0025
5	0.035	0.2	0.29	...	0.001
6	0.05	0.25	0.31	...	0.01
7	0.061	0.3	0.35	...	0.31
8	0.063	0.1	0.063	...	0.35
9	0.065	0.16	0.065	...	0.063
10	0.07	0.2	0.07	...	0.065
...	...	...	...	...	0.07
n-8	0.1	0.02	0.1	...	0.063
n-7	0.15	0.025	0.15	...	0.065
n-6	0.2	0.035	0.2	...	0.07
n-5	0.25	0.05	0.25	...	0.1
n-4	0.3	0.061	0.3	...	0.15
n-3	0.35	0.4	0.35	...	0.2
n-2	0.4	0.42	0.4	...	0.25
n-1	0.42	0.48	0.42	...	0.3
n	0.45	0.51	0.45	...	0.35

جدول 7 : مثال يحاكي كيفية توزيع معاملات مجال معين على كل المعاني ضمن مشروع مجالات ووردنت الموسعة XWND

ليكن هذا التمثيل البياني لتوزيع معاملات المعاني الممكنة لكلمة معينة في كل مجال



رسم توضيحي 21 :مثال على توزيع معاملات بعض المجالات على معاني كلمة معينة ضمن مشروع مجالات ووردنت الموسعة XWND

من اجل حساب **العتبتين** :

العتبة الأولى متعلقة فقط بالمجال وليست متعلقة بمعاني الكلمة، من اجل كشف الكلمات المجازية	$min(d) + \boldsymbol{\alpha 1} * (max(d) - min(d))$
العتبة الثانية متعلقة بالمجال وكذلك بمعاني الكلمة، من اجل تحديد المعاني	$min(d,w) + \boldsymbol{\alpha 2} * (max(d، w) - min(d,w))$

min(d) هي القيمة الدنيا لمجموعة معاملات المجال d لكل المعاني الموجودة.

max(d) هي القيمة الاعظمية لمجموعة معاملات المجال d لكل المعاني الموجودة.

$\alpha 1$ هي نسبة متغيرة يتم تثبيتها عند بداية كل اختبار

min(d,w) هي القيمة الدنيا لمجموعة معاملات المجال d لمعاني الكلمة w.

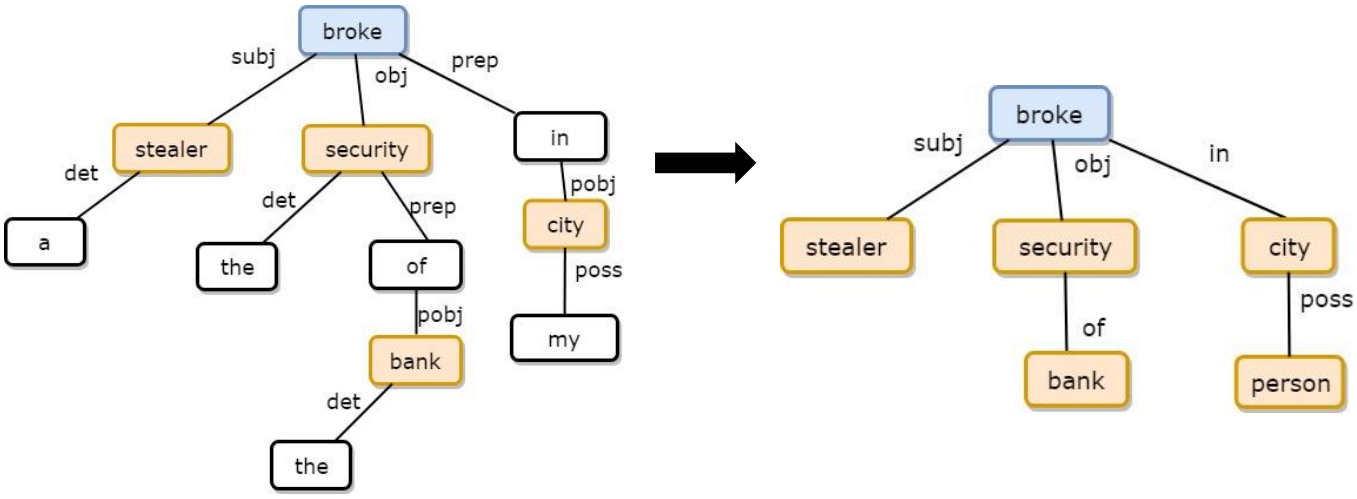
max(d,w) هي القيمة الاعظمية لمجموعة معاملات المجال d لمعاني الكلمة w.

$\alpha 2$ هي نسبة متغيرة يتم تثبيتها عند بداية كل اختبار

د. 2 تقنية تصفية المعاني حسب السياق المحلي

نتعامل مع عملية تحديد المعنى على انها مهمة تصنيف classification حيث يعتمد هذا المصنف في عمله أساسا على التعلم بالامثلة، فالمعاني هي الأصناف والسياق هو العنصر المطلوب تصنيفه الى أي معنى، ولهذا قاعدة بيانات التدريب تحتوي على امثلة (المعاني و كل سياقاتها الممكنة), وعندما يطلب تصنيف مثال جديد (أي سياق جديد) يتم ذلك بالبحث عن اكثر السياقات المشابهة دلاليا ثم تحديد المعاني المرافقة لها في مجموعة التدريب ثم اختيار معنى الكلمة الأكثر تشابها الى هذه المعاني المختارة.

بما ان هذه الوحدة تستقبل مخرجات الوحدة السابقة (التحليل التركيبي) فإن الجمل تكون على شكل شجري كما توضحه الصورة ادناه والتي تمثل التحليل التركيبي التبعي (dependency parsing) لجملة a stealer broke the security of the bank in my city.

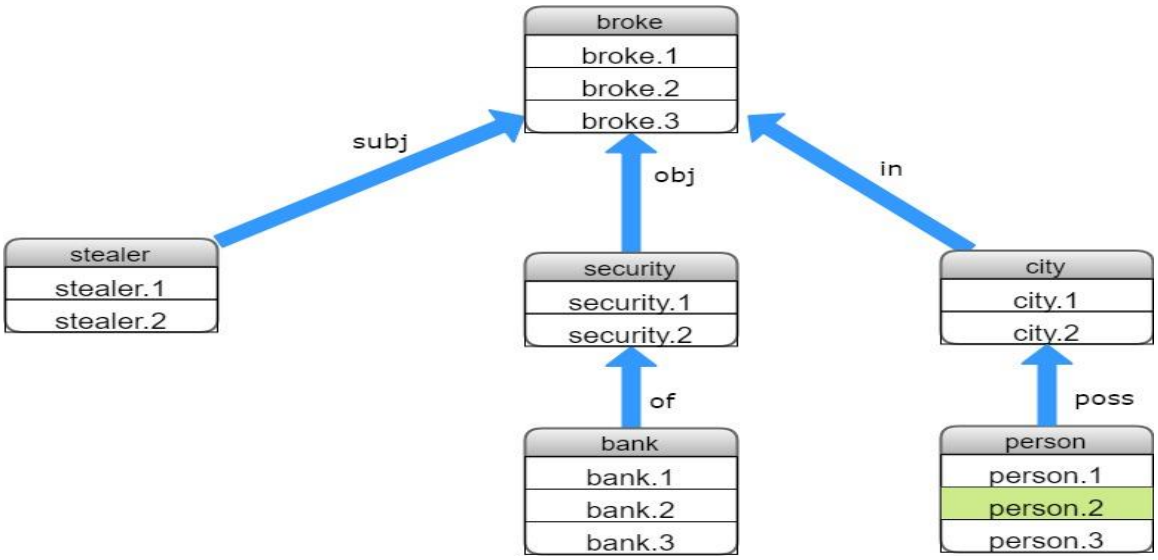


رسم توضيحي 22: خطوة اختزال شجرة التركيب

هذا التمثيل الشجري مناسب جدا لكي نستخرج منه علاقات النحوية الممكنة بين الكلمات، لكن قبل ذلك نغير فيه تغيير بسيط وذلك :

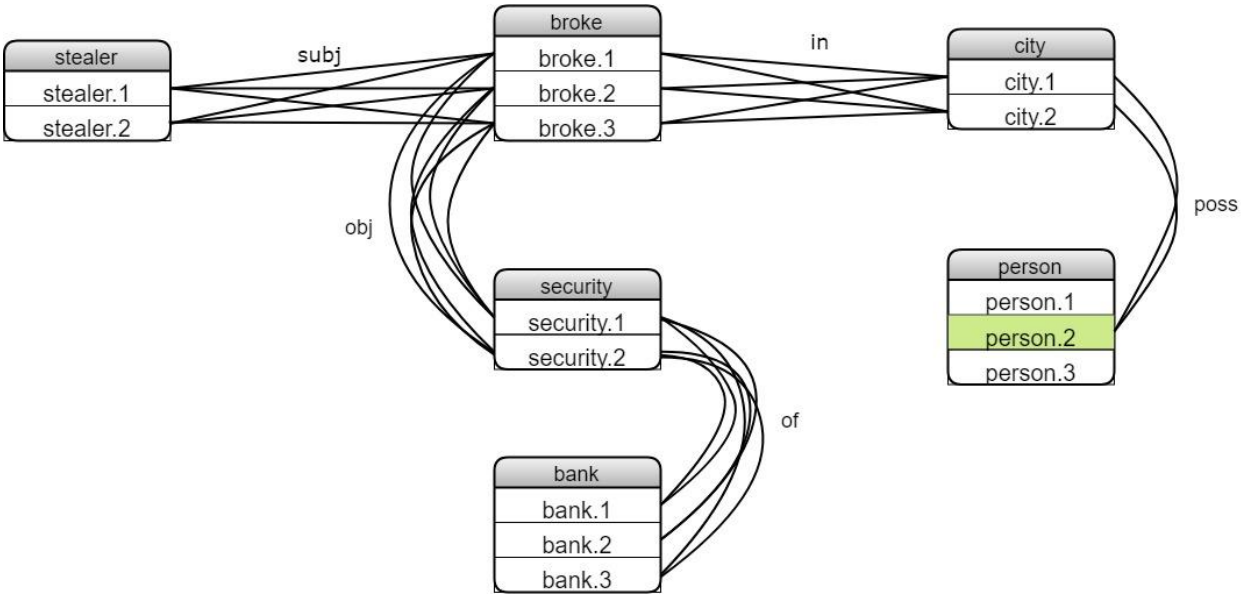
- بتجاهل بعض الأدوات التي لا تؤثر في المعاني الأخرى كالادوات a – the و صفات الملكية my – your – his
- اما الضمائر ... I – you – he فتستبدل بمفهوم ووردنت person.2 الذي يعنى "شخص"
- اما أدوات الجر مثل : ... in – of – on – with فيتم حذفها و استبدالها بعلاقات تبعية طويلة المدى تحمل اسمها

بعد عملية التعديل هذه يكون التمثيل الشجري للجملة السابقة هكذا :



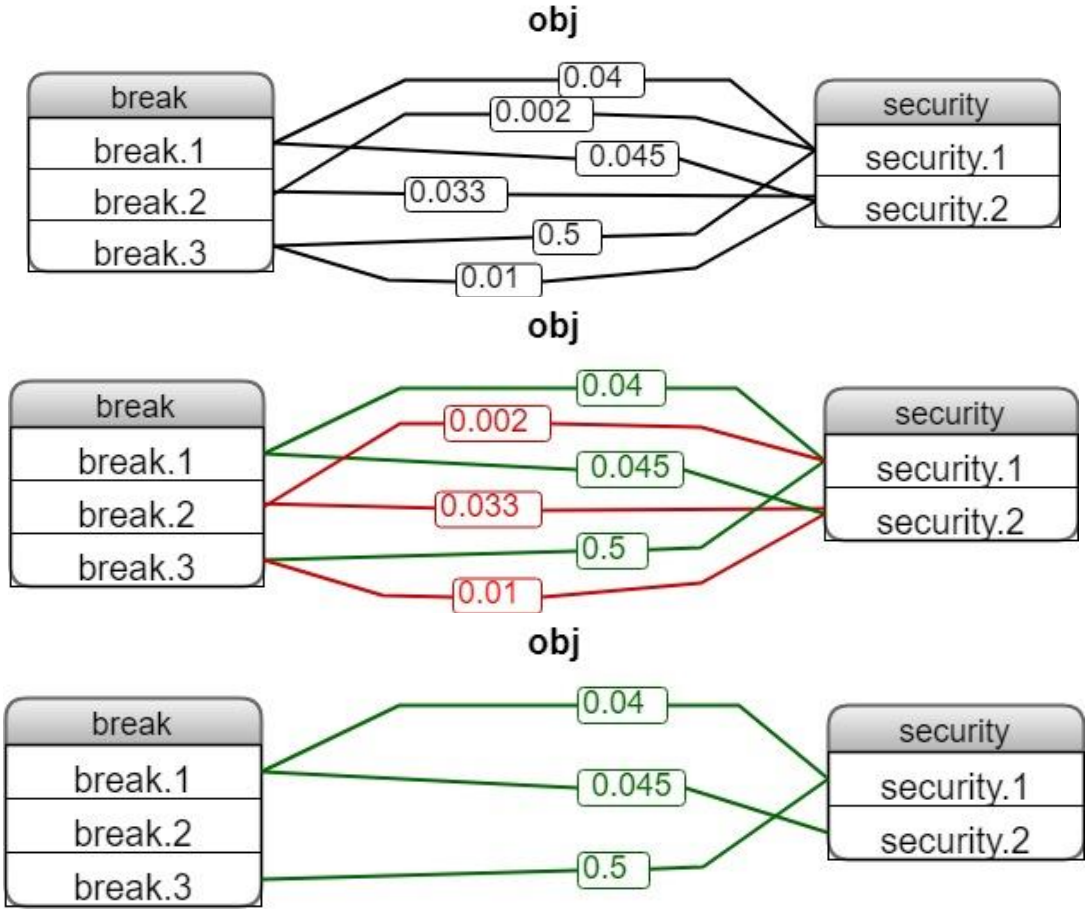
رسم توضيحي 23 : خطوة تحميل المعاني من القاموس

بعد ان يتم تحميل معاني كل كلمة، ننشر العلاقات النحوية بين الكلمات لنستنتج كل العلاقات النحوية بين المعاني

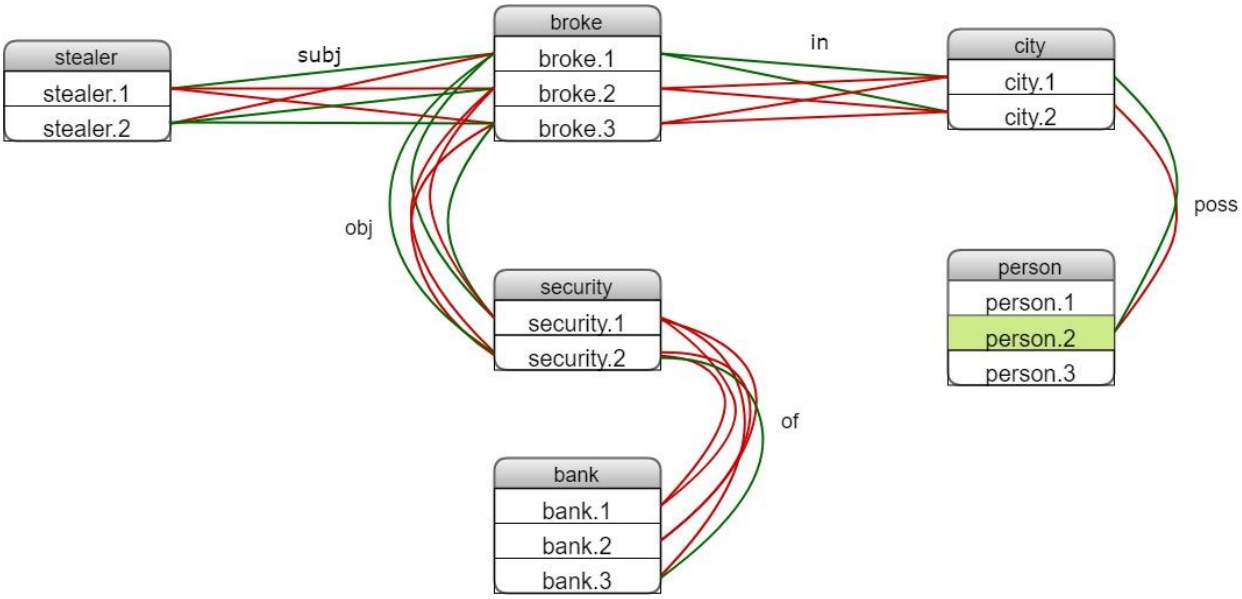


رسم توضيحي 24 : خطوة نشر و توزيع العلاقات النحوية بين المعاني

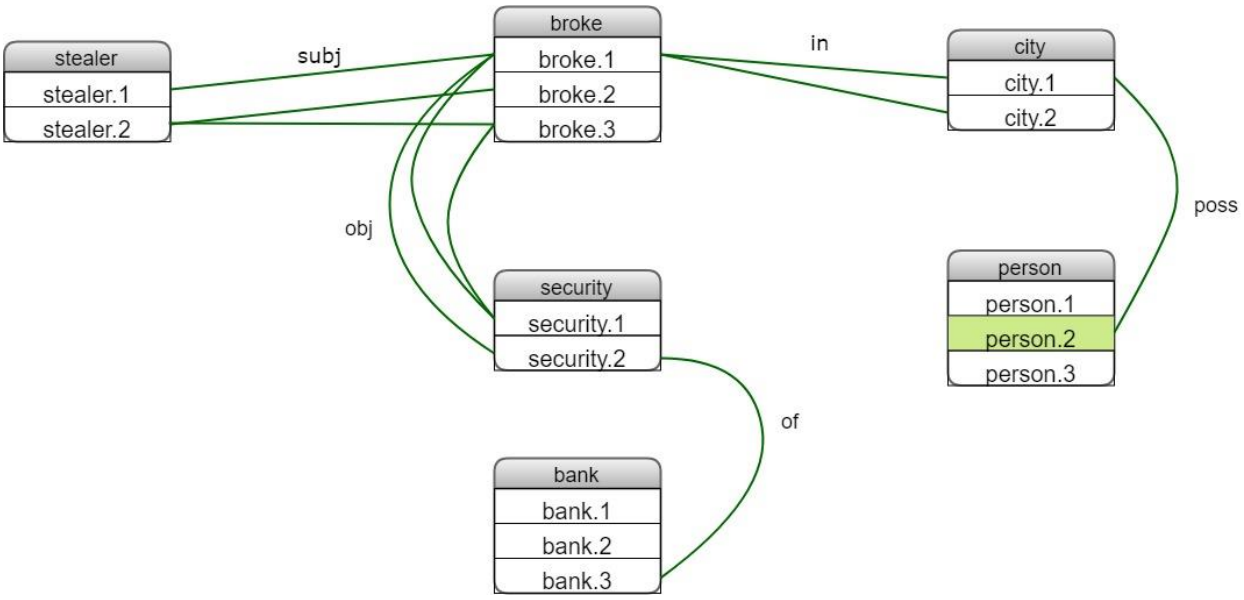
بعد هذا يتم **تقييم العلاقات النحوية** بين المعاني scoring لكي يعطى **معامل سياق** لكل علاقة نحوية بين المعاني، وبعد **حساب عتبة السياق** يتم قبول كل علاقة لها **معامل سياق** اكبر من هذه العتبة وترفض كل علاقة اصغر منها.



رسم توضيحي 25: تقييم العلاقات النحوية بين معاني كلمتين:

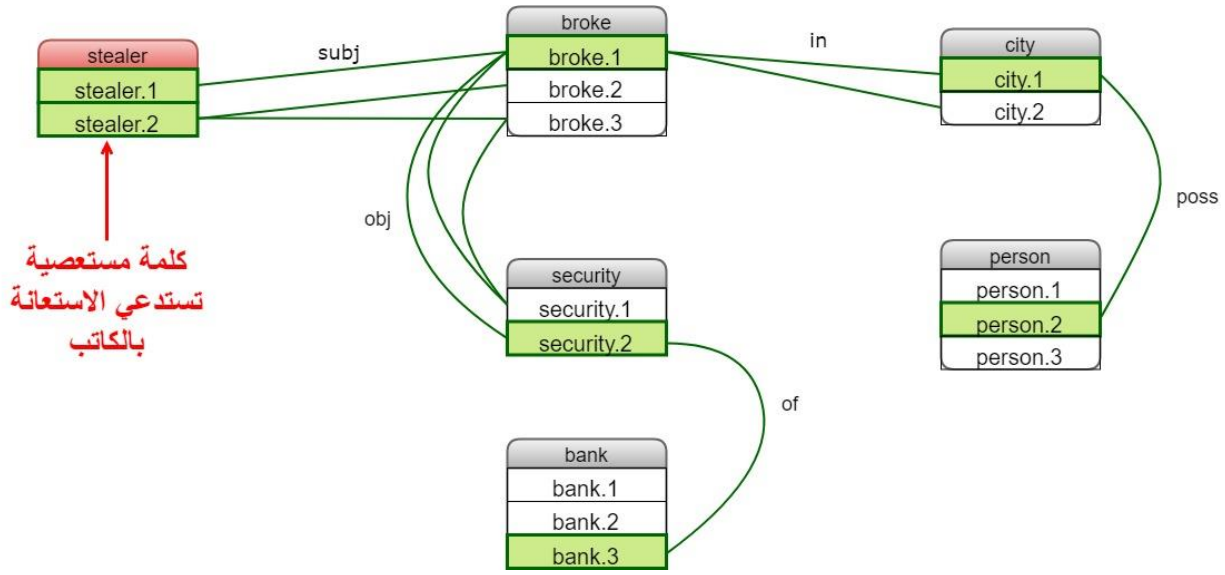


رسم توضيحي 26 : تقييم العلاقات النحوية بين معاني كل الكلمات



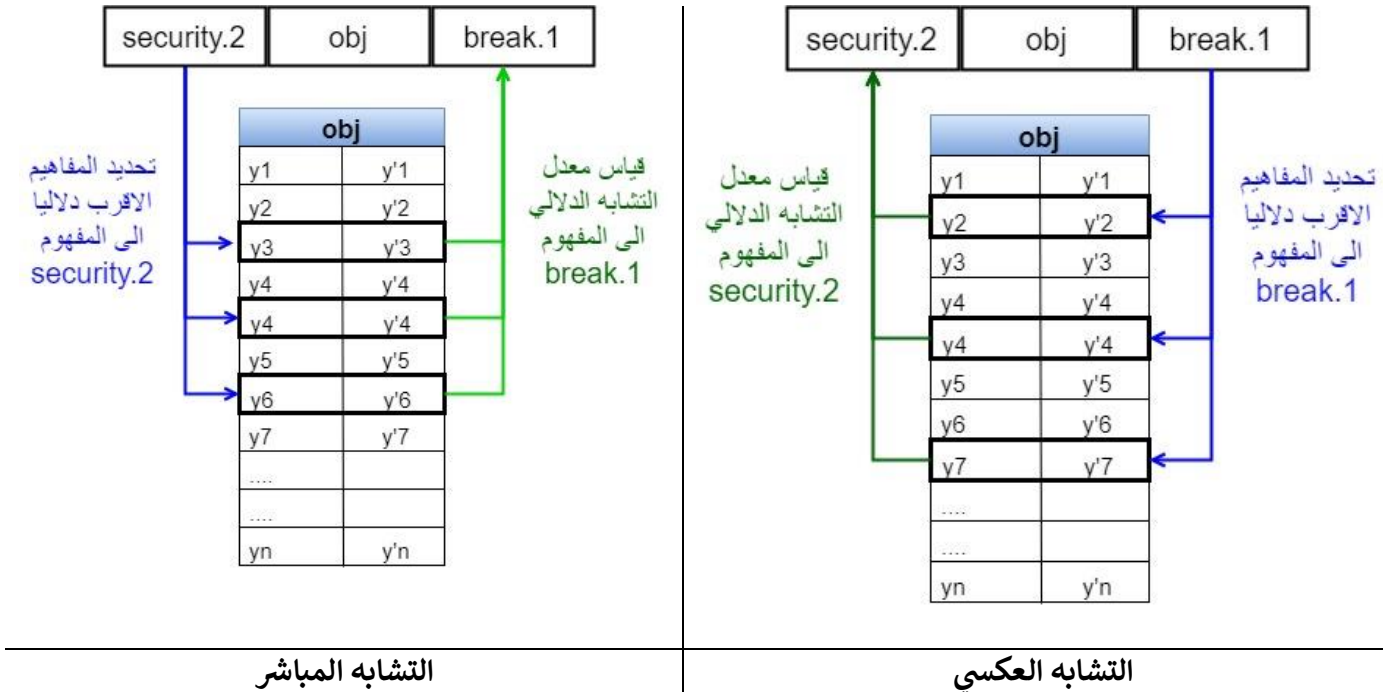
رسم توضيحي 27 : تجاهل العلاقات المرفوضة بين المعاني

وأخيرا نأتي الى العملية الأخيرة وهي تقييم المعاني : حيث نرفض كل المعاني ما عدا تلك التي لديها علاقة مقبولة على الأقل نحو كل الكلمات المجاورة لها.



رسم توضيحي 28 : تصفية المعاني

والان وصلنا الى مرحلة تقييم العلاقات، وهذه المرحلة تعتمد على مجموعة بيانات التدريب السياقية المجهزة سابقا و تعتمد بشكل أساسي على التشابه الدلالي بين مفاهيم ووردنت، وأخيرا على عتبة التقييم.



رسم توضيحي 29 : حساب نسبتي التشابه المباشر و العكسي

لتكن علاقة التبعية التالية (ch، dep، hd) حيث :

- ch هو المعنى التابع
- dep هو نوع التبعية
- hd هو المعنى المتبوع

صيغة التقييم:

نسبة التشابه العكسي ويمثل قيمة احتمال ان يكون المعنى المتبوع هو hd علما ان المعنى التابع هو ch

نسبة التشابه المباشر ويمثل قيمة احتمال ان يكون المعنى التابع هو ch علما ان المعنى المتبوع هو hd

$$Y = \frac{P(child = ch \mid head = hd) + P(head = hd \mid child = ch)}{2}$$

نسبة التشابه المباشر

max_sim(hd))،P(child = ch | head = hd) = similarity_elt_list(child_list(max_sim(hd))

نسبة التشابه العكسي

max_sim(ch))،P(head = hd | child = ch) = similarity_elt_list(head_list(max_sim(ch))

قياس معدل التشابه الدلالي

$$\text{similarity_elt_list(list، elt)} = \frac{\sum_{x \epsilon list} similarity(x, elt)}{lenght(list)}$$

قياس التشابه الدلالي بين مفهومين

$$\text{similarity(x، y)} = \frac{\text{wup_sim(x، y)} + \text{path_sim(x، y)}}{2}$$

حيث:

- Y هو خرج تقييم العلاقة
- child هو عمود المعاني التابعة في جداول بيانات التدريب الخاصة بالسياق
- head هو عمود المعاني المتبوعة في جداول بيانات التدريب الخاصة بالسياق
- child_list هي دالة تعطي قائمة المعاني من جداول بيانات التدريب التابعة لمعنى معطى hd
- head_list هي دالة تعطي قائمة المعاني من جداول بيانات التدريب المتبوعة من طرف معنى معطى ch
- max_sim هي دالة ترجع المفهوم الأكثر تشابه مع المفهوم المعطى من بين كل مفاهيم العمود المتعلق بها في مجموعة التدريب
- similarity_elt_list هي دالة قياس معدل التشابه الدلالي بين مفهوم و قائمة مفاهيم
- similarity هي دالة قياس التشابه الدلالي بين مفهومين من مفاهيم ووردنت
- wup_sim و path_sim هما دالتا Wu & Palmer و The Shortest Path based على الترتيب لقياس التشابه الدلالي بين المفاهيم

حساب قيمة عتبة السياق :

لنفرض ان هذه مجموعة العلاقات بين معاني كلمتي break و security

معنى الكلمة التابعة	نوع التبعية	معنى الكلمة المتبوعة	معامل السياق	المقارنة مع العتبة	القبول	
معنى se1	obj	معنى br1	**0.75**	اكبر من العتبة	مقبول	max(w1,w2)
معنى se1	obj	معنى br2	0.05	اصغر من العتبة	مرفوض	
معنى se1	obj	معنى br3	**0.6**	اصغر من العتبة	مقبول	
معنى se2	obj	معنى br1	0.35	اكبر من العتبة	مرفوض	min(w1,w2)

... وهكذا الى اخر كل العلاقات

جدول 8 : تقييم العلاقات النحوية بين معاني كلمتين

من اجل هذه العملية وقع اختيارنا على عبارة **العتبة الثالثة** التالية :

$$min(w1, w2) + \boldsymbol{\alpha 3} * (max(w1, w2) - min(w1, w2))$$

حيث :

- $\boldsymbol{\alpha 3}$ هي نسبة متغيرة يتم تثبيتها عند بداية كل اختبار
- $\boldsymbol{min(w1, w2)}$ هي القيمة الدنيا لمجموعة معاملات تقييم كل العلاقات بين الكلمتين w1 و w2
- $\boldsymbol{max(w1, w2)}$ هي القيمة العظمى لمجموعة معاملات تقييم كل العلاقات بين الكلمتين w1 و w2

3.د عملية البحث عن ترجمة الكلمات :

تستغل هذه العملية وجود العديد من الالفاظ العربية المعبرة عن كل معنى من اجل تحديد اللفظ المعبر عن كل المعاني المتبقية بعد عمليتي التصفية السابقتين.

مثال: هذه مختلف معاني كلمة window والالفاظ العربية المعبرة عنها

الالفاظ العربية المعبرة عليه	التعريف بالانجليزية	رمز المعنى	رقم المعنى
نافذة، شباك	a framework of wood or metal that contains a glass windowpane and is built into a wall or roof to admit light or air	A	1
نافذة، شباك	a transparent opening in a vehicle that allow vision out of the sides or back; usually is capable of being opened	B	2
نافذة	a transparent panel (as of an envelope) inserted in an otherwise opaque material	C	3
نافذة، فتحة	an opening that resembles a window in appearance or function	D	4
نافذة	the time period that is considered best for starting or finishing something	E	5
زجاج، زجاج النافذة	a pane of glass in a window	F	6
نافذة	an opening in a wall or screen that admits light and air and through which customers can be served	G	7
نافذة، إطار	(computer science) a rectangular part of a computer screen that contains a display different from the rest of the screen	H	8

جدول 9 : معاني كلمة window

فإذا كان يراد بالكلمة معنى "نافذة" فإنه يكفي فقط استثناء المعنى السادس حتى يمكننا إيجاد الترجمة العربية والتي تمثل تقاطع جميع مجموعات الالفاظ العربية الخاصة بكل معنى من المعاني المتبقية بعد الاستثناء. وبالتالي هذا سيكلفنا عناء إيجاد المعنى الدقيق للكلمة من اجل ترجمتها.

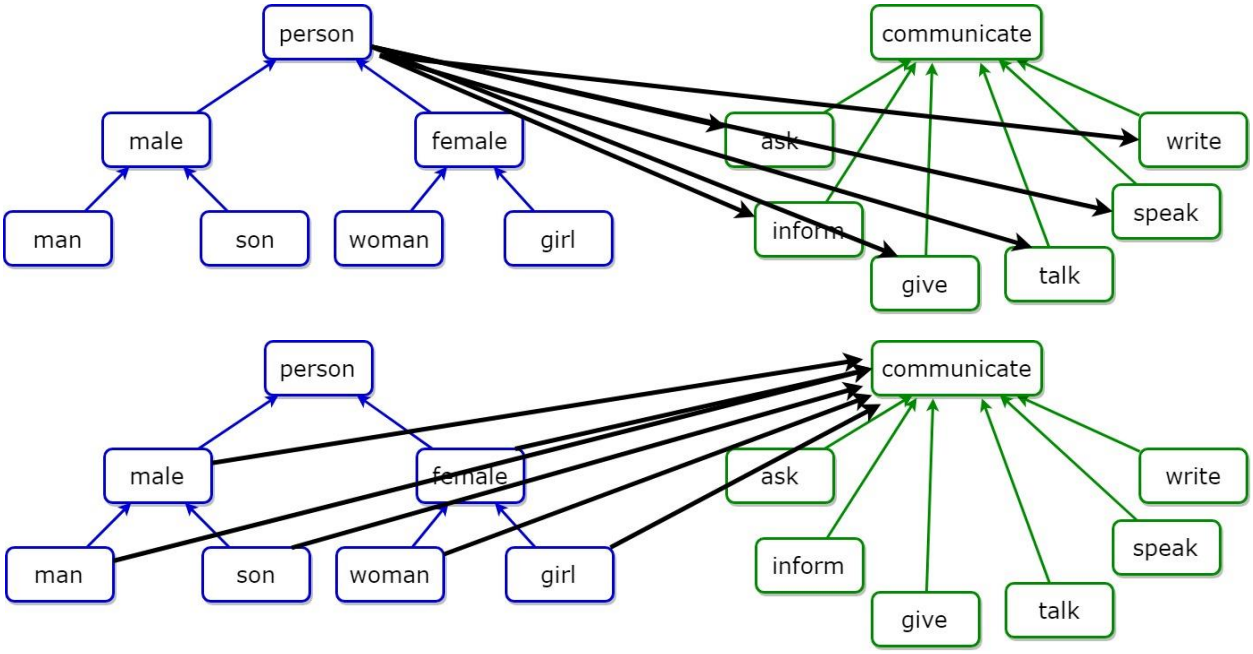
د.4 عملية إضافة المعاني وسياقها الى بيانات التدريب:

تأتي هذه العملية بعد الانتهاء من تحديد معاني كل الجملة وعند حفظ الملف, حيث يتم إضافة الخصائص السياقية للمعاني في قاعدة بيانات التدريب، والجدير بالذكر هنا ان هذه العملية لايتم فيها فقط إضافة الخصائص السياقية المطلوب اضافتها ولكن كذلك يستنتج خصائص أخرى بفضل علاقات الشمول التي تتميز بها مفاهيم ووردنت.

مثال: لو أردنا من النظام ان يضيف علاقة (person, nsubj, communicate) الى قاعدة البيانات فإنه يستنتج و يضيف كذلك كل هذه العلاقات:

- (male , nsubj, communicate)
- (female, nsubj, communicate)
- (man , nsubj, communicate)
- (woman, nsubj, communicate)
- (son , nsubj, communicate)
- (girl , nsubj, communicate)

- (person, nsubj, ask)
- (person, nsubj, write)
- (person, nsubj, talk)
- (person, nsubj, speak)
- (person, nsubj, give)
- (person, nsubj, inform)



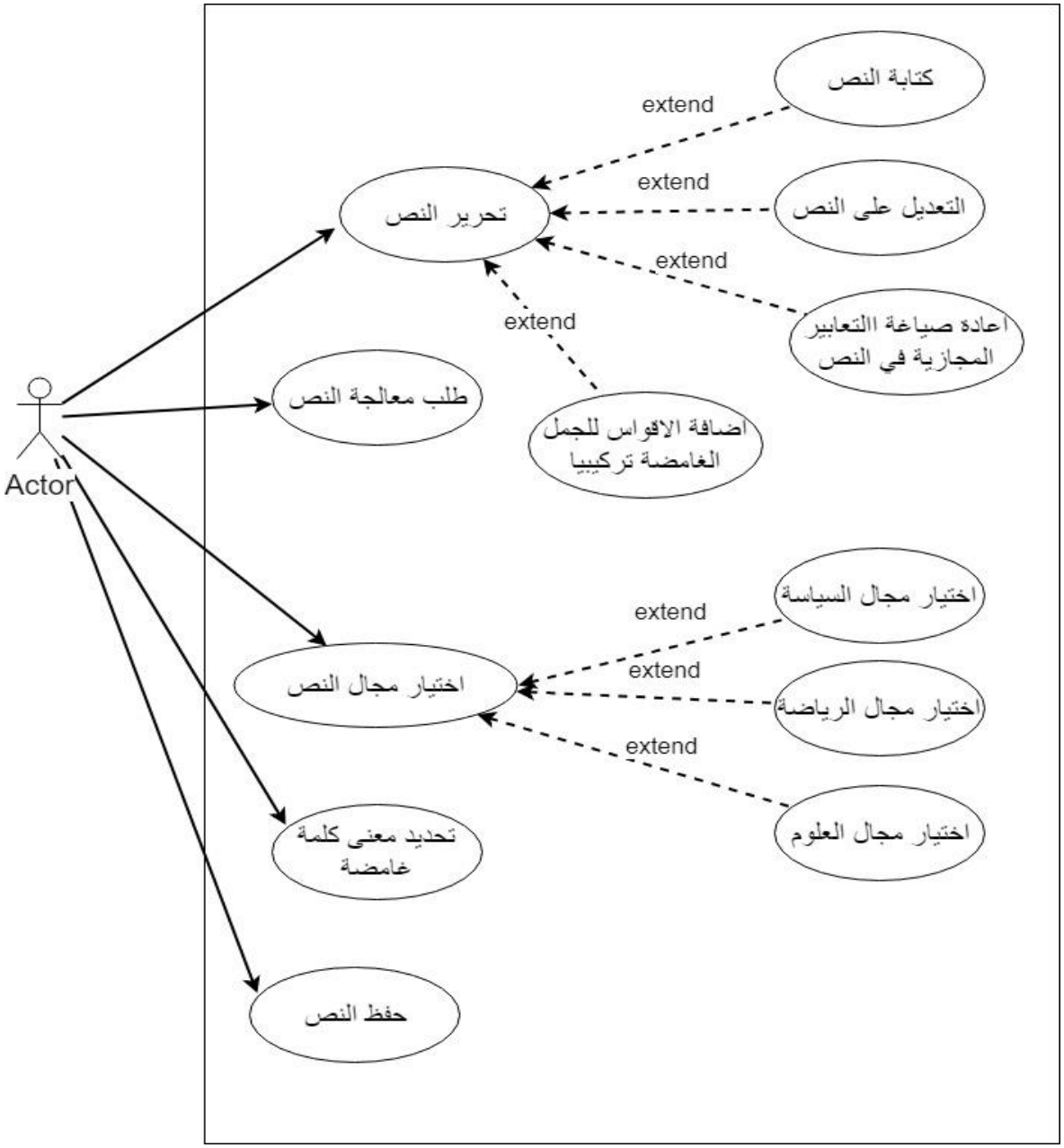
رسم توضيحي 29 : استنتاج خصائص سياقية جديدة بفضل علاقات الشمول في ووردنت

4.4 التفاعل بين النظام و كاتب النص

كما لاحظنا في نتائج الدراسة التي قمنا بها في الفصل السابق، هناك العديد من حالات الغموض لا يمكن تجاوزها الا بإشراك الكاتب في عملية ازالتها، وبالتالي سمحنا للكاتب بان يقوم بعدة مهام وتفاعلات مع النظام، من بين هذه

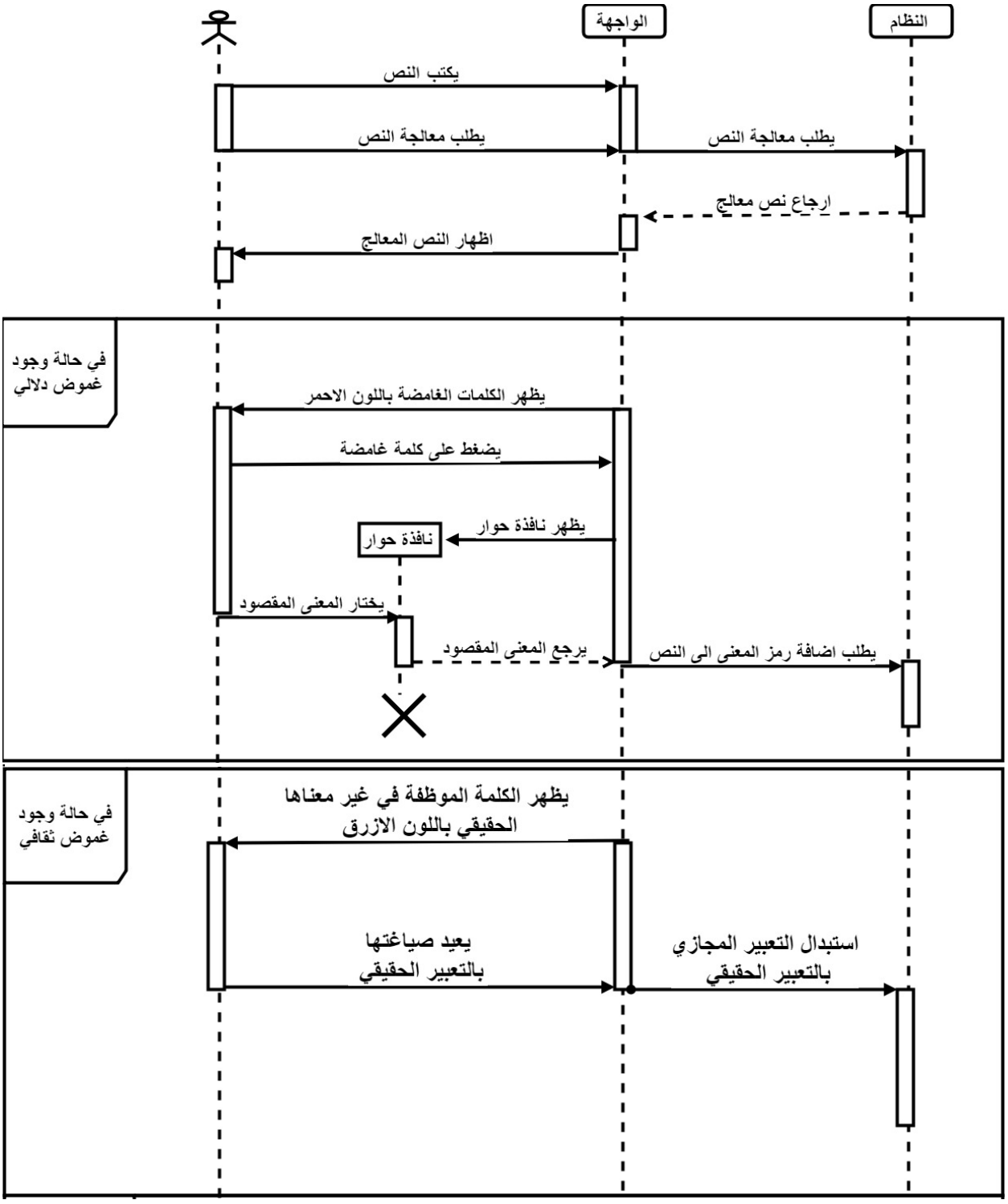
المهام : يختار مجال النص، يحدد معنى كلمة غامضة، يضيف الاقواس للجمل الغامضة تركيبيا، يعدل على النص ...الخ

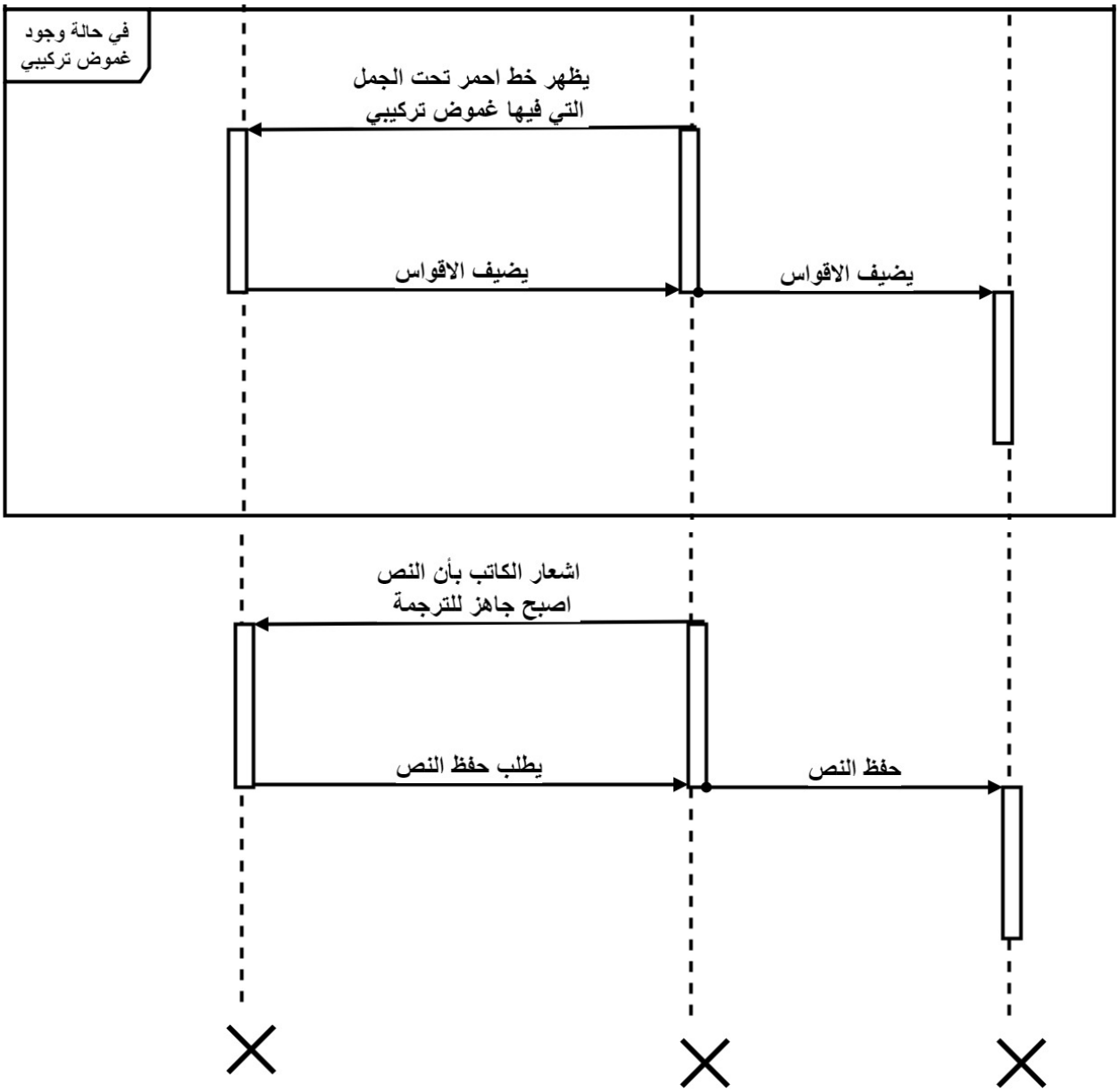
يوضح الرسم التخطيطي ادناه مختلف المهام التي يمكن للكاتب القيام بها.



رسم توضيحي 30 : مختلف المهام التي يمكن للكاتب القيام بها

مختلف التفاعلات و الرسائل المتبادلة بين الكاتب و النظام و الواجهة، او يمكن ان نسميها سيناريو تحرير النص، مسرود بالتفصيل في الرسم التخطيطي ادناه، بما في ذلك سيناريو فك الغموض المعجمي والتركيبي و الثقافي.


الواجهة
النظام
يكتب النص
يطلب معالجة النص
يطلب معالجة النص
ارجاع نص معالج
اظهار النص المعالج
في حالة وجود غموض دلالي
يظهر الكلمات الغامضة باللون الاحمر
يضغط على كلمة غامضة
يظهر نافذة حوار
نافذة حوار
يختار المعنى المقصود
يرجع المعنى المقصود
يطلب اضافة رمز المعنى الى النص
في حالة وجود غموض ثقافي
يظهر الكلمة الموظفة في غير معناها الحقيقي باللون الازرق
يعيد صياغتها بالتعبير الحقيقي
استبدال التعبير المجازي بالتعبير الحقيقي



رسم توضيحي 31 : التفاعل بين النظام و الكاتب

5.4 خاتمة

شرحنا في هذا الفصل تصميم نظام الترميز الدلالي حيث قدمنا بنيته المكونة من : واجهة – متحكم – نموذج، وفصلنا في كل مكون على حدة حيث يحتوي المتحكم على 4 وحدات : وحدة التجزيء، وحدة التحليل الصرفي، وحدة التحليل التركيبي، وحدة التحليل الدلالي. بينما يحتوي النموذج على 3 قواعد بيانات : قاعدة بيانات معجمية وقاعدة بيانات التدريب وقواعد التحليل التركيبي. ثم وضحنا كيفية تفاعل النظام مع كاتب النص ومختلف الرسائل المتبادلة بينهما.

سيتم في الفصل القادم وهو فصل التجريب و التنفيذ، تجسيد و تحقيق هذا النظام والقيام بتجارب لتقييم اداءه في مختلف المهام المطلوبة منه.

الفصل الرابع

التنفيذ و التجريب

5. الفصل الرابع : التنفيذ و التجريب

1.5 المقدمة

بعد اتمام العمل النظري ومرحلة التصميم، نأتي الي مرحلة التجسيد العملي للنظام ،في هذا الفصل سنتطرق إلى الأدوات المستخدمة و لغات البرمجة المطلوبة في تنفيذ و تحقيق هذا النظام، ثم نقوم بعدة تجارب لاختبار وتقييم اداءه في مختلف الوظائف المطلوبة منه (فك الغموض التركيبي، المعجمي و الثقافي) بالإضافة الى تجربة أخرى من اجل تقييم عملية ترجمة الكلمات بالاعتماد على سياقها في النص، بعد هذا تتم مناقشة هذه النتائج.

2.5 أدوات، واجهات البرمجة و بيئات التطوير المستخدمة

لغة بايثون Python

تكتب بايثون باللغة العربية و هي لغة برمجة عالية المستوى إبتكرها Guido Van Rossum أثناء عمله في مركز أبحاث Centrum Wiskunde & Informatica عام 1986. عام 1991 تم نشر أول إصدار منها لتصبح في متناول الجميع.

إستمر تطوير هذه اللغة و إضافة الكثير من المزايا عليها في كل إصدار جديد منها إلى يومنا هذا حتى اصبحت إحدى أهم لغات العصر و التي يمكن إستخدامها لبناء برامج سطح المكتب, تطبيقات الويب, الألعاب, سكريبتات إلخ..

بايثون تعمل على جميع و أهم أنظمة التشغيل مثل Windows, Mac OS, Linux, Unix إلخ.. و تعتبر من أشهر لغات البرمجة على الإطلاق.

بيئة طوني Thonny

هي بيئة تطوير متكاملة ومجانية لبرمجة التطبيقات بلغة بايثون، بالإضافة الى كونها مجانية تمتاز بخصائص أخرى مثل البساطة وسهولة الاستخدام، ما يجعلها خيار جيد للبرمجة بلغة بايثون.

حزمة NLTK

NLTK هي عبارة عن منصة رائدة لبناء برامج بيثون للعمل مع بيانات اللغة البشرية. توفر واجهات سهلة الاستخدام لأكثر من 50 من الموارد المعجمية مثل WordNet، جنبًا إلى جنب مع مجموعة من مكتبات معالجة النصوص من أجل التصنيف، والترميز، ووضع العلامات، والتحليل، والاستدلال المنطقي، ومنتدى مناقشة نشط.

تعتبر NLTK "أداة رائعة لتعلم اللغويات الحاسوبية باستخدام Python والعمل بها" و "مكتبة مذهلة للتعامل مع اللغة طبيعية".

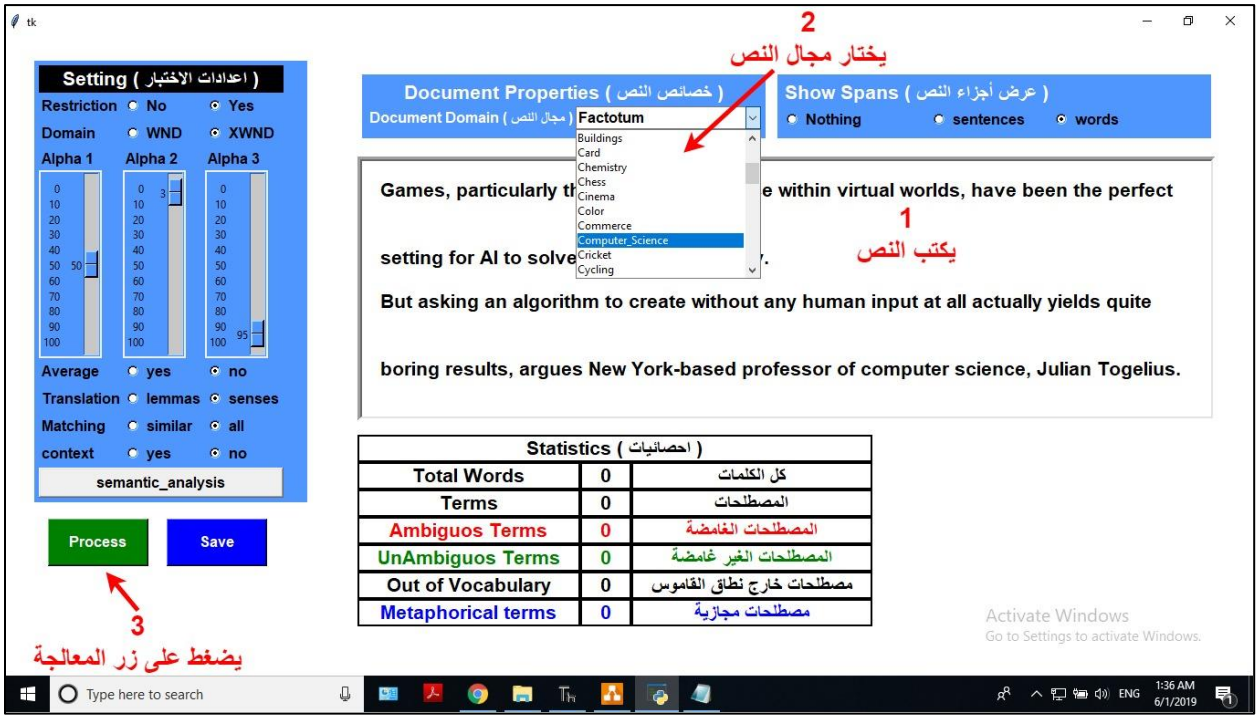
حزمة Tkinter

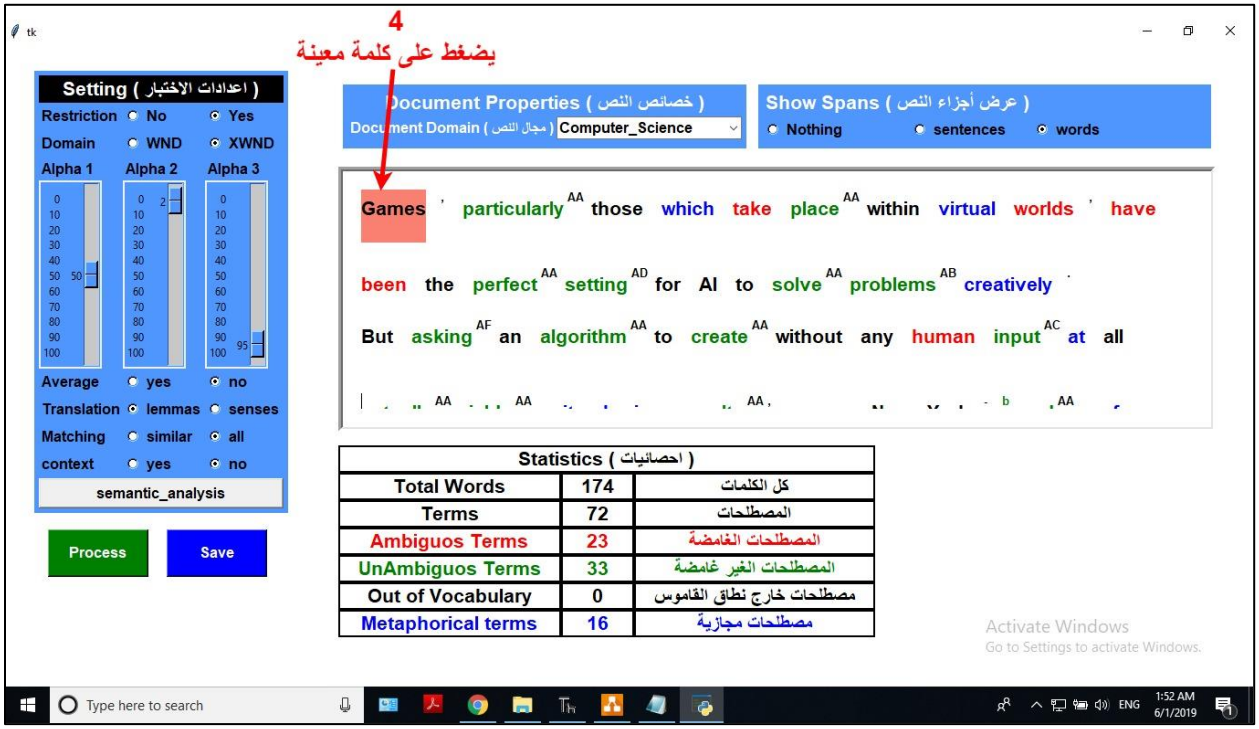
الـ tkinter عبارة عن واجهة رسومية GUI مشتقة من مكتبة Tk GUI الرسومية للعمل معها بواسطة لغة Python. وهي الواجهة الرسومية القياسية والرسمية للغة Python, مما يجعلها حزمة رسومية متاحة للعمل عند تثبيت بيئة Python. بالإضافة إلى أنها الواجهة الرسومية المستعملة لنوافذ بيئة python.

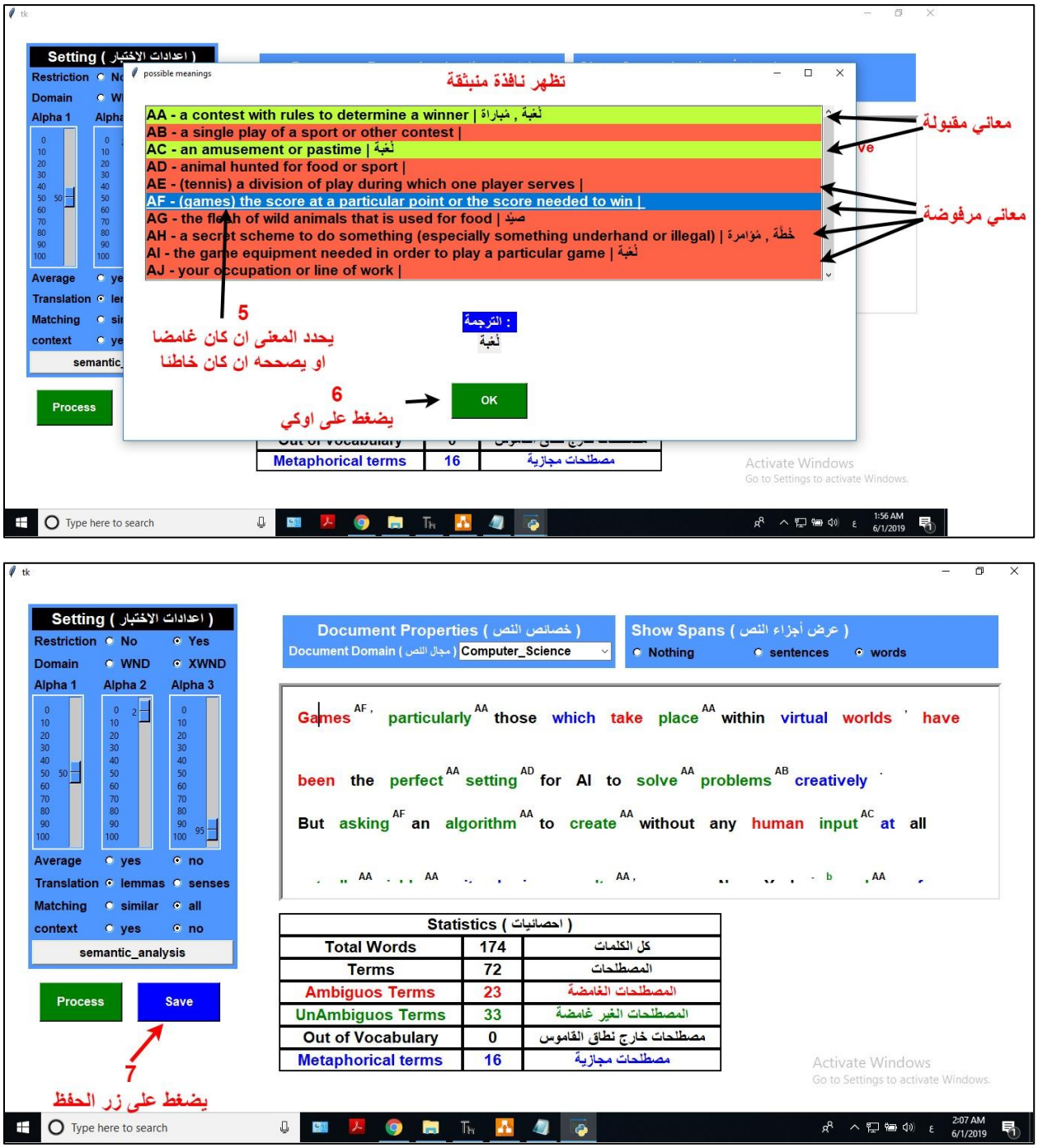
3.5 الواجهة الرسومية

صورة 4 : الواجهة الرسومية للنظام

خطوات تحرير النص :







صورة 5 : خطوات تحرير النص

4.5 التجارب، النتائج و المناقشة

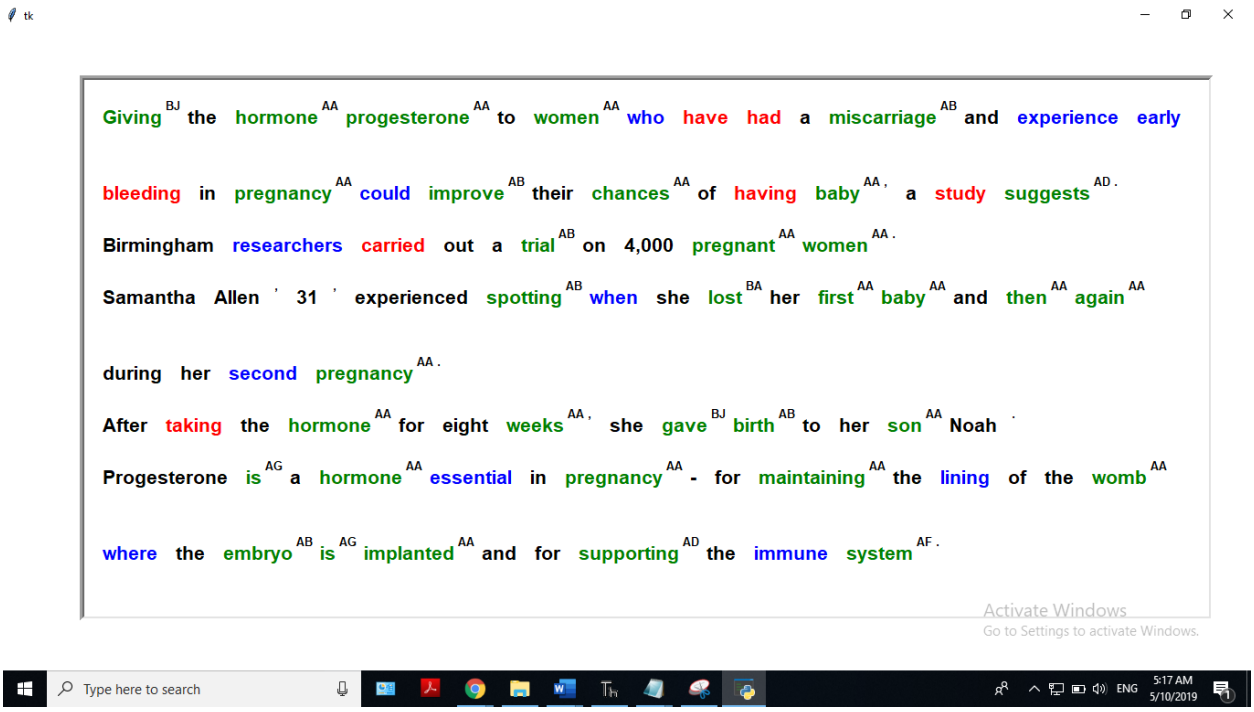
بما ان نظام الترميز الدلالي المقترح يتعامل مع ثلاثة مسائل أساسية (الغموض المعجمي، الثقافي و التركيبي) فاننا قسمنا التجارب حسب هذه المسائل، تجربة خاصة بتقييم عملية فك الغموض المعجمي و أخرى خاصة بعملية فك الغموض الثقافي و أخرى لتقييم عملية فك الغموض التركيبي.

1.4.5 تقييم عملية فك الغموض المعجمي

- التجربة

خلال هذه التجربة، جمعنا 3 نصوص انجليزية من موقع BBC News في مجالات (الصحة – الرياضة – علم الحاسوب) بهدف تقييم أداء النظام في عملية تحديد المعاني الصحيحة للكلمات، ثم قمنا بمعالجتها بواسطة النظام المقترح

ملاحظة : تمت هذه التجربة بالقيم التالية للعتبات : $\alpha 3 = 0$, $\alpha 2 = 0.02$, $\alpha 1 = 0.5$



صورة 6 : لقطة شاشة توضح نتيجة اختبار فك الغموض المعجمي

فكانت النتيجة كالتالي :

- النتيجة

اجمالي الكلمات	277	كلمات غير غامضة	160		
		كلمات غامضة	117	كلمات اصاب في تحديد معناها	72
				كلمات اخطأ في تحديد معناها	45

جدول 10 : نتيجة اختبار فك الغموض المعجمي

نسبة الدقة : $\text{precision} = \frac{72}{117} = \mathbf{61.15\%}$

- المناقشة

من الجدير بالذكر ان دقة هذه العملية مرهون بثلاثة نقاط أساسية وهي :

- دقة المشروع الذي استخدمناه في ربط المفاهيم بمجالاتها وهو مشروع مجالات ووردنت الموسعة eXtended WordNet Domains
- دقة العتبة الخاصة باقصاء المفاهيم المجازية
- دقة العتبة الخاصة باقصاء المفاهيم الغير متعلقة بالمجال المختار من طرف الكاتب
- دقة العملية السابقة لهذه العملية و هي عملية التحليل الصرفي و تحديد أصول و أنواع الكلمات

ولكن على الرغم من هذا، فتحديد معنى 72 كلمة بشكل صحيح من اصل 117 كلمة يعتبر امر جيدا على الرغم من الأخطاء التي وقع فيها المحلل الصرفي الذي استخدمناه، كمثال على ذلك كلمة who التي استخدمت في النص بنوع :

- صيغة سؤال تعنى "من الفاعل ؟"

ولكنه اعتبر نوعها :

- "اختصار" لكلمة منظمة الصحّة العالمية World Health Organization

والعديد من الحالات التي تم تحديد معناها بشكل خاطئ يرجع الى احد الأسباب الأربعة المذكورة أعلاه، فاختيار العتبتين كان امر صعب للغاية حيث يجب عدم اختيارهما كبيران لكي لايتم اقصاء المفاهيم المقصودة ولا يجب اختيارهما صغيران بالقدر الذي يسمح باختيار المفاهيم الغير مقصودة.

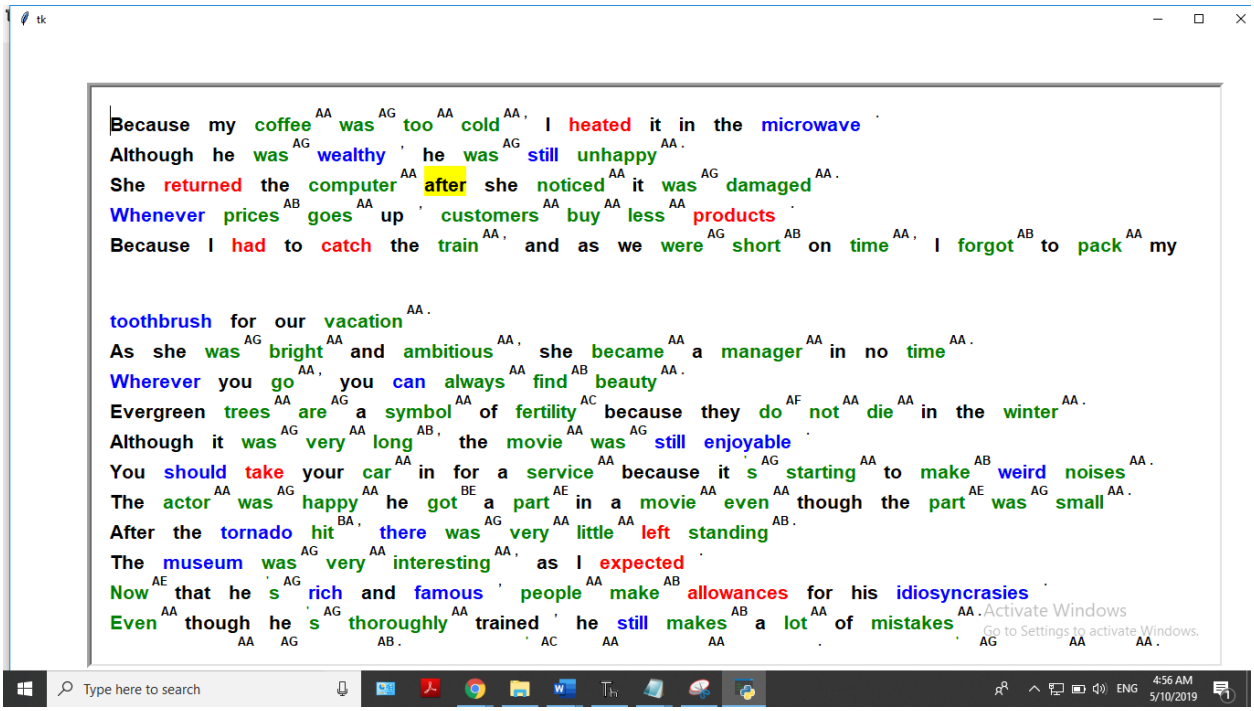
2.4.5 تقييم عملية فك الغموض الثقافي

- التجربة

من اجل تقييم اداء نظامنا في عملية كشف العبارات المجازية والاصطلاحية، حضرنا مجموعة من جمل إنجليزية تحتوي على مجاز من مواقع متفرقة على الانترنت، و قمنا بمعالجتها بواسطة النظام المقترح

ملاحظة : تمت هذه التجربة بالقيم التالية للعتبات : $\alpha 3 = 0$, $\alpha 2 = 0.02$, $\alpha 1 = 0.5$

فكانت النتائج كالتالي :

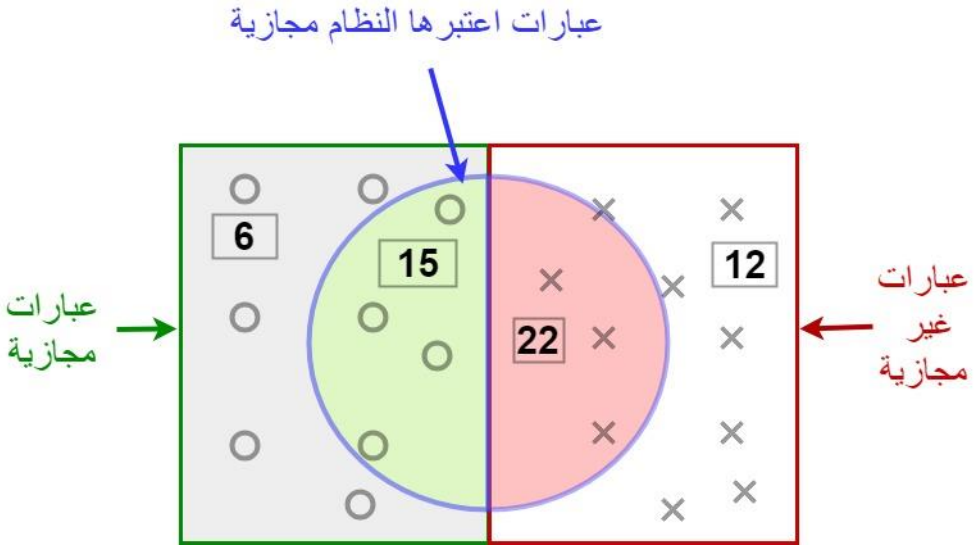


صورة 7 : لقطة شاشة توضح نتيجة اختبار فك الغموض الثقافي

- النتيجة

اجمالي العبارات	55	عبارات مجازية	21	عبارات اعتبرها مجازية	15
				عبارات اعتبرها غير مجازية	6
		عبارات غير مجازية	34	عبارات اعتبرها مجازية	22
				عبارات اعتبرها غير مجازية	12

جدول 11 : نتيجة اختبار فك الغموض الثقافي



رسم توضيحي 32 : نتيجة اختبار فك الغموض الثقافي

حساب مقياسي الدقة (precision،recall)

$$\textbf{Precision} = \frac{relevent\ selected\ items}{relevent\ items} = \frac{15}{21} = 71\%$$

$$\textbf{Recall} = \frac{relevent\ selected\ items}{selected\ items} = \frac{15}{37} = 41\%$$

- المناقشة

نفس الملاحظة التي ذكرناها الخاصة بالنقاط التي تتحكم في دقة عملية فك الغموض المعجمي تنطبق كذلك على عملية كشف التعابير المجازية باستثناء النقطة الثالثة، وبالتالي هذه هي النقاط المتحكمة :

- دقة المشروع الذي استخدمناه في ربط المفاهيم بمجالاتها وهو مشروع مجالات ووردنت الموسعة eXtended WordNet Domains
- دقة العتبة الخاصة باقصاء المفاهيم المجازية
- دقة العملية السابقة لهذه العملية و هي عملية التحليل الصرفي و تحديد أصول و أنواع الكلمات

من بين 21 عبارة مجازية تم اعتبار 15 منها مجازية وهذا امر مشجع للغاية، خاصة واننا نعرف أسباب الأخطاء التي وقع فيها، فمعظم التعابير المجازية التي لم يتم اكتشافها، لم يتم اكتشافها لشيئين :

- اما بسبب سوء اختيار عتبة اقصاء التعابير المجازية (كبيرة اكثر مما ينبغي)
- اما بسبب عدم دقة الربط بين المفاهيم و المجالات

بينما في المقابل من بين 37 عبارة تم اعتبارها مجازية 15 فقط كانت مجازية، فهذا تقريبا راجع الى نفس الأسباب :

- اما بسبب سوء اختيار عتبة اقصاء التعابير المجازية (صغيرة اكثر مما ينبغي)
- اما بسبب عدم دقة الربط بين المفاهيم و المجالات

3.4.5 تقييم عملية ترجمة الكلمات من خلال السياق المحلي

- التجربة

نظرا لضيق الوقت ولتعقيد هذه العملية اقتصرنا في هذه التجربة فقط على ترجمة الأفعال.

اخترنا فعل من اكثر الأفعال غموضا في الإنجليزية وهو فعل break حيث يحمل هذا الفعل حوالي 59 معنى حسب ووردنت، ثم جمعنا 57 جملة (بعضها من امثلة ووردنت والبعض الاخر من انشائنا) يظهر فيها هذا الفعل ب 6 معاني مختلفة. وهي :

- كسر – حطم شيء ما ----- متعدي
- اخترق - انتهك – خالف (قانون او قاعدة)
- انهى، قطع، أوقف (نشاط مستمر) ----- (متعدي)
- انكسر، تكسّر، تحطّم ----- (لازم)
- عطّل، اعطب، أوقف (عمل جهاز) ----- (فعل متعدي)
- تعطّل، توقّف (عن العمل) ---- (فعل لازم)

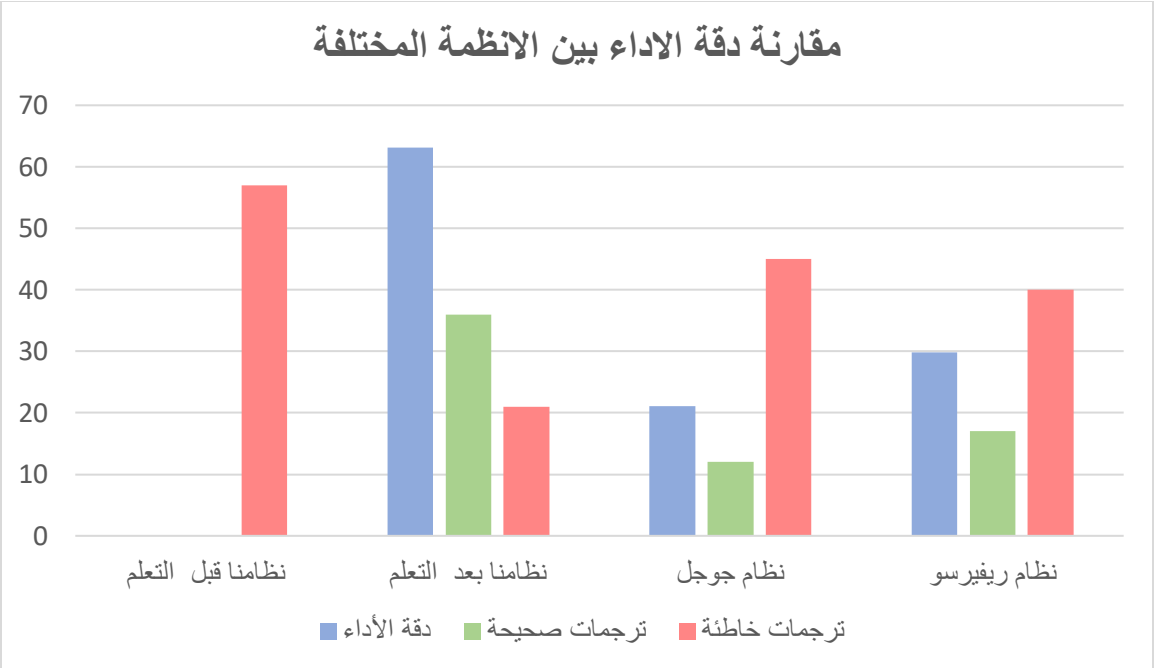
قبل التعلم : عالجنا هذه المجموعة بدون ان نزود النظام بأي معلومة عن المعاني.

بعد التعلم : عالجنا هذه المجموعة بعد تزويد النظام بمعاني كلمة break في 6 جمل فقط.

ملاحظة : تمت هذه التجربة بالقيم التالية للعتبات : $\alpha3 = 0.95$, $\alpha2 = 0.02$, $\alpha1 = 0.5$

نظام الترجمة	ترجمات خاطئة	ترجمات صحيحة	نسبة الدقة
نظامنا قبل التعلم	57	0	%0
نظامنا بعد التعلم	**21**	**36**	**%63.15**
نظام جوجل	45	12	%21.05
نظام ريفيرسو	40	17	%29.82

جدول 12 : نتيجة اختبار ترجمة كلمة break *من خلال السياق المحلي لها*



رسم توضيحي 33 : مقارنة دقة الاداء بين الانظمة المختلفة

- **المناقشة**

نلاحظ ارتفاع دقة أداء النظام بعد عملية التعلم مقارنة بدقة اداءه قبلها وهذا يثبت مدى جدوى هذه العملية في تحسين اداءه خاصة في تحديد معاني الأفعال الغامضة، ونلاحظ أيضا تفوق كبير للنظام مقارنة بالانظمة الأخرى وهذا راجع الى عدة أسباب منها :

- محدودية الاختبار حيث انه مخصص لفعل واحد فقط ول6 معاني له فقط ومجموعة اختبار تحتوي على 57 جملة فقط، فلو كان الاختبار أوسع نطاق لكانت النتائج بشكل اخر حيث نتوقع انخفاض دقة أداء النظام عن هذه النسبة.
- اعتماد النظام على عاملين فعّالين في تحديد معاني الكلمات وهما : عامل المجال العام للنص و عامل السياق المحلي للكلمة الغامضة

جدول13: التفاصيل الكاملة للنتائج المتحصل عليها

الرقم	الجمل	نظامنا	ترجمة جوجل	ترجمة ريفيرسو	المعنى
1	the child broke the window.	كسر	كسر الطفل النافذة.	قام الطفل بكسر النافذة.	كسر، حطّم (شيء ما) (فعل متعدي)
2	this person broke my cup of coffee.	خالف	كسر هذا الشخص فنجان قهوة.	قام هذا الشخص بكسر فنجان	
3	my friend broke his big wall.	كسر	كسر صديقي جداره الكبير.	لقد كسر صديقي حائطه الكبير.	
4	he broke the glass plate.	كسر	كسر لوحة الزجاج.	كسرت اللوح الزجاجي.	
5	they were breaking their gears.	كسر	كانوا يكسرون التروس.	كانوا يكسرون قواتهم.	
6	we finally broke the old chair	كسر	لقد حطمنا الكرسي القديم تمامًا.	لقد كسرنا أخيرا الكرسي القديم	
7	he broke figurine .	/	كسر التمثال.	هو [بروك] [فيغبول].	
8	he broke the freshly baked loaf.	/	كسر الرغيف الطازج.	وقد كسر الرغيف المخبوز حديثا	
9	this man breaks the license.	كسر	هذا الرجل يكسر الترخيص.	هذا الرجل يكسر الرخصة	خالف، انتهك، اخترق (قانون او وعد او قاعدة او نظام) (فعل متعدي)
10	my friend breaks his law.	كسر	صديقي يكسر قانونه.	صديقي يخالف قانونه	
11	the stealer broke the security of the	/	كسر السارق أمن البنك.	وقد كسر المتسلل أمن البنك	
12	she breaks the written agreement.	خالف	انها كسر الاتفاق المكتوب.	هي تكسر الإتفاق المكتوب	
13	they breaks the assignment.	كسر	يكسرون المهمة.	فيكسرون المهمة	
14	they were breaking all laws of	/	كانوا يخرقون جميع قوانين الإنسانية	إنهم ينتهكون جميع قوانين	
15	they break the basic laws or human	/	يخرقون القوانين الأساسية أو الحضارة	إنها تنتهك القوانين الأساسية أو	
16	he breaks the law	/	انه يخرق القانون	فهو ينتهك القانون	
17	she broke her promise	خالف	كسرت وعدها	لقد خالفت وعدها	
18	I have broken a lucky streak	انهى	لقد كسرت سلسلة محظوظ	أنا قد كسرت خط محظوظة	انهى، قطع، أوقف (نشاط مستمر) (فعل متعدي)
19	we are breaking the cycle of poverty	/	نحن نكسر حلقة الفقر	إننا نكسر حلقة الفقر	
20	she broke her pregnancy	انهى	كسرت حملها	هي [بروك] حالت حملها	
21	she breaks the show	انهى	إنها تكسر العرض	هي تكسر العرض	
22	they will broke the contest	/	سوف يكسرون المسابقة	هم سيكسرون المسابقة	
23	he broke the affair	انهى	كسر العلاقة	لقد كسر هذا الأمر	
24	he broke the party	انهى	كسر الحفل	هو [بروك] الحزب	
25	he broke the celebration	انهى	كسر الاحتفال	وقد كسر الاحتفال	
26	he broke the game	/	كسر اللعبة	هو كسر اللعبة	
27	he broke the movement	/	كسر الحركة	لقد كسر الحركة	
28	he broke the episode	انهى	كسر الحلقة	هو كسر الحلقة	

الرق	الجمل	نظام	ترجمة جوجل	ترجمة ريفيرسو	المعنى
29	he broke the success	انهى	كسر النجاح	هو [بروك] النجاح	انهى، قطع، أوقف (نشاط مستمر) (فعل متعدي)
30	he broke the miracle	انهى	كسر المعجزة	لقد كسر المعجزة	
31	he broke the match	عطّل	انه كسر المباراة	وقد كسر المباراة	
32	the window broke	تعطّل	نافذة مكسورة	تم كسر النافذة	انكسر، تكسّر، تحطّم (وقع عليه فعل الكسر) (فعل لازم)
33	my cup of coffee broke	تعطّل	بلدي فنجان من القهوة كسر	فنجان القهوة مكسور	
34	his big wall broke	انكسر	كسر جداره الكبير	وقد انهار حائطه الكبير	
35	the glass plate broke	تعطّل	كسر لوحة الزجاج	كسرت الصفيحة الزجاجية	
36	their gears was breaking	تعطّل	كسر التروس بهم	وقد انكسرت تروس تلك القوات	
37	the old chair finally broke	تعطّل	الكرسي القديم كسر أخيرا تماما	وقد انكسر الكرسي القديم تماما في النهاية	
38	the figurine broke	تعطّل	كسر التمثال	[فيغبول] [بروك]	
39	the freshly baked loaf broke	انكسر	كسر الرغيف الطازج	وقد كسرت الرغيف الطازج	
40	you broke the alarm clock when you	عطّل	لقد كسرت المنبه عندما انفصلت	لقد حطمت ساعة المنبه عندما أخذتها بعيدا	عطّل، اعطب، أوقف (عمل جهاز) (فعل متعدي)
41	he broke the engine finally	عطّل	كسر المحرك في النهاية	لقد كسر المحرك أخيرا	
42	they broke the car on the road	عطّل	كسروا السيارة على الطريق	وكسروا السيارة على الطريق	
43	he broke the bus we travelled in on	عطّل	كسر الحافلة التي سافرنا إليها في	لقد كسر الحافلة التي سافرناها في الطريق إلى	
44	he broke the coffee maker	كسر	كسر صانع القهوة	وقد كسر آلة صنع القهوة	
45	he broke the engine on the way to	عطّل	كسر المحرك في الطريق إلى	وكسر المحرك في الطريق إلى البلدة	
46	he broke her eyesight after the	انهى	كسر بصرها بعد الحادث"	كسر بصرها بعد الحادث"	
47	he broke my radio	عطّل	كسر راديو بلدي	هو [بروك] إذاعتي	
48	they finally broke The lawn mower	عطّل	أنها كسرت أخيرا جزازة العشب	هم أخيرا كسروا المرج جزازة	
49	the alarm clock broke when you took	تعطّل	المنبه كسر عندما تفرقت!	لقد انقطعت ساعة المنبه عندما أخذتها بعيدا!	تعطّل، توقّف (عن العمل) (فعل لازم)
50	the engine finally broke	تعطّل	المحرك كسر أخيرا	وأخيرا انكسر المحرك	
51	the car broke on the road	تعطّل	اندلعت السيارة على الطريق	وقد كسرت السيارة على الطريق	
52	the bus we travelled in broke down	تعطّل	الحافلة التي سافرنا إليها تعطلت في	تعطلت الحافلة التي سافرناها في الطريق إلى	
53	the coffee maker broke	انكسر	كسر صانع القهوة	كسر آلة صنع القهوة	
54	the engine broke on the way to town	تعطّل	كسر المحرك في الطريق إلى	وانشق المحرك في طريقه إلى البلدة	
55	her eyesight broke after the	انكسر	كسر بصرها بعد الحادث"	انكسر بصرها بعد الحادث"	
56	my radio broke	تعطّل	كسر الراديو الخاص بي	لقد انكسر الراديو الخاص بي	
57	the lawn mower finally broke	تعطّل	جزازة العشب كسر أخيرا	كسر جزازة المرج أخيرا	

4.4.5 تقييم عملية فك الغموض التركيبي

- التجربة

من اجل تقييم اداء نظامنا في هذه المهمة، اخترنا عدة جمل انجليزية تحتوي على غموض تركيبي، ثم اخضعناها للمعالجة :

I said that i would see you on Tuesday . The old men and women stayed at home . They fed her rat poison . He saw a man eating fish . Students hate annoying Professors . They saw the woman with the telescope . Joyce said that Tom left you on yesterday . We saw her duck. He fed her cat food. Look at the dog with one eye. Mano met a woman with a wooden leg named Aminah. The lady hit the man with an umbrella. They are looking for good French and German and Japanese teachers. Mary saw the dog in the window. she tell me that he went there on Monday .

some juice to drink? The pizza smells delicious. There is a fly in the car with us. Look on top of the refrigerator for the key. I am out of paper for the printer. Will you help me with the math homework? The music is too loud for my ears.Because my coffee was too cold, I heated it in the microwave. Although he was wealthy, he was still unhappy. She returned the computer after she noticed it was damaged. Whenever prices goes up, customers buy less products. Because I had to catch the train, and as we were short on time, I forgot to pack my toothbrush for our vacation. As she was bright and ambitious, she became a manager in no time. Wherever you go, you can always find beauty. Evergreen trees are a symbol of fertility because they do not die in the winter. Although it was very long, the movie was still enjoyable. You should take your car in for a service because it's starting to make weird noises.

Activate Windows

صورة 8 : لقطة شاشة تظهر نتيجة اختبار فك الغموض التركيبي

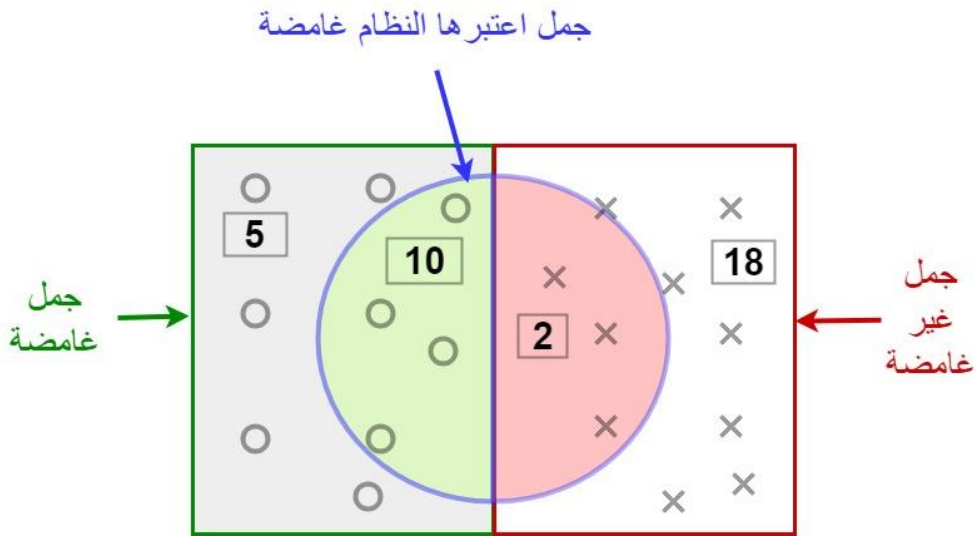
فكانت النتيجة كالتالي :

- النتيجة

اجمالي العبارات	35	عبارات غامضة تركيبيا	15	عبارات اعتبرها غامضة تركيبيا	10
				عبارات اعتبرها غير غامضة تركيبيا	5
			20	عبارات اعتبرها غامضة تركيبيا	2

		عبارات غير غامضة تركيبيا		عبارات اعتبرها غير غامضة تركيبيا	18

جدول 14 : نتيجة اختبار فك الغموض التركيبي



رسم توضيحي 34 : نتيجة اختبار فك الغموض التركيبي

حساب مقياسي الدقة (precision،recall)

$$\textbf{Precision} = \frac{relevent\ selected\ items}{relevent\ items} = \frac{10}{12} = 83\%$$

$$\textbf{Recall} = \frac{relevent\ selected\ items}{selected\ items} = \frac{10}{15} = 66\%$$

- المناقشة

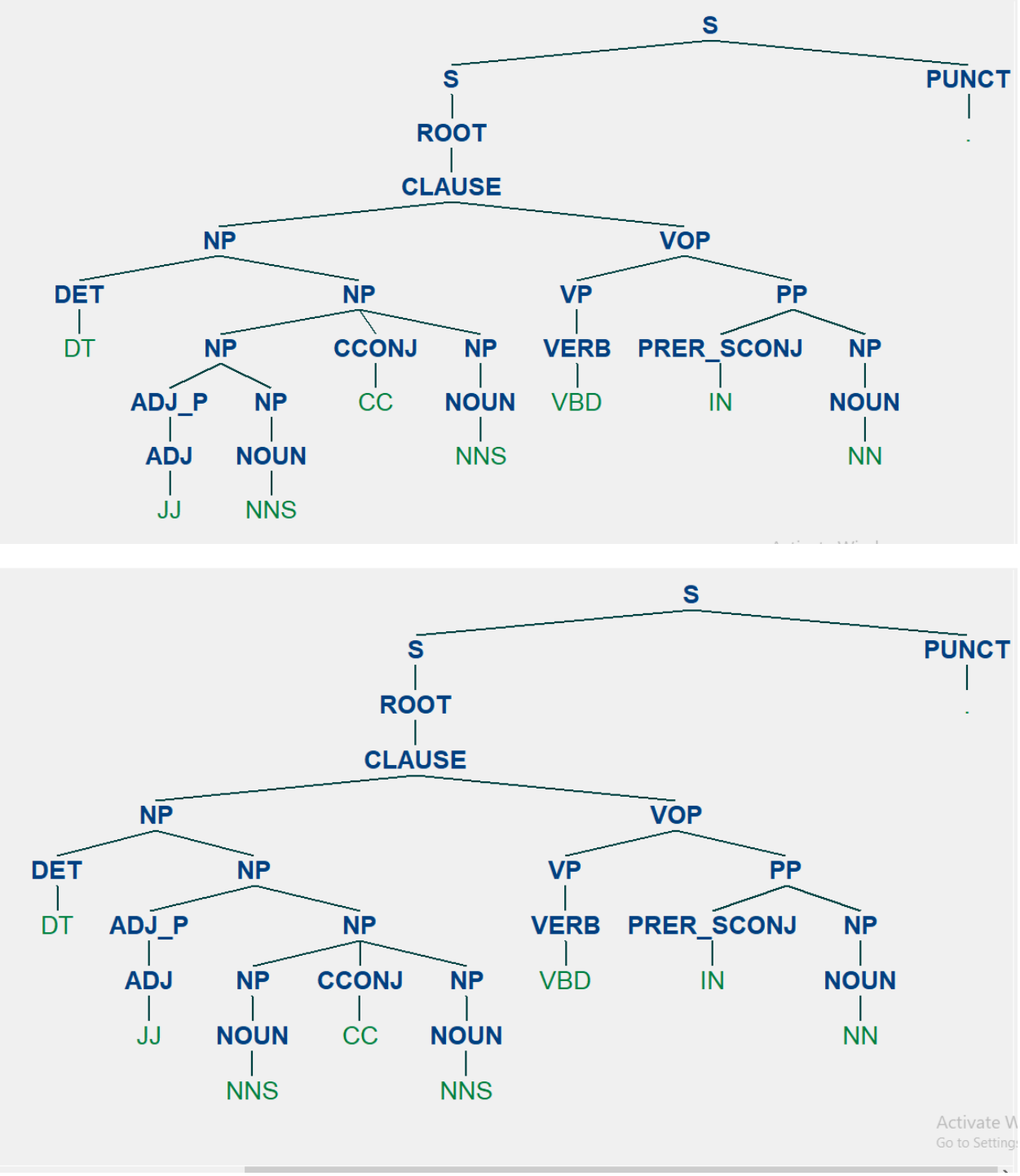
اذن حسب النتائج المتحصل عليها، حقق النظام نسبتي دقة جيدة (83% بالنسبة لمقياس precision و 66% بالنسبة لمقياس recall)، حيث يمثل مقياس precision نسبة العناصر المختارة بشكل صحيح على اجمالي العناصر الصحيحة (أي في حالتنا هذه يمثل نسبة العبارات المعتبرة غامضة تركيبيا شكل صحيح على اجمالي العبارات الغامضة تركيبيا)، اما مقياس recall فيمثل نسبة العناصر المختارة بشكل صحيح على اجمالي العناصر المختارة (وفي حالتنا هذه يمثل نسبة العبارات المعتبرة غامضة تركيبيا بشكل صحيح على اجمالي العبارات المعتبرة الغامضة تركيبيا), وهذه نسبة جيدا نظرا الى ان التحليل التركيبي الذي اعتمدناه كان تحليلا سطحيا يتعامل فقط مع أنواع الكلمات (اسم، فعل، صفة ... الخ) وليس تحليل عميقا يتعامل كذلك مع الخصائص الصرفية للكلمات (العدد، الجنس ..الخ).

بالإضافة الى ذلك فان معظم الجمل الغامضة و التي لم يتم اكتشافها كانت تحمل تراكيب معقدة لم نعتبرها في مجموعة القواعد الخاصة بالتحليل التركيبي، و هذه الجملة مثال عليها :

جملة : They are looking for good French and German and Japanese teachers.

وهذا مثال اخر على جملة تم كشف غموضها مرفقة بشجرتي التحليل التركيبي الخاصة بها :

جملة : The old men and women stayed at home.



رسم توضيحي 35 : شجرتي التحليل التركيبي لجملة غامضة

5.5 الخاتمة

في هذا الفصل قمنا بتجسيد وتحقيق نظام الترميز الدلالي المقترح الذي يهدف الى تحديد دلالة الجمل بالاستعانة الى الكاتب و مجال النص وسياق الكلمات، وذلك من اجل تسهيل عملية ترجمتها من طرف النظم الخاصة بالترجمة الى اللغة العربية. كما نود الإشارة الى ان هذا العمل سيبقى مفتوح على تحديثات وتحسينات واضافات أخرى في المستقبل بسبب العديد من النواقص التي تحتاج اتمامها والعديد من الحالات التي يجب اعتبارها في المستقبل.

6. خاتمة عامة

لقد تم في هذه المذكرة تقديم نظام ترميز دلالي مقترح من طرفنا لترميز او تشفير المعنى من اجل استعمال هذه الرموز الدلالية فيما بعد بواسطة النظم المختصة في الترجمة الالية، يأتي عمل هذا النظام المقترح عند عملية تحرير النص من طرف الكاتب وذلك بهدف رفع الغموض عنه بأنواعه الثلاثة (المعجمي، الدلالي والتركيبي) من اجل انتاج نص جاهز للترجمة في الأخير.

تعتمد مقاربتنا في تحليلها للنص الاصلي على المنهج القواعدي بمساعدة الكاتب في نفس الوقت، حيث يتعمق التحليل ليتجاوز التحليل الصرفي والتركيبي للجملة الى التحليل الدلالي وذلك بفضل قاعدة بيانات معجمية دلالية (ووردنت) وباستعمال عوامل وتقنيات مشروحة بالتفصيل (عامل المجال العام للنص، تقنية إزالة الغموض المعجمي بواسطة السياق المحلي للكلمة الغامضة، العلاقات الدلالية والتشابه الدلالي التي يتميز بهما قاموس ووردنت).

وفي الأخير بعد تقييم اداء النظام وذلك بتقييم اداء كل عملية على حدى (عملية فك الغموض الدلالي بنوعيه، عملية فك الغموض التركيبي، عملية ترجمة الكلمات التي تختلف معناها باختلاف سياقها). وقد أظهر هذا التقييم نتائج مشجعة وذلك باعتبار ان نطاق الاختبار لم يكن واسع بالقدر الكافي ليشمل كل الحالات والامثلة، ومع هذا فان إضافة بعض التعديلات والزيادات الى النظام من شأنها ان تجعله يغطي نطاق أوسع وامثلة اكثر وبالتالي كفاءة وأداء احسن وافضل.

المراجع

[1] J. Hutchins، (1 ed.) The Oxford Handbook of Computational Linguistics، 2005.

[2] M. A. Chéragui, Theoretical Overview of Machine translation, Adrar, Algeria: African University.

[3] أ. بلحوتس، تقييم أداء برامج الترجمة الالية : فك الغموض في الترجمة إلى اللغة العربية، جامعة مولود معمري بالجزائر.

[4] T. D. Hedden, "MACHINE TRANSLATION: A BRIEF INTRODUCTION," [Online]. Available: http://ice.he.net/~hedden/intro_mt.html. [Accessed 8 4 2019].

[5] Available: [متصل]. "Pre-editing and post-editing (MT)"، [تاريخ الوصول 8 4 2019]. ./https://www.unige.ch/fti/en/faculte/departements/dtim/recherches/ta

[6] Delta State, Nigeria: M. D. Okpor، "Machine Translation Approaches: Issues and Department of Computer Science, Delta State Polytechnic, Ozoro.

[7] "Wikipedia," [Online]. Available: https://en.wikipedia.org.

[8] Nanchang330013, China : L. Peng، "A Survey of Machine Translation Methods Foreign Language School East China Jiaotong University.

[9] Siemens Information Systems W. BENNETT، "The Place of Semantics in MT Systems. Inc., Austin, Texas, USA.

[10] [تاريخ الوصول 20 4 2019]. Available: http://wndomains.fbk.eu/index.html. [متصل] "Wordnet Domains"،

[11] Available: [متصل]. "WordNet، WordNet Princeton University، [تاريخ الوصول 20 4 2019]. ./https://wordnet.princeton.edu