الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي و البحث العلمي

Université Abou Bekr Belkaid
Tlemcen Algérie
جامعة أبي بكر بلقايد
تلمسان الجزائر

كلية الآداب واللغات

قسم اللغة العربية وآدابها

تصميم طرق معالجة لغوية لتلخيص النصوص العربية

العرض باللغة العربية والفرنسية

رسالة مقدمة لنيل شهادة الدكتوراه في الدراسات اللغوية

من إعداد الطالب :

عيجولي حسين

تحت إشراف الأستاذ الدكتور :

محمد عباس

لجنة المناقشة :

أ.د. هشام خالدي	أستاذ التعليم العالي	جامعة تلمسان	رئيسا
أ.د. محمد عباس	أستاذ التعليم العالي	جامعة تلمسان	مشرفا مقررا
أ.د. قريش أحمد	أستاذ التعليم العالي	جامعة تلمسان	عضوا مناقشا
أ.د. فتحي محمد	أستاذ التعليم العالي	جامعة سيدي بلعباس	عضوا مناقشا
د. سعيد عكاشة	أستاذ محاضر - أ -	جامعة سيدي بلعباس	عضوا مناقشا
د. أحمد دواح	أستاذ محاضر - أ -	المركز الجامعي بمغنية	عضوا مناقشا

1439 / 1438

2018 / 2017

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي و البحث العلمي

Université Abou Bekr Belkaid
Tlemcen Algérie

جامعة أبي بكر بلقايد
تلمسان الجزائر

كلية الآداب واللغات

قسم اللغة العربية وآدابها

تصميم طرق معالجة لغوية لتلخيص النصوص العربية

العرض باللغة العربية والفرنسية

رسالة مقدمة لنيل شهادة الدكتوراه في الدراسات اللغوية

من إعداد الطالب :

عيجولي حسين

تحت إشراف الأستاذ الدكتور :

محمد عباس

1439 / 1438

2018 / 2017

بسم الله الرحمن الرحيم

شـــكـــــــر

أتقدم بجزيل الشكر والامتنان والعرفان إلى الأستاذ الدكتور محمد عباس الذي شرفني بقبول إدارته البحث وتوجيهه لي لاستكمال هذا العمل والذي لم يكن ليرى النور لولا مساندته وتوجيهاته الحكيمة ونصائحه وعنايته وصبره معي، والذي لم يأل جهدا في سبيل إخراج هذه الرسالة وإتمام فصولها وأبوابها .

كما أشكر أعضاء لجنة المناقشة الموقرة، الذين تحملوا عناء القراءة والتمحيص والتصحيح لإثراء موضوع هذا العمل.

وإلى كل من مد لي يد العون والمساعدة وأسهم في أن تخرج هذه الرسالة إلى النور.
ليجد الجميع هنا أسمى عبارات الشكر والامتنان .

الإهداء

أهدي هذا العمل إلى روحَيْ
فقيدَيَّ في 2015 :
الوالدة والزوجة

مقدمـــــة

تعتبر اللغة العربية جزءا من اللغات الصعبة الفهم في مجال معالجة المعلومات ونظمها. ومع ذلك، فإن الحاجة إلى نظم المعلومات أصبحت ضرورية بصورة متزايدة مع انفجار استخدام الإنترنت حيث لا تزال لغة الإنسان ناقلا رئيسيا للمعلومات، وخاصة ما يرتبط باللغة العربية ويتعلق بمكوناتها الطبيعية فيها، بحروفها وأصواتها ورسمها وتراكيبها، هذه اللغة التي يتجاوز عدد المتحدثين أو المتكلمين بها أربع مائة وعشرون مليونا نسمة (420 مليون نسمة).

فمنذ عدة سنوات، وبعد الانتشار العالمي للانترنت وتوسع وسائل الاتصال باللغة العربية، تم تصميم عدد من نظم المعلومات المعربة مثل نظم الترجمة، نظم استعـــــلام قواعد البيانات ونظم البحث المعلوماتي، والأخذ بالتلخيص الآلي وتحليل النصـــــــوص باللغة العربية.

والهدف من السعي إلى معالجة اللغة الطبيعية هو تصميم برامج قادرة على معالجة البيانات في اللغة الطبيعية التي نقوم من خلالها بعدة خطوات تحليل ضرورية لاستخراج المعلومات التي تحويها.

مع ظهور الوثائق الإلكترونية، يتم إنشاء كميات هائلة من المعلومات المتزايدة.

وهذه الزيادة في حجم النصوص يتطلب إنتاج الأدوات القوية لتكنولوجية المعلومات كفيلة بإيجاد واستخراج المعلومات المهمة في شكل مكثف.

إن التلخيص الآلي للنصوص يبدو حلا جيدا يقف على مفترق طرق تخصصين :

أ- أحدهما يختص بالمعالجة الآلية للغة (TAL)

ب- والآخر يختص باسترجاع المعلومات (RI).

فالتلخيص الآلي للنصوص هو إنتاج تمثيل نصي قصير مع الاحتفاظ بأهم ما جاء في النص الأصلي من معلومات .

لتطوير أنظمة التلخيص الآلي للنصوص، يعتمد معظم الباحثين على الأنظمة المستندة إلى المعرفة اللغوية. و تستخدم هذه الأنظمة أساسا تقنيات الاستخراج حيث المبدأ هو إبراز المعلومات المهمة عن طريق اختيار الفقرات التي تميزها، وقد برزت مجموعة متنوعة من هذه الأساليب كما يرى ذلك الباحث هوراسيو ساكجيون من جامعة مونتريال بكندا.

يمكن جمع عدة طرق من أجل الحصول على ملخصات أفضل كما أنه تم احصاء عدد قليل جدا من الاعمال لانتاج ملخص جيد للنصوص العربية .

إن الغرض من هذه الدراسة يتضمن أمرين :

أولا : اختيار الطريقة المناسبة لتحليل اللغة العربية، ويجب أن يكون لديها القدرة على التكيف مع خصائص اللغة العربية.

ثانيا : تصميم نظام التلخيص الآلي للنصوص العربية بواسطة تكييف تقنيات الاستخراج المختلفة التي استخدمت لعلاج النصوص باللغة الإنجليزية.

وكإشكالية رئيسية يمكننا القول أن :

كل المعطيات السابقة وغيرها دفعتنا إلى التساؤل عن إمكانية **تصميم طرق معالجة لغوية مناسبة لتلخيص النصوص العربية بالخصوص** وذلك من خلال الإشكالية التالية :

كيف يمكن تصميم طرق معالجة لغوية مناسبة لتلخيص النصوص العربية؟

تساعدنا على الإجابة عن هذه الإشكالية مجموعة من **التساؤلات الفرعية**، يمكن صياغتها كالتالي :

- ✓ **كيف يستخدم علم الاعلام الالي في خدمة اللغة العربية؟**
- ✓ **ما مفهوم التلخيص وماهي آليات التلخيص في النصوص غير العربية؟**
- ✓ **ماهي طرق المعالجة الالية للغات ومالنحو المناسب لذلك؟**
- ✓ **ماهي مميزات النصوص العربية وخصوصيات تلخيصها؟**
- ✓ **ماهي طرق المعالجة اللغوية لتلخيص النصوص العربية وسبل تصميمها؟**

وضعت لهذه **الدراسة عدة فرضيات** يمكن صياغتها كالتالي :

1. حققت اللغة العربية قفزة نوعية في عدة مجالات بعد استخدام علم الاعلام الآلي
2. مميزات النصوص العربية وخصوصياتها لاتشكل عائقا في تطبيق المعالجة اللغوية بالحاسوب.
3. يمكن إيجاد طرق وتصميمات معالجة لغوية لتلخيص النصوص العربية

تحديد إطار الدراسة: نظرا لشساعة الموضوع وتشعب إشكالاته فقد حاول الباحث وضع إطار محدد للدراسة يتيح له الوصول إلى نتائج محددة للإشكاليات المطروحة سابقا، وذلك وفق التحديد التالي:

الإطار المكاني: تتمحور هذه الدراسة حول النصوص العربية بشكل عام (العالم العربي خصوصا والاسلامي عموما)، وسنبين في منهجية الدراسة كيف تستنبط طرق التلخيص الخاصة باللغات الأجنبية ومحاولة ملاءمتها مع اللغة العربية .

الإطار الزماني: تم **التركيز** على المرحلة الممتدة من سبعينيات القرن الماضي إلى وقتنا الحالي وهي فترة شهدت تطورا كبيرا في علم الإعلام الآلي وتطبيقاته وخاصة تلخيص النصوص، وبعض الأحيان، للضرورة العلمية والتاريخية نخرج عن هذا الإطار بعض الشيء

دراسات سابقة:

فيما يخص الدراسات السابقة فإن ماتم الحصول عليه لا يمثل حقيقة كل الدراسات التي تمت في هذا المجال، فالإمكانيات المحدودة تحول دون الاطلاع على كل ما كتب، ولعل أبرز هذه الدراسات :

باللغة الأجنبية:

(1) : Atefeh Farzindar, Développement d’un système de résumé automatique de textes juridiques, Université de Montréal 2001.

قام الكاتب في هذا البحث بتصميم نظام تلخيص آلي للنصوص القانونية

(2) : Horacio Saggion, Génération automatique de résumé par analyse sélective, thèse de doctorat, Université de Montréal 2000.

رسالة دكتوراه قدمها الباحث وتتمحور في التوليد الآلي للملخصات باستعمال التحليل الانتقائي

باللغة العربية :

قمت بجمع إسهامات الباحثين العرب في هذا المجال وأفردت له فصلا كاملا وهو الفصل الأول

دوافع اختيار الموضوع :

إن اختيار هذا الموضوع نابع من أسباب وعوامل، يمكن ذكر أهمها في ما يلي :

- ✓الرغبة في دراسة طرق التلخيص الآلي للنصوص العربية؛
- ✓محاولة البحث عن أحسن الطرق للاستفادة من علم الإعلام الآلي بما يخدم مصالح الشعوب العربية وغاياتها ؛
- ✓دراسة التصاميم وطرق المعالجة الآلية للنصوص العربية بما يوضح حدود هذه المناهج وأهم أدواتها والناجعة منها .

المنهج المتبع وأدوات الدراسة :

إن طبيعة هذه الدراسة تطلبت الاعتماد على المنهج الوصفي التحليلي، وذلك في سرد المفاهيم العامة والتعاريف المتعلقة بالموضوع، وتوصيف المعالجة الآلية للغات. وقد استعنا في هذا الصدد بالدراسات السابقة سواء باللغات الأجنبية أو باللغة العربية.

كما تطلب في جزء منها المنهج التاريخي والمنهج المقارن، وذلك في الفصل الأول، عند محاولة فهم التطور التاريخي الذي تزامن مع ثورة الإعلام الآلي، وكذا مقارنة بعض البحوث المنجزة في بعض البلدان العربية، وبيان ايجابيات وسلبيات كل منها وكذا أوجه الشبه والاختلاف فيما بينها.

محتويات الدراسة :

لدراسة هذا الموضوع ارتأى الباحث أن يسير على خطة تتوزع على خلال خمسة فصول زيادة على المقدمة العامة والخاتمة ومكتبة البحث حيث كان الهيكل على النحو التالي:

إستعرض الباحث في **الفصل الاول** إستخدام علم الإعلام الآلي في خدمة اللغة العربية وتحدث عن نشأة اتجاه استخدام جهاز الاعلام الآلي في دراسة علوم اللغة العربية، والظروف والملابسات التي أسهمت في تكوينه بوساطة الجهود الفردية، أو الجهود المؤسساتية والرسمية، والمشكلات التي واجهته في ضوء خصوصية اللغة العربية، والبرمجيات، والحاسوبات، وما قدم من حلول لمعالجة تلك المشكلات .

كما تناول هذا الفصل نتائج استثمار هذا المنحى في مجال تعليم العربية لأبنائها، وللناطقين بغيرها من اللغات، وفي مجال الترجمة الآلية، والتعريب، والإحصاء اللغوي، والمعالجة الآلية للأصوات، والصرف، والنحو، والمعجم والدلالة. ويخلُص إلى تحديد ملامح هذا الاتجاه الحديث في اللسانيات العربية المعاصرة، وأثره في تطوير اللغة العربية وتنميتها في العصر الحديث .

وقد خصص **الفصل الثاني** لدراسة كم هائل من القسم النظري للملخصات من تعاريف ونبذة تاريخية عن المنشأ وأنواع الملخصات وأهميتها كما تم التطرق إلى طرق التلخيص اليدوي وكذا الآلي وتم ختم الفصل بعرض ميادين التلخيص الآلي وإيجابيات وسلبيات طرقه.

وشمل **الفصل الثالث** المعالجة الآلية للغات وأهم أنواع النحو المستعمل، إضافة على التعاريف النظرية كما تعرض لأهم ميادين تطبيقات المعالجة الآلية للغات وقائمة عن بعض مشاريع التلخيص الآلي ثم تم سرد أهم أنواع النحو المستعمل في هذا المجال.

وجاء **الفصل الرابع** ليعطي تصور الباحث واقتراحه لنهج مدمج لتلخيص النصوص العربية وكانت النتائج تتمثل في تصميم نظام تلخيص آلي قوامه تطبيق عدة طرق للتلخيص، ويجمع نتائجها تلخيصا متميزا للنص الأصلي من حيث المضمون.

وتضمن **الفصل الخامس** تطبيقات نظرية لما يمكن أن يعطيه لنا النهج المدمج لتلخيص النصوص العربية باستعمال برامج إعلام آلي تعتمد على تصميمنا.

أما **الخاتمة** فقد احتوت على حصيلة إشكالية هذه الدراسة استجابة للاسئلــة الفرعيـــــة المكونة لها، مع اختبار صحة الفرضيات من عدمها والتي قد طرحت في المقدمة العامة، كما تضمنت عددا من النتائج المتوصل إليها وتوصيات تخص هذا الموضوع، بالإضافة إلى الآفاق المتاحة للبحث وعدد من الاشكاليات القريبة من موضوع البحث، تجلت خلال سير هذه الدراسة، والتي من الممكن أن تكون مشاريع بحثية قادمة.

وإن كان لابد من كلمة تقال في هذا المضمون، فإن البحث في مثل هذه القضايا المتعلقة بمستقبل اللغة العربية، أصبح ضرورة حتمية في المعالجة الآلية لكينونة هذه اللغة وأسرارها.

وفي الأخير، أغتنم هذه السانحة لأقول أن الفضل كل الفضل و جزيل الشكر والامتنان والعرفان بعد الله سبحانه كان للأستاذ الدكتور محمد عباس بجامعة ابي بكر بلقايد بتلمسان الذي شرفني بقبول إدارته البحث وتوجيهه لي لاستكمال هذا العمل والذي لم يكن ليرى النور لولا مساندته وتوجيهاته الحكيمة ونصائحه وعنايته، والذي لم يأل جهدا في سبيل إخراج هذه الرسالة واتمام فصولها، فقد وجدت فيه الأب الروحي في هذه الرعاية بالنصيحة العلمية التي ارتسمت في التوجيه والتصحيح والقراءة فجزاه الله خير الجزاء.

عيجولي حسين

تلمسان يوم الجمعة 18 جمادى الثانية 1438 هـ

الموافق لـ 17 مارس 2017 م

الفصل الأول

إستخدام علم الإعلام الآلي في خدمة اللغة العربية

يمكن القول إن وعي الدول العربية كان غائبا بأهمية الطفرة الإعلامية التي تحققت دوليا في مجـــــال التخطيط لعصر المعلومات، أي تصاميم الحواسيب، وأنظمتها, ووضع برامجها، ونظم تشغيلها، وتطويع تقنياتها [1] حيث تؤدي اللغة العربية فيها حجر الزاوية.

يقول د. نبيل علي : " إن نقص الدراسات اللسانية الحاسوبية سيكون تحديا قاسيا لدى التصدي لمشكلة قصور الدراسات اللغوية وضمورها في العربية، إذ أثبتت المنجزات التي تمت على صعيد اللغات الأخرى ما لتزاوجها مع الحاسوب من قدرة فريدة على إكساب هذه اللغة مزيد ارتقاء وكفاءة وحيوية ومرونة، وخصوبة، ومنطقية وصمود للزمن، قياسا على ذلك لنا أن نتصور ما يمكن أن يؤديه الحاسوب في تعويض تخلفنا اللغوي: تنظيرا وتقعيدا، واستخداما " [2] .

---

[1] ينظر : عبد الغني ، أبو العزم

http://www.aljabriabed.net/n31_04abualazm.%282%29.htm

شوهد يوم 29 افريل 2016 على 11سا40د

[2] د. نبيل ، علي ، "اللغة العربية والحاسوب"، دار تعريب، 1998 ص. 9

لم يقتصر الأمر في الدعوة إلى العناية بحوسبة اللغة العربية وعلومها على المبادرات الفردية للباحثين وبعض المهندسين العرب الذين اشتهروا بحبهم للغة العربية وغيرتهم عليها ، إذ سرعان ما بدأت المجامع العلمية المهتمة باللغة العربية وعلومها تعقد الندوات العلمية لمقاربة موضوع اللغة العربية والمعلوميات، نذكر منها على سبيل المثال ندوة الجزائر الدولية في ديسمبر 2002 حول "اللغة العربية وتكنولوجيا المعلومات" [1] .

وقد راهنت هذه الندوة على إبراز دور اللغة العربية وقدرتها على مسايرة التطورات التكنولوجية الحديثة.

يتناول هذا الفصل جهود الباحثين المعاصرين العرب – بصفة عامة، واللغويين بوجه خاص – في تطويع تقنيات الحاسوب لخدمة الدراسات اللغوية العربية، أصواتاً، وصرفاً، ونحواً، ومعجماً، ودلالة، ومدى إفادتها منه في معالجة قضاياها المختلفة .

يقول د. وليد أحمد العناتي " لعل اللسانيات الحاسوبية تكون أحدث فروع اللسانيات، فهي فرع بينيّ ينتسب نصفه إلى اللسانيات وموضوعها اللغة، ونصفه الآخر حاسوبي وموضوعه ترجمة اللغة إلى رموز رياضية يفهما الحاسوب، أو تهيئة اللغة الطبيعية لتكون لغة تخاطب

---

[1] ينظر : اليوسفي، أحمد شعيب ، كلية الآداب، تطوان، نحو نظام جديد لإدارة المعلومات 2015 ، ص 89

وتحاور مع الحاسوب، بما يفضي إلى أن يؤدي الحاسوب كثيراً من الأنشطة اللغوية التي يؤديها الإنسان، مع إقامة الفرق في الوقت والكلفة " [1]

وكما هو معلوم أن "اللسانيات الحاسوبية علم بيني يجمع بين علم اللغة وبين علوم الحاسوب الإلكتروني لاستغلال اللغات الطبيعية في البرامج الحاسوبية عبر طرق ذكية آلية عديدة منها الذكاء الاصطناعي، والدلالة الاصطناعية، وهو ما يعرف جملة بالنظم الخبيرة." [2]

وكما هو معلوم فإن العلاقة بين جهاز الاعلام الآلي واللغة العربية تقوم على محورين أساسيين : أولهما المحور النظري، والآخر التطبيقي .

---

[1] العناتي، وليد أحمد، اللسانيات الحاسوبية العربية ، ص34، المعرفة – رؤية ثقافية 2015

[2] أوراق ملتقى فعالية اللسانيات الحاسوبية العربية، ص 13، جامعة ورقلة ، 2014

في ضوء هذه المقدمة نستعرض نشأة اتجاه استخدام جهاز الاعلام الآلي في دراسة علوم اللغة العربية، والظروف والملابسات التي أسهمت في تكوينه بوساطة الجهود الفردية، أو الجهود المؤسساتية والرسمية، والمشكلات التي واجهته في ضوء خصوصية اللغة العربية، والبرمجيات، والحاسوبات، وما قدم من حلول لمعالجة تلك المشكلات .

كما يتناول هذا الفصل نتائج استثمار هذا المنحى في مجال تعليم العربية لأبنائها، وللناطقين بغيرها من اللغات، وفي مجال الترجمة الآلية، والتعريب، والإحصاء اللغوي، والمعالجة الآلية للأصوات، والصرف، والنحو، والمعجم والدلالة. ويخلُص إلى تحديد ملامح هذا الاتجاه الحديث في اللسانيات العربية المعاصرة، وأثره في تطوير اللغة العربية وتنميتها في العصر الحديث .

1ـ نبذة تاريخية :

على المستوى الغربي :

تمَّ اختراع جهاز الاعلام الآلي – كما تذكر المصادر – في أَواخر النصف الأول من القرن المنصرم (القرن العشرين)، وتحديداً عام 1948م، وأصبح منذ ذلك التاريخ متاحاً للإفادة منه في جميع مجالات الحياة، ومختلف العلوم والمعارف الإنسانية [1].

---

[1] ينظر: العارف، عبدالرحمن بن حسن، جامعة أم القرى، توظيف اللسانيات الحاسوبية في خدمة الدراسات اللغوية العربية " جهودٌ ونتائج "، مجمع اللغة العربية الاردني 2016، ص 40

وتطورت تقنية هذا الجهاز عبر السنوات تطوراً مذهلاً، منذ ظهور الجيل الأول من الحواسيب الآلية سنة 1951م، وحتى ظهور الجيل الخامس منه سنة 1991م .

أما بدء استخدام علم وجهاز الاعلام الآلي في دراسة اللغة على مستوى العالم، فمن الصعوبة بمكان وضع تأريخ زمني محدد له ؛ وذلك لأنه لم يحدث دفعة واحدة، بل تمَّ نتيجةً لمحاولات متفرقة، وعلى مراحل زمنية مختلفة، وفي دول متعددة.

يذكر أستاذ علم الدلالة ومنظم البرمجة اللسانية الآلية بجامعة جورج تاون بالولايات المتحدة الامريكية الدكتور مايكل زار تشناك (M.Zarechnak) أن العمل في اللسانيات الآلية بدأ في قسم اللسانيات بجامعة جورج تاون سنة 1954م، وذلك في حقل الترجمة الآلية من اللغات الأخرى إلى الإنجليزية. وهذا يعني أن بداية الخمسينيات من القرن الماضي شهدت ولادة المعالجة الآلية للغات البشرية [1].

أما على المستوى الأوروبي فتذكر المصادر أن أقدم محاولة لدراسة اللغة بوساطة الحاسوب تمَّت سنة 1961م، بجامعة قوتبرغ (Goteborg) السويدية، لكن هذه المحاولة ظلت ذات طابع محلي، ولم ترق إلى مستوى الذيوع والانتشار والتأثير في محيطها الأوروبي .

---

[1] ينظر : المرجع نفسه : ص.43

والبداية الفعلية لهذا الاتجاه كانت – كما تقر المصادر – لمركز التحليل الآلي للغة بمدينة (قالارات Gallarat) بإيطاليا، الذي كان يشرف عليه الاستاذ روبارتوبوزا (Roberto Busa)، حيث وضع سنة 1962م الدعائم الأولى لاستخدام الحاسوب في دراسة اللغة [1] .

ثم توالى بعد ذلك افتتاح المراكز الحاسوبية للغة في أوروبا والاتحاد السوفياتي، كما هي الحال في المركز الحسابي لدراسة الأدب واللغة في جامعة كامبردج سنة 1964م، والمركز المعجمي بمجمع دالاكروسكا (Dellacrusca) بإيطاليا سنة 1964م، ومعهد الألسنية التابع لمجمع العلوم بكيف في أوكرانيا (الاتحاد السوفييتي سابقاً) سنة 1964م – أيضاً– .

[1] ينظر : المرجع نفسه : ص.49

على المستوى العربي :

أما بالنسبة للعلوم النظرية عند العرب في العصر الحاضر فقد كانت العلوم الشرعية من أسبق العلوم الإنسانية استخداماً لتقنية الحاسبات الإلكترونية ونظم المعلومات، حيث بُدئ بالعمل بها والإفادة منها في السبعينيات من القرن الماضي. وظلت علوم اللغة العربية في منأى عن الانتفاع بها بعض الوقت، حتى قيض الله لها من رأى أنه يمكن لهذه العلوم أن تفيد من الحاسوب فائدة كبرى .

2ـ ميادين تطبيقات الاعلام الآلي :

الميدان الأول : الإحصاء اللغوي

يقول الدكتور إبراهيم أنيس بأنه انتهز فرصة زيارته لجامعة الكويت سنة 1971م للعمل بها أُستاذاً زائراً، وهناك التقى بالدكتور علي حلمي موسى، أستاذ الفيزياء النظرية في جامعة الكويت، وطرح عليه فكرة الاستعانة بالحاسوب في إحصاءات الحروف الأصلية لمواد اللغة العربية، بُغْية الوقوف على نسج الكلمة العربية. وقد رحب بهذه الفكرة واستحسنها، وبدأ بالتخطيط لها وتنفيذها في النصف الأول من عام 1971م، وكان من ثمرة ذلك صدور للجوهري (324هـ) [1] الدراسة الإحصائية للجذور الثلاثية وغير الثلاثية لمعجم الصحاح .

[1] ينظر: موسى ، علي حلمي ، دراسة إحصائية لجذور معجم الصحاح باستخدام الكمبيوتر، ص9، جامعة الكويت 1973

أما خطوات العمل في هذا الإحصاء فتوزعت على ثلاث مراحل : الأولى إدخال المواد اللغوية في ذاكرة الكمبيوتر، والثانية وضع برنامج له بإحدى لغات الكمبيوتر، والثالثة التنفيذ الفعلي لهذا البرنامج .

وجاءت نتائج هذه الدراسة في صورة جداول إحصائية لجذور اللغة، وحروفها، وتتابع أصواتها، وخصائص حروفها، مقرونةً بدراسة تحليلية موجزة عن التفسير اللغوي لما ورد بتلك الجداول .

واستقبل الباحثون والعلماء هذا العمل العلمي بقبول حسن، رغم وجود فئة حاولت أن تُشَكِّك وتُهَوِّن من جدوى هذه الدراسة، وفائدتها على الدرس اللغوي .

ومما لاشك فيه أن اللغة العربية بعلومها المختلفة، كالأصوات، والبلاغة، والعروض والقافية، أفادت أيما فائدة من نتائج هذه الإحصائيات الدقيقة.

وتبع ذلك صدور دراسة ثانية [1] لإحصاء جذور معجم لسان العرب لابن منظور (711هـ)، وكان هذا عام 1972م، ودراسة ثالثة لإحصاء جذور معجم تاج العروس للزُّبيدي

[1] ينظر: العارف، عبدالرحمن بن حسن، توظيف اللسانيات الحاسوبية في خدمة الدراسات اللغوية العربية، مرجع سابق، ص 53.

(1205هـ)، واشترك في هذا العمل الأخير الدكتور عبد الصبور شاهين، وكان هذا عام 1973م .

وقد صدرت هذه الأعمال جميعها عن جامعة الكويت، وكانت بحق ابتكاراً جديداً لم يسبق إليه من قبل، بل هي المرة الأولى في العالم العربي التي تجرى فيها هذه الإحصائيات على أسس علمية حديثة ودقيقة .

كما تَمَّ – ربما لأول مرة أيضاً – تعاون الفيزيائيين واللغويين حول إحصاء كلمات اللغة العربية الواردة في أشهر المعاجم اللغوية، وتحليل ما نتج عن ذلك من جداول تحليلاً لغوياً قوامه استخراج مادة اللغة (جذورها)، سواء كانت ثلاثية أو رباعية أو خماسية، وتردد الحروف، وتتابعها، ومقارنة نتائج هذه المعاجم الثلاثة بعضها ببعض .

ويذكر الدكتور علي حلمي موسى [1] أنه بدأ عام 1974م بالبحث في ألفاظ القرآن الكريم بقصد حصرها، ومن ثمَّ تحليلها ومقارنتها بألفاظ معجم الصحاح، كما أنه أخذ بالبحث عن دراسة العلاقة بين الحروف والحركات في القرآن الكريم، ومقارنة السُّور المكية بالسُّور المدنية، مستعيناً في ذلك بالآلات الحاسبة الإلكترونية، ومشيراً في هذا الصدد إلى أنه قَدَّم أجزاء من هذه البحوث في مؤتمرات علمية عالمية .

---

[1] ينظر: موسى، علي حلمي، دراسة إحصائية لجذور معجم الصحاح باستخدام الكمبيوتر، مرجع سابق، ص17

ولعل الباحث يرى أن هذا التوجُّه في الفكر العربي المعاصر قد فتح الباب واسعاً للباحثين في الدراسات اللغوية والأدبية للولوج من خلاله إلى عالم الكمبيوتر، وتسخيره لخدمة البحث اللغوي والأدبي.

وأقرب مثال لهذا ما قامت به الباحثة –آنذاك– وفاء محمد كامل في رسالتها للماجستير [1] عن كعب بن زهير بن أبي سلمى – دراسة لغوية، من الاستعانة بالحاسوب في دراسة شعر هذا الشاعر، وذلك للمرة الأولى – كما يذكر الدكتور حسين نصار – في الدراسات اللغوية في مصر .

وهكذا كان حقل الإحصاء اللغوي هو الميدان الأول لتطبيق اللسانيات الحاسوبية على اللغة العربية .

إن استخدام الإحصاء الرياضي في اللغة يحقق تقييماً كمياً لبعض الخواص النوعية للغة، كمعدلات استخدام الحروف، والكلمات، والصيغ الصرفية، والموازين الشعرية، وأنواع الأساليب النحوية، أو التوزيع النسبي للأفعال المعتلة والصحيحة، أو للإفراد والتثنية والجمع، أو لحالات الإعراب المختلفة. كما يحقق توصيفاً كمياً لبعض العلاقات اللغوية، كالعلاقة

---

[1] ينظر: العارف، عبدالرحمن بن حسن، توظيف اللسانيات الحاسوبية في خدمة الدراسات اللغوية العربية، مرجع سابق، ص 57.

بين طول جذر الكلمة وعدد مرات تكراره، والعلاقة بين طول الكلمة ومعدل استخدامها داخل النصوص .

ويقوم الإحصاء بتفسير بعض الظواهر اللغوية وتحليلها. ليس هذا فحسب، بل هناك إحصاء جديد يستطيع أن يتعامل مع البنية المعقدة للسياق اللغوي، حتى يكشف لنا عن علاقات الترابط والتماسك بين فقراته وجمله وألفاظه، وتلك التي تربط بين ظاهر العبارات وما تبطنه من معان وإشارات .

ومن المشاريع العلمية القيمة في هذا المجال – بالإضافة إلى ما ذكر سابقاً – ما قام به كل من الدكتور يحيى مير علم، والدكتور محمد حسان الطيان والأستاذ مروان البواب (سورية)، تحت إشراف الدكتور محمد مراياتي [1]، من دراسات إحصائية لدوران الحروف في الجذور العربية، وللمعجم العربي، ولدوران الحروف العربية المشكولة، ولحروف اللغة العربية. وهناك دراسات إحصائية أخرى صدرت باللغة الإنجليزية في الجامعات الأمريكية والأوروبية لجوانب لغوية متعددة، كالأصوات، والصرف، والنحو للغة العربية .

[1] ينظر : المرجع نفسه، ص 59.

الميدان الثاني : علم اللغة الحسابي

لقد كانت هذه الإرهاصات بداية لظهور فرع جديد من فروع علم اللغة، يطلق عليه (علم اللغة الحسابي) أو (اللسانيات الحاسوبية) Computational Lingustics أو (اللسانيات الإعلامية) .

وإذا أردنا تعريف هذا العلم بشكل مختصر قلنا إنه العلم الذي يبحث في اللغة البشرية كأداة طيّعة لمعالجتها في الآلة (الحاسبات الإلكترونية = الكمبيوتر)، وتتألف مبادئ هذا العلم من اللسانيات العامة بجميع مستوياتها التحليلية : الصوتية، والنحوية، والدلالية، ومن علم الحاسبات الإلكترونية (الكمبيوتر)، ومن علم الذكاء الاصطناعي، وعلم المنطق، ثم علم الرياضيات [1].

وكانت البداية الحقيقية لهذا العلم لدى الغرب قد جاءت بعد بزوغ فجر النظرية التوليدية التحويلية، حيث قامت بتطبيق الأسس والمعادلات الرياضية على التحليل اللغوي، ومن ثَمَّ صياغة اللغة صياغة رياضية من أجل برمجتها في الحاسوب، وذلك بغرض استنباط قواعد مقننة ودقيقة . وإن كان هذا لا يمنع من القول إن المدرسة البنوية قد مهدت الطريق أمام العلماء لربط الدراسات اللغوية بالحاسوب، لكنها لم تستطع بعد ذلك تطوير أفكارها لتساير ذلك المدّ التكنولوجي المتنامي .

---

[1] ينظر : المرجع نفسه، ص 61.

وتقوم اللسانيات الحاسوبية على جانبين رئيسين هما : الجانب النظري، والجانب التطبيقي. فأما الجانب الأول (النظري) فيبحث «في الإطار النظري العميق الذي به يمكننا أن نفترض كيف يعمل الدماغ الإلكتروني لحل المشكلات اللغوية»، وأما الجانب الآخر (التطبيقي) فهو يُعنَى « بالناتج العملي لنمذجة الاستعمال الإنساني للغة، وإنتاج برامج ذات معرفة باللغة الإنسانية » .

والواقع أن جهود العلماء العرب المعاصرين والمؤسسات العلمية في هذه المجال يمكن نظم عقدها في أربعة صور: الأولى تتمثل في مؤلفات خُصِّصت للعربية والحاسوب، أو الحاسوب والعربية، وجاءت الثانية على هيئة مقالات وبحوث نشرت في المجلات والدوريات العلمية، أو ضمن أعمال المؤتمرات، ووقائع الندوات والملتقيات العلمية، أما الثالثة فكانت خاصة بالبرامج والنظم التي وضعت لحوسبة العربية، أو لعوربة الحاسوب، سواء ما كان منها فردياً محضاً، أو نتاجاً مشتركاً، أو عملاً تجارياً عاماً .وأما الصورة الرابعة فتمثلت في إنشاء بعض الكليات الجامعية قسماً خاصاً لعلم اللغة الحاسوبي [1]، كما هي الحال في جامعة الأمير سلطان الأهلية بالرياض (المملكة العربية السعودية). وسوف نعرض بالتفصيل لكل ذلك ما أمكننا، في إطار الهدف الموضوع والخطة المرسومة لهذه الدراسة .

[1] ينظر : المرجع نفسه، ص 62.

وإذا أردنا عرض مراحل التطور لعلم اللغة الحاسوبي في الدراسات العربية المعاصرة أمكن القول بأن كتاب الدكتور نبيل علي (اللغة العربية والحاسوب) [1] يُعَدُّ أول مؤلَّف يتناول موضوع اللسانيات الحاسوبية مطبقةً على أنظمة اللغة العربية، صوتاً، وصرفاً، ونحواً، ومعجماً، مع المعالجة الآلية لهذه النظم اللغوية جميعها.

وكان تأريخ صدوره لأول مرة سنة 1988م. وقد حالف التوفيق المؤلف في كثير من القضايا المتصلة بالحاسوب واللغة، وذلك حينما انطلق في عمله هذا من وضع دراسات تقابلية بين العربية والإنجليزية شاملة لكل النظم اللغوية، بالنظر إلى أن الإنجليزية هي اللغة الأم لتقنيات نظم الحاسوب والمعلومات، وهذا ما نتج عنه معرفة أوجه الاختلاف والاتفاق بين اللغتين، وكان هذا النهج بمثابة الأرض الصُّلْبة والقاعدة المتينة التي هيأت للمؤلِّف منهجية وموضوعية، ومكَّنته من الإسهام الإيجابي في جهود تعريب الحاسوب من جهة، والمعالجة الآلية للغة العربية من جهة أخرى .

إن هذا الكتاب يمثل حجر الزاوية في مسيرة البحث اللغوي العربي في اللسانيات الحاسوبية، بل إنه كما وصفه الدكتور نهاد الموسى «خطوة واسعة واثقة، تنتظم مشروعاً مستوعباً لتأسيس اللسانيات الحاسوبية في العربية، على أساس نظري وتطبيقي في آن واحد معاً» [2].

---

[1] ينظر: نبيل، علي، اللغة العربية والحاسوب، دار تعريب، 1998، ص 11.

[2] ينظر : الموسى، نهاد؛ العربية – نحو توصيف جديد في ضوء اللسانيات الحاسوبية، دار الفكر، 2000م،ص 25

صحيح أنه لم يستوعب جميع قضايا اللغة باستعمال الحاسوب، إلا أن هذا أمر متوقع فيمن يفتتح التصنيف، أو يرد الطريق لأول مرة في أي فن غالبا .

وبعد نشر هذا الكتاب بسنوات ثمان – أي سنة 1996م – صدر كتاب الدكتور عبد ذياب العجيلي (الحاسوب واللغة العربية) ، وهو – كما يقول الدكتور نهاد الموسى – [1]: «خطوة جزئية إيجابية نحو معالجة مسائل متنوعة من العربية بلغة برولوج Prolog .وهو يمثل جهداً حميداً في هذا الاتجاه البيني (اللسانيات العربية الحاسوبية)» .

[1] ينظر: الموسى، نهاد ؛ العربية – نحو توصيف جديد في ضوء اللسانيات الحاسوبية، مرجع سابق، ص 27

ومن هذه المؤلفات في اللسانيات الحاسوبية كتاب الدكتور نهاد الموسى (العربية – نحو توصيف جديد في ضوء اللسانيات الحاسوبية)، الذي صدر سنة 2000م .

والكتاب – كما يذكر مؤلفه – «محاولة في الانتقال من وصف العربية إلى توصيفها، وذلك في ضوء الأطروحة العامة للسانيات الحاسوبية» [1] .

وقد اشتمل الكتاب على رؤى حاسوبية حاول المؤلف إسقاطها على أنظمة العربية، وخاصة النحو (الإعراب)، والصرف (البنية)، والمعجم، إضافة إلى التصويب اللغوي (الأخطاء النحوية، والصرفية، والإملائية) .

إن هذه الجهود التي تمت ضمن هذا الإطار كانت – كما يلاحظ – فردية الطابع، لكن ذلك لم يدم طويلاً، إذ سرعان ما أصبحت متعددة الأطراف، بعد أن احتضنتها المراكز والمعاهد التقنية، والجمعيات الحاسوبية في الوطن العربي وخارجه، والمؤسسات والشركات التجارية المحلية والعالمية، وذلك عقب حدوث ثورة المعلوماتية Infomatization))، والتفجر المعرفي في عالم اليوم، وشعور الجميع أفراداً وجماعات بأنهم أمام تحدٍّ حضاري كبير، وإيمانهم بضرورة نقل هذا الصراع العلمي الثقافي – إن صح التعبير – إلى حوار منهجي

[1] المرجع نفسه، ص 36.

وتكامل معرفي، يؤدي في نهاية المطاف إلى ردم الهوَّة، أو تقليص مسافة الفجوة – على أقل تقدير – بين الغرب والشرق العربي، وذلك ما سينتج عنه تصحيح لتلك المفاهيم الخاطئة، والتصورات المغرقة في التشاؤم، عن العلاقة بين اللغة العربية والحاسوب، ومن ثم بلورة صياغة لغوية تقنية لاستخدام الحاسوب وتوظيفه في خدمة علوم العربية .

أما البحوث والمقالات الخاصة باللسانيات الحاسوبية، فمنها ما نشر في مجلات علمية، ومنها ما أُلقي أو قُدِّم في الندوات والمؤتمرات التي خُصِّصت أصلاً للغويات الحاسوبية، أو اللسانيات التطبيقية، أو لتكنولوجيا الحاسوب ومجالات استخدامه في العلوم الإنسانية، ثم نشرت هذه البحوث ضمن أعمال تلك المؤتمرات والندوات .

والجدير ذكره في هذا المقام أن هذه البحوث والمقالات قد أضحت تمثل تياراً واضحاً في الجهود اللسانية الحاسوبية، وهذا ما جعل كلاً من الدكتور وليد العناتي وزميله الدكتور خالد الجبر، من جامعة البترا الأهلية (الأردن)، يقومان بوضع دليل بيبليوغرافي لها أسمياه (دليل الباحث إلى اللسانيات الحاسوبية العربية) [1]، حاولا فيه أن يستقصيا جميع ما وقفا عليه من

[1] ينظر : العناتي، وليد، و الجبر، خالد، دليل الباحث إلى اللسانيات الحاسوبية العربية، دار الشروق، 2015 ص 144

أعمال علمية تنتظم في هذا الميدان. وبلا شك فإن هذا الكتاب سَيَسُدُّ ثغرة واضحة في المكتبة اللغوية عامة، واللسانيات الحاسوبية خاصة.

وقبل أن أبدأ الحديث عن الصورة الثانية من صور جهود العرب المعاصرين في ميدان اللسانيات الحاسوبية، يجدر بي أن أثبت حقيقة تأريخية، وهي أن بحوث الدكتور إبراهيم أنيس التي كتبها بآخر عمره، تُعَدُّ من أوائل الأعمال التي وجهت الأنظار إلى الاستعانة بتقنية الحاسوب، وتوظيفها لخدمة البحث اللغوي. ليس هذا فحسب، بل إنه دلف إلى هذا الميدان واستثمر نتائج تلك الجداول الإحصائية اللغوية التي كان يخرجها له الكمبيوتر الموجود بمعهد الإحصاء –آنذاك – بجامعة القاهرة لصالح تفسير إحدى الظواهر اللغوية، وهي ظاهرة القلب المكاني [1]، وهذه سابقة علمية في مجال اللغة تُحسب للدكتور أنيس، وتُذكر له في مضمار الحاسوب واللغة، أو اللسانيات الحاسوبية العربية .

الميدان الثالث : التعريب

ففي مجال التعريب، ونعني به هنا تعريب الحاسوب من حيث أنظمته وبرامجه ومصطلحاته، فقد اتجهت جهود التعريب فيه إلى إعداد أنظمة وتصميمها لكي تكون قادرة على العمل باللغة العربية بدلاً من اللغة الإنجليزية، إضافة إلى إصدار المؤلفات الخاصة بعلوم الحاسبات وتقنيتها باللغة العربية، وترجمة ما كان مؤلفاً بغير العربية .

[1] ينظر : إبراهيم، أنيس، القلب المكاني بتقنية الحاسوب، دار الفكر، 1998، ص 113

ولعل من أهم الإنجازات في هذا المجال ما قامت به الشركات العربية والأجنبية العاملة في مجال الحاسوبات، كالشركة العالمية للبرامج (صخر)، وشركة (آي. بي. إم)، والجريسي للتقنية، من تطوير الحواسيب الشخصية (PC) باللغة العربية، ووضع معالج النصوص (عربستار 2001) بالعربية أيضاً، وتعريب نظام قواعد المعلومات الخاص بتخزين المعلومات واسترجاعها، وتعريب البرامج اللاتينية...إلخ، علاوة على الهيئات العلمية العربية، كالمنظمة العربية للتربية والثقافة والعلوم (الأليكسو)، ومعهد الكويت للأبحاث العلمية، ومعهد الدراسات الإحصائية بجامعة القاهرة، ومعهد العلوم اللسانية والصوتية بالجزائر، ومعهد الدراسات والأبحاث للتعريب بالمغرب [1].

أما مصطلحات الحاسوب – وهي مسألة لا تقل أهمية عن سابقتها – فقد أسهم فيها الأفراد، والمؤسسات، والشركات .وقد طُرحت في هذا المقام اقتراحات عدة من قبل خبراء الحاسوبيات، وكذلك اللغويين، وقام عدد من المتخصصين في المدرسة الوطنية للمهندسين بجامعة تونس بتعريب المصطلحات الخاصة بالحاسوبات الصغروية، كما قامت بعض المؤسسات العلمية، كمدينة الملك عبد العزيز للعلوم والتقنية (الرياض)، ومعهد الدراسات والأبحاث للتعريب (الرباط)، ومجمع اللغة العربية الأردني، والمعهد القومي للمواصفات

---

[1] ينظر: العارف، عبدالرحمن بن حسن، توظيف اللسانيات الحاسوبية في خدمة الدراسات اللغوية العربية، مرجع سابق، ص 71.

والملكية الصناعية بتونس، بإنشاء بنوك للمصطلحات، تهدف إلى توفير المصطلحات المعرَّبة وتوثيقها، وتنميطها وتقييسها وتوحيدها .

والواقع أن موضوع التعريب والمصطلح كان وما زال من أهم القضايا التي تشغل الأمة العربية وحضارتها المعاصرة، ورغم ما بذل من جهود في هذا الإطار فإن النتائج لم تكن على مستوى التقدم التقني الهائل في عصر المعلوماتية والعولمة !!.

الميدان الرابع : العلاج الآلي للغة العربية

أما في مجال المعالجة الآلية للغة العربية، فقد شملت الجهود مستويات اللغة كافة، كالمستوى الصوتي، والصرفي، والنحوي، والمعجمي، والدلالي، يضاف إليها الترجمة الآلية، والكتابة العربية .

فالمستوى الصوتي تمت معالجته آلياً بوساطة تحليل طيف الصوت [1]، وتوليد (إنتاج) الكلام، وتخزين الأنماط الصوتية للشخص المتكلم. وتبعاً لهذا تمَّ تصميم أجهزة تخليق الكلام وتحليله، وتوليد الكلام المنطوق آلياً بتحويل النصوص المدخلة في جهاز الحاسوب إلى مقابلها الصوتي، وعلاج عيوب النطق.

[1] ينظر : المرجع نفسه، ص 72.

وقد أنجزت دراساتٌ عدة في هذا المجال، من بينها دراسة الدكتور منصور الغامدي (السعودية) عن الإدراك الآلي للتضعيف [1]، وهي محاولة لكيفية حل مشكلة التفريق بين الأصوات اللغوية الطويلة والقصيرة في اللغة العربية، قد تعين مبرمجي الحاسوب على الإدراك الآلي للأصوات اللغوية. كما تأتي دراسة الدكتور محمد مرياتي (معالجة الكلام – تطبيق على اللغة العربية) ضمن هذا التوجُّه في تمثيل النظام الصوتي للغة العربية آلياً. ويلحق بهذه الدراسات أيضاً ما كتبه الدكتور سالم غزالي عن (المعالجة الآلية للكلام المنطوق، التعرف والتأليف) .

ويوجد ببعض الجامعات العربية، والمعاهد العلمية، والمؤسسات التقنية أقسام خاصة للصوتيات، أو مراكز للسمع والنطق، أو معالجة الكلام، تجري فيها أبحاث صوتية تعتمد في المقام الأول على أجهزة الحاسوب، ويتولى الإشراف عليها أساتذة متخصصون في علم الأصوات، كالدكتور محمد صالح الضالع (جامعة الإسكندرية)، والدكتور سمير استيتية (مدير مركز السمع والنطق بجامعة اليرموك)، والدكتور سالم غزالي (مدير مخبر معالجة الكلام العربي بالمعهد الإقليمي لعلوم الإعلامية والاتصال عن بُعْد I.R.S.I.T. بتونس) والدكتور منصور الغامدي بمركز علوم وتقنية الأصوات بمدينة الملك عبد العزيز للعلوم

[1] نشرت هذه الدراسة ضمن (السجل العلمي لندوة استخدام اللغة العربية في تقنية المعلومات) مطبوعات مكتبة الملك عبد العزيز العامة، الرياض، ص83–90.

والتقنية، والدكتور محمد صالح بن عمر[1] (معهد بورقيبة للغات الحية بتونس)، والدكتورة تغريد السيد عنبر(كلية الألسن بجامعة عين شمس)، والدكتور سلمان العاني (جامعة انديانا). وهناك إنجازات أخرى صدرت باللغة الإنجليزية عن معالجة الكلام العربي آلياً، ومع كل ما ذكر من جهود فما زال العمل في هذا المجال ينتظر بذل المزيد من الجهود، لا على مستوى الأفراد بل على مستوى الفريق البحثي المتكامل «نظراً للطبيعة الخاصة لمعالجة الكلام الآلي» .

أما المستوى الصرفي، فقد تمت المعالجة الآلية له في ضوء أهمية الصرف العربي بالنسبة لنظام اللغة ككل. وقد تناولت هذه المعالجة الآلية بعض جوانب الصرف العربي، كالخاصية الثلاثية للجذور العربية، وأصل الاشتقاق، والأنماط الصرفية، وثنائية الصيغة الصرفية والميزان الصرفي، والإنتاجية الصرفية، والفائض الصرفي، واللبس الصرفي .

وفي هذا الصدد قَدَّم الدكتور نبيل علي إطاراً عاماً لمعالجة الصرف العربي آلياً، وأورد عدة نماذج للتحليل الصرفي الآلي بوجه عام، مبيناً مدى ملائمتها لمطالب الصرف العربي. وتبعاً لهذا قام بعرض نموذج وضعه لمعالجة الكلمات العربية صرفياً في أطوار التشكيل المختلفة،

---

[1] ينظر : بن عمر، محمد صالح، الثورة التكنولوجية واللغة، ط1، دار الشؤون الثقافية العامة – وزارة الثقافة والإعلام، بغداد – العراق، 1986م، ص32–35.

وهو نموذج التحليل بالتركيب، وأطلق عليه اسم (المعالج الصرفي المتعدد الأطوار). ويشتمل هذا النموذج المبتكر على عناصر أربعة هي : المعالج الصرف – نحوي، والمعالج الاشتقاقي، والمعالج الإعرابي، ومعالج التشكيل .

ويذكر صاحب هذا النموذج أنه قام بمعاونة إحدى المتخصصات في اللسانيات الحاسوبية (أمل الشامي) [1] بتطوير هذا المعالج الصرفي على ضوء أصول الصرف العربي وخصائصه المميزة، وعمد بعد ذلك إلى إخضاعه لاختبارٍ قاسٍ في مجالين اثنين هما : تحليل النص القرآني كاملاً، مع إعادة توليده آلياً، ومفردات المعجم الوسيط، وبعد اجتيازه هذا الاختبار بنجاح – كما يقول – تم استخدامه في عدة تطبيقات أساسية، شملت ضغط النصوص، واسترجاعها، واكتشاف الأخطاء الإملائية، وتحليل النصوص صرفياً، وميكنة المعجم العربي، مع دمج هذه التطبيقات في قاعدة للنصوص العربية الكاملة .

وقد أسهمت إدارة البحوث والتطوير بشركة العالمية للبرامج في هذا النموذج التطويري للمعالج الصرفي، وتُعَدُّ موسوعة الحديث الشريف – وهي من إنتاج هذه الشركة – من أهم أنظمة استرجاع النصوص التي استخدمت تقنية التحليل الصرفي .

---

[1] وقد ذكر الدكتور نبيل علي أن له وزميلته أمل الشامي بحثاً عن التحليل الصرفي للقرآن باستخدام الحاسوب، قيد النشر.

ينظر : نبيل، علي، اللغة العربية والحاسوب، مرجع سابق، ص332، 571 .ولا أدري هل صدر هذا الكتاب أم لا؟

وليس هذا هو الأُنموذج الأوحد أو المحاولة الأولى لتطوير نظام آلي للتحليل والتركيب الصرفيين، بل هناك نماذج ومحاولات أخرى، قدمها على حدة كل من الدكتور يحيى هلال، والدكتورة نادية حجازي بالاشتراك مع عبد الفتاح الشرقاوي، والدكتور مأمون الخطاب بالاشتراك مع الدكتور حسان عبد المنان .

وبطبيعة الحال كانت هناك إشكالات كثيرة واجهت معالجة الصرف العربي آلياً، جُلُّها يندرج تحت طبيعة الصرف العربي، وفي مقابل ذلك أمكن التغلب على تلك المصاعب، وإجراء عمليتي التوليد والتحليل الصرفيين الآليين ضمن الإطار العام للمعالجة الآلية للصرف العربي وقد أنجز العديد من الدراسات المعالجة للصرف العربي آلياً، كما هو الشأن في النماذج السابقة، وقريباً منه ما قدمه الدكتور نهاد الموسى في كتابه (العربية...)[1] ، وكذلك دراسة الأستاذ مروان البواب وزملائه عن نظام اشتقاق الكلمة العربية بالحاسوب .

أما المستوى النحوي فتمت معالجته آلياً بوساطة تشخيص أزمة النحو العربي أولاً، ثم إدراك خصائص هذا النحو وتحديد أنسب النماذج النحوية التي تتلاءم مع هذه الخصائص ثانياً، والكشف عن موقع هذا النحو بإزاء النظريات النحوية الحديثة ثالثاً، وخاصة نظرية تشومسكي التوليدية التحويلية.

---

[1] ينظر : الموسى؛ نهاد، العربية – نحو توصيف جديد في ضوء اللسانيات الحاسوبية، مرجع سابق، ص 55

وتبعاً لذلك جاءت معالجة النحو العربي آلياً ذات جانبين: أحدهما تحليلي، والآخر توليدي. فعلى الجانب الأول يقوم المحلل النحوي الآلي بتفكيك الجملة إلى عناصرها الأولية (أي تحليلها إعرابياً)، واستظهار العلاقات النحوية المختلفة .أما على الجانب الآخر فيقوم المولد النحوي بتكوين الجمل على صورتها الأصلية، وبعد ذلك تُجرى عليها عمليات التحويل النحوي المختلفة، كالحذف والإضمار، والتقديم والتأخير .

والجهود في هذا المجال كثيرة على المستويين النظري والتطبيقي، فنجدها – على سبيل المثال لا الحصر – في بحث الدكتور عبد الرحمن الحاج صالح عن (منطق النحو العربي والعلاج الحاسوبي) [1]، وبحث الدكتور نبيل علي عن (الحاسوب والنحو العربي)، وهو بحث لا يكاد يخرج عما أورده في كتابه (اللغة العربية والحاسوب) .

ويضاف إلى هذه الأعمال العلمية في ميدان المعالجة الآلية للنحو العربي بحثٌ للدكتور مازن الوعر بعنوان (التوليد الصوتي والنحوي والدلالي لصيغ المبني للمجهول في اللغة العربية – معالجة لسانية حاسوبية) ، وكذا ما أورده الدكتور نهاد الموسى عن تمثيل النظم، وتمثيل الإعراب، ضمن كتابه (العربية...).

[1] ينظر : حاج صالح، عبد الرحمان، منطق النحو العربي والعلاج الحاسوبي للغات، مجلة المبرز، مجلة دورية أكاديمية، المدرسة العليا للآداب والعلوم الانسانية، ص34، العدد الثالث 1993

كما أقيمت دراسات متعددة للمعالجة الآلية في ترتيب عناصر الجملة العربية باستخدام برامج ذات شبكات موسعة، ومن أبرزها نظام المعمدني، ونظام بن حماد وسعيدي، ومحاولة الدقاشي.

ويذكر الدكتور نبيل علي[1] في معرض تناوله للعناصر الأساسية المكونة للمعالج النحوي الآلي متعدد الأطوار للجمل العربية المكتوبة، أنه بصدد تطوير معالج آلي للنحو العربي، يقوم بالمهمة الأساسية للتحليل النحوي الآلي وهي توفير المعطيات اللازمة للتحليل اللغوي الأعمق، التي تتمثل في المصحِّح الآلي للأخطاء النحوية، والتخاطب مع قواعد البيانات باللغة الطبيعية، والترجمة الآلية من وإلى العربية، وتعليم النحو بواسطة الحاسوب، وإعراب الجملة العربية آلياً. وهذه – على وجه العموم – هي جملة ما تفيده العربية (النحو) من استخدام المعالج النحوي .

أما المستوى المعجمي فمساحة استفادته من الحاسوب واسعة جداً، وأدى هذا إلى ظهور ما يسمى بالمعاجم الحاسوبية أو المعاجم الآلية، بل إنه بدأ يأخذ بالبروز بوصفه علماً مستقلاً، أو فرعاً من فروع علم اللغة الحاسوبي يطلق عليه علم المعجم الحاسوبي (MRD Machine Dictionary Readable) وبظهوره بدأت الصناعة المعجمية تتحول من المعاجم اليدوية أو الورقية إلى المعاجم الآلية أو الإلكترونية.

---

[1] ينظر: نبيل، علي، اللغة العربية والحاسوب، دار تعريب، 1998، ص 101.

والمعجم الحاسوبي قطاع عام يضم معاجم لا حصر لها، سواء أكانت هذه المعاجم للناطقين بالعربية، أم معاجم للمصطلحات العلمية، أم معاجم من أنواع خاصة، أم معاجم مفهرسة، أم معاجم نصية. ويتميز هذا المعجم بميزات هائلة لا تتوافر في المعاجم التقليدية، كالشمول، والانتظام، والاطراد، والدقة والوضوح، والقابلية للتوسع والتعديل .

ونظراً للقضايا الشائكة التي تحيط بهذا النوع من المعجم، من حيث مستوياته، وحقله المعجمي، ومحتوياته، فقد عقدت له ندوة خاصة نظمتها مدرسة الملك فهد العليا للترجمة بطنجة (المغرب) عام 1995م، وكان عنوانها (التقنيات الحاسوبية في خدمة المصطلح العلمي والمعجم المختص). كما خُصص له جلسة نقاش بعنوان (بناء المعجم حاسوبياً) ضمن ندوة استخدام اللغة العربية في تقنية المعلومات، ولا تكاد تخلو ندوة من الندوات أو مؤتمر من المؤتمرات في مجال اللسانيات الحاسوبية من بحوث عن المعاجم الآلية .

ويقف الدكتور محمد الحناش (المغرب)[1] في صدارة اللغويين العرب المعاصرين الذين يولون المعجم الحاسوبي عناية خاصة، وجهداً كبيراً. وقد تمثل هذا في دراساته المتعددة عن بناء المعاجم الآلية في اللغة العربية، والمعجم الإلكتروني، والمعجم التركيبي للغة العربية.

وهو صاحب مشروع علمي كبير عمل عليه لسنوات طويلة من البحث اللساني الحاسوبي،

---

[1] ينظر : الحناش، محمد، مشروع نظرية حاسوب لسانية في بناء معاجم آلية للغة العربية، مجلة التواصل اللساني، المجلد2، العدد2، 1990م، ص43

تُوِّج بإصداره كتاب (المعجم التركيبي للغة العربية – مقدمات في المعالجة الحاسوبية للغات الطبيعية) .

وأسهم في هذا الإطار نظرياً كل من الدكتور نهاد الموسى، والدكتور أحمد مختار عمر، والدكتور عبد القادر الفاسي الفهري، بما كتبوه من بحوث عن حوسبة المعجم العربي، كما أسهم فيه عملياً الدكتور محمد مراياتي وزملاؤه بوساطة نظام خبير للغة العربية .

ومن المشاريع العلمية ضمن هذا التوجه أيضاً ما قام به معهد اللغة العربية بجامعة الملك سعود بالرياض (السعودية) من إصدار المكنز Corpus الوجيز (معجم في المترادف والمتوارد)، ومعجم التعبيرات الاصطلاحية، بإشراف – الدكتور محمود إسماعيل صيني[1] .

وهناك جهود أخرى تبذل لميكنة المعجم العربي من قبل المراكز العلمية والمؤسسات التجارية، كما هي الحال في مشروع الشركة العالمية للبرامج (صخر) بالكويت، ومشروع الشركة العالمية لبرامج الحاسوب بالقاهرة عن المكنز الآلي أو قاعدة بيانات المادة المعجمية العربية، ومشروع المركز العلمي لشركة آي.بي.إم بالقاهرة لتطوير قاعدة بيانات معجمية .

---

[1] ينظر: صيني، محمود إسماعيل، بنوك المصطلحات الآلية، مجلة اللسان العربي، العدد 48، 1999م، ص211–221

وينبغي أن أشير هنا إلى أن استخدام الحاسوب في الصناعة المعجمية، رغم كل تلك الميزات والإيجابيات التي تتحقق للمعجم ومستخدميه، هناك بعض السلبيات التي تنتج عنه جراء ذلك، لعل من أبرزها التكاليف المادية الباهظة التي يتطلبها هذا النوع من المعاجم، وما يقتضيه من مهارات خاصة قد لا تتوافر لكثير من مستخدميه .

على أن ذلك بطبيعة الحال لا يمكن أن يقلل من هذا التوجه المعاصر في حوسبة المعجم العربي، ولا ريب أن صياغته وفق أهداف واضحة، وغايات محددة، ومنهج علمي، ومن ثمَّ توظيفه التوظيف الصحيح، سيحقق للعربية، والمعجمية بوجه خاص ما كانت – وما زالت – تصبو إليه من شمولية، ومرونة، ودقة، ومعاصرة، كما سيحقق لعلمائها وباحثيها ما كانوا يظنون أنه من الأحلام والرؤى، بل المستحيل عينه .

أما المستوى الدلالي فيعد من أعقد الأنظمة اللغوية، وأشدها تعصِّياً على جهاز الحاسوب ؛ وذلك عائد إلى أن الدلالة من أقل المستويات اللغوية فيما يخص التباين اللغوي – كما يقول الدكتور نبيل علي [1]، كما أنه يشيع فيها عدة ظواهر تُخرجها من واقع الاستخدام اللغوي وحقيقته إلى المجاز، كالاستعارة، والكناية، والتشبيه، وهذا أمر يتطلب تحديد تلك التعابير غير الحقيقية وتصنيفها دلالياً بما يساعد النظام الحاسوبي على تمثلها، ومن ثم معالجتها آلياً

[1] ينظر: نبيل، علي، اللغة العربية والحاسوب، مرجع سابق، ص 103

ويمثِّل المعنى مشكلة كبرى بالنسبة للنظم الآلية، فتعدد المعنى للكلمة الواحدة، وحساسية السياق في تحديد دلالة الكلمة، واختلاف الدلالة باختلاف الثقافات، كل ذلك يجعل المعالجة الآلية للدلالة تنطوي على مفارقات يصعب بسببها تمثيل هذا المستوى أو توصيفه حاسوبياً، وبسبب من هذا تجاوزت أول دراسة صادرة عن اللسانيات الحاسوبية العربية الحديث عن المعالجة الآلية لعنصر الدلالة في العربية !.

على أن هذا لا يعني أن المعالجة الآلية لجانب الدلالة في اللغة العربية قد أُغفلت تماماً، بل إنه كان لها حضورها ضمن المستويات اللغوية الأخرى، كالمستوى الصوتي، والصرفي، والنحوي، والمعجمي، وضمن قضايا لغوية ذات صلة وثقى بالدلالة، كالترجمة الآلية. وهذا ما نلمسه في الجهود التي بذلت لتغطية هذا الجانب من اللسانيات الحاسوبية، سواءً أَكان ذلك في صورة بحوث نظرية، أو برامج تطبيقية .

فمن تلك البحوث ما كتبه الدكتور محمد غزالي خياط[1] – وهو متخصص في الهندسة – عن تمثيل الدلالة الصرفية في النظم الآلية لفهم اللغة العربية، وقد خصَّصه صاحبه لأوزان الأفعال في العربية، معتمداً في ذلك التمثيل الدلالي على استخدام نظم القواعد الشرطية، والجمل الإخبارية، والأنماط التقليدية، وقَدَّم في ضوء هذا طريقة مقترحة لتمثيل الدلالة الصرفية لأوزان الأفعال .

[1] ينظر: خياط، محمد غزالي، تمثيل الدلالة الصرفية في النظم الآلية لفهم اللغة العربية، دار تعريب، 2001، ص65

ومن ذلك أيضاً ما ذكره الدكتور محمد عز الدين (تونس) أثناء حديثه عن تصميم برنامج للترجمة الآلية أُطلق عليه (الناقل العربي) [1]، من أن هذا البرنامج يعمل على مستويات خمسة، من ضمنها مستوى التحليل الدلالي. وقد أوضح الدكتور عز الدين أن التمثيل الدلالي للجملة في هذا البرنامج يهدف إلى تحديد معنى كل كلمة في الجملة حسب السياق، مستعيناً في ذلك بمعطيات معجمية ودلالية، وبقواميس التعبير الاصطلاحية. وللسيد نصر الدين السيد بحث عن التحليل الدلالي للجملة الخبرية العربية باستخدام الحاسوب.

ومجمل القول في هذا إن المعالجة الآلية للدلالة العربية ما زالت في مراحلها الأولى، وهي تحتاج إلى بذل المزيد من الجهود لتصل إلى مرحلة متقدمة من مراحل التنظير والتطبيق والبرمجة .

[1] ينظر : عز الدين، محمد، الناقل العربي : تصميم برنامج للترجمة الآلية، دار الهدى، تونس، 2008، ص 112

الميدان الخامس : الترجمة الآلية

تعد الترجمة الآلية Machine Translation من أقدم مجالات استخدام الحاسوب في اللغة وقد نال هذا الجانب من اللسانيات الحاسوبية العربية حيزاً كبيراً ومساحةً واسعةً من الجهود المبذولة، وذلك بالنظر إلى أنها الأنموذج الآلي للمنظومة اللغوية .

وكانت فكرة الاستعانة بالحاسوب في الترجمة قد طرحت عام 1949م بأمريكا، ثم تحولت إلى مشروع علمي عام 1951م في معهد ماساشوستس للتقنية (M.I.T)، وكان عام 1954م ميلاد الترجمة الآلية الفعلي، التي سرعان ما انتقل الاهتمام بها إلى المراكز البحثية والجامعية في أوروبا والاتحاد السوفييتي، ودخل القطاع الخاص (التجاري) في السبعينيات منافساً لتلك المراكز العلمية في العناية بالترجمة الآلية [1].

أما واقع الترجمة الآلية في الوطن العربي فقد حدث في التسعينيات من القرن المنصرم (العشرين) تطور نوعي في مشروعات الترجمة الآلية على المستويين النظري والعملي (البرامج) .

---

[1] ينظر: العارف، عبدالرحمن بن حسن، توظيف اللسانيات الحاسوبية في خدمة الدراسات اللغوية العربية، مرجع سابق، ص 74.

فعلى المستوى النظري نجد طائفة من البحوث التي تعنى بهذه المسألة، سواء أكانت مؤلفة أو مترجمة. وقد اقتصرت الأبحاث النظرية على تحديد الصعوبات التي تعترض الترجمة الآلية، وكيفية التغلب عليها، وميزات هذا النوع من الترجمات، مع ذكر عمليات هذه الترجمة وأنظمتها المختلفة.

أما المستوى الآخر، وهو برامج الترجمة الآلية العلمية منها والتجارية، فقد أسهمت فيها جهود ذاتية (فردية)، وشركات تجارية، منها ما هو عربي، ومنها ما هو غربي، ومنها ما هو مشترك بينهما. ومن ذلك برنامج شركة صخر المسمى (القاموس Dictionary)، وبرنامج الناقل العربي، وبرنامج ترجمان، وبرنامج المترجم، وبرنامج عرب ترانز Arabtrans [1] ..إلخ وقد بدأت هذه البرامج بالعمل الفعلي، ومازال التطوير يلاحقها يوماً بعد يوم، وتتم الترجمة فيها من الإنجليزية إلى العربية، والعكس، وهذا تطور إيجابي يعكس الرغبة في نقل الثقافة العربية إلى الآخرين، مما يعد تحولاً ذا دلالة عميقة في الانتقال بالترجمة من الاستيراد إلى التصدير، بلغة الاقتصاديين .

[1] ينظر : المرجع نفسه، ص 76.

ومن الملاحظ أن الترجمات الآلية كانت في بدء أمرها مقتصرة على ثنائية اللغات، أو ما يعرف باللسانيات التقابلية Contrastive Linguistics، أما الآن وبعد النقلة الكبيرة في هذا المضمار فقد أصبحت تقوم بترجمة عدد كبير من اللغات في وقت واحد، وهو ما يعرف بالترميز الرقمي Digital Coding .

وبعد، فلئن كانت هناك في الماضي صيحات تحاول أن تُهوِّن من شأن الترجمة الآلية، وتُشَكك في نجاحاتها، فإنها الآن بدأت تطالب بإلحاح بتطوير أنظمة هذه الترجمة، بعد أن حققت نتائج ملحوظة وخاصة في ميدان النصوص العلمية. واللغويون والحاسوبيون العرب مطالبون أكثر من أي وقت مضى بالإفادة القصوى من معطيات الترجمة الآلية المعاصرة لدى الغرب، وتسخيرها لخدمة اللغة العربية .

الميدان السادس : الكتابة العربية ومعالجتها آليا

أما الكتابة العربية ومعالجتها آلياً، فتعد الكتابة العربية من ضمن أهم المشكلات التي واجهت التحليل الحاسوبي[1]، حيث تتعدد الأشكال البصرية للحرف الواحد تبعاً لموقعه من الكلمة، كما أن اتجاه الكتابة العربية هو من اليمين إلى اليسار، يضاف إلى ذلك أن حروفها متصلة وليست منفصلة .

[1] ينظر : المرجع نفسه، ص 79.

وتبعاً لهذا قامت عدة محاولات لتلافي مشاكل الكتابة العربية في الحاسوب، وكان من بينها مشروع الأستاذ أحمد الأخضر غزال، الذي أطلق عليه الطريقة المعيارية للطباعة العربية، أو العربية المعيارية المشكولة – الشفرة العربية، والتي تعرف اختصاراً بـ(العمم–شع) [1]، وقد تم هذا عام 1954م، ثم طُورت لتتلاءم مع التقدم التقني في الحاسوبات عام 1974م .

ويذكر الدكتور مازن الوعر أنه مع تطور الإلكترونيات أصبح هذا التصميم غير واقعي، مما جعل بعض الباحثين يضع تصميماً آخر عرف بـ(معالجة السياق)، أي استنباط الحرف من سياق الحروف، وليس من سياق المعنى .

وهناك أيضاً الشفرة العربية الموحدة للكتابة العربية التي تعرف بـ(الشفرة سباعية العزوم) التي أقرتها سنة 1983م المنظمة العربية للمواصفات والمقاييس. وفي الحقيقة أنني لا أعلم الآن ماذا تم بشأن تطوير هذه الشفرة خاصة أنه مضى عليها زمن طويل .

ومن تلك المحاولات ما قدمه الدكتور عاصم عبد الفتاح نبوي، والدكتور صبري عبد الله محمود، من تطوير نظام للتعرف على حروف العربية باستخدام شبكة عصبية ذات انتشار رجوعي .

---

[1] (العمم – شع) هو مصطلح مختصر منحوت من أوائل الحروف الأولى من النموذج الذي وضعه الأستاذ أحمد الأخضر غزال، وأطلق عليه اسم (العربية المعيارية المشكولة – الشفرة العربية .(وللمزيد حول هذا المشروع ينظر: د.مازن الوعر، قضايا أساسية في علم اللسانيات الحديث، الوفاء لدنيا الطباعة والنشر مصر 2009، ص418 فما بعدها.

الميدان السابع : تعليم اللغة

ومن الجوانب الأخرى التي أمكن للغة العربية الاستفادة منها من الحاسوب تعليم اللغة، سواء للناطقين بها من أبنائها، أو للناطقين بغيرها من اللغات. لقد استطاع الحاسوب أن يقدم للناطقين بالعربية نظماً حاسوبية وبرامج لإكساب المتعلمين المهارات اللغوية المتعددة، كالقراءة، والكتابة، والمحادثة، والاستماع، إضافة إلى معالجة الخطوط العربية معالجة حاسوبية، والتدقيق الإملائي والنحوي، ووضع معاجم لغوية حاسوبية لمراحل التعليم العام، وتعليم الأطفال الأرقام والحروف والكلمات .

كما استطاع الحاسوب أن يسهم في مجال تعليم اللغة العربية لغير الناطقين بها، بإمكاناته وقدراته الهائلة في التعليم المبرمج، ويوجد بمعهد اللغة العربية بجامعة أم القرى (مكة المكرمة) حالياً معمل حاسوبي لتعليم العربية للناطقين بغيرها [1]، وهي تجربة فريدة ذات أبعاد علمية وحضارية، نأمل أن تتوسع دائرتها ومجال تطبيقاتها، وأن يكتب لها النجاح .

ومن البديهي القول إن الإفادة من الحاسوب في هذا المجال (المجال التعليمي) مرهونة بالنجاح في المعالجة الآلية للغة العربية أولاً .

---

[1] ينظر: العارف، عبدالرحمن بن حسن، توظيف اللسانيات الحاسوبية في خدمة الدراسات اللغوية العربية، مرجع سابق، ص 81.

هذا ما يتعلق بالمعالجة الآلية للنظم اللغوية العربية، أما البرامج الموضوعة لذلك، وهي كما يقول الدكتور نهاد الموسى : «منجزات تطبيقية تستثمر التوصيف في وجوه من التوظيف»[1] فإنها أكثر من أن تحصر، وخاصة أن الشركات التجارية العاملة في مجال الحاسوبيات تدفع يومياً إلى السوق برامج علمية وتعليمية، تتفاوت فيما بينها دقةً ومنهجيةً وأهدافاً، وهي تحتاج من الباحثين تقييماً وتقويماً.

وأما الصورة الأخيرة لرصد الجهود في ميدان اللسانيات الحاسوبية العربية فتتمثل – كما تقدم – في إنشاء قسم خاص لعلم اللغة الحاسوبي (اللسانيات الحاسوبية) في جامعة الأمير سلطان الأهلية بالرياض (السعودية)، وكذا إنشاء مركز البحث العلمي والتقني لتطوير اللغة العربية بجامعة الجزائر 2 ببوزريعة، ولا شك أن إنشاءهما جاء تتويجاً للجهود الحاسوبية العربية التي أصبحت تشكل اتجاهاً عاماً في الدراسات اللغوية المعاصرة .

الميدان الثامن : التلخيص

ومن المجالات الهامة التي ساهم فيها ودرسها الباحثون العرب التلخيص ـ الذي سنفرد له فصلا ـ الذي هو ضرورة من ضرورات الحياة التي يحتاجها الباحث سعيا إلى الاقتصاد في

---

[1] ينظر: نهاد الموسى ؛ العربية – نحو توصيف جديد في ضوء اللسانيات الحاسوبية، مرجع سابق، ص 59

الجهد وتوفير الوقت والمال [1] ففي ظل الثورة المعلوماتية الهائلة التي يعيشها العالم في وقتنا المعاصر والمسؤوليات الجسام التي تقع على عاتق المسئولين في كثير من المواقع لا يجد المرء ،في أحيان كثيرة، الوقت الكافي لمتابعة كل ما يدور حوله من تطور معرفي وتقدم علمي وأحداث متسارعة .إذ نجده كثيرا ما يكتفي بمتابعة أهم الأنباء في نشرات الأخبار والعناوين الرئيسة في الصحف والمجلات حتى إذا لقي ما يلفت انتباهه بحث عن بقية المقال كي يستزيد بما فيه.

كما قامت شركة صخر بانشاء الملخص الآلي الذي يتميز بخاصية التلخيص الآلي للوثائق، والبريد الإلكتروني وصفحات الإنترنت بالإضافة إلى استعراض الأفكار الرئيسية التي يتضمنها أي نص باللغتين العربية والأنجليزية.

[1] ينظر: دينا عويس 2014 http://mawdoo3.com شوهد يوم 02 ماي 2016 على 09 سا 05 د

لقد اعتمد الملخص الآلي على ما تم إنجازه من تقنيات صخر في مجال البرمجة المعلوماتية، أي المحلل الصرفي والمفهرس الآلي، والمشكل الآلي، ويمكن تلخيص خصائصه في المهام التالية :

أ – قدرته الفائقة على استخراج الأفكار الرئيسية من الوثيقة وتقديمها بطريقة مختصرة مما يؤدي إلى اختصار الوقت اللازم للقراءة، الأمر الذي يستلزم توظيف العديد من التقنيات المتقدمة في الإحصاء، بالإضافة للفهم الدلالي للنص الذي يتطلب قاعدة لغوية تغطي معظم مفردات اللغة.

ب – تحديد حجم النص الأصلي المرغوب إيجاده في النسخة الملخصة.

ج – تحديد عدد الجمل الأصلية التي تظهر في النسخة الملخصة.

د – تحديد عدد كلمات النص الأصلي التي يراد إدراجها في النسخة الملخصة.

تمثل هذه الخصائص مجرد حلقة أولى، إذ أن الملخص الآلي حسب تقدير مهندسي صخر مهيأ ليصبح إحدى تقنيات المستقبل اللازمة إلى أدوات المكتب الأنترنت والأنترانت نظرا لدوره الحيوي في توفير الوقت عند التعامل مع الكميات الضخمة من المعلومات [1].

[1] ينظر: عبد الغني أبو العزمhttp://www.aljabriabed.net/n31_04abualazm.%282%29.htm
شوهد يوم 29 افريل 2016 على 11سا40د

3ـ أعلام واتجاهات :

وباستعراض سريع لتلك البحوث نجد أنها كُتبت بالعربية، والإنجليزية – أيضاً –، وجاءت عناوينها شاملة للمستويات اللغوية كافَّة، أصواتاً، وتراكيبَ، وبنيةً، ودلالة، ومعجماً، ولبعض قضايا اللغة من المنظور الحاسوبي، كالترجمة الآلية، وبنوك المصطلحات، وتعليم اللغات، والذكاء الاصطناعي .

أما أصحابها فنجد أن جُلَّهم من اللغويين الأكثر حضوراً وفاعليةً على الساحة اللغوية، من أمثال الدكتور محمد الحناش (المغرب)، والدكتور محمود إسماعيل صيني (السعودية)، والأستاذ أحمد الأخضر غزال (المغرب)، والدكتور عبد القادر الفاسي الفهري (المغرب)، والدكتور مازن الوعر (سورية)، والدكتور محمود فهمي حجازي (مصر)، والدكتور عبد الرحمن الحاج صالح (الجزائر)، والدكتور سالم الغزالي (تونس)، والدكتور داود عبده (الأردن)، وبعضهم من المتخصصين في الحاسوب أو الهندسة الحاسوبية، كالدكتور يحيى هلال (المغرب)، والدكتور محمد مراياتي (سوريا)، والدكتور نبيل علي (مصر)، والدكتورة نادية حجازي (مصر) . ويلاحظ على أغلب هذه البحوث أنها انتقلت باللسانيات الحاسوبية من مجالها النظري أو التنظيري إلى الجانب التطبيقي، وهو تطور إيجابي يُحسب لأصحاب هذا الاتجاه [1] .

ومن المعروف أن الجانب التطبيقي – وهو الجانب الأهم في اللغويات الحاسوبية – يتمثل في تسخير العقل الإلكتروني لحل القضايا اللغوية، وهنا يبرز الدور الرئيس والأثر الفاعل لالتقاء اللغويين والحاسوبيين، والتعاون فيما بينهم، وما يثمر عنه من نتائج تسهم إلى حد

---

[1] ينظر: العارف، عبدالرحمن بن حسن، توظيف اللسانيات الحاسوبية في خدمة الدراسات اللغوية العربية، مرجع سابق، ص 83.

كبير في تذليل العقبات وحل المشكلات التي تواجه التحليل الحاسوبي للغة العربية، هذه العقبات والمشكلات بعضها يتصل بطبيعة اللغة العربية، أصواتاً، وبنيةً، وتركيباً، ودلالة، ومعجماً، وبعضها يتعلق بنظام الكتابة العربية، وبعضها يتصل بالمصطلح العلمي التكنولوجي للسانيات العربية، كما أن هناك مشكلات أخرى تتعلق بالبرمجيات، إعداداً، واختياراً للمادة اللغوية العربية (أنموذج لساني عربي)، وتعريباً للبرمجة. وثالث هذه المشاكل يكمن في الجهاز الحاسوبي (الكمبيوتر)، وأنظمة تمثيل المعرفة على الحاسوب باللغة العربية وقد بُذلت جهود كبيرة من الأطراف المعنية كافة بهذه القضية للتغلب على تلك الإشكالات، ومن ذلك ما قدمه الدكتور عبد الرحمن الحاج صالح (الجزائر) من تصور حول وضع أنموذج لساني للعلاج الآلي للغة العربية، وما طرحه الدكتور محمد عبد المنعم حشيش (مصر) من تصميم قاعدة للمعلومات بغرض تغطية الثروة اللفظية للغة العربية، والمشروع الذي تبنته مدينة الملك عبد العزيز للعلوم والتقنية (الرياض) حول إنشاء وتطوير بنك آلي للمصطلحات أطلق عليه (باسم)، وما وضعه الأستاذ أحمد الأخضر غزال (المغرب) من تصميم طريقة تكنولوجية آلية لتعريب الحاسوب، ووضع اللغة العربية في الحاسوبات الإلكترونية وفق هويتها وخصوصية ومارفهما ورسومها[1].

إن معالجة اللغة العربية حاسوبياً أصبحت اليوم أمراً لا حيدة عنه ولا مفرّ منه، وخاصة أن استثمار الدراسة الحاسوبية والمعلوماتية – بصفة عامة – يحقق نتائج كبيرة للغة العربية، في مجال التعريب، والإحصاء اللغوي، والمعالجة الآلية، وتعلم اللغات، والترجمة الآلية، وفي مجال التربية والتعليم . لقد تبين لنا بهذا أن الحاسوب يمكن تطويع آلياته وأنظمته خدمة للدراسات اللغوية والأدبية العربية.

[1] ينظر : المرجع نفسه ص 85.

4- أمثلة من تطبيقات حوسبة اللغة :

-معالجات لغوية آلية في الصرف والنحو.

-الترجمة الآلية وأدواتها.

-البحث الذكي وخياراته.

-نظم معالجة الكلام آليا.

-التلخيص الآلي للنصوص. [1]

---

[1] ينظر: اليوسفي، أحمد شعيب، نحو نظام جديد لإدارة المعلومات، مرجع سابق، ص114

خاتمة ومقترحات

تحت عنوان : « منطلقات لتهيئة اللغة العربية لعصر الإنترنت » قدم الأستاذ نبيل علي نماذج من هذه المنطلقات منها:

1- منطلقات عامة: وهي بلورة سياسية لغوية على مستوى الوطن العربي يساهم فيها اتحاد المجامع العربية ويدعى للمشاركة في وضعها بجانب اللغويين غيرهم من علماء التربية وعلماء النفس وعلماء الاجتماع والبيولوجيين على أساس أن اللغة هي مسؤولية النخبة المفكرة قبل أن تكون مسؤولية الساسة وأهل الاختصاص.

2- منطلقات خاصة بالتنظير واستغلال توسطية اللغة العربية :من أجل الإسراع في سد فجوة التنظير من خلال الاسترشاد بما تم في اللغات التي تتشابه مع العربية في بعض خصائصها.

3- منطلقات خاصة بتطوير المعجم: مراجعة شاملة لآلية توليد الكلمات في العربية.

إدراج علم المعجم في عمل المجامع ومناهج الجامعات، خاصة فيما يتعلق بالدلالة المعجمية وظاهرة المجاز.

4- منطلقات خاصة بتعليم اللغة العربية وتعلمها ذاتيا: التأسيس النظري لتعليم اللغة العربية

وتعلمها على ضوء الإنجازات الحالية لعلم تعليم اللغة. والتركيز على الجوانب الوظيفية وتنمية المهارات اللغوية الأربع بصورة متوازنة (الكتابة والقراءة والتحدث والاستماع.

5- منطلقات خاصة بالترجمة: وضع دليل المترجم العربي للترجمة العلمية تتضمن قائمة بالمشكلات التي تواجه الترجمة من وإلى العربية وكيف تم التعامل معها من قبل ثقات المترجمين. دعم الجهود الحالية في الترجمة الآلية.

6- منطلقات خاصة بمعالجة اللغة العربية آليا: استغلال ما يعرف حاليا بـ "أزمة البرمجيات" للحاق بالموجة الثانية لمعالجة اللغات الطبيعية آليا وتهدف إلى التوسع في تطبيق أساليب الذكـــــاء الاصطناعي [1].

نستعرض في الفصل القادم تعريفات التلخيص وأنواعه وتطبيقاته وكيف نستخدمه لتطوير اللغة العربية خاصة في علم الإعلام الآلي وفي مختلف مجالاته.

(1) ينظر: الأستاذ نبيل علي في موقع شذرات

http://www.shatharat.net/vb/showthread.php?t=12168

شوهد يوم 29 افريل 2016 على 17سا30

الفصل الثاني

LES RESUMES

ET LES RESUMES AUTOMATIQUES

Dans ce chapitre, nous aborderons le domaine des résumés par un ensemble de définitions et de concepts liés à la réduction des textes que ce soit manuel ou automatique.

A/ Les résumés :

1. Les définitions :

Dans la littérature, Il existe différentes définitions du terme « résumé », parmi ces différents définitions on trouve celle du colloque international du résumé des sciences qui s'est tenu à Londres entre le 20 et le 25 du mois de juin 1949 et qui stipule que :

« Le résumé est une contraction d'un ouvrage ou de documents accompagnée d'une description sur un support comportant la facilité d'accès au document d'origine » [(1)].

Alain conte stipule que «le résumé est une contraction d'un texte accompagné d'une bibliographie permettant l'accès à ce texte» [(2)].

---

(1) النوايسة، غالب عوض، الاستخلاص لأغراض استرجاع المعلومات، مجلة رسالة المكتبة، 1999 ص154، الأردن

(2) Conte, Alain, La révolution d'informations : l'utilisation des ordinateurs pour le stockage et la récupération d'informations, traduction Hechmat Kacem et Chawki Salem 2ème édition, Kuwait : l'agence des publications, p172, 1979.

Mohamed Fathi Abdelhadi définit le résumé comme étant " une présentation restreinte et exacte du contenu d'un document dont le style ressemble à celui-ci, accompagné d'une description bibliographique permettant l'accès au premier document " [1].

Horacio Saggion de l'université de Montréal , définit le résumé comme étant " un texte concis qui rend compte du contenu « essentiel » d'un document " [2].

Plus généralement, Gergisse et Abdeljabar Abderrahmane définissent le résumé comme " une sorte de forme bibliographique qui donne le même sens qu'un texte ou un document, et qui permet d'émettre la même idée sur le secteur de traitement de ces documents facilitant ainsi la recherche suivie " [3].

Résumer c'est donc transformer un texte en texte plus court mais de sens équivalent (fidélité aux idées du texte original).

---

[1] محمد فتحي، عبد الهادي، العمليات الفنية في المكتبات ومراكز البحث المجلة العربية للمعلومـات، مج4، ع2، ص44 (1983)

[2] Horacio, Saggion, Génération automatique de résumé par analyse sélective Thèse de doctorat, p45-p46, Université de Montréal 2000.

[3] جرجيس، جاسم محمد، عبد الجبار عبد الرحمان، المراجع والخدمات المرجعية، مركز التوثيق الإعلامي لدول الخليج العربي ص32، (1985)

2. Apparition des résumés :

Les résumés sont apparus selon Bradford " au commencement d'édition des premiers périodiques relatifs à ceux-ci à partir de Janvier 1665 "[1]. Toutefois les résumés sont apparus bien avant [2] mais l'apparition réelle revient à l'édition des résumés chimiques de l'association américaine de la chimie au 19ème siècle.

Selon Nouaissa, " l'apparition des résumés était une nécessité due à :

- L'important développement humain dans plusieurs domaines tant sur le volet documentaire, périodiques, conférences, …
- La pluralité des sources de l'information
- La complexité des besoins des utilisateurs et leurs demandes en service rapides
- La complexité des liens entre les sujets et leurs relations les uns des autres
- La pluralité des langues d'édition "[3] .

---

[1] حشمت،قاسم، خدمات المعلومات : مقوماتها وأشكالها، القاهرة : مكتبة غريب، ص209 (1984)

[2] ينظر : القزويني، جلال الدين محمد بن عبد الرحمن، كتاب التلخيص في وجوه البلاغة ، ضبط وشرح عبد الرحمان البرقوقي ، ص 187-199 ، دار الفكر العربي 1999

[3] النوايسة، غالب عوض، الاستخلاص لأغراض استرجاع المعلومات، ص 45، مجلة رسالة المكتبة، الأردن 1999

3. L'importance des résumés :

L'importance des résumés comme outil d'accès à l'information réside dans ce qui suit :

- Les résumés sont considérés comme un style évolué permettant aux chercheurs le suivi des évolutions des sciences sans rentrer dans les détails, vu l'explosion du secteur de l'information.
- Le bon résumé aide à l'économie de l'effort et du temps du lecteur du fait qu'il fait gagner environ 9/10 du temps qui serait normalement dégagé pour la lecture du document d'origine.
- Le résumé aide à l'économie des frais de recherches sur un sujet.
- Le résumé aide à éviter les problèmes de traduction du fait de l'existence d'environ 70 langues au niveau des sciences et de la technologie. Un chercheur qui maîtrise en moyenne deux langues peut éviter ce genre de problème en utilisant la langue la plus répondue au niveau des résumés.
- Les résumés aident à faire le choix des documents à utiliser relativement à un certains sujets, du fait que certain titres de documents peuvent être très généralistes ceci peut être du à la sécurité par exemple ou le mauvais choix de titre [(1)].

---

(1) ينظر : Horacio, Saggion, Génération automatique de résumé par analyse sélective, مرجع سابق, p49

- Les résumés aident aussi à la recherche d'informations pertinentes sur un sujet.
- Les résumés aident au développement des sciences du fait de leur pertinence dans le style et de l'économie de l'effort, du temps et des frais de recherches.
- Les résumés aident à faciliter la révision et le contrôle de la documentation scientifique du fait de leur simplicité par rapport aux documents d'origines [1].

Additionnement à cela, Cleber Guenitcha et Michel Manou stipulent que les résumés ont trois fonctions importantes :

1. "La diffusion des informations
2. L'acquisition de l'information pertinente
3. Restitution de l'information" [2]

De son côté Rafik El-halimi stipule l'importance des résumés dans ce qui suit :

1. " Développe en nous la lecture critique et la compréhension totale de ce qu'on lit.
2. Nous aide à éviter la perte du temps et de l'effort en lisant des détails non trop important " [3].

---

[1] ينظر : النوايسة، غالب عوض، الاستخلاص لأغراض استرجاع المعلومات، ص47 مرجع سابق

[2] Guenitcha, Cleber, et Manou, Michel, Introduction générale des sciences et techniques d'informations et d'archivage, UNISCO, p171, Tunis, 1987

[3] الحليمي، رفيق حسن، فن كتابة التلخيص والمختصرات، مركز المخطوطات والتراث والوثائق الكويت، ص 129، ط1 2002

3. On distingue les idées principales et secondaires.
4. On ingurgite rapidement un nombre important d'idées dans un monde d'explosion de l'information.
5. Nous permet de voir le sens esthétique des textes et leurs différents types.
6. Nous aide à être sûrs de nous-mêmes car c'est un effort individuel.
7. Nous est utile lors des études, des conférences ou dans la recherche scientifique.

4. Les méthodes des résumés :

Une préparation pour faire un résumé est indispensable, elle consiste en :

1. Une lecture itérative du texte original
2. Comprendre le contenu
3. L'analyser et dégager avec précision les idées et leurs enchaînement

Dans les cours du journalisme scientifique [(1)], on peut trouver l'éclaircissement des points précédents comme suit :

1) Survoler le texte

a) Prendre appui sur la disposition du texte (titres, sous-titres, énumération, nombre de paragraphes)

[(1)] ينظر : Cours « Information scientifique et technique » 1ère année magister Journalisme scientifique, faculté des sciences, Université de Blida, 2006/2007

b) Parcourir l'introduction et lire la conclusion

c) Lire la première phrase de chaque paragraphe (indication du contenu de chaque

paragraphe, rapports logiques : ensuite, enfin, mais, par contre, d'autre part)

d) Saisir des mots ou des phrases de transition entre les différentes parties.

2) Éclairer toute la signification du texte

Il faut faire plusieurs lectures en deux temps :

i. Quelques lectures attentives
ii. Souligner les difficultés et les complications

3) Dégager les idées essentielles (idée principale et idées complémentaires de chaque partie)

4) Faire apparaître les idées sous forme d'un plan [1].

Toutefois, les règles les plus connues et les plus utilisées en résumé sont les suivantes :

1ere règle : La suppression

On peut supprimer totalement les phrases qui ne contribuent pas à la compréhension du texte. A titre d'exemple : la description des objets, d'individus ou de tâches secondaires.

---

[1] ينظر : Horacio, Saggion, Génération automatique de résumé par analyse sélective, مرجع سابق, p51

2ème règle : La fusion

On peut fusionner une phrase au niveau d'autres qui sont des conditions essentielles de celle-ci ou des conséquences.

3ère règle : La construction

On peut construire une phrase à la place d'autres à condition qu'elle soit le résultat naturel de celles-ci.

4ère règle : La généralisation

On peut substituer des phrases par une autre plus globale ayant le même sens [1].

Le fait de résumer n'est pas la seule caractéristique des matières scientifiques, littéraires ou écrites, c'est aussi nécessaire dans l'ère de l'information et de l'explosion des connaissances.
En effet, on peut en avoir besoin pour reformuler des matières audio-visuelles en effaçant ce qui est superflus et en ne gardant que l'essentiel par un traitement aboutissant à des idées simples et claires .

---

[1] ينظر : الشاعر، شفيق سالم،-معلم بوكالة الغوث- فن التلخيص، مجلة المعلم تربوية ثقافية 2006

5. Les types de résumés :

Il existe plus d'un critère pour la typologie des résumés, en effet, il est possible de classer ces dernières selon le but de leur élaboration en : résumés informatifs, indicatifs, descriptifs ou critiques; ou encore selon la méthodologie suivie pour résumer en résumés unifiés, télégraphiques ou adaptés.

Alain Conte classe les résumés comme suit :

1. " Les résumés traditionnels qui englobent les résumés descriptifs et les résumés informatifs
2. les résumés adaptés
3. les résumés de style unifié "[(1)].

Alors que Mohamed Mohamed Amane les classe de la façon suivante [(2)]:

1. Le résumé informatif
2. Le résumé descriptif

---

(1) Conte, Alain, La révolution d'informations : l'utilisation des ordinateurs pour le stockage et la récupération d'informations, traduction Hechmat Kacem et Chawki Salem 2ème édition, Kuwait : l'agence des publications, p172, 1979.

(2) ينظر : أمان، محمد محمد، خدمات المعلومات والإحاطة الجارية، الرياض، 1985 دار المريخ، ص67

3. Le résumé mixte
4. Le résumé prévisionnel
5. Le résumé d'auteur

Atefeh Farzindar de l'université de Montréal, cite les types de résumé suivants :

" Résumé indicatif : fournit une idée du texte sans donner le contenu spécifique, il signale les thèmes du document et son style peut être télégraphique.

Résumé informatif : renseigne sur les informations essentielles contenues dans le document.

Résumé sélectif : qui néglige les aspects très généraux d'un document et développe les parties spécialisées.

Résumé cible : se concentre sur le(s) topique(s) d'intérêt pour l'utilisateur.

Résumé générique : reflète le point de vue de l'auteur " [1].

---

[1] Atefeh, Farzindar, Développement d'un système de résumé automatique de textes juridiques, Université de Montréal 2001.

Horacio Saggion de l'université de Sheffield [1], de son coté, énumère les types suivants :

1. L'extrait : un ou plusieurs passages tirés d'un texte.
2. Le sommaire : énumération des points principaux d'un discours.
3. L'abrégé : écrit réduit au points essentiels .
4. Le précis : un exposé exact et succinct d'un texte.
5. Le paraphrase : interprétation des idées de l'auteur du document source.
6. Le digest : un condensé d'un livre ou d'un article journalistique.
7. Le highlight : un commentaire ajouté dans des parties spécifiques d'un document pour attirer l'attention du lecteur sur les points importants.
8. Le synopsis : appliqué en cinématographie, est un récit très bref constituant un schéma de scénario.

---

[1] ينظر : Horacio, Saggion, Génération automatique de résumé par analyse sélective, , مرجع سابق, ص 104.

6. Les documents concernés par le résumé :

D'après Owlley Jennifer [1], les documents concernés par le résumé sont les suivants :

1. Les documents ciblés par les utilisateurs
2. Les documents qui apportent du nouveau dans les différents domaines critiques
3. Les rapports finaux qui sont argumentés
4. Les documents qui contiennent des informations difficiles à obtenir ou des informations rares
5. Les présentations importantes
6. Les magazines de renommée

En outre, Hechmat Kacem [2] ajoute les classes suivantes :

1. Les manuels d'utilisation ainsi que les brevets
2. Les recherches scientifiques aboutissants à de nouveaux résultats
3. Les thèses de recherches

---

[1] ينظر : Owlley, Jennifer, Abstacting and indexing, London : Clive Bingley, p12-20, 1982.

[2] حشمت،قاسم، خدمات المعلومات : مقوماتها وأشكالها، القاهرة : مكتبة غريب، ص209 (1984)

Les grandes entreprises, les grandes administrations (ministères, services publics, etc.), les laboratoires et les services de développement sont confrontés à un défi : gérer la masse des documents textuels saisis sur des supports électroniques. Comment les classer ? Comment les stocker pour y retrouver rapidement les informations qu'ils contiennent ? Comment diffuser ces informations à ceux qui sauront les utiliser ? Comment filtrer une information pertinente parmi toutes les informations contenues dans les documents stockés ? Ces tâches sont d'autant plus complexes que ce qui est jugé pertinent pour l'un ne l'est pas nécessairement pour l'autre.

Les critères traditionnels, plus ou moins efficaces pour les textes saisis et manipulés par les supports imprimés, comme l'emplacement physique du document dans les archives ou la mémoire du documentaliste, ne sont pas applicables aux documents électroniques. D'autres critères, plus rigoureux, doivent être trouvés pour s'adapter aux possibilités du traitement informatique.

(1) Minel, J. Luc, Filtrage sémantique, P163, Edition Lavoisier 2001.

Les techniques automatiques de recherche d'informations ne sont pas toujours très satisfaisantes et ne répondent qu'imparfaitement aux besoins des utilisateurs dans la mesure où elles sont souvent trop bruitées (trop d'informations non pertinentes sont présentes dans la réponse) et fondés sur des techniques purement quantitatives.

Les moteurs de recherche actuels, à l'œuvre dans des grandes bases documentaires ou sur le web, sélectionnent un très grand nombre de documents en rapport avec une requête (par exemple une requête sur «le résumé automatique» avec le moteur de recherche Google fournissait 1430000 références, site consulté le 25/03/2014). Dans ce contexte, il est intéressant d'offrir des outils de visualisation rapide des textes sélectionnés afin que l'utilisateur puisse évaluer leur pertinence par rapport à sa requête. Il y a deux manières de répondre à cette problématique : produire un résumé statique du texte et le proposer à l'utilisateur ou bien permettre à ce dernier de visualiser seulement les points qui l'intéressent par une navigation interactive

(1) Minel, J. Luc, Filtrage sémantique, P164, المرجع نفسه.

navigation interactive dans les documents, ce qui s'apparente alors à la notion de résumé dynamique, construit en interaction avec l'utilisateur.

En ce qui concerne le résumé statique, pourquoi ne pas se contenter d'un résumé rédigé à la main par un résumeur professionnel ? D'abord parce que tous les textes ne sont pas systématiquement accompagnés d'un résumé, notamment les textes qui circulent sur le réseau Internet. Et surtout parce que le coût de production d'un résumé par un résumeur professionnel est très élevé et que la productivité de ce même professionnel est faible. A titre d'exemple, pour un texte source d'une dizaine de pages, un résumeur professionnel, lorsqu'il est spécialiste du domaine, produit un résumé en une dizaine de minutes, mais il lui faut presque une heure lorsque le domaine traité ne relève pas de sa compétence. Ensuite, parce que les travaux menés en collaboration avec les résumeurs professionnels ou en comparaison avec les résumés produits par ces professionnels ont montré la difficulté à réaliser des résumés standard, c'est-à-dire construits sans tenir compte des besoins des utilisateurs.

(1) Minel, J. Luc, Filtrage sémantique, P165, المرجع نفسه.

En effet, il n'existe pas de critères précis pour déterminer ce que serait un « bon » résumé ou un « résumé idéal ». Par exemple, le résumé scolaire, qui vise à tester les capacités de paraphrasage, de reformulation et de synthèse des élèves, n'est pas conçu et organisé de la même façon que le résumé d'auteurs.

Par ailleurs, l'activité résumante des humains a été encore fort peu étudiée par les psychologues et les résumés sont très différents selon les utilisateurs auxquels ils sont destinés. On arrive alors à la nécessité d'élaborer un résumé à la demande, convenant à l'utilisateur.

Les résumés dépendent également des types de texte. On ne résume pas de la même façon un texte narratif, un article scientifique relatif à une science expérimentale, un article d'une science théorique ou d'un domaine spéculatif, des articles juridiques, etc. Il n'y a donc pas de façon général et standard pour faire un résumé idéal qui serait indépendant des demandes des utilisateurs et des types de texte.

(1) Minel, J. Luc, Filtrage sémantique, P166, المرجع نفسه.

" L'une et l'autre solution, résumé statique ou dynamique, reposent sur une analyse du contenu des documents afin d'en extraire les thèmes abordés (les points les plus importants et leur structuration), dans le but de fournir des résumés structurés en fonction des besoins et objectifs d'un utilisateur et de ses préférences " [(1)].

A l'heure actuelle, avec la masse astronomique d'information qui circule sur Internet et dans les fonds documentaires électroniques et les différents services électroniques, personne n'a le temps de lire et assimiler tous ces données d'où la nécessité de produire une technologie de résumé automatique efficace - en terme de temps et d'argent- capable de résoudre ce problème.

(1) Minel, J. Luc, Filtrage sémantique, P167, المرجع نفسه.

B/ Les résumés automatiques

" Le traitement automatique des langues (TAL) est une branche de l'intelligence artificielle, son objectif est la conception de logiciels ou programmes, capables de traiter de façon automatique des données linguistiques, c'est-à-dire des données exprimées dans une langue dite 'naturelle' " [(1)] .

Parmi les diverses applications possibles qui peuvent être rattachées au TAL, on peut citer principalement [(2)] :

- ***La traduction automatique*** : impliquant compréhension d'une langue, transfert dans une autre langue et génération dans cette dernière.
- ***Le résumé automatique des textes*** : l'analyse est suivie d'une étape de sélection des éléments considérés comme essentiels et leur génération.
- ***La réalisation d'interface en langue naturelle*** : permet l'interrogation de bases de données, l'utilisation de systèmes experts ou la constitution de leur base de connaissances.

---

(1) Mourad, Amine, Glossaire de linguistique computationnelle, P101, Lavoisier, 1995.

(2) ينظر : Sabah, Gerard, Le TALN & les modèles de représentation des connaissances, P114, Hachette, 1988.

On a alors surtout besoin de connaissances sur le domaine (souvent assez restreint, alors que les deux types d'applications précédents, le domaine est ouvert) plus que de connaissances générales et de procédures de gestion de dialogue élaborés.

- ***L'indexation automatique de documents, et la recherche documentaire*** : Plus que de compréhension, il s'agit ici de projection des documents traités sur un certain nombre de concepts clefs plus ou moins fixés à priori, en vu de retrouver les plus pertinents à partir d'une question donnée.
- ***La commande de robot*** en langue naturelle demandera bien entendu une compréhension bien détaillée des ordres donnés, ainsi qu'une bonne représentation du monde dans lequel se trouve le rebot. Des problèmes de planification devront être abordés, et selon le type de tâche envisagée, des procédures de dialogue pourront aussi être nécessaires [(1)].
- ***L'enseignement assisté par ordinateur*** : Là, une représentation fine de l'utilisateur est nécessaire, afin que le programme puisse s'adapter au profil particulier de l'élève qu'il a devant lui. Des procédures de gestion de dialogue et un module pédagogique élaborés seront alors essentiels, alors que les connaissances sur le monde seront souvent assez réduites .

---

(1) ينظر : Mourad, Amine, Glossaire de linguistique computationnelle, P106, مرجع سابق.

1. Le résumé automatique

Un résumé est un texte concis qui rend compte du contenu « essentiel » d'un document. "Comment faire pour qu'un ordinateur puisse calculer le contenu essentiel d'un document et l'exprimer sous la forme d'un nouveau texte cohésif et cohérent ? C'est la problématique du résumé automatique" [(1)].

La notion de ***résumé automatique*** devient un des grands thèmes du Traitement Automatique des Langues. Plutôt que de diffuser les documents entiers, n'est-il pas préférable de diffuser seulement les résumés qui contiendraient les informations vraiment pertinentes ? En effet, il est plus facile de lire quelques lignes ou quelques pages susceptibles d'apporter l'information cherchée que de lire des centaines de pages pour s'apercevoir qu'aucune information nouvelle ne s'y trouve.

Quelle que soit la méthode utilisée pour produire un résumé automatique, on peut considérer que le processus de production automatique d'un résumé est composé des étapes suivantes :

[(1)] Sabah, Gerard, Le TALN & les modèles de représentation des connaissances, P116, مرجع سابق

*''**Interprétation du texte*** : qui transforme la chaîne de caractères en une représentation syntaxique, rhétorique ou conceptuelle, dépendant du cadre théorique utilisé.

Sélection des unités : qui utilise la représentation pour décider quelles sont les unités de «contenu» les plus représentatives du texte que ce soit des mots, des phrases, des sections ou des propositions.

Condensation des unités : qui élimine des propositions répétées et produit des généralisations.

Génération du résumé : qui, à partir du contenu «propositionnel» essentiel, génère un nouveau texte grammatical et respectant les caractéristiques du genre textuel" [(1)].

[(1)] Horacio, Saggion, Génération automatique de résumé par analyse sélective, , مرجع سابق, ص 117.

2. L'histoire du résumé automatique des textes

L'historique ci-dessous est un extrait de la présentation d'Udo Hahn à l'atelier de résumé automatique, Seattle, 2000, (1)

L'annexe A reporte le texte original en anglais

Le résumé automatique a connu ses importants progrès dans le dernier siècle selon la chronologie suivante:

1- Les premières extraction 1955 - 1973

- Luhn, 1955, statistique lexicale d'occurrences
- Edmundson, 1969, indicateurs de position, sélections lexicale, expression indicative, expressions de sélection
- Mathais, 1973, amélioration de cohésion

2- Usage des approches linguistiques 1961 - 1979

3-Utilisations de l'approche de l'intelligence artificielle des années 80

- 1982. DeJong. ROMBIÈRE, en utilisant des manuscrits
- 1985. SUSY, logique et règles de production
- 1988. Armatures de Reiner et de Hahn.TOPIC et réseaux sémantiques
- 1989. Rau et autres. Représentations hybrides

4- La renaissance : les années 90 et les années 2000

- Répétition des techniques statistiques
- approches hybrides .

---

(1) ينظر : Udo, Hahn, l'atelier de résumé automatique, P98, Seattle, USA 2000

3. Les approches de résumé automatique :

Il existe deux approches principales pour produire un résumé automatique :

1. Produire un ***extrait*** du document source
2. Utiliser des systèmes fondés sur la « ***compréhension*** » du texte source pour produire un vrai texte.

3.1 Les méthodes d'extraction de phrases

L'objectif des méthodes d'extraction des phrases est "de repérer dans le texte source les phrases les plus importantes. Le résultat obtenu est alors un extrait du texte source" [(1)] .

3.1.1 Méthodes à base de mots clés

Cette méthode est basée sur le fait que " l'auteur se sert (pour exprimer ses idées principales) de quelques mots-clés qui ont tendance à être récurrents dans le texte " [(2)] . Le résumé automatique est alors produit en recherchant dans le texte source les unités de texte minimales réunissant ses mots-clés. Ce principe est souvent appliqué en différentes variantes présentées dans les sous-sections qui suivent.

---

(1) Monod, Morad, Le résumé automatique, un petit état de l'art, P55, LAB, 2004.

(2) Pardo, T.A.S. L.H.M. Rino, M.G.V. Nunes, Extractive summarization: how to identify the gist of a text, *International.Bibliographie* 63 *Information Technology Symposium - I2TS 2002, Florianópolis-SC, Brazil, pp.245-260, 01-05 October 2002.*

3.1.1.1 Mots-clés prédéfinis

Pour calculer le score de chaque phrase S selon les mots-clés qu'elle contient, on peut calculer le score suivant :

$$\text{Score}_{\text{mot-cle}}(S) = \sum a(w) \times F(w) \qquad w \in S$$

Où $a(w) = \begin{cases} A & \text{si } w \in \text{liste de mot-clés } (A>1) \\ 1 & \text{sinon} \end{cases}$

et $F(w)$ est la fréquence du terme w dans la phrase S

La liste de mots-clés peut être introduite par l'utilisateur (domaine d'intérêt) ou composée des mots-clés établis par l'auteur. L'importance du poids du terme w est donné par $A \times F(w)$, avec A>1.

3.1.1.2 Titres

Étant donné que le titre est l'expression la plus significative et qui résume le mieux un document en quelques mots, on peut dire que la phrase qui ressemble le plus au titre est la plus marquante du document. Par conséquent, on peut attribuer à chaque phrase un poids en fonction de sa ressemblance avec le titre [(1)] .

Dans ce cas on considère les mots du titre du texte comme des mots-clés et on produit le résumé en sélectionnant les phrases qui couvrent certains mots apparaissant dans un titre.

[(1)] ينظر : Ishikawa, K., Ando, S., Okumura, A.: Hybrid Text Summarization Method based on the TF Method and the Lead Method. *Proceedings of the Second NTCIR Workshop Meeting on Evaluation of Chinese & Japanese Text Retrieval and Text Summarization. Tokyo.Japan. p219, March 2001.*

3.1.1.3 Méthode de distribution de termes (DT)

Le pionnier dans le domaine de la génération de résumés est Luhn qui a introduit la méthode de distribution de termes. "L'idée de cette méthode est de considérer comme « importantes » les phrases qui contiennent des mots « importants » du texte. Un mot est considéré important s'il est employé assez fréquemment dans le texte" [(1)].

Le processus d'*interprétation* s'effectue en deux étapes. Dans la première, le texte source est traité pour calculer la fréquence de chaque mot de « contenu » du texte et dans la deuxième les fréquences sont utilisées pour associer un poids à chaque phrase.

Pour le calcul des fréquences on considère généralement les mots qui appartiennent à des classes non fermées de la langue tels que les noms et les verbes. On considère comme un même mot les mots dérivés de la même racine (par exemple, résumé, résumés et résumer).
Une fois la fréquence de chaque mot calculée, une liste triée par fréquence est obtenue, il s'agit de la liste de distribution de termes.
Pour mesurer le poids d'une phrase, on utilise le texte source et la liste de distribution de termes. Plusieurs critères peuvent être appliqués

---

[(1)] Luhn, P. H. The Automatic Creation of Literature Abstracts, IBM Journal, pp. 159-165, April 1958.

pour calculer le poids d'une phrase : ainsi on peut considérer que le poids d'une phrase est la somme des fréquences des mots dans la phrase, la quantité de mots importants qu'elle contient, ou le fait que plusieurs mots co-occurrent dans la même phrase. Ensuite, le processus de *sélection* choisit les phrases les plus « pesantes». La sélection peut être faite en terme d'un pourcentage du texte original, en nombre de phrases ou en nombre de mots. Le processus de *génération* consiste à juxtaposer les unités sélectionnées en ordre d'apparition dans le texte source. La Table 1 résume les étapes de la méthode de distribution de termes. Les avantages de la méthode sont sa robustesse (n'importe quel texte aura un résumé) et sa facilité d'implantation. Les limitations sont toutefois nombreuses. Comme on ne prend pas en considération les relations entre les différents éléments du texte, le résultat risque d'être incohérent et même d'omettre de l'information importante [(1)].

Méthode de distribution de termes	
Étape	Action
Interprétation	Attribuer un poids à chaque phrase en fonction de la distribution des mots dans le document source
Sélection	Les phrases les plus pesantes
Génération	Juxtaposition

Tab. 01: Résumé des étapes de la méthode de distribution de termes

[(1)] ينظر : Horacio, Saggion, Génération automatique de résumé par analyse sélective, , مرجع سابق, ص 118.

3.1.2 Méthode de la position (P)

Cette méthode a été introduite par Edmunson pour compléter la méthode de distribution de termes qu'il a appelé «key method». Elle est utilisée en combinaison avec d'autres méthodes d'attribution de poids pour faire augmenter ou diminuer le poids d'une phrase lors de son interprétation. La méthode de la position considère que "les premières et dernières phrases de chaque paragraphe sont importantes car elles sont considérés comme thématiques" [(1)], c'est-à-dire elles «résument» le contenu du paragraphe, donc ces phrases auront leurs poids augmentés (cette affirmation est appuyée par les expériences de Baxendale [(2)]). La méthode considère aussi des phrases positionnées dans certaines sections conceptuelles importantes, par exemple dans «Introduction» et «Conclusion». Lin and Hovy [(3)] ont récemment développé des algorithmes capables d'identifier les positions thématiques dans un genre textuel quelconque (i.e., article technique,

---

(1) Edmundson, H. P. : New methods in automatic abstracting, *Journal of the Association for Computing Machinery (ACM), vol. 16 N°2* pp. 264-285, April 1969.

(2) ينظر : J. Baxendale, Man-made Index for Technical Litterature - an experiment. IBM J. Res. Dev., P 2(4) :354-361, 1985.

(3) ينظر : Lin and Hovy, Identifying Topics by Position. In Fifth Conference on Applied Natural Language Processing, pages 283-290. Association for Computational Linguistics 1997.

journalistique) à condition d'avoir un corpus de textes accompagnés d'une liste de mots clés et/ou de résumés. Les positions constituent une liste appelée Optimal Position Policy.

Leurs expériences constituent une validation empirique de la méthode de la position.

L'inconvénient de cette méthode est qu'elle dépend de la nature du texte à résumer ainsi que du style de l'auteur [1] .

Méthode de la position	
Étape	Action
Interprétation	Augmenter le poids de la phrase si elle se trouve au début ou à la fin d'un paragraphe ou dans une section considéré importante (i.e., «Introduction», «Conclusion»)
Sélection	Les phrases les plus pesantes
Condensation	
Génération	Juxtaposition
Commentaire	
Utilisée avec d'autres méthodes	

Tab. 02: Les étapes de la méthode de la position

[1] ينظر : Horacio, Saggion, Génération automatique de résumé par analyse sélective, , مرجع سابق, ص 119.

3.1.3 Expressions indicatives

Cette méthode a été introduite par Paice avec l'objectif de produire des résumés indicatifs. Elle est aussi connue avec le nom de «cue method».

Dans la littérature scientifique, on trouve souvent des expressions qui, indépendamment du domaine particulier du texte, font référence à des catégories conceptuelles. Ainsi, si dans l'introduction d'un article on trouve une phrase avec l'expression «L'objectif de cet article est», on peut être presque sûr que ce qui suit dans la phrase est l'information sur les objectifs de l'article et lorsqu'on trouve une phrase qui commence par «Pour en conclure», on a la certitude qu'il s'agit d'une conclusion [(1)] .

L'idée de la méthode des expressions indicatives est de sélectionner dans tout le texte des phrases contenant ces types d'expression.

[(1)] ينظر : Paice, D. The Automatic Generation of Literary Abtracts : An Approach based on Identication of Self-indicating Phrases. In Norman, O., Robertson, S., van Rijsbergen, C., and Williams, P., editors, Information Retrieval Research, Butterworth, London 1981.

Chaque expression indicative pourra avoir un poids associé et, alors, la sélection des phrases sera basée sur leur poids tel que fait Lehmam [1].

Méthode des expressions indicatives	
Étape	Action
Interprétation	Repérer les phrases contenant des expressions indicatives préalablement définies
Sélection	Sélectionner un sous-ensemble des phrases repérées
Condensation	
Génération	Juxtaposition
Commentaire	
Les phrases peuvent avoir un poids	

Tab. 03: Les étapes de la méthode des expressions indicatives

3.1.4 Cohésion lexicale

Les propriétés cohésives du texte ont été prises en considération dans les travaux de Benbrahim and Ahmad [2] et Barzilay and Elhadad [3] pour faire des systèmes d'extraction de phrases. Ils se basent sur l'utilisation des thesaurus pour établir d'abord des liens entre les noms d'un texte et ensuite, des relations entre les phrases (*interprétation*).

[1] ينظر : Lehmam, A. Automatic Summarization on the WEB. A System for Summarizing using Indicating Fragments : RAFI. In Proceedings of Computer-Assisted Information Searching on Internet Conference. RIAO'97, pages 112-122, McGill University, Quebec, Canada, 1997.

[2] ينظر : Benbrahim and Ahmad, Text Summarisation : the Role of Lexical Cohesion Analysis. The New Review of Document & Text Management, pages321-335, 1995.

[3] ينظر : Barzilay and Elhadad, Using Lexical Chains for Text Summarization. In Proceedings of the ACL/EACL'97 Workshop on Intelligent Scalable Text Summarization, pages 10-17, Madrid, Spain 1997.

Dès qu'on met en rapport le mot «ordinateur» de la phrase S1 avec le mot «calculatrice» de la phrase S2, une relation est créée entre les phrases S1 et S2. Ces relations sont la base pour la sélection des phrases.

Dans l'approche de Benbrahim and Ahmad, les phrases sont classifiées en «début de thème», «continuation de thème», «clôture de thème» et «marginale» selon la quantité et type de connections (i.e. en arrière, en avant). Le système peut ainsi sélectionner des phrases qui introduisent, continuent et terminent les thèmes (*sélection*).

Dans celui de Barzilay and Elhadad des **chaînes lexicales** qui représentent chaque «concept» traité dans le texte sont construites et les chaînes les plus «longues» sont retenues pour le résumé (*sélection*). Les particularités de ces méthodes sont présentées dans les Tables 4 et 5.

Salton et al. appliquent les techniques de recherche d'information pour construire des résumés par **extraction des paragraphes** plutôt que des phrases isolées. "Leur méthode se base sur l'identification de segments du texte et sur l'importance des paragraphes"[(1)]. Un segment est un groupe de paragraphes bien connectés, les rapports entre les paragraphes sont calculés en utilisant des mesures de similarité issues de recherche d'information (*interprétation*).

---

(1) Salton, G., Singhal, A., Mitra, M., and Buckley, C., Automatic Text Structuring and summarization. Information Processing & Management, P 33(2) :193-207, 1997.

Les paragraphes avec plus de connections sont sélectionnés pour construire le résumé (*sélection*). Quelques heuristiques sont appliquées pour garantir la couverture de tous les aspects du texte, c'est ici que les segments jouent un rôle majeur dans cette approche. La Table 6 résume les étapes de cette méthode [(1)].

Les auteurs prétendent que tout en sélectionnant des paragraphes les chances d'obtenir des résumés plus cohérents augmentent.

Méthode de cohésion lexicale	
Étape	Action
Interprétation	Construire des relations de cohésion lexicale entre les phrases et classifier les phrases selon «début de thème», «continuation de thème», «clôture de thème» et «marginale»
Sélection	Sélectionner un sous-ensemble de phrases qui introduisent, continuent et terminent les thèmes
Condensation	
Génération	Juxtaposition

Tab. 04: Description de la méthode de cohésion lexicale

Méthode des chaînes lexicales	
Étape	Action
Interprétation	Construire des chaînes lexicales qui lient des phrases contenant des mots liés par des relations de cohésion lexicale
Sélection	Sélectionner un sous-ensemble de chaînes et ensuite un ensemble de phrases de chaque chaîne
Condensation	
Génération	Juxtaposition

Tab. 05: Les étapes de la méthode des chaînes lexicales

[(1)] ينظر : Horacio, Saggion, Génération automatique de résumé par analyse sélective, , مرجع سابق, ص 121.

Méthode d'extraction de paragraphes	
Étape	Action
Interprétation	Calculer les connections entre les paragraphes d'un texte en utilisant des mesures de similarité
Sélection	Sélectionner les paragraphes les plus connectés
Condensation	
Génération	Juxtaposition

Tab. 06: Les étapes de la méthode d'extraction de paragraphes

3.1.5 Classification des éléments

Dans les textes de science et de technique il y a des phrases qui font référence à des catégories conceptuelles telles que : Connaissances Antérieures, Contenu, Méthode et Résultat, on peut également constater que dans les résumés de science et technique des informations relatives à ces catégories sont souvent retenues pour le résumé.

"Cette approche essaie de classifier sémantiquement les phrases d'un texte tout en oubliant le contexte" [(1)] .

[(1)] Larkey, L. S., Ballesteros, L. and Connell M., improving Stemming for Arabia Information Retrieval: Light Stemming and co-occurrence Analysis, In proceeding of the 25th annual International conference on Research and development in information Retrieval (SIGIR 2002), Tampere, Finland, P.89, August 2002.

Méthode de classification sémantique	
Étape	Action
Interprétation	Attribuer à chaque phrase une étiquette sémantique
Sélection	Les phrases ayant certaines étiquettes
Condensation	
Génération	Juxtaposition
Commentaires	
Les phrases peuvent avoir un poids	

Tab. 07: Les étapes de la méthode de classification sémantique.

3.2. Approches hybrides

Les méthodes présentées dans les sections précédentes utilisent des traits (fréquence, position, expression indicative, etc.) qui ne peuvent isolément garantir des résultats optimaux. "On combine souvent ces traits"[(1)] par exemple avec l'équation suivante :

$$Score_{hybride}(S) = a_1 * Score_{DT}(S) + a_2 * Score_{P}(S) + a_3 * Score_{exp.ind}(S) + a_4 * Score_{Titre}(S) \quad (1)$$

Fonction de calcul du score global

Les poids a_i peuvent être fixés arbitrairement ou déterminés de manière expérimentale (par apprentissage par exemple).

[(1)] Horacio, Saggion, Génération automatique de résumé par analyse sélective, , مرجع سابق, ص 127.

Certaines expériences de Edmundson sur un corpus hétérogène de 200 documents ont montré que "si on combine les méthodes expression indicative, titre et position (poids zéro pour la méthode mot-clés), on obtient de meilleurs résultats que si on les combine avec la méthode mot-clés" [(1)].

Dans le cas de textes journalistiques, Strzalkowski et al. ont combiné les méthodes de distribution de termes, du titre et de la position, en considérant la spécificité du texte. Ils ont fait ressortir que "les phrases qui commencent par des nominaux ou contiennent des mots du titre semblent être plus pertinentes que des phrases n'ayant pas ce caractère. De plus, les mots ou les phrases n'apparaissant que dans quelques paragraphes sont plus importants que ceux mentionnés dans tous les paragraphes" [(2)]. Pour garder la cohérence du texte, le résumé est composé d'une sélection de paragraphes pertinents.

---

(1) Edmundson, H. P.: New methods in automatic abstracting, *Journal of the Association for Computing Machinery (ACM), vol. 16 N°2* pp. 264-285, April 1969.

(2) Strzalkowski, T., Wang, J. and Wise, B., Summarization-based Query Expansion in Information Retrieval, *Proceedings of 36th Annual Meeting of the ACL, V. 2, pp. 1258-1264, Montreal 1998.*

3.3. Méthodes de compréhension et génération

A la différence des mesures «quantitatives» attribuant un poids à chaque phrase, les méthodes de compréhension et génération essaient de découvrir comment chaque phrase contribue à l'organisation du texte, quelle est la fonction de chaque phrase dans le tout.

Pour produire un bon résumé il n'est pas suffisant de repérer l'information importante mais aussi de la régénérer.

Dans cette section nous nous concentrons sur deux approches : d'abord l'utilisation des scénarios pour le processus de compréhension, et ensuite la fouille de patrons et la génération.

3.3.1 Scénarios

Les scripts ou scénarios sont des structures de connaissance sur des suites stéréotypées d'actions dans une situation particulière.

Dans cette approche, les processus d'interprétation et de sélection peuvent être considérés ensemble. Le texte original est examiné pour obtenir certains «cues» qui permettent d'activer un script [(1)] .

Pour le processus de génération, on utilise un modèle de résumé qui est associé au sketchy-script qui sont des structures de connaissances pour représenter les évènements les plus intéressants d'une histoire, les structures instanciées lors du traitement sont passées au générateur.

[(1)] ينظر : Edmundson, H. P.: New methods in automatic abstracting, p. 289, مرجع سابق

"Il y a condensation de l'information lors de l'interprétation du texte : certaines informations d'une phrase complexe pourront être directement ignorées. Il y a aussi condensation lors de la génération quand une proposition plus conceptuelle est générée à la place des informations spécifiques sur l'évènement" (1) .

Méthode des scénarios	
Étape	Action
Interprétation et sélection	Identification d'un script et traitement en fonction desAttentes
Condensation	Propositions déduites à partir des propositions du texte
Génération	Modèle associé au script identifié

Tab. 08: Résumé des étapes de la méthode des scénarios

3.3.2 Instanciation de patrons et génération

Une étude menée par Liddy sur les résumés du domaine de la recherche empirique a montré que "certains concepts abstraits à inclure dans un résumé et leurs relations doivent toujours apparaître dans le résumé lorsqu'ils se trouvent dans le texte source" (2).

Si l'on est capable de faire un inventaire des modèles qui sont utilisés pour exprimer les concepts et les relations dans le domaine, on pourra les utiliser pour chercher l'information dans le texte source et ensuite l'exprimer dans un résumé. L'inventaire de modèles doit être fait à partir de l'analyse d'un corpus.

(1) Horacio, Saggion, Génération automatique de résumé par analyse sélective, , مرجع سابق, 134 ص.

(2) Liddy, E. The discourse-level structure of empirical abstracts : an exploratory study. Information processing & management,27(1):55 81, 1991.

On peut identifier un seul processus qui est en charge de l'interprétation et de la sélection de l'information. Le texte est traité pour chercher tous les modèles possibles. Lorsqu'un modèle est trouvé, ses «slots» sont instanciées avec le matériel textuel contre lequel on a fait l'appariement. En cas d'ambiguïté, des heuristiques sont appliquées pour décider l'instanciation d'un slot. Pour le processus de génération, on utilise un ensemble de patrons avec lesquels on décrit les relations entre les concepts. Cette approche permet d'obtenir des résumés indicatifs, car la sélection de matériel de type résultat et conclusion est beaucoup plus difficile [(1)] . Ceci est l'une des seules approches qui essaient de générer un nouveau texte, mais elle reste très limitée tant dans le type de résumé généré que dans le domaine traité. La limitation du domaine est nécessaire afin de pouvoir définir les concepts et les relations entre les concepts. L'implantation du modèle est assez simple une fois qu'on a une description formelle du domaine et l'inventaire de modèles.

Méthode des patrons	
Étape	Action
Interprétation et sélection	Fouille des patrons préalables dans les textes
Condensation	Informations répétées ignorées et extraction des fragments de phrases
Génération	Plusieurs modèles fixes de résumé associé au domaine textuel

Tab. 09: Résumé des étapes de la méthode des patrons

(1) ينظر : Paice, D. The Automatic Generation of Literary Abtracts, p. 169, مرجع سابق

4. Les domaines d'applications des résumés automatiques :

Les applications du résumé automatique de documents textuels sont nombreuses. On peut citer, parmi d'autres, les suivantes [(1)]:

- augmenter la performance des systèmes classiques de recherche et d'extraction d'information (système de résumé couple aux systèmes de question-réponse);
- résumé de nouvelles de presse
- résumé de fils RSS (flots d'informations)
- résumé de dépêches
- résumé de blogs, de pages web et de courriels
- résumé de rapports pour des hommes d'affaires, politiciens, chercheurs, etc.
- résumé de réunions, de comptes-rendus, etc.
- extraits biographiques
- extraction et génération automatique de titres
- résumé de documents spécialisés (médecine, chimie, juridiques, etc.)
- résumé d'opinions etc.
- dans des moteurs de recherche pour présenter les descriptions compressées des résultats de recherche.
- pour chercher dans des langues étrangères et obtiennent un résumé automatiquement traduit du texte automatiquement résumé.

[(1)] ينظر : Minel, J. Luc, Filtrage sémantique, P173, Edition Lavoisier 2001.

5. Avantages et inconvénients :

De l'analyse précédant des méthodes existantes, nous pouvons affirmer qu'il est dur de fixer une approche de résumé automatique.

En effet, nous disposons principalement de méthode d'extraction selon plusieurs approche souvent à caractère quantitatives, et d'autre part, d'une autre approche visant la compréhension du texte source et la reformulation d'un résumé qui essaye être fidele au idées de l'auteur. La table 10 présente les avantages et les inconvénients de chaque méthodes [1]

Approche	Avantages	Inconvénients
Le résumé par extraction de phrase et par couper/coller	Mise en œuvre facile Rapidité de traitement Indépendance des traitements par rapport à la langue Compression paramétrable en modifiant les seuils de sélection	Gestion de la cohérence * structure rhétorique partielle * Anaphores incomplètes Synthèse/Reformulation de situations
Le résumé par reformulation	*Synthèse/Reformulation de situations * Gestion de la cohérence	Mise en œuvre extrêmement difficile

Tab. 10 : Avantages et inconvénients des méthodes de résumé automatique

[1] ينظر : Monod, Morad, Le résumé automatique, مرجع سابق,P111

CONCLUSION

Dans ce chapitre, nous avons passé en revue, de ce qui a trait au domaine du résumé automatique, ses définitions et ses techniques.

Nous avons répertorié toute les méthodes d'extraction, hybrides et les méthodes de compréhension.

Les méthodes de compréhension ou par reformulation semblent très difficile vu qu'elles ne peuvent être généralisées vu ses restriction aux modèles de scenarios et patrons.

Les méthodes d'extraction offrent beaucoup d'avantages. Certaines méthodes semblent offrir de meilleurs résultats que d'autres, cela est dû en grande partie à la nature du texte et au style de l'auteur.

L'approche mixte (hybride) s'apprête mieux utilisée : vu que les coefficients des paramètres peuvent être déterminés expérimentalement, on arrive à obtenir de bons résultats.

Notre méthode est basée essentiellement sur cette approche (l’approche hybride) qui a prouvé son efficacité pour d'autres langues et semble donner des résultats satisfaisants à la langue arabe au niveau informationnel.

الفصل الثالث

TRAITEMENT AUTOMATIQUE DES LANGUES ET GRAMMAIRES

Dans ce chapitre on va présenter l'objectif du traitement automatique de la langue avec ces différents niveaux de traitement, ainsi qu'une liste de projets de résumé automatique et on va finir par une citation des grammaires et les formalismes grammaticaux existants.

1. Présentation du Traitement Automatique des Langues (TAL)

1.1. Préliminaires

"Le traitement automatique des langues (TAL) a pour objectif de traiter des *données linguistiques (textes*) exprimées dans une *langue* dite *"naturelle" "* [(1)].

L'objectif des traitements automatiques des langues est la conception de logiciels ou programmes, capables de *traiter* de façon *automatique* des *données linguistiques*, c'est-à-dire des données exprimées dans une *langue* (dite "naturelle").

[(1)] Lionel Delafosse, Présentation du TAL, page consulté le 29/03/2014http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/cours/parcours/introtal.htm

Ces données linguistiques peuvent être des *textes écrits*, ou bien des *dialogues écrits* ou *oraux*, ou encore des *unités linguistiques* de taille inférieure à ce que l'on appelle habituellement des textes (par exemple des phrases, des énoncés, des groupes de mots ou simplement des mots isolés). Le traitement, dit *automatique* par opposition à un traitement manuel ou instrumental opéré par l'humain, utilise un ordinateur c'est-à-dire une machine conçue pour effectuer des calculs. Un traitement automatique est une suite d'actions ou calculs à faire effectuer par la machine dans un certain ordre chronologique, c'est-à-dire un programme. Traiter un objet linguistique de façon automatique, implique un certain nombre de *contraintes* dans la description même de cet objet : il faut pouvoir arriver à formuler de façon totalement *explicite* et *cohérente* des ensembles de règles caractérisant le fonctionnement du texte [(1)].

Le TAL met en oeuvre *des outils* et des *techniques* de traitement qui sont de trois ordres :

- linguistiques
- formels
- informatiques

---

(1) ينظر : Guenitcha, Cleber, et Manou, Michel, Introduction générale des sciences et techniques d'informations et d'archivage, UNISCO, p173, Tunis, 1987.

L'élaboration de systèmes performants passe donc par le détour de recherches fondamentales, en matière notamment de *compréhension de texte* et de *génération de texte*. Dans ces deux perspectives, le traitement de la langue porte non seulement sur les formes, mais aussi sur le contenu ; il doit mettre en oeuvre des *connaissances linguistiques* très complètes (relevant des niveaux de la *morphologie*, de la *syntaxe*, de la *sémantique* et de la *pragmatique*), ainsi que des connaissances d'univers. De telles recherches revêtent nécessairement un caractère pluridisciplinaire, et doivent associer étroitement *linguistes* et *informaticiens* [(1)].

1.2. Les niveaux de traitement :

La mise en œuvre d'une véritable analyse linguistique nécessite la capacité, pour le système de pouvoir :

- **reconnaître (niveau morpholexical)**
- **structurer (niveau syntaxique)**
- **comprendre (niveau sémantique)**
- **contextualiser (niveau pragmatique)**

1.2.1. Analyse morphologique et lexicographique : types problèmes

- Qu'est ce qu'un mot ? Comment segmenter un texte ?
- Quels types de séparateurs faut-il prendre en compte ?

(1) ينظر : Minel, J. Luc, Filtrage sémantique, P165, Edition Lavoisier 2001.

- Comment traiter les formes inconnues .....

La morphologie :

Les problèmes précités relèvent principalement de la **morphologie** (étude de la structure des mots) et de la **lexicographie** (recensement et classification des formes d'une langue). Ces domaines s'intéressent plus particulièrement à la dimension paradigmatique de la langue (par opposition à sa dimension syntagmatique)

Un exemple :

Pour illustrer les problèmes abordés ici, nous reprenons ici l'exemple de Delafosse [(1)].

Soit la chaîne de caractères :

Jean a mangé des pommes.

La segmentation se fera de la manière suivante :
U1 = *Jean*, **U2** = *a mangé*, **U3** = *des*, **U4** = *pommes*, **U5** = . (point)

Maintenant, on pourra associer toutes sortes d'informations aux **Ui** (**i** = **1**, **2**, **3**, ...), comme par exemple :

U1 = *Jean* :

[(1)] Lionel Delafosse, Présentation du TAL, 13ص ، مرجع سابق

- ***Informations morpho-syntaxiques*** : nom propre, masculin, singulier.
- ***Informations sémantiques*** : animé humain, prénom ...

U2 = *a mangé* :

- ***Forme lemmatisée*** : *manger*
- ***Informations morpho-syntaxiques*** : verbe, passé composé, indicatif, 3ème personne, singulier, constructions : transitif, ...

Idem pour **U3**, **U4**, ...

Remarque : il y a des phénomènes (concernant le choix et le statut des unités) qui sont répertoriés de longue date par les linguistes : qui conduisent à s'interroger sur la notion de mot : élision, amalgames, flexions, dérivations, composition, …

1.2.2. Analyse syntaxique

La syntaxe vise à l'étude des contraintes entre les catégories morpho-syntaxiques devant être prises en compte pour la description des séquences de mots "acceptables" dans une langue donnée. Les contraintes peuvent être de nature sélectionnelle (règles d'accord) ou positionnelle [(1)]. La description des contraintes caractéristiques d'une langue se fait par le biais d'une grammaire.

---

(1) ينظر : Guenitcha, Cleber, et Manou, Michel, Introduction générale des sciences et techniques d'informations et d'archivage, p175, مرجع سابق

La syntaxe permet de résoudre (ou réduire) des ambiguïtés au niveaux inférieurs :

- au niveau phonétique :
  il mange
  île m'en jeu
  yle m'ange ->...
- au niveau lexical :
  il dst parti
  il est dit ...

Elle peut aussi faciliter la formulation des connaissances sémantiques ou pragmatiques. Par exemple, elle impose des contraintes sélectionnelles pour le verbe " manger " : sujet animé, objet comestible.

1.2.3. Analyse sémantique

Au sens littéral, la sémantique vise à l'étude du sens hors contexte.

Le niveau sémantique [1] est encore beaucoup plus complexe à décrire et à formaliser que les niveaux de traitements précédents, par conséquent les réalisations qui sont opérationnelles sont peu nombreuses, et elles concernent des applications très **limitées** où l'analyse sémantique se réduit à un domaine parfaitement circonscrit

[1] ينظر : p176, المرجع نفسه

par contre, on est encore loin de savoir construire en grandeur réelle des analyseurs sémantiques **généraux** qui couvriraient la totalité de la langue et seraient indépendants d'un domaine d'application particulier.

Le traitement sémantique prend comme unité d'analyse la phrase, et conduit à représenter sa partie significative. Ces phrases, dont l'analyseur sémantique doit décrire le sens, se composent d'un certain nombre de **mots** identifiés par l'analyse morphologique, et regroupés en **structures** par l'analyse syntaxique. Ces mots et ces structures constituent autant d'**indices** pour le calcul du sens : on pourrait dire, que " le sens résulte de la double donnée du sens des mots et du sens des relations entre mots " (1).

1.2.4. Analyse pragmatique

La pragmatique vise à l'étude du sens en contexte

L'analyse sémantique de la phrase isolée, traitée hors contexte, ne conduit à représenter que la partie de la signification des mots dans cette phrase, elle n'épuise donc pas ce que l'on peut appeler la **signification complète** d'un texte, telle que l'humain l'appréhende lors d'un processus de compréhension. C'est la raison pour laquelle une analyse **pragmatique** est nécessaire , et qui consiste à trouver

(1) Lionel Delafosse, Présentation du TAL, 19ص ، مرجع سابق

la signification "réelle" des phrases liées aux conditions **situationnelles** et **contextuelles** d'utilisation des mots.

1.2.5. Exemples de difficultés

Difficultés au niveau lexical

- Problème de la correction lexicale (mots erronés) :

  Exemple :

  la disprition du coupable

  disprition-> disparition : facile...

- Les formes erronées ne sont pas toujours évidentes à repérer :

  Exemple :

  la lasse du chien

  lasse -> laisse (liasse?)

- Et la distance lexicographique à la forme correcte n'est pas toujours un critère fiable [(1)] :

  Exemple :

  ils ont été painné par cette nouvelle

  painné -> peinés

[(1)] ينظر : Sabah, Gerard, Le TALN & les modèles de représentation des connaissances, P115, Hachette, 1988.

Difficultés aux niveaux sémantique et pragmatique

ambiguïté du sens des mots

Exemple :

La mousse aux fraises est sur la table de l'avocat.

mousse : dessert ou végétal

fraise : fruit ou outil

table : meuble ou ensemble de données

avocat : juriste ou fruit

- Le problème de la résolution des anaphores

Exemple :

le bus a renversé un passant ...

... je l'ai entendu freiner

... je l'ai entendu crier

le professeur a envoyé l'élève chez le proviseur car ...

... il faisait trop de bruit

... il était excédé

... il l'avait convoqué

- traitement des ellipses [1].

[1] ينظر : المرجع نفسه , P116.

Exemple :

Jacques a évité le singe, le tigre aussi... Jacques a évité le tigre, le singe aussi...

- rattachement des adjectifs

  Exemple :

  Elle a épousé un professeur de droit anglais.

- rattachement des groupes nominaux

  Exemple :

  Il regarde l'homme avec un téléscope.

- rattachement des relatives

  Exemple :

  La maison de la femme que nous avons vue hier...[(1)].

---

(1) ينظر : Pardo, T.A.S. L.H.M. Rino, M.G.V. Nunes, Extractive summarization : how to identify the gist of a text, *International.Bibliographie* 63 *Information Technology Symposium - I2TS 2002, Florianópolis-SC, Brazil, p286, 01-05 October 2002.*

1.3. Les domaines du Traitement Automatique des Langues (TAL)

Le TAL apparaît comme une composante identifiée dans plusieurs grands domaines d'application : **le résumé automatique**, la traductique , la bureautique , la gestion électronique de documents , les Interfaces Homme-Machine, on peut citer aussi les domaines :

- Traitement de la parole
- Aide à la rédaction : correction de textes, génération de textes
- Traduction automatique
- Compréhension automatique des textes
- Génération automatique de textes
- Dialogue homme-machine
- Gestion électronique de l'information et des documents existants (GEIDE)
- Recherche documentaire
- Filtrage, classification d'information
- Traitement du courrier électronique, dépêches d'agences, brevets, ...
- Interfaces en langage naturel : commande vocale (poste de pilotage, tri de bagages, chirurgie, automobile, ...)
- Serveurs vocaux
- Dialogue (interactif) en langage naturel

---

(1) ينظر : Sabah, Gerard, Le TALN & les modèles de représentation des connaissances, P117, Hachette, 1988.

2. Une liste de projets de résumé automatique

Depuis trois décennies, une panoplie d'outils et de projets de résumé automatique ont vu le jour, un répertoire de ces projets fait par un site scientifique est disponible à : http://www.ics.mq.edu.au/~swan/summarization/projects_full.htm[(1)]

voir l'annexe B où on peut trouver le texte original en anglais

On peut distinguer deux grands axes de projets :

- Des projets dans le domaine de recherche
- Et d'autres projets à caractère commercial

2.1 Projets et outils du domaine de recherche

2.1.1 Extraction de phrase

- **Université de Surrey : Applet de sommet**

Ce système de résumé automatique extraire les phrases en utilisant la cohésion lexicale.

- **Institut de technologie royal (Suède) : SweSum**

SweSum extrait des phrases pour produire un type résumé d'extrait. On le lie étroitement pour travailler sous ISI. Des résumés sont créés des textes suédois ou anglais dans les domaines de presse ou d'université.

---

(1) شوهد الموقع يوم 15 أوت 2013 على 16سا35

http://www.ics.mq.edu.au/~swan/summarization/projects_full.htm

Des phrases sont extraites par l'arrangement de phrases selon le dispositifs pesés de niveau de mot et ont été formées sur un corpus suédois étiqueté de presse. L'outil de résumé automatique peut être accroché jusqu'aux résultats du moteur de recherche.

- **Université d'Ottawa : Le projet de résumé automatique des textes [(1)]**

Pas beaucoup est disponible au sujet de ce projet de recherche excepté leur proposition de projet. Dans ce projet, ils ont proposé d'employer des techniques d'étude de machine pour identifier des mots-clés. L'identification de mot-clé peut alors être employée pour choisir des phrases pour l'extraction. Elles ont projeté employer les statistiques de niveau extérieures telles que les featurs linguistiques de niveau d'analyse et de surface de fréquence tels que la position de phrase.

- **Université de Colombie : FociSum**

Le système de FociSum adopte une approche de questions et réponses au résumé automatique. Des phrases qui répondent aux questions principales concernant des participants, des organismes et d'autres wh-questions sont extraites. Le résultat est une concaténation des fragments de phrase et les clauses ont trouvé dans le document original. Le système utilise d'abord un extracteur appelé d'entité pour

(1) ينظر : http://www.site.uottawa.ca/tanka/ts.html 2013 أوت 23 شوهد يوم

trouver les centres du document. Un générateur de question est utilisé pour suggérer des rapports entre ces entités.

Le document est analysé pour trouver des réponses de candidat pour ces dernières questions sur la base de la forme syntactique. Des fragments et les clauses de phrase sont retirés des phrases choisies.

- **Université de la Californie de sud : ISI Summarist [(1)]**

Summarist est produit des résumés des documents de Web. Il a été accroché jusqu'au système de traduction de Systran pour fournir un outil gisting pour des articles de nouvelles dans n'importe quelle langue. Summarist identifie d'abord les matières principales du document en utilisant des techniques statistiques sur des dispositifs tels que la position, et le mot compte. La recherche courante est en cours d'employer les expressions de sélection et la structure de discours. Ces concepts doivent être interprétés de sorte que d'une chaîne des phrases lexically reliées, la phrase avec le concept le plus général soit choisie et extraite. Le travail suivant prendra ces phrases extraites pour construire un résumé plus logique.

[(1)] ينظر : https://www.isi.edu/natural-language/people/hovy/papers/98hovylin-summarist.pdf 2013 أوت 26 شوهد يوم

2.1.2 Arrangement profond

- **Le CHEVALET d'université de Sheffield [(1)]**

Ce projet produit des résumés dans le domaine de nouvelles. Il emploie MUC pour extraire les concepts principaux du texte qui alors vraisemblablement est employé aux résumés produits. Malheureusement, pas beaucoup d'information est disponible sur le site Web officiel concernant l'architecture de système.

- **Université de Colombie : SUMMONS**

SUMMONS est un système de résumé de multi-document dans le domaine de nouvelles. Elle commence par les résultats d'un processus d'extraction de l'information de MUC-modèle, à savoir un calibre avec instanciation des fentes de sémantique prédéfinie. De ceci, il peut produire d'un résumé en employant une étape sophistiquée de génération de langage naturel. Cette étape a été précédemment développée sous d'autres projets et inclut un sous étape content de choix, un sous étape de planification de phrase et une étape extérieure de génération. Puisque les calibres ont la sémantique bien définie, le type de résumé a produit les approches qui des résumés humains. C'est eux sont plus logique et lisible. Cependant, cette approche est détail de domaine, se fondant sur la disposition des articles de nouvelles pour

(1) ينظر : http://nlp.shef.ac.uk/trestle/ 2013 شوهد يوم 30 أوت

l'étape d'extraction de l'information.

2.1.3 L'approche hybride (ces approches combinent les techniques d'extraction avec des techniques plus traditionnelles de NLP)

- **Université de Colombie : MultiGen [(1)]**

MultiGen est un système de multi-document dans le domaine de nouvelles. Il extrait les fragments de phrase qui représentent les informations principales dans l'ensemble de documents relatifs. Ceci est fait en employant l'étude de machine pour grouper ensemble divisent en paragraphes les morceaux classés du texte dans des faisceaux des matières relatives. Des phrases de ces faisceaux sont analysées et les arbres résultants sont fusionnés ensemble pour former, établissant les représentations logiques des propositions contenant les concepts généralement de occurrence. Cette représentation logique est transformée en phrase en utilisant la grammaire de FUF/SURGE. L'assortiment des concepts emploie la connaissance linguistique telle que des classes de refouler, de partie du discours, de synonymity et de verbe. Le fusionnement des arbres se sert des règles identifiées de paraphrase.

[(1)] ينظر : http://www.cs.columbia.edu/~hjing/sumDemo/2013 شوهد يوم 03 سبتمبر

- **Copier et coller**

Le système de copier et coller est un résumeur de document simple qui est indépendant de domaine. Il est conçu pour prendre les résultats d'un résumeur d'extraction de phrase et pour extraire les concepts principaux à partir de ces phrases. Ces concepts sont alors combinés pour former de nouvelles phrases. Le système, les copies ainsi la forme extérieure de ces concepts principaux et les colle dans les nouvelles phrases [(1)].

Ceci est fait en réduisant d'abord la phrase enlevant n'importe quelle information étrangère. Cette étape emploie des probabilités apprises d'un corpus de formation, et des liens lexicologiques.

Les phrases réduites sont fusionnés en employant des règles telles qu'ajouter des informations supplémentaires sur des haut-parleurs, ajouter des conjonctives et fusionner les éléments communs.

(1) ينظر : http://www.cs.columbia.edu/~hjing/sumDemo/2013 شوهد يوم 03 سبتمبر

2.2 Projets et outils du domaine commercial

Ces projets sont la plupart du temps des projets basés sur les méthodes d'extraction

Datahammer par Glucose [(1)]

Datahammer est un produit à conçu pour récapituler les textes en ligne et fonctionne dans la conjonction le web browser de l'utilisateur. Il extrait des phrases en employant un algorithme appelé « le règlage d'arbre de Microword » qu'ils ont créé. Une version de démo est fournie par leur site Web.

- **Analyste des textes par Megaputer**

L'analyste des textes extrait des phrases des documents sur l'ordinateur de l'utilisateur. Le site Web officiel de l'analyste des textes décrit le processus de résumé automatique qu'ils emploient. Un réseau sémantique est construit du document de source en utilisant un réseau neurologique. Ils déclarent que la construction du réseau sémantique ne dépend pas de la connaissance antérieure de détail de domaine. Une représentation graphique des concepts et des rapports dans le document de source est montrée à l'utilisateur pour le choix. Des phrases avec des concepts et des rapports assortis sont extraites.

---

(1) ينظر : http://www.mactech.com/content/md1-glucose-releases-data-hammer

شوهد يوم 17 سبتمبر 2013

- **Produits de résumé automatique d'IBM**

IBM Japon incorpore des outils de résumé automatique dans deux de ses produits : Roi d'Internet de la traduction (japonaise) et du Word de lotus pro (version japonaise). Le type de résumés produits sont des extraits de phrase, choisis en utilisant des relations et la position rhétoriques dans le document. L'extraction est faite statistiquement et peut utiliser des dispositifs de détail de genre. IBM a également un outil appelé analyse de texte, qui a en tant qu'un de ses composants, un outil de résumé automatique. Cet outil fait partie du produit intelligent de mineur des textes. Il produit des résumés en extrayant des phrases. Comme avec la plupart des produits commerciaux, ceci est fait en rangeant les phrases par une mesure d'importance et puis en choisissant l'arrangement le plus élevé. Le rang est réalisé par des dispositifs de niveau de mot et l'utilisateur peut choisir la longueur d'extrait.

- **Websumm par Mitre**

WebSumm de la mitre exécute les documents simples ou multiples d'excédent d'extraction de phrase en même temps qu'un moteur de recherche. Le résumé résultant est un extrait des phrases basées sur une question d'utilisateurs [1].

[1] ينظر: مرجع سابق , شوهد الموقع يوم 15 أوت 2013 على 16سا35

http://www.ics.mq.edu.au/~swan/summarization/projects_full.htm

Ceci est fait en représentant les documents de source comme réseau des phrases.

En utilisant les termes de question pour choisir les noeuds qui sont connexes, les phrases sont extraites. L'outil de résumé automatique peut manipuler les phrases semblables et contrastantes à travers les documents multiples.

- **Moteur de recherche d'InText**

Ce kit de développement de moteur de recherche est employé pour faire un moteur de recherche pour un site Web particulier. De cette manière, on le cense permettre la création dynamique des résumés des documents dans le site Web une fois demandé par des utilisateurs. Il n'y a aucune information quant à la façon dont ceci est fait sur leur site Web.

- **InText par Island Soft**

InText [1] extrait les phrases principales en employant des mots clés, bien que l'IS-IS exact de technique non mentionné sur leur site Web. Leur description mentionne que l'utilisateur peut choisir une de plusieurs techniques d'extraction. InText est un produit que l'utilisateur installe et emploie sur des documents résidant déjà sur l'ordinateur.

[1] ينظر : http://www.islandsoft.com/2013 شوهد يوم 15 سبتمبر

- **British Telecom (ProSum) ou NetSum [1]**

Il est difficile d'arriver à l'emplacement officiel, ainsi ce paragraphe est basé sur des descriptions d'occasion. British Telecom a produit un outil de résumé automatique pour les textes en différé et en ligne qui fonctionne par la sélection des phrases principales et de les extraire comme résumé. Cependant, depuis ProSum et NetSum sont les produits disponibles dans le commerce, les mécanismes internes derrière le processus d'extraction. Alternativement, au lieu d'extraire les phrases, l'outil les accentue dans le document original. L'utilisateur peut changer la longueur du résumé produit. Selon des chercheurs à l'université d'Ottawa, ProSum fonctionne mieux avec les documents effectifs d'un thème simple. Ceux-ci incluent des genres tels que des nouvelles, des articles article et des journaux techniques.

- **inXight (LinguistX)**

le serveur du résumé des inXight est une application qui cré les résumés par extraction en différé. Les utilisateurs voient les résumés quand ils déplacent leur souris au-dessus d'un lien hypertexte à un document qui a été précédemment récapitulé. Il emploie des techniques statistiques d'extraction basées sur des dispositifs tels que la position de phrase, la longueur de phrase et les mots-clés.

---

[1] ينظر : https://www.lesechos.fr/british-telecom.htm, 2013 شوهد الموقع يوم 28 أوت

L'application permet à l'utilisateur d'indiquer la longueur et la prépondérance de certains mots-clés. Elle fournit également la capacité pour davantage de formation sur les documents structurés d'autres genres.

- **Logiciel de Tetranet (extracteur)**

L'extracteur sélectionne des mots-clés d'un document. À partir du site, l'extracteur a produit la liste suivante :

- Langage naturel
- Les sciences de communication
- Technologie interactive intelligente
- CMIS
- Groupe de technologie
- Université de Macquarie
- Honore l'étudiant.

L'extracteur trouve simplement les expressions principales et les phrases qui les emploient. L'algorithme d'extracteur emploie un ensemble de paramètres (tels que la longueur de tige) qui sont accordés par un algorithme génétique (GenEx). Les expressions extraites sont alors assorties avec leurs occurrences dans le document et la phrase correspondante est extraite [1].

---

[1] ينظر: مرجع سابق , شوهد الموقع يوم 15 أوت 2013 على 16سا35

http://www.ics.mq.edu.au/~swan/summarization/projects_full.htm

Il y a un simple refouler et l'analyse morphologique des mots clés extraits pour marquer des expressions de non-nom plus bas, cependant, là n'est aucune détection de synonyme. La liste de phrases extraites est filtre basé sur l'heuristique concernant la présentation.

- **Projet d'IBM Japon de résumé automatique** [(1)]

« L'importance d'une phrase est déterminée par certains indices extérieurs tels que le nombre de mots-clés importants, le type de phrase (fait, conjecture, opinion, etc.), des relations rhétoriques dans le contexte, et l'endroit dans lequel une phrase existe dans un document ».

On peut aussi citer les projets commerciaux de résumé automatique suivants :

- **Copernic Summarizer**
- **Sinope Summarizer**
- **Pertinence Summarizer**
- **Apple (Project)**
- **Microsoft (MS Word Summarization automatique)**
- **Produits discontinués de Summarisation**
- **Webcompass par Quarterdeck**
- **Le résumeur automatique de Sakhr**

(1) ينظر: المرجع نفسه

3. Grammaires et formalismes grammaticaux :

3.1 Grammaires classiques

La plupart des théories linguistiques sont d'accord sur le fait qu'il existe une organisation entre les mots de la phrase et qu'il est possible de la présenter sous forme de structure. Les deux principales structures de représentation sont la structure syntagmatique et la structure de dépendance. L'idée d'une grammaire de dépendance (DG dorénavant, de l'anglais dependency Grammar) est que les mots dans la phrase dépendent les uns des autres ; et l'idée d'une grammaire syntagmatique (PSG dorénavant, de l'anglais *Phrase Structure Grammar*) est que les mots se rassemblent en constituants et que chaque constituant doit avoir une tête.

L'usage de dépendances dans la description des langues naturelles est beaucoup plus ancien que celui de la structure syntagmatique. Il remonte au moins à Panini en Inde, il y a quelques siècles, qui distingue trois types de dépendance : sémantique, syntaxique et morphologique.

" Les grammairiens arabes du 8ème siècle, comme Sibawayh, distinguent gouverneur et gouverné en syntaxe pour formuler des règles d'ordre des mots et de régime" (1).

---

(1) OWENS Jonathan, The foundations of Grammar: An Introduction to Medieval Arabic Grammatical Theory, éd. J. Benjamins (SiHoLS, P.45), Amsterdam, (1988).

Les grammairiens latins ont aussi reconnus les structures de dépendance et de détermination. Au 19ème siècle, les grammaires scolaires de l'anglais ont enseigné l'analyse de la phrase sous forme de diagramme basé sur la dépendance.

Dans les années 60, Hays développe la première grammaire de dépendance. Les DG se sont ensuite développées partout en Europe et aux Etats Unies; citons entre autres *Case Grammar*, développée principalement par Anderson, *Word Grammar*, développée par Hudson, *Dependency unification Grammar*, de Hellwig, *Functional-Generative Description*, de Sgall, *Lexicase*, de Starosta, ainsi que le Modèle Sens-Texte de Mel'čuk.

La représentation syntaxique d'une phrase par une structure syntagmatique, quant à elle, ne s'est développée qu'au 20ème siècle avec Bloomfield (1) et les travaux des distributionnalistes. Les années 60 et 70 témoignent de la montée de la grammaire générative et transformationnelle de Chomsky (2) (3).

---

(1) ينظر : Bloomfield, Leonard. .Language. Revised from 1914 edition. New York: Holt. 564 pages. (Fischer-Jorgensen 1975.) Location: Dallas SIL Library 410 B655l. Interest level: academic, 1933

(2) ينظر : Chomsky Noam , Syntactic Structures, P.65-76 Mouton France 1957

(3) ينظر : Chomsky N.,Aspects of the theory of syntax. P.17-33 MIT Press USA 1965

Les années 80 voient l'introduction de l'usage de la dépendance syntaxique dans des grammaires syntagmatiques telles LFG, GPSG, HPSG et TAG sous la forme plus ou moins variée de cadres de sous-catégorisation ou de régimes.

Une structure de dépendance est constituée de noeuds représentant les mots avec leurs catégories grammaticales, tandis que les arcs orientés précisent la fonction du dépendant syntaxique par rapport à son gouverneur. Les arcs ne sont pas linéairement ordonnés.

La structure syntaxique comprend uniquement l'information relative au lien entre la tête et ses dépendants. Le rassemblement en constituants n'est pas explicité par les noeuds non terminaux [(1)].

A l'opposé, les règles syntagmatiques donnent à la fois l'information sur la dépendance (la tête syntagmatique constitue le gouverneur syntaxique des autres membres du syntagme) et l'information sur le regroupement des éléments en constituants ainsi que leur ordre linéaire.

Une structure syntagmatique pure encodant uniquement le regroupement en constituants est impossible.

---

(1) ينظر : OWENS Jonathan, The foundations of Grammar: An Introduction to Medieval Arabic Grammatical Theory, p.50, مرجع سابق

Il est vrai que les règles de réécriture constituent un outil important grâce à leurs propriétés mathématiques et computationnelles bien connues ainsi que grâce à la disponibilité des différents algorithmes, mais elles présentent de grands inconvénients à cause de leur forte dépendance à la langue en question et de la particularité des constructions. Le résultat des règles de réécriture peut être aussi représenté sous la forme d'une structure arborescente dont les noeuds terminaux représentent les mots de la phrase dans l'ordre [1].

Cette grammaire assure d'une part la bonne formation de la structure et d'autre part indique l'ordre des mots. Mais, dans la pratique, les règles font face à beaucoup d'exceptions, dont en voici quelques unes :

- Une phrase canonique ne contient pas nécessairement un verbe fini comme tête. Ces phrases n'ont pas de verbe : « Heureusement que tout n'est pas parfait », « Voilà un autre exemple ».
- Un verbe fini n'a pas nécessairement un sujet, comme c'est le cas pour l'impératif «partez». Il y a aussi les langues pro-drop où le sujet n'est pas systématiquement réalisé.
- Les actants du verbe ne sont pas nécessairement des groupes nominaux, comme dans les exemples suivants : « *Voler* est interdit », « Je veux *partir* », « *Que tu partes* m'inquiètes ».

[1] ينظر : المرجع نفسه , p.52.

- Quant à la linéarisation, le sujet ne se place pas nécessairement avant le verbe et le complément après, comme dans la phrase : « Le livre que lit *Marie* à sa soeur ».

Conformément à une grammaire syntagmatique, on a ici un constituant discontinu ; la description d'une langue à ordre libre avec de telles règles s'avère d'une grande difficulté.

Le brouillage et les mouvements ont fait couler beaucoup d'encre dans les divers cadres; cependant, il s'agit d'un problème de formalisme avant d'être un phénomène linguistique.

Il est vrai qu'il y a toujours la possibilité d'ajouter d'autres règles de réécriture afin de gérer les exceptions ; mais cela ne va pas sans une complexification des catégories.

Résumons : " un arbre de dépendance a pour seule tâche d'encoder et d'étiqueter le lien entre les mots de la phrase. Par contre, un arbre syntagmatique doit encoder toutes les informations pertinentes de la phrase, c'est pourquoi l'arbre syntagmatique est par principe ordonné et projectif " [(1)].

[(1)] Dina El Kassas, thèse de doctorat 'Une étude contrastive de l'arabe et du français dans une perspective de génération multilingue', P. 169, Université Paris 7, 2005

3.2 Grammaires d'unifications :

On assiste ces toutes dernières années à l'abandon progressif de la notion de règle pour décrire les phénomènes syntaxiques. En syntaxe, on assiste à l'élaboration par contraintes, il s'agit de proposer des contraintes hiérarchisées dont l'interaction limite les représentations de sorties possibles.

Les formalismes dits d'unification tendent à intégrer au long de l'analyse les différentes facettes de l'analyse linguistique. Ces formalismes ne favorisent guère, voire ne permettent pas, la séparation en différents modules explicites, destinés à interagir.

Dans ce qui suit, nous allons donner une brève description de quelques formalismes d'unification

3.2.1 La grammaire lexicale fonctionnelle (Lexical Functional Grammar, LFG) :

La grammaire lexicale fonctionnelle propose un modèle se voulant valide et universel basé sur le concept des fonctions grammaticales. C'est une grammaire d'unification mais dont les principes sont radicalement différents. L'information sur la construction syntaxique est directement codée dans le lexique où elle est décrite en terme de fonction (sujet, verbe, etc.) et non en terme de catégorie. " Il existe des règles de dérivation ne fournissant qu'un cadre dans lequel va pouvoir s'opérer l'analyse guidée et filtrée par des informations lexicales " [(1)].

[(1)] Anne Abeillé, Cours DEA, www.llf.cnrs.fr/Gens/Abeille/hpsg-04.doc , 2014, ص ، 23Consulté le 06/04/2014

3.2.2 La grammaire syntagmatique généralisée (Generalized Phrase Structure Grammar, GPSG) :

La grammaire syntagmatique généralisée, source d'inspiration de nombreux systèmes de TAL et d'autres théories linguistiques, telles que HPSG, est une théorie linguistique issue des travaux de Gerald Gazdar à la fin des années soixante-dix. Cette grammaire s'est rapprochée du concept d'algorithme de caractérisation des langues naturelles.

La théorie a évolué d'un modèle dynamique aux composantes strictement ordonnées vers un modèle déclaratif, défini comme un ensemble de conditions de bonne formation dont l'ordre d'application n'est contraint que par des conditions d'ordre intrinsèque.

" Bien que GPSG ait été motivée au départ par un regain de préoccupation pour le lexique, et surtout pour la composante sémantique, la syntaxe y occupe une place prépondérante " [(1)] .

---

(1) Mammeri Mahmoud Fawzi, Une approche pour l'analyse syntaxique de l'arabiyya basé sur la TNK et sur la grammaire syntaxique guidée par les têtes, P. 88, Mémoire de magister CRSTDLA Alger, 2003

A la base de la syntaxe de GPSG, on trouve d'abord les règles de dominance immédiate (en abrégé, règle de DI). Ces règles expriment la relation de dominance qui lie un syntagme à ses constituants immédiats (CI), indépendamment de l'ordre linéaire de ces derniers. Les règles d'ordre linéaire (en abrégé, OL) sont dissociées des relations de dominance hiérarchique et appartiennent à une sous-composante grammaticale distincte. La dissociation entre les relations de dominance et d'ordre linéaire est appelée format DI/OL.

GPSG recoure à des techniques contemporaines, telles que l'unification des structures de traits.

Les éléments essentiels de la théorie des traits syntaxiques de GPSG sont [1]:

- D'une part, la conception des catégories syntaxiques comme ensembles de trait.
- Et d'autre part, les principes basés sur l'unification qui contraignent l'instanciation des traits syntaxiques sur les catégories partiellement spécifiées qui apparaissent dans les règles de DI.

[1] ينظر : المرجع نفسه , p.90

- Enfin, les métarègles sont des mécanismes qui s'appliquent aux règles de la grammaire pour produire de nouvelles règles. Formellement, une métarègle comporte deux parties, un modèle et une cible. Elle s'applique à toute règle de DI lexicale qui satisfait la description du modèle, et définit à partir d'elle une règle modifiée d'après la description donnée par la cible. Les métarègles sont donc des fonctions partielles de l'ensemble des règles de DI lexicales vers lui-même. Elles permettent d'exprimer des généralisations d'un niveau supérieur à celles qu'expriment les règles elles-mêmes.

Du point de vue du lexique, la grammaire GPSG se limite à certains aspects des interfaces de cette composante avec la syntaxe et avec la sémantique. La représentation des entrées lexicales se fait à l'aide du quadruplets dont le premier élément contient l'information phonologique, le deuxième l'ensemble des traits syntaxiques, le troisième l'information morphologique, et la quatrième la forme sémantique [1]. Chaque trait syntaxique est une paire de la forme <nom de trait, valeur de trait> ou, en d'autres termes, <attribut, valeur>.

[1] ينظر : المرجع نفسه , p.91

En ce qui concerne les cadres de sous-catégorisation, GPSG les définit par des règles syntaxiques appelées «règles de DI lexicales» et le puissant mécanisme de généralisation que sont les métarègles syntaxiques.

3.2.3 Les grammaires syntagmatiques guidées par les têtes (Head-driven Phrase Structure Grammar, HPSG) :

Les grammaires syntagmatiques guidées par les têtes sont un formalisme grammatical récent, actuellement utilisé en linguistique informatique dans différents grands projets. Elles font parties des grammaires dites d'unification, s'appuyant sur un riche langage logique de description et une unification étendue par la notion de structures de traits typées.

Ce modèle, le plus utilisé en Europe et aux Etats Unies dans la majorité des applications informatiques, " insiste sur une formalisation mathématique rigoureuse et qui élimine complètement les transformations au profit des règles syntagmatiques dont la puissance a été augmenté " [(1)].

---

(1) Anne Abeillé, Cours DEA, 47 ص ،مرجع سابق

3.2.4 Les grammaires d'arbres adjoints (Tree Adjoining Grammar, TAG) :

Les grammaires d'arbres adjoints sont un formalisme arborescent qui manipule des structures élémentaires plus étendues que les grammaires basées sur des règles de réécriture hors contexte. Une grammaire d'arbres adjoints est un ensemble d'arbres élémentaires. Ces arbres peuvent être de deux types : les arbres initiaux et les arbres auxiliaires. Ces arbres sont de profondeur quelconque et ils ont à leurs feuilles des terminaux ou des nœuds à substituer. Les arbres auxiliaires possèdent une feuille (appelée pied) portant un nœud non terminal et de même catégorie que leur racine. Le langage TAG est produit par la combinaison de ces arbres au moyen de deux opérations : la substitution et l'adjonction. Les arbres obtenus sont appelés les arbres dérivés.

" L'opération de substitution insère un arbre initial (ou dérivé d'un arbre initial) à la frontière d'un arbre élémentaire ou dérivé.

L'opération d'adjonction insère un arbre auxiliaire (ou dérivé d'un arbre auxiliaire) à un nœud quelconque de même catégorie dans un arbre élémentaire ou dérivé " [(1)].

---

(1) Mammeri Mahmoud Fawzi, Une approche pour l'analyse syntaxique de l'arabiyya basé sur la TNK et sur la grammaire syntaxique guidée par les têtes, مرجع سابق ص115

Conclusion

L’aspect théorique que nous venons de donner et les définitions des différents formalismes grammaticaux ainsi que l’état de l’art des projets de résumé automatique, peuvent nous aider a bien aboutir à notre but de réaliser un resumeur des texte arabes.

Nous présentrons dans le chapitre suivant notre système de résumé automatique des textes arabe que nous avons conçus.

الفصل الرابع

نهج مدمج لتلخيص النصوص العربية

خصص هذا الفصل لوصف نهجنا المقترح لتلخيص النصوص العربية – المعتمد أساسا على أساليب استخراج الجمل.

نبدأ أولا بملخص موجز لما تمت دراسته في الفصول السابقة وهي: أساليب معالجة اللغة العربية وطرق التلخيص ثم سنقدم تصميما لنهجنا وسنقوم بالتفصيل والشرح لكل من عناصر ووحدات هذا التصميم.

1. تلخيص نص

التلخيص لغة:

ورد على لسان العرب [1] بأن معناه التبيين و الشرح ويقال لخصت الشيء لحصته بالخاء والحاء. والتلخيص التقريب والاختصار. يقال لخصت القول آي اقتصرت فيه واختصرت منه

التلخيص اصطلاحا:

- هو تقرير موجز لتقرير مطول.
- عملية اختصار موضوعية للنص ، تحافظ على أبرز الأفكار ، و بأسلوب الشخص.
- إعادة صياغة النص في عدد قليل من الكلمات مع المحافظة على الأفكار ، وإعادة صياغتها.

[1] ابن منظور، محمد بن مكرم، لسان العرب، الناشر: دار صادر – بيروت، 2010، ص 113

– مهارة لغوية تقوم على استخلاص الأفكار ، وإعادة صياغتها بإيجاز وبدقة وترابط لغوي، وسلامة الأفكار دون إضافة أو تحريف أو نقد أو تعليق.

– عملية فكرية تتضمن القدرة على إيجاد لب الموضوع ، استخراج الأفكار الرئيسية فيه، أي أن التلخيص هو إعادة بناء للموضوع بهيكلية جديدة لها وقواعد جديدة.

التلخيص ضرورة حيوية يحتاجها الإنسان سعيا إلى توفير الجهد والوقت والاقتصاد في المـال [1] ، ففي ظل ثورة المعلومات الهائلة التي يعيشها العالم في وقتنا المعاصر والمسؤوليات الجسام للمسئولين في كثير من المواقع لا يجد الباحث الوقت الكافي لمتابعة كل ما يدور حوله من تطور معرفي وتقدم علمي وأحداث متسارعة. إذ نجده كثيرا ما يكتفي بمتابعة أهم الأنباء في نشرات الأخبار والعناوين الرئيسة في الصحف والمجلات حتى إذا لقي الباحث ما يلفت انتباهه بحث عن بقية المقال كي يستزيد بما فيه. وقد عرف العرب التلخيص ونقلوا لنا الكثير من التلخيصات للعديد من مصنفات الأدب والعلوم. كما نبهوا فيه وتفننوا في طرقه وتقنياته وتركوا بعضا من التعريفات له التي تمكننا من الوقوف على معناه وآلياته، فعرفه ابن المسعود بأنه " إيراد الأصول وحذف الفضول "[2] وعرفه ابن عبد البر بأنه : " الاكتفاء بالدرر والفوائد " [3]

---

[1] ينظر : دينا عويس ، http://mawdoo3.com 2014 شوهد يوم 02 ماي 2016 على 09 سا 05

[2] ابن منظور، محمد بن مكرم، لسان العرب، دار صادر، مرجع سابق ، ص 115

[3] المرجع نفسه، ص 245

ولا تختلف تعريفات العرب للتلخيص عن تعريفات الأعاجم من المحدثين، فيرد تعريفه في الاتحاد الأمريكي للمكتبات بأنه ":تحويل نص معين إلى شكل مختصر يتم خلاله التكثيف والحذف مع الحفاظ على المعنى العام، وعلى طريقة العرض المستخدمة في الأصل ".

من التعريفات السابقة يمكننا أن نعرف أن التلخيص هو " اقتضاب للنص وان صنعته تقوم على الاكتفاء بالعناصر الرئيسة لمعاني النص وحذف الزائد منها مع الحفاظ على المعنى العام للنص وطريقة عرضه " [1].

كما يعرف التلخيص أنه إستخلاص الأفكار المهمة أو النقاط الرئيسة لنصٍ ما وإستخراج العناوين المهمة الواردة فيه ، وهو التعبير بالشرح والكلمات البسيطة عن أفكار الموضوع الأساسية دون إحداث خللٍ في مضمون النص أو تغيير الصياغة ، ويكون تلخيص النصوص بصفحةٍ واحدةٍ من التلخيص لكل 10 صفحات من النص الأصلي (تقريباً).

[1] موقع http://faculty.ksu.edu.sa/montaser/DocLib8 شوهد يوم 2015/12/22 على الساعة 13:28

كما يمكن أن يكون التلخيص إبرازا للنص الأصلي في عدد قليل من الكلمات مع الحفاظ على صلب النص المكتوب دون إخلال بالمضمون أو إبهام في الصياغة، فنحن حين نلخص عبارة فإننا نستخلص منها الفكرة الأساسية التي تتضمنها.

إن تلخيص النص هو إختزاله بمعنى أن نضغط ما جاء في النص الأصلي من كلمات وأفكار بعدد كلمات لا يزيد عن عدد كلمات النص الأصلي ، مع مراعاة الحفاظ على نفس الأفكار الموجودة في النص الأصلي دون تغيير أو حذف أيٍ منها أو من مضمونها.

مهارة التلخيص تتضمن وتفترض اتقان مهارات أخرى مهمة جدا: هي مهارات الفهم والاستيعاب (للكلام والنصوص) أي إتقان مهارة الإصغاء والاستماع والقراءة النشطة، وكذلك مهارة التعبير (الكلامي والكتابي) التي بواسطتها سنقوم بصياغة تلخيصنا. وكلها مهارات لا تكتسب الا بالممارسة والتمرن والصبر والمثابرة [1].

لذا فالتلخيص أداة فعالة ومضمونة لمراقبة الفهم والاستيعاب. لما يقال ويسمع ولما يكتب ويقرأ .

[1] ينظر: صيني، محمود إسماعيل، بنوك المصطلحات الآلية، مرجع سابق، ص201

فالتلخيص تمرين على إعادة صياغة المقروء أو المسموع بشكل موجز ومركز في أسلوب شخصي سليم وواضح، مع المحافظة على الأفكار الرئيسية ويحتاج الطالب والطالبة في دراستهما إلى تلخيص المحاضرات والدروس، كما يحتاجان إليه عند الإجابة عن أسئلة الاختبارات بجمع أفكار الإجابة في سطور قليلة.

2. خصائص التلخيص:

- عملية فكرية – مهارة لغوية راقية.
- صورة مصغرة لأصل مطول – ضرب من الإيجاز يراعى فيه الحجم المطلوب.
- بدل دلاله كاملة على الموضوع والأفكار الرئيسية فيه.
- نسق من الجمل والعبارات المترابطة [1].

3. فوائد التلخيص:

- توفير الوقت والجهد.
- تمييز الأفكار الرئيسية من الأفكار الثانوية.
- القدرة على اكتشاف ما يريد الكاتب وماذا يقول بين السطور للتعرف على أهدافه.
- الاستغناء عن قراءة المادة والاكتفاء بالرجوع إلى الملخص.

---

[1] ينظر: المرجع نفسه، ص211

- التذوق السليم لأسلوب الكاتب وطريقة عرض الأفكار.

- تدريب الطلبة على تنظيم وترتيب الأفكار.

- تدريب الطلبة على الدقة السرعة و الإتقان لما يقرأ ويسمع [1].

4. كيفية تلخيص النصوص:

للقيام بعملية تلخيص النص يجب إتباع ما يأتي :

1. قراءة النص بتمعن أكثر من مرة لفهم الأفكار الواردة فيه.

2. الإحتفاظ بأفكار النص الأساسية وإلغاء الأفكار الثانوية ، وأيضاً حذف الصفات والمرادفات التي لا تغير من مضمون أفكار النص.

3. الإبتعاد عن النقاط الفرعية التي تعد هامشية بالنسبة لنص الموضوع ، وذلك للحصول على ملخص فعال.

4. الوفاء عند إعداد التلخيص ؛ أي علينا الإلتزام بما ورد به بغض النظر عما ورد به ، فلا يجوز لنا أن نغير أي فكرةٍ أساسية أو منه لإنها تتعارض مع رأينا الشخصي.

---

[1] ينظر : نبيل ، علي ، "اللغة العربية والحاسوب"، مرجع سابق، ص. 34

5. الإبتعاد عن التأويل ، التحريف ، التعليقات أو الإنتقادات ، وعدم إطلاق الأحكام على ما ورد في النص.

6. عدم التعبير عن الرأي أثناء التلخيص ، إنما الإلتزام وإحترام رأي المؤلف الذي تعب في الأساس لتأليف هذا النص.

7. مراعاة ترابط الجمل أثناء التلخيص لكي لا يكون نصك ضعيفاً ومفككاً.

8. التحدث بلسان المؤلف وبنفس الضمائر إن أمكن.

9. التخلص قدر الإمكان من الهوامش الواردة في النص والأمثلة الكثيرة التي لا تعطي هدفاً في عملية التلخيص ، وذلك لا يعني حذف كل الأمثلة ، لأن ذلك يترتب عليه جموداً في النص ، كما أنه هناك أمثلة تؤثر وتغير في المضمون إن تم حذفها.

10. يمكن كتابة التلخيص أو إدراجه في شكل رسم بياني.

11. تجهيز ملخصات بسيطة يسهل حملها مثل بطاقاتٍ صغيرة يسهل التلخيص فوقها بدلاً من الورق لسهولة الإستخدام والحفاظ عليها لمدة أطول بدون تلف.

12. إعداد مقارنة للتلخيص بالنص الأصلي ، للحرص على أنه لم يتم الإخلال بمضمون النص أو تسلسل أفكاره.

13. إعادة قراءة التلخيص وتصحيح الأخطاء إن وردت ، كالأخطاء الإملائية أو النحوية أو في علامات الترقيم.

14. وأخيراً نسخ التلخيص بخطٍ واضح ، ويفضل أن يتم طباعته عن طريق جهاز الحاسوب [1].

وهذه هي أهم النقاط الواجب مراعاتها أثناء تلخيص أي موضوع.

5. قواعد لتلخيص النصوص :

هناك قواعد أخرى لتلخيص النصوص يمكن ذكرها كما يلي :

1.5. قاعدة الحذف:

يمكن حذف كل الجمل التي لا تساهم في فهم النص مثل : تحديد الزمان والمكان ، ووصف الأشياء والأشخاص والأعمال الثانوية.

2.5. قاعدة الدمج:

يمكن دمج الجملة في جمل أخرى تشكل شرطا لازما أو نتيجة للجملة.

3.5. قاعدة البناء:

يمكن بناء جملة من جمل وإحلالها محلها شرط أن تكون الجملة المبنية الناتج الطبيعي للجمل.

[1] ينظر: صيني، محمود إسماعيل، بنوك المصطلحات الآلية، مرجع سابق، ص 213

4.5. قاعدة التعميم:

يمكن استبدال مجموعة من الجمل بجملة تعميمية تحمل في ذاتها المعاني التي حملتها الجمل المستبدلة.

مبادئ أساسية يجب أن تراعى في التلخيص :

1. الاستغناء عن التفاصيل والمناقشات المتعددة الواردة في الأصل .

2. عدم تحريف أو تشويه المعلومات الواردة في الأصل .

3. عدم إهمال المراجع والأدلة التي اعتمد عليها النص.

4. لا بأس من ذكر تعليقات أو تنويهات يضيفها الكاتب إذا رأى في ذلك إثراء للنص الأصلي.

5. إدراك أن نسبة طول الملخص إلى طول الموضوع الأصلي تختلف باختلاف تكثيف النص الأصلي .فقد يكون النص الأصلي مركزا تركيزا واضحا لا تستطيع أن تختصره كثيرا ، مثل التعريفات أو الخلاصات النحوية [1] .

[1] ينظر : اليوسفي، أحمد شعيب ، مرجع سابق، ص 127

6. خطوات التلخيص :

1. قراءة النص قراءة استكشافية : لإدراك الفكرة الأساسية التي يتضمنها النص .

2. قراءة النص مرة ثانية قراءة متمعنة وتحديد الكلمات المفتاحية : هي الكلمات التي تتضمن الفكرة التي تتناولها كل فقرة أو كل جزء من النص المقروء أو المسموع

3. كتابة التلخيص : وأسلم طريقة لكتابة التلخيص هي أن نضع النص الأصلي جانبا بعد اتمام الخطوتين السابقتين ،ثم نكتب التلخيص من استيعابنا للفقرة . هذه الطريقة تجنبنا الوقوع في خطأ وضع النص الأصلي أمامنا والتقاط بعض الجمل بنصها منه ، ثم ربط بعضها ببعض فنخرج بهذه الطريقة غير السليمة باقتباس وليس تلخيصا ،بل ربما ينتج عن ذلك إفساد للمعنى الذي يقصده الكاتب ،وبالتالي نخرج بتلخيص مهلهل وغير سليم .

4. مراجعة التلخيص بعد كتابته بالطريقة التي ذكرناها؛ وذلك للتحقق من صحة التلخيص للأصل وما تقتضيه المراجعة من تعديلات على التلخيص نحوا وإملاء وأسلوبا [1].

[1] ينظر : المرجع نفسه، ص 131

المرحلة الأولى: قراءة النص الأصلي قراءة متأنية فاحصة واستيعاب مضمونه وأهدافه.

المرحلة الثانية: تدوين الأفكار الرئيسية " أثناء القراءة " في مذكرات مختصرة خارجية او في الهامش.

المرحلة الثالثة: إعادة صياغة الفكرة أو الأفكار الرئيسية بأسلوب خاص، بإيجاز محكم دون إضافة أو تعليق [1] مع مراعاة السمات التالية:-

أ) تنقيح الكتابة للتأكد من وجود التتابع المنطقي وتسلسل الأفكار كما وردت في النص الأصلي.

ب) التأكد من سلامة التلخيص من الأخطاء اللغوية والنحوية.

جـ) تجنب التعديل أو التحريف المخل الذي يشوه ويغير المعنى الأصلي.

د) أن لا يتجاوز التلخيص للموضوع الثلث من النص الأصلي ونادراً ما تأتي في نفس الحجم الأصلي من حيث الطول.

هـ) الاستعانة بالمعاجم اللغوية عن فهم بعض المفردات والاصطلاحات العلمية أو الفنية أو اللغوية في الموضوع.

[1] ينظر : العناتي، وليد أحمد، اللسانيات الحاسوبية العربية ، ص54، مرجع سابق

وهذه طريقة أخرى لتلخيص النصوص [1]:

1- إحصاء عدد الكلمات أو الأسطر ، وتعيين الحد المطلوب " الربع _ النصف "

2- قراءة النص قراءة صامتة أكثر من مرة لفهم الفكرة الرئيسية فيه.

3- يكتب جملة واحدة فقط ليحدد فكرة الموضوع بلغته الخاصة ، ثم يضع خطا تحت الجملة الدالة على هذه الفكرة ، ثم يقارن بين الجملتين ليعيد النظر في الانسجام بينهما.

4- قراءة النص مرة أخرى ووضع خطوط تحت الجمل الدالة على الفكرة الفرعية التي تدعم الفكرة الرئيسية

5- وضع خطوط تحت الجمل الأساسية التي تشير إلى كيفية ترابط أجزاء النص مع بعضها وحذف العبارات الزائدة.

6- إعادة بناء النص " تلخيص النص "

• تبدأ بكتابة اسم الكاتب ، وعنوان الموضوع أو النص والفكرة الرئيسية له ، والاستمرار في كتابة التلخيص دون حذف أية فكرة باستخدام أدوات الربط.

• التعبير بإيجاز • إنهاء التلخيص بجملة نهائية تشير إلى أهمية الموضوع.

• تلخيص ما يقوله الكاتب وعدم تحريف معنى النص الأصلي.

7- مراجعة النص بعد الكتابة الأولى له وتدقيقه.

---

[1] ينظر : المرجع نفسه، ص 59

8- قراءة النص الملخص مرة أخرى من قبل شخص آخر قراءة نقدية.

9- تحرير النص لغويا بتصحيح الأخطاء الإملائية والنحوية ووضع علامات الترقيم ثم كتابته بشكله النهائي مع ملاحظة :

- الصياغة السليمة
- الإيجاز
- الأسلوب اللغوي الصحيح
- الدقة
- الترابط اللغوي
- سلامة الأفكار الرئيسية والفرعية
- شمول الأفكار ودقتها
- صحة عرض الأفكار
- تجنب الإضافة أو التحريف أو التعريف أو النقد للموضوع [1] .

[1] ينظر: الشاعر، شفيق سالم، وكالة الغوث ، فن التلخيص ، مجلة المعلم تربوية ثقافية جامعة 2006 ، ص75

7. الأسس الفنية للتلخيص

تتمثل الأسس الفنية للتلخيص [1] في النقاط التالية :

1- الفهم الواعي والاستيعاب الدقيق لموضوع التلخيص.

2- فهم الفكرة الأساسية في الفقرة، والأفكار الجزئية التي تندرج تحتها.

3- معرفة الجمل الأساسية ، والجمل المفسرة الشارحة، والجمل المؤكدة والمعللة.

4- معرفة الجمل المفتاحية التي تبدأ بها الفقرة.

5- إعادة صياغة الفقرة مع الحفاظ على الأفكار الأساسية، وحذف الجمل المترادفة والتكرار والحشو مع بقاء جمل التعليل.

6- تحديد الفكرة المحورية للموضوع.

7- معرفة المساحة التي سيعرض فيها التلخيص، فقد تكون مساحة محدودة جداً لا تتسع لأكثر من الفكرة المحورية، وقد تتسع لإيراد الفكرة الرئيسية ، وقد تزداد اتساعاً فتسمح بإيراد بعض التفاصيل.

---

[1] ينظر : مسعود عمشوش التاريخ : 17 ماي, 2013 12:18 صباحًا. التصنيف : دروس, مقالات, مهارات www.mahaarat.com ، ص 13

8- التعبير يكون بألفاظ المُلَخِص، وقد يسمح بالاستعانة ببعض ألفاظ الموضوع وخاصة المصطلحات العلمية.

9- لابد من الالتزام بالأمانة العلمية عند التلخيص فتكتب أفكار الكاتب وآراؤه، كما هي دون تدخل من المُلخِّص إلا في حالة التعليق النهائي إذا تطلب التلخيص ذلك.

10- إذا كان الموضوع يتضمن بعض الاستشهادات تذكر بمعناها ولا بد من توخي الدقة إذا كان الاستشهاد بمعنى آية قرآنية كريمة. وكذلك الشأن بالنسبة للحوار يثبت مضمون الحوار فقط.

كما يراعى في تقدير درجات التلخيص [1] المهارات الآتية :

1. أن تهتم بالأفكار الأساسية .
2. أن تراعي التسلسل المنطقي في عرض الأفكار .
3. التحرر من لغة الموضوع الأصلية .
4. مراعاة الوقف والترقيم .
5. تجنب التكرار .

---

[1] ينظر : المرجع نفسه، ص 15

وتتكون بنية التلخيص، في الأغلب، من ثلاثة عناصر أساسية هي :الافتتاحية وصلب الموضوع والخاتمة .وتتكون الافتتاحية من جملتين يعرض فيهما كاتب التلخيص موضوع النص وغرض الكاتب من كتابته .بينما يتكون صلب التلخيص من جملة يفتتح بها كاتب التلخيص كيف عرض الكاتب موضوعه، ثم ينتقل بعد ذلك إلى عرض ما جاء في النص من أفكار رئيسة حول الموضوع .أما الخاتمة فتشتمل على رأي كاتب النص الأصلي أو النتيجة التي توصل إليها أثناء عرض موضوعه.

وتجدر الإشارة إلى ضرورة أن يلتزم كاتب التلخيص بما جاء في النص الأصلي من آراء حول موضوع النص وأن يبتعد عن استخدام كل ما يدل على رأيه الشخصي أو انطباعه الذاتي في الموضوع.

" وقد يضطر كاتب التلخيص أحيانا إلى استخدام أدوات الربط في لغة كتابة التلخيص للربط بين الأفكار الرئيسة في صلب التلخيص، ومن هنا ننصح بالا يكثر من استخدام الأداة الواحدة بل عليه أن ينوع من أدوات الربط كي يضفي على تلخيصه ثراء وتنوعا"[1].

[1] ينظر : المرجع نفسه، ص 18

8. عرض وتلخيص :

تمتلك اللغة العربية، إلى جانب الخصائص الأساسية المشتركة مع لغات أخرى مثل اللاتينية، خصائص خاصة جدا مثل عدم وجود التشكيل وفواصل الفقرات، بشكل منتظم، ووجود التثنية والتراص. وبالتالي فإن اختيار طريقة تحليل النصوص العربية يجب أن تكون لها القدرة على التكيف مع خصائص اللغة العربية.

في الواقع، يجب أن نأخذ بعين الاعتبار كلا من قواعد وخصائص اللغة العربية أثناء التحليل، ومن أجل وصف بنيات اللغة مع عدد صغير من القواعد، يمكن أن تكون ممثلة في مداخل المعجمية، ومعلومات صوتية، صرفية، نحوية، دلالية وذرائعية أوواقعية. وهذه الثروة تسمح لنا بتجنب الحلول الزائفة، على سبيل المثال، التراص وغياب التشكيل من خلال استغلال الميزات النحوية والدلالية. توفر هذه الميزات القرائن إلى أنواع من الكلمات المتوقعة في الجملة دون اللجوء إلى قواعد إعادة الكتابة .

نذكر أيضا، كما قدمناه في فصل سابق، أن هناك بعض التقنيات لتلخيص النصوص العربية:

أساليب الاستخراج، والأساليب الهجينة وأساليب الفهم والتوليد.

أ. طرق استخراج الجمل :

الهدف من استعمال طرق استخراج الجمل هو تحديد العبارات الأكثر أهمية في النص الأصلي. والملخص هو مقتطف من النص الأصلي.

أ.1. طرق الكلمات المفتاحية :

وتستند هذه الطريقة على حقيقة أن المؤلف يعتمد (للتعبير عن أفكاره الرئيسية) على بعض الكلمات الرئيسية التي تتكرر بشكل ملحوظ في النص. ثم يتم إنتاج الملخص من خلال البحث عن الوحدات النصية الدنيا التي تحوي هذه الكلمات. وغالبا ما يطبق هذا المبدأ في أنواع مختلفة :

أ.1.1. كلمات مفتاحية معرفة مسبقا :

يمكن إدخال قائمة من الكلمات المفتاحية من قبل المستخدم أو كلمات منشأة من قبل المؤلف.

أ.2.1. العناوين :

على اعتبار أن العنوان هو التعبير الأبرز الذي يلخص مستند في بضع كلمات، يمكننا القول أن الجملة التي تشبه أكثر من غيرها العنوان هي العبارة الأكثر أهمية. ولذلك، فإننا يمكن تعيين وزن لكل جملة وفقا لقربها من حيث الشبه بالعنوان [1].

[1] ينظر : جرجيس، جاسم محمد، عبد الجبار عبد الرحمان، ص39، مرجع سابق

في هذه الحالة نعتبر الكلمات المكونة لعنوان النص ككلمات رئيسية ويتم إنتاج الموجز عن طريق اختيار العبارات التي تغطي بعض الكلمات التي تظهر في العنوان.

أ.3.1. طرق توزيع الكلمات :

فكرة هذه الطريقة هي اعتبار الجمل "هامة" إذا احتوت على كلمات في النص "مهمة". وتعتبر كلمة مهمة إذا تم استخدامها في حالات كثيرة جدا في النص أي اذا تكرر استخدامها بشكل ملفت.

أ.2. طريقة الموضع :

ويعتبر أسلوب الموضع الجملتين الأولى والأخيرة من كل فقرة مهمة. وتعتبر الطريقة أيضا العبارات المتمركزة في بعض المقاطع هامة، على سبيل المثال، في "المقدمة" و "الخلاصة". وعيب هذه الطريقة هو أن ذلك يعتمد على طبيعة النص وكذا أسلوب المؤلف [1].

---

[1] ينظر : Horacio, Saggion, Génération automatique de résumé par analyse sélective, ، مرجع سابق

أ.3. التعابير الإرشادية :

في الكتابات العلمية، كثيرا ما نجد التعبيرات التي، بصرف النظر عن النطاق المحدد للنص، تشير هاته التعبيرات الى فئات مفاهيمية خاصة. وهكذا، إذا وجدنا مقدمة مقال من الشكل "الهدف من هذا المقال "، يمكنك أن تكون متأكدا أن الجملة الموالية تتضمن معلومات عن أهداف المقال وعندما نجد جملة تبدأ "ولنلخص ما قيل"، فمن المؤكد أن هذا هو استنتاج أو خاتمة شاملة.

فكرة طريقة التعابير الإرشادية هي لتحديد الجمل في النص التي تحتوي على هذه الأنواع من التعابير.

أ.4. التماسك المعجمي :

إنشاء علاقات تماسك معجمي بين الجمل وترتيب هذه الأخيرة وفقا لتصنيف "بداية الموضوع"، "استمرار الموضوع"، "إغلاق الموضوع" و "الهوامش".

تحدد مجموعة فرعية من الجمل التي تبدأ بعرض، تواصل وتنهي المواضيع [1].

[1] ينظر : حشمت،قاسم، خدمات المعلومات ، مرجع سابق، ص179

أ.5. سلاسل معجمية :

بناء سلاسل معجمية التي تربط الجمل التي تحتوي على كلمات ذات علاقات تماسك معجمي.

تحدد مجموعة فرعية من السلاسل ومن ثم مجموعة من الجمل في كل سلسلة.

أ.6. استخراج الفقرات :

حساب اتصالات بين فقرات النص باستخدام تدابير التشابه.

يتم تحديد الفقرات الأكثر صلة ببعضها.

أ.7. تصنيف العناصر :

في نصوص العلوم والتكنولوجيا هناك عبارات تشير إلى فئات مفاهيمية مثل: المعرفة السابقة، المحتوى، والطريقة والنتيجة لذلك، يمكن للمرء أيضا أن يرى هذه الفئات في ملخصات النصوص العلمية والتقنية [1].

[1] ينظر:النوايسة، غالب عوض، الاستخلاص لأغراض استرجاع المعلومات، ص 50، مرجع سابق

ب. النهج المدمج :

الأساليب المعروضة في الأقسام السابقة تستخدم الميزات (التردد، والموضع، والتعابير الإرشادية، الخ) التي لا يمكن منفردة ضمان تحقيق أفضل النتائج.

في كثير من الأحيان تطبق هذه الطرق جنبا إلى جنب للحصول على نتائج أفضل.

جـ. طرق الفهم والتوليد :

وخلافا للاجراءات "الكمية" التي تعين وزنا لكل جملة، فإن أساليب الفهم والتوليد تحاول أن تكتشف كيف تساهم كل جملة لتنظيم النص، وما هي وظيفة كل جملة في النص إجمالا. لإيجاد ملخصا جيدا ليس كاف تحديد المعلومات المهمة ولكن الأهم كيفية إنتاجها.

وعيب هذه الأساليب أن تنفيذها لا يزال صعبا للغاية [1] .

9. مزايا طرق الاستخراج :

- البساطة في التنفيذ،
- معالجة سريعة،
- استقلال العلاج بغض النظر عن اللغة،
- حجم التلخيص متحكم فيه بتغيير عتبات الاختيار أو النسبة المئوية المرغوبة.

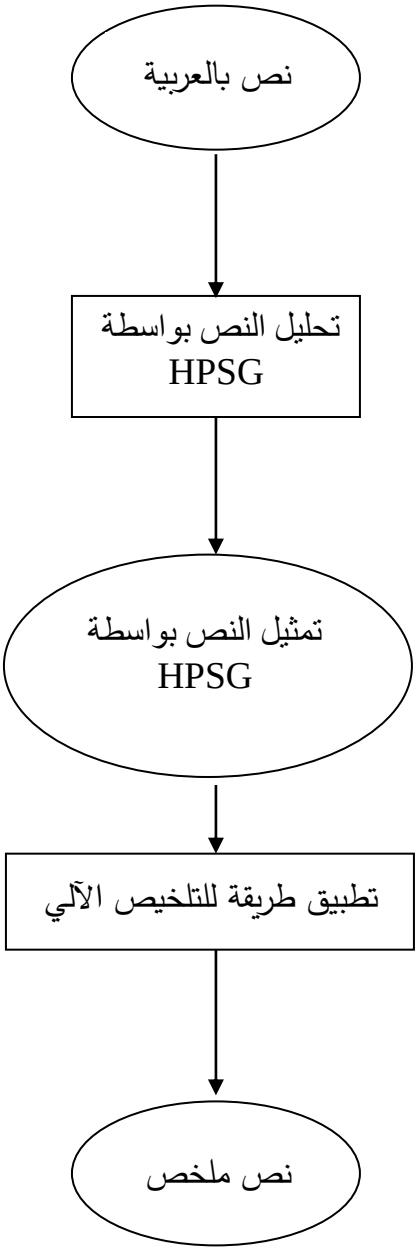
[1] الحليمي، رفيق حسن، فن كتابة التلخيص والمختصرات، ص 131، مرجع سابق

إن استخدام طريقة واحدة فقط قد لا يعطي النتائج المرجوة، وغالبا ما يتم استخدام نهج مختلط (هجين) الذي يجمع بين عدة أساليب للتلخيص.

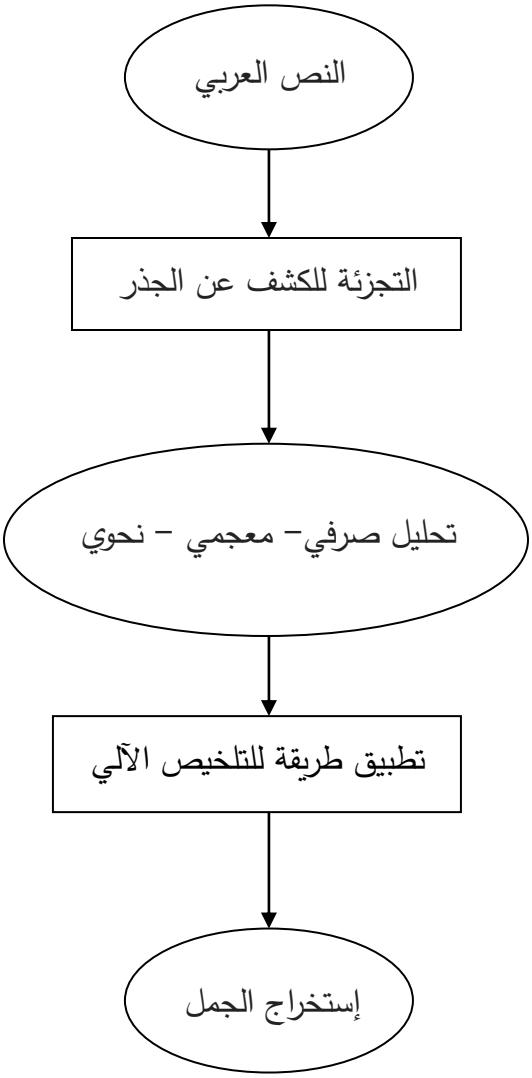
لهذا اقترحنا طريقة سميناها " أوجز " (AWDJIZ) لتلخيص النصوص العربية تقوم أساسا على فكرتين:

- العمل بقواعد اللغة العربية لتحديد وتمثيل الوحدات اللغوية،
- استخدام نهج هجين للتلخيص تعطي نتائج مرضية في اللغة العربية.

إن التصميم الذي نقترحه للتنفيذ الوظيفي لأوجز يوجد في الشكل الموالي، وهو يقوم على تحديد مكونات الجمل وكذلك الجمع بين عدة طرق للوصول إلى تلخيص جيد .



نمـوذج 1 لأوجـــــز



نمـوذج 2 لأوجـــــز

10. وصف الوحدات الرئيسة المكونة لأوجز :

1.10. التجزئة للكشف عن الجذر :

التجزئة هي الكشف عن الفقرات والكلمات والجذور.

لتقسيم النص إلى جمل، وارسيغني [1] يستخدم كل من:

• التجزئة الصرفية بناء على علامات الترقيم (Ponctuation)،

• أومن خلال تقسيم على أساس معرفة العلامات الصرفية والنحوية أو الكلمات الوظيفية مثل أو، و، أي، لكن حتى.

ومع ذلك، قد تؤدي هذه الأدوات دورا غير الفصل بين العبارات.

في حالتنا، نتعرف على الكلمات عن طريق إجراء تجزئة تقوم على مؤشرات الفصل أو الفراغات بين الكلمات والجمل.

---

[1] وارسيغني ، رياض Modélisation des expressions figées en arabe en vue de la constitution d'une base de données lexicale. Thèse de doctorat., Université Lyon 2. 2001

وفي نهاية المطاف يستخدم الجدول 11 لإزالة السوابق واللواحق التي يمكن لصقها على الجذر (تحتوي الجذور العربية عموما 3-4 حروف).

نستخدم قائمة السوابق واللواحق التي اقترحها درويش [1] أنظر الجدول رقم 11. العديد منهم استخدمها " تشان" [2] للوصول إلى جذور الكلمات العربية :

السـوابـق							
والـ	بتـ	وتـ	بمـ	كمـ	للـ	فيـ	لا
فالـ	يتـ	ستـ	لمـ	فمـ	ليـ	وا	با
بالـ	متـ	نتـ	ومـ	الـ	ويـ	فا	
اللـواحـق							
ات	وه	ته	هم	ية	ين	ة	ا
وا	ان	تم	هن	تك	يه	ه	
ون	تي	كم	ها	نا	ية	ي	

جدول 11: قوائم معظم السوابق المشتركة واللواحق

---

[1] ينظر: درويش . ق Building a Shallow Arabic Morphological Analyzer in One Day. Proceedings of the workshop on Computational Approaches to Semitic Languages in the 40th Annual Meeting of the Association for Computational Linguistics (ACL-02), Philadelphia, PA, USA. pp. 47-54.

[2] ينظر : A. Chen and F. Gey : Building an Arabic Stemmer for Information Retrieval. Proceedings of the Eleventh Text Retrieval Conference (TREC 2002). National Institute of Standards and Technology, Nov 18-22, 2002, pp631-640.

2.10. تحليل صرفي- معجمي :

تأخذ مدخلات هذه الخطوة نتائج التجزئة ونتحصل بعد هذا التحليل الصرفي والمعجمي على الجذور وما يرافقها من خصائص صرفية ونحوية ودلالية.

هذه الخطوة أمر مهم جدا لمعرفة خصائص كل عنصر [1]

3.10. تحليل نحوي :

يحدد هذا التحليل ما إذا كانت الجملة (أو سلسلة من الكلمات) تنتمي أو لا تنتمي إلى اللغة، وبالتالي الامتثال لقواعد اللغة العربية.

تم استنباط هذه الوحدة على أساس مشروع SYNTAX [2] والذي يسمح بتحليل ومعرفة الجمل الاسمية والفعلية واللفظية وتنتج أشجار عبارات إذا قبلت بالكامل من قبل قواعد اللغة أو شجرة جزئية من العبارات التي يعترف بها المحلل.

---

[1] ينظر : Mourad Loukam, une plate-forme d'analyse basée sur le formalisme HPSG pour l'Arabe standard, Université de Chlef 2008

[2] ينظر : نور الدين لوكيل، قيس حدار، عبد المجيد بن حمادو
Laboratoire de Recherche en Informatique et Multimédia, Sfax, Tunisie
Normalisation de la représentation des lexiques syntaxiques arabes pour les formalismes d'unification www.miracl.rnu.tn 2009 Consulté le 31/03/2014

4.10. تطبيق طريقة مدمجة للتلخيص (الاستخراج) :

وتعتمد هذه الطريقة على الاختيار من بين الأساليب المذكورة أعلاه تلك التي تعطي أفضل النتائج وهي: طريقة الكلمات المفتاحية، وطريقة العنوان، وطريقة الموضع، وطريقة التعبير الإشاري.

إذا كان النص علميا أو تقنيا يمكن إضافة طريقة تصنيف العناصر.

الطريقة المدمجة = طريقة الكلمات المفتاحية + طريقة العنوان + طريقة الموضع + طريقة التعبير الاشاري.

والنتيجة هي اتحاد الجمل التي نتحصل عليها بتطبيق كل طريقة مع الأخذ بعين الاعتبار ترتيب ظهور هذه العبارات في النص الأصلي ودون تكرار (على سبيل المثال إذا جاءت نفس الجملة من تطبيق طريقتين تؤخذ مرة واحدة فقط في الملخص).

لنأخذ بعض الأمثلة:

طريقة الكلمات المفتاحية : وتستند هذه الطريقة على حقيقة أن المؤلف يعتمد (للتعبير عن أفكاره الرئيسية) على بعض الكلمات الرئيسية التي تتكرر بشكل ملحوظ في النص. ثم يتم إنتاج الملخص من خلال البحث عن الوحدات النصية الدنيا التي تحوي هذه الكلمات.

على سبيل المثال الذكاء الاصطناعي، العلاج الآلي للغة، التلخيص الآلي

أما بالنسبة لطريقة العنوان : على اعتبار أن العنوان هو التعبير الأبرز الذي يلخص مستند في بضع كلمات، يمكننا القول أن الجملة تشبه أكثر من غيرها العنوان هي العبارة الأكثر أهمية. ولذلك، فإننا يمكن تعيين وزن لكل جملة وفقا لقربها من حيث الشبه بالعنوان.

مثلا: التلخيص الآلي

أما طريقة الموضع : يعتبر أسلوب الموضع الجملتين الأولى والأخيرة من كل فقرة مهمة. وتعتبر الطريقة أيضا العبارات المتمركزة في بعض المقاطع هامة، على سبيل المثال، في "المقدمة" و "الخلاصة".

طريقة التعابير الإرشادية : في الكتابات العلمية، كثيرا ما نجد التعبيرات التي، بصرف النظر عن النطاق المحدد للنص، تشير هاته التعبيرات الى فئات مفاهيمية خاصة. وهكذا، إذا وجدنا مقدمة مقال من الشكل "الهدف من هذا المقال هو ..."، يمكنك أن تكون متأكدا أن الجملة الموالية تتضمن معلومات عن أهداف المقال وعندما نجد جملة تبدأ "ولنلخص ما قيل..."، فمن المؤكد أن هذا هو استنتاج أو خاتمة شاملة.

فكرة طريقة التعابير الإرشادية هي لتحديد الجمل في النص التي تحتوي على هذه الأنواع من التعابير.

مثال : وأخيرا، وهنا كنتائج .....

وأخيرا طريقة تصنيف العناصر: في نصوص العلوم والتكنولوجيا هناك عبارات تشير إلى

فئات مفاهيمية مثل: المعرفة السابقة، المحتوى، والطريقة والنتيجة لذلك، يمكن للمرء أيضا أن يرى هذه الفئات في ملخصات النصوص العلمية والتقنية.

إذا جمعنا نتائج كل طريقة، بترتيب ظهور الجمل في المستند الأصلي وبدون تكرار، نصل إلى خلاصة جيدة جدا (وذات دلالة) من النص الأصلي.

5.10. إستخراج الجمل :

يسمح لنا بالحصول على النتيجة النهائية حسب اختيار نسبة الضغط (اختيار المستخدم).

وتمثل هذه النسبة عدد الجمل المستخرجة بالمقارنة مع عدد الجمل في المستند الأصلي.

خاتمة

في هذا الفصل، وصفنا نهجنا لإنتاج التلخيص الآلي للنصوص باللغة العربية، ووصف بنية أوجز، تستند إلى مرحلة أولية نحلل في الخطوة الأولى من خلالها النص ثم في الخطوة الموالية نطبق أساليب الاستخراج المدمج (الجمع بين عدة أساليب للاستخراج لإعطاء نتائج أفضل).

تستند قوة هذا النهج من النقاط التالية :

- تحليل النصوص العربية بطرق علمية.
- يتم الحصول على التلخيص الآلي من خلال تطبيق طريقة مدمجة تتكون من مزيج من عدة طرق للاستخراج
- بساطته من حيث التنفيذ

الفصل الخامس

تطبيقات للنهج المدمج لتلخيص النصوص العربية

يتضمن هذا الفصل التطبيق اليدوي للبرنامج المقترح في الفصل السابق والذي عنوناه بالنهج المدمج لتلخيص النصوص العربية والنتائج المحصل عليها هي مانأمل الوصول إليه بتطبيق برنامج إعلام آلي متقن.

النص الأول :

نصوص عربية قديمة تصف الفايكينغ [1]

(إنهم الأكثر قذارة بين مخلوقات الله: لا يتطهرون بعد التغوط أو التبول، لا يغتسلون عندما لا يكونون طاهرين طقسياً بعد الجماع، ولا يغسلون أيديهم حتى بعد تناول الطعام). هذا ما دوّنه الكاتب العربي أحمد بن فضلان بعد لقائه مسافرين من الفايكينغ قبل نحو ألف سنة.

(1) ينظر : مجلة (الجريدة) الالكترونية ، شوهدت يوم 09 جويلية 2017 في 15سا 32د

http://www.aljarida.com/articles/1462272103588350200

درس المؤرخ الأيسلندي ثورير جونسون هروندال مجموعة من التعليقات عمَن ندعوهم {الفايكينغ} في نصوص أصلية وضعها مؤرخون وجغرافيون عرب. تصف هذه النصوص لقاء العرب بمجموعات من الاسكندينافيين في مناطق حول بحر قزوين ونهر الفولغا.

لكن وصفهم يختلف كثيرًا عن صور الغزاة الفايكينغ المخيفين التي وصلتنا من الجزر البريطانية وفرنسا في تلك الحقبة.

الإسكندينافيون الأقوياء

يشير جونسون هروندال: (الاختلاف شاسع بين الاسكندينافيين الذين سافروا شرقًا ومَن أبحروا غربًا. ويعود ذلك إلى أن المجتمعات التي صادفوها في الشرق كانت أكثر تقدّمًا ورقيًّا من المجتمعات الغربية).

قدّم هذا المؤرخ أخيرًا رسالة دكتوراه أعدها في جامعة بيرجن، وتناول فيها مَن يدعون {الروس}، أي التجار والمحاربين الاسكندينافيين الذين سافروا إلى أوروبا الشرقية، وسط آسيا، والشرق الأوسط.

يوضح: (كان الاسكندينافيون، على ما يبدو، شعبًا مرنًا أجاد التكيّف مع الحياة في مناطق متعددة وشارك في عدد من بُنى القوة).

من شبه الجزيرة الإسكندينافية إلى بغداد

منذ منتصف القرن التاسع حتى عام 1000 بعد الميلاد تقريبًا، شهدت الشعوب الاسكندينافية ما يُعتبر موجة كبيرة من السفر والترحال. فقد انطلق الفايكينغ ليستكشفوا العالم ويتاجروا مع شعوب أخرى أو يحاربوها. ويرتبط اسم النرويجيين بالغارات التي شُنّت على أوروبا الوسطى وبالرحلات إلى أيسلندا وغرينلاند وأميركا الشمالية.

لكن جونسون هروندال يخبر أن الفايكينغ اعتادوا التوجه شرقًا أيضًا، مستخدمين الأنهار الأوروبية الكبرى. يذكر هذا الباحث: (يفوق عدد القطع الاسكندينافية التي عُثر عليها في أوروبا الشرقية ما وُجد في الغرب، ما يغطي مساحة جغرافية أوسع).

كذلك كانت الرحلات نحو المشرق مكثّفة. يشير مصدر تناوله جونسون هروندال أن الفايكينغ اتبعوا الأنهار الأوروبية حتى بحر قزوين، حيث عبروا المياه بسفن ليتابعوا بعد ذلك طريقهم إلى بغداد مع قوافل الجمال. وهكذا يكونون قد اجتازوا مسافة 5 آلاف كيلومتر.

زيارة بلدان حكمها الترك

لم يكن حضور الاسكندينافيين في المناطق الشرقية من القارة نادرًا طوال قرون. فكانوا محاربين، تجارًا، وعمالاً عاديين. حتى إن بعض المصادر يشير إلى أن الفايكينغ تولوا حراسة القوافل، أي أنهم كانوا جنودًا مرتزقة.

يقول جونسون هروندال: (أظهرت المصادر الأثرية وجود الفايكينغ في مناطق أبعد إلى الغرب، إلا أن العرب كانوا يلتقون بهم غالبًا قرب بحر قزوين، وشكّل هذا إطارًا ثقافيًّا مختلفًا، إطارًا حكمه الترك).

الترك مجموعة إثنية تشمل شعب تركيا المعاصرة، إلا أنها تضم شعوبًا أخرى أيضًا. وعلى غرار المغول، ينحدر الترك من مجموعة أورال آلتاي الإثنية.

يذكر هذا المؤرخ: (كان الترك، خصوصًا الخزر والبلغار، قوة مهيمنة في المنطقة مع وصول الروس. تُظهر النصوص مدى قوة الترك، لذلك لم يكن بإمكان الروس المجيء حاملين السيوف للسيطرة على المنطقة).

وسيمون ومجانين

لكن الكاتب المذكور أعلاه، أحمد بن فضلان، لم ينفر من الاسكندينافيين تمامًا: لم أرَ بنية جسدية كاملةً مثل بنيتهم: طوال كشجر النخل، حسنو المنظر، وشعرهم أحمر، فضلاً عن أنهم لا يلبسون رداء أو قفطانًا.

وفي احدى النصوص العربية التي تتحدث عن الفايكينغ، يصف أحمد بن فضلان طقوس دفن تتلاءم عمومًا مع طقوس الفايكينغ في اسكندينافيا، مع أنها كانت غريبة بالنسبة إلى رجل متعلّم مسلم: إن كان الرجل ثريًّا، يجمعون ممتلكاته ويقسّمونها إلى ثلاثة أقسام :ثلث لأسرته، ثلث يشترون له به لباس الدفن، وثلث يخمّرون به الكحول ليشربوها في اليوم الذي تقتل فيه أمته نفسها وتُحرق مع سيدها .

أوصاف عربية معاصرة

يعرف الباحثون والمؤرخون منذ زمن أن الفايكينغ سافروا شرقًا، لكن علماء الآثار والمؤرخين لا يملكون موارد كثيرة تفصّل هذه الرحلات. في النرويج، مثلاً، تعود المواد المكتوبة عن هذه الرحلات إلى حقبة لاحقة كانت هذه الرحلات قد انتهت قبلها بزمن طويل، ما يقوّض مصداقية النصوص كمصادر تاريخية.

نتيجة لذلك، ركّز علماء الآثار على التحف من بيزنطة في تركيا اليوم ومن مناطق سلافية، مثل هنغاريا.

يذكر جونسون هروندال: «لكن الكتّاب المسلمين قاموا أيضًا بأسفار والتقوا تجارًا اسكندينافيين، فدوّنوا هذه الحوادث في الفترة التي وقعت فيها». لذلك يُعتبرون مصادر ممتازة للحصول على معلومات دقيقة عن الروس.

قطع فضية تؤكد التواصل

ثمة أسباب عدة جعلت الدارسين يتجاهلون الفايكينغ الذين توجهوا شرقًا ويركزون على مَن سافروا غربًا. فقد أعاقت المشاكل السياسية عمل علماء الآثار الغربيين طوال عقود، وخلال الجزء الأكبر من القرن العشرين، صعب على الباحثين في أوروبا الغربية الاطلاع على قطع أثرية جُمعت خلف الستار الحديدي.

يوضح يان بيل، بروفسور متخصص في علم الآثار في جامعة أوسلو: (حصلنا على معلومات أكبر من الغرب بسبب الثقافة اللغوية والكتابة التي انتشرت في تلك المنطقة. لا يعني ذلك أن الاتصال بالشرق لم يكن مهمًّا، إلا أننا لم نحظَ بفرصة كبيرة لدراسته).

ولكن ثمة استثناءات. يشير بيل إلى العثور على بعض قطع من الفضة العربية وغيرها من تحف من كازاخستان في هيمدالسيورديت، سوق سابقة لا تبعد كثيرًا عن تلة سفينة الفايكينغ غوكستاد في ساندفيورد جنوب غرب أوسلو. يذكر بيل: (أتت من طريق الحرير وتؤكد أن الفايكينغ كانوا على اتصال بالمناطق الإسلامية).

تكمن عقبة أخرى في أن النصوص العربية لا تذكر أسماء الاسكندينافيين الذين تصفهم، بخلاف كثير من المصادر الغربية. فمن الواضح أن العرب نظروا إليهم كمجموعة ولم يعتبروا اسماء {الكفار} الذين قابلوهم صدفة مهمّة لليوميات التي دونوها عن رحلاتهم. نتيجة لذلك، يعتبر المؤرخون الشخصية الرئيسة مجهولة الهوية، ما يحدّ من حماستهم لدراستها.

يأمل جونسون هروندال أن تحفّز رسالة الدكتوراه التي أعدّها إلى مزيد من الدراسات عن الفايكينغ في الشرق. يقول: (يُذكر الاسكندينافيون الذين سافروا شرقًا بشكل موجز في الكتب الدراسية والمدارس . وأخيرا، أعتقد أنه من الضروري أن نسلط مزيدًا من الضوء على هذا الجزء من تاريخنا).

تطبيق النهج المدمج لتلخيص النص :

الملخص هو ناتج جمع فقرات تطبيق طريقة الكلمات المفتاحية، وطريقة العنوان، وطريقة الموضع، وطريقة التعبير الإشاري

تطبيق الطريقة الأولى :

الكلمات المفتاحية في النص هي : نصوص عربية + الفايكينغ

نتحصل على الجمل التالية :

1) درس المؤرخ الأيسلندي ثورير جونسون هروندال مجموعة من التعليقات عمَن ندعوهم {الفايكينغ} في نصوص أصلية وضعها مؤرخون وجغرافيون عرب. تصف هذه النصوص لقاء العرب بمجموعات من الاسكندينافيين في مناطق حول بحر قزوين ونهر الفولغا.

2) يقول جونسون هروندال: " أظهرت المصادر الأثرية وجود الفايكينغ في مناطق أبعد إلى الغرب، إلا أن العرب كانوا يلتقون بهم غالبًا قرب بحر قزوين، وشكّل هذا إطارًا ثقافيًّا مختلفًا، إطارًا حكمه الترك ".

تطبيق الطريقة الثانية أي طريقة العنوان : نبحث في النص عن الفقرات التي تحوي العنوان أو جزء منه، فنجد 15 فقرة تشمل كلمة (فايكينغ) ولكن أقرب فقرة للعنوان هي :

(وفي احدى النصوص العربية التي تتحدث عن الفايكينغ، يصف أحمد بن فضلان طقوس دفن تتلاءم عمومًا مع طقوس الفايكينغ في اسكندينافيا، مع أنها كانت غريبة بالنسبة إلى رجل متعلّم مسلم: (إن كان الرجل ثريًّا، يجمعون ممتلكاته ويقسّمونها إلى ثلاثة أقسام :ثلث لأسرته، ثلث يشترون له به لباس الدفن، وثلث يخمّرون به الكحول ليشربوها في اليوم الذي تقتل فيه أمته نفسها وتُحرق مع سيدها)) .

ثم نطبق الطريقة الثالثة وهي طريقة الموضع والتي –للتذكير– تأخذ الجمل الأولى والأخيرة من كل فقرة

1) منذ منتصف القرن التاسع حتى عام 1000 بعد الميلاد تقريبًا، شهدت الشعوب الاسكندينافية ما يُعتبر موجة كبيرة من السفر والترحال.

2) ويرتبط اسم النرويجيين بالغارات التي شُنّت على أوروبا الوسطى وبالرحلات إلى أيسلندا وغرينلاند وأميركا الشمالية.

وأخيرا نطبق الطريقة الرابعة والتي تدعى التعبير الإشاري وتتمثل في البحث في النص الأصلي عن جمل تشمل تعابير مثل : لنلخص ما قيل، وأخيرا، مجمل القول ....

فنتحصل على :

وأخيرا، أعتقد أنه من الضروري أن نسلط مزيدًا من الضوء على هذا الجزء من تاريخنا.

وبدمج النتائج المحصل عليها –وبدون تكرار– نتحصل على الملخص التالي :

1) درس المؤرخ الأيسلندي ثورير جونسون هروندال مجموعة من التعليقات عمَن ندعوهم {الفايكينغ} في نصوص أصلية وضعها مؤرخون وجغرافيون عرب. تصف هذه النصوص لقاء العرب بمجموعات من الاسكندينافيين في مناطق حول بحر قزوين ونهر الفولغا.

2) يقول جونسون هروندال: (أظهرت المصادر الأثرية وجود الفايكينغ في مناطق أبعد إلى الغرب، إلا أن العرب كانوا يلتقون بهم غالبًا قرب بحر قزوين، وشكّل هذا إطارًا ثقافيًّا مختلفًا، إطارًا حكمه الترك).

3) وفي احدى النصوص العربية التي تتحدث عن الفايكينغ، يصف أحمد بن فضلان طقوس دفن تتلاءم عمومًا مع طقوس الفايكينغ في اسكندينافيا، مع أنها كانت غريبة بالنسبة إلى رجل متعلّم مسلم: (إن كان الرجل ثريًّا، يجمعون ممتلكاته ويقسّمونها إلى ثلاثة أقسام : ثلث

لأسرته، ثلث يشترون له به لباس الدفن، وثلث يخمّرون به الكحول ليشربوها في اليوم الذي تقتل فيه أمته نفسها وتُحرق مع سيدها) .

4) منذ منتصف القرن التاسع حتى عام 1000 بعد الميلاد تقريبًا، شهدت الشعوب الاسكندينافية ما يُعتبر موجة كبيرة من السفر والترحال. ويرتبط اسم النرويجيين بالغارات التي شُنّت على أوروبا الوسطى وبالرحلات إلى أيسلندا وغرينلاند وأميركا الشمالية.

5) وأخيرا، أعتقد أنه من الضروري أن نسلط مزيدًا من الضوء على هذا الجزء من تاريخنا.)

انتهى الملخص.

وبهذا نكون قد تحصلنا على نتيجة جيدة جدا، فقد استطعنا أن نلخص نصا من حوالي ست صفحات الى نص ذو معنى يشمل خمس فقرات فقط.

النص الثاني : النمر العربي [1]

النمر العربي من الحيوانات الثديية اللاحمة آكلات اللحوم والتي تتبع فصيلة السنوريات، وهي إحدى سُلالات النمور. يُعتبر النِمر العربي أكبر وأقوى أنواع القطط العربية، ولكنه يبقى الأصغر حجماً بين سُلالات النمور جميعها، والتي تنتشر في قارتي آسيا وأفريقيا، وتعتبر هذه السلالة مهددة بصورة حرجة عبر موطنها بأكمله الذي يشمل اليوم جنوب سوريا، فلسطين، السعودية، الإمارات، اليمن، وعمان، حيث أن أعدادها مازالت تتناقص شيئا فشيئا.

المعيشة والحياة الاجتماعية

النمور العربية حيوانات تقطن الجبال العالية على عكس ما يظنه العامّة من أنها حيوانات قاطنة للصحراء ومناطق الآجام الغالبة على الدول التي تستوطنها. فهي توجد بصورة رئيسيّة في جبل سمحان في مُحافظة ظفار العمانية ومنطقة بلجرشي و عسير الجبليتان في السعودية وفي محمية عين جدي بالقرب من البحر الميت بفلسطين.

---

[1] موسوعة ويكيبيديا ، شوهدت يوم 17 مارس 2017 على 13سا19د

https://ar.wikipedia.org/wiki/

تقتات النمور العربية على الوعول النوبية، الماعز البري، الطهر العربي، الثعالب، وغيرها من الحيوانات الجبلية، ويمتلك كل نمر منطقة خاصة به ويقوم بالدفاع عنها ضد النمور الأخرى من الجنس نفسه. تكون منطقة الذكر أكبر حجما من منطقة الأنثى وتتقاطع دائما مع مناطق إناث متعددة، وتقوم النمور داخل هذه المناطق بالصيد والتزاوج وتربية جرائها، وفي المناطق الجافة والقاحلة تكون مساحة حوز النمور أكبر من ذاك في المناطق الأقل جفافا وذلك عائد إلى ندرة الطرائد في المناطق الجافة مما يرغم النمور أن توسّع من نطاق بحثها عن الغذاء، وهذا يدل أيضا على أن النمور العربية لم تكن يوما وافرة العدد في هكذا مناطق.

النِمر العربي يتميز بنشاطهِ نهاراً وليلاً ولكنه يبقى حذرا من الوجود البشري، وهذه الحيوانات انعزالية لا تلتقي غيرها من النمور إلا في فترة التزاوج والتي تدوم تقريباً 5 أيام، يتم خلالها التزاوج عدة مرات. وبعد فترة حمل تدوم ما بين 98 إلى 100 يومِ، تلد الأنثى من جرو إلى أربعة جراء في أحد الكهوف أو الشقوق الصخرية. وتكون الأشبال عمياء عند الولادة، وتفتح أعينها بعد تسعة إلى عشر أيامِ، وعندها يبدأون باستكشاف بيئتهم المحيطة. وهم عادة لا يغادرون العرينِ لوحدهم حتى يصلوا إلى عمر أربعة أسابيع على الأقل .أثناء هذه الأسابيع الأولى تقوم الأم بنقل الأشبال من عرين إلى آخر عدّة مرات للتقليل من فرص عثور

الضواري الأخرى عليهم. وتصل أعمار النمور في البرية ما بين 10 إلى 15 سنة، بينما تصل أعمارها في حدائق الحيوان حتى 52 سنة.

الصفات الخارجية

تكون النمور العربية فاتحة اللون بشكل كبير حيث أن اللون الذهبي المصفرّ والذي يتواجد في العادة بين البقع في معظم أنحاء جسم باقي السلالات لا يتواجد عند هذه السلالة إلا على طول ظهورها، ومن ثم يبهت إلى الأصفر الشاحب أو الأبيض على باقي الجسد. وتتميز بلون عيونها الزرقاء خلافاً لمثيلاتها الأفريقية .تزن الأنثى البالغة 20 كيلوغراماً (45 رطلا)، بينما يزن الذكر البالغ ما يقارب الـ 30 كيلوغراماً (65 رطلا). وبالمُقارنة فإن النمور الأفريقية الجنوبية يمكن أن يتراوح وزنها ما بين 50 إلى 70 كيلوغراماً، وبالتالي يمكن اعتبار النمور العربية أصغر من باقي سلالات النمر الإفريقية والآسيوية. يتشابه الذكر مع الأنثى، ولكن الذكور تكون أكبر من الإناث، كما يتم التمييز بينهما بوجود كيس الصفن الواضح لدى الذكور، وتمتاز النمور العربية بذيولها الطويلة، والذي تستعمله في التوازن أثناء التسلق.

الغـــــذاء

تقتات النمور العربية على الماشية المستأنسة (الماعز عادة) بشكل كبير وذلك عائد إلى أن طريدتها الطبيعية أي الطهر العربي وغزال الجبل قد أصبحت نادرة لدرجة كبيرة أو حتى

اختفت في بعض الدول التي تقطنها، وهذا التصرف بدوره يضعها في مواجهة مباشرة مع الإنسان بالرغم من أن هذا يحدث في اليمن ومسندم والسعودية غالبا. تفترس النمور العربية الثعالب وأي نوع أخر من الثدييات الصغيرة بالإضافة إلى الطيور والجيفة، وهي تصطاد أثناء فترة الغسق والفجر إجمالا وتنشط أثناء الليل بشكل أكبر من النهار الذي تمضيه في أماكن ظليلة مخفيّة عن العيون هربا من الحرارة المرتفعة.

الوضع الحالي والتعداد

تعتبر هذه السلالة من النمور معرّضة بصورة حرجة، وقد أدى الصيد المكثّف في أوائل التسعينات من القرن العشرين إلى إطلاق عدّة برامج للحفاظ على هذه النمور وعلى مسكنها الجبلي وجميع أصناف الحياة البرية التي تقطنه. يبلغ عدد النمور العربية حوالي 15 أو 18 نمرا في صحراء النقب ووادي عربة في فلسطين، وقرابة 200 نمر في شبه الجزيرة العربية بأكملها وخصوصا على امتداد جبال مدين والحجاز والسروات حتى اليمن وكذلك في شرق الإمارات، ولا تزال الأعداد تتناقص في المناطق البريّة الغير محمية بسبب اضطهادها وصيدها على اعتبار أنها من الحيوانات المضرّة والمؤذية.

كان النمر العربي في السابق أكبر السنوريات وأكثرها انتشاراً في شبه الجزيرة العربية، فقد امتد موطنه على مسافة 1700 كيلومتر عبر جبال الحجاز وفي شرق الجزيرة في عمان

والإمارات وغيرها. لكنه الآن مُهدد وَمَوضوع على القائمة الحمراء للأنواع المهددة بالانقراض، ويُعتقد أنه بهذا المُعدل متجه إلى الانقراض تماماً من مناطقه الواسعة التي تمتد عبر مُعظم أفريقيا وغرب آسيا .تفيد الدراسات الحديثة بأن انتشار النمر العربيّ في المنطقة العربية أصبح مَحدوداً جداً، وقد اختفى من العديد من بعض المناطق مثل جبال مدين، بينما تضاءل وُجوده كثيراً في مناطق أخرى مثل جبال السروات والحجاز عموماً، وقتل آخر فرد منه في جبال الطويق عام 1970م. وفي عام 2006.

ومجمل القول أنه توصلت دراسة من قسم علم الحيوان في جامعة الملك سعود قامت بمسح أكثر من 50 موقعاً في السعودية إلى وُجود 65 فرد فقط من النمر العربي في كافة المناطق التي مسحتها عبر جبال الحجاز.

تطبيق النهج المدمج لتلخيص النص :

الملخص هو ناتج جمع فقرات تطبيق طريقة الكلمات المفتاحية، وطريقة العنوان، وطريقة الموضع، وطريقة التعبير الإشاري

تطبيق الطريقة الأولى :

الكلمات المفتاحية في النص هي : النمور ، النمر العربي ، الحيوانات الثديية

نتحصل على الجمل التالية :

1) النمر العربي من الحيوانات الثديية اللاحمة آكلات اللحوم والتي تتبع فصيلة السنوريات، وهي إحدى سُلالات النمور.

2) النمور العربية حيوانات تقطن الجبال العالية على عكس ما يظنه العامّة من أنها حيوانات قاطنة للصحراء ومناطق الآجام الغالبة على الدول التي تستوطنها، فهي تتواجد بصورة رئيسيّة في جبل سمحان في مُحافظة ظفار العمانية ومنطقة بلجرشي و عسير الجبليتان في السعودية وفي محمية عين جدي بالقرب من البحر الميت بفلسطين.

تطبيق الطريقة الثانية أي طريقة العنوان : نبحث في النص عن الفقرات التي تحوي العنوان أو جزء منه، فنجد أن أقرب فقرة للعنوان هي الفقرة الأولى وتجنبا للتكرار لا تؤخذ في الحسبان لأنه تم أخذها بتطبيق الطريقة الأولى.

ثم نطبق الطريقة الثالثة وهي طريقة الموضع والتي –للتذكير– تأخذ الجمل الأولى والأخيرة من كل فقرة فنتحصل على الجمل التالية :

1) يُعتبر النِمر العربي أكبر وأقوى أنواع القطط العربية، ولكنه يبقى الأصغر حجماً بين سُلالات النمور جميعها، والتي تنتشر في قارتي آسيا وأفريقيا، وتعتبر هذه السلالة مهددة بصورة حرجة عبر موطنها بأكمله الذي يشمل اليوم جنوب سوريا، فلسطين، السعودية، الإمارات، اليمن، وعمان، حيث أن أعدادها مازالت تتناقص شيئا فشيئا.

2) النمور العربية حيوانات تقطن الجبال العالية على عكس ما يظنه العامّة من أنها حيوانات قاطنة للصحراء ومناطق الآجام الغالبة على الدول التي تستوطنها.

3) وتصل أعمار النمور في البرية ما بين 10 إلى 15 سنة، بينما تصل أعمارها في حدائق الحيوان حتى 52 سنة.

وأخيرا نطبق الطريقة الرابعة والتي تدعى التعبير الإشاري وتتمثل في البحث في النص الأصلي عن جمل تشمل تعابير مثل : لنلخص ما قيل، وأخيرا، مجمل القول ....

فنتحصل على :

1) ومجمل القول أنه توصلت دراسة من قسم علم الحيوان في جامعة الملك سعود قامت بمسح أكثر من 50 موقعاً في السعودية إلى وُجود 65 فرد فقط من النمر العربي في كافة المناطق التي مسحتها عبر جبال الحجاز.

وبدمج النتائج المحصل عليها –وبدون تكرار– نتحصل على الملخص التالي :

1) النمر العربي من الحيوانات الثديية اللاحمة آكلات اللحوم والتي تتبع فصيلة السنوريات، وهي إحدى سُلالات النمور.

2) النمور العربية حيوانات تقطن الجبال العالية على عكس ما يظنه العامّة من أنها حيوانات قاطنة للصحراء ومناطق الآجام الغالبة على الدول التي تستوطنها، فهي تتواجد بصورة رئيسيّة في جبل سمحان في مُحافظة ظفار العمانية ومنطقة بلجرشي و عسير الجبليتان في السعودية وفي محمية عين جدي بالقرب من البحر الميت بفلسطين.

3) يُعتبر النِمر العربي أكبر وأقوى أنواع القطط العربية، ولكنه يبقى الأصغر حجماً بين سُلالات النمور جميعها، والتي تنتشر في قارتي آسيا وأفريقيا، وتعتبر هذه السلالة مهددة بصورة حرجة عبر موطنها بأكمله الذي يشمل اليوم جنوب سوريا، فلسطين، السعودية، الإمارات، اليمن، وعمان، حيث أن أعدادها مازالت تتناقص شيئا فشيئا.

4) النمور العربية حيوانات تقطن الجبال العالية على عكس ما يظنه العامّة من أنها حيوانات قاطنة للصحراء ومناطق الآجام الغالبة على الدول التي تستوطنها.

5) وتصل أعمار النمور في البرية ما بين 10 إلى 15 سنة، بينما تصل أعمارها في حدائق الحيوان حتى 52 سنة.

6) ومجمل القول أنه توصلت دراسة من قسم علم الحيوان في جامعة الملك سعود قامت بمسح أكثر من 50 موقعاً في السعودية إلى وُجود 65 فرد فقط من النمر العربي في كافة المناطق التي مسحتها عبر جبال الحجاز.

وبهذا نكون قد تحصلنا على نتيجة جيدة جدا، فقد استطعنا أن نلخص نصا من حوالي أربع صفحات الى نص ذو معنى يشمل ست جمل فقط.

النص الثالث : هل تسير تركيا إلى "دكتاتورية منتخبة"؟ [1]

أدى الاستفتاء على التعديلات الدستورية في تركيا إلى إحداث حالةٍ من الاستقطاب الحاد داخل المجتمع التركي بين مؤيديه ومعارضي نتائجه. واتسعت دائرة الخلاف، لتتجاوز التعديلات الدستورية إلى الخلاف حول شخص رجب طيب أردوغان، باعتباره من كان وراءها، والمستفيد الأول منها .

وبالنسبة لمعارضي هذه التعديلات، فإنهم يخشون أن تؤدي إلى ترسيخ حكم الفرد الواحد، وتؤسّس منظومة ديكتاتورية فعلية في بلادٍ لم تتعاف بعد من حكم العسكر، ولم تتخلص بعد من خطر الانقلابات العسكرية. أما مؤيدو هذه التعديلات فينظرون إليها على أنها خطوةٌ إلى الأمام لإحداث تغييراتٍ جذرية في السياسة التركية، من خلال وجود حكم تنفيذي قوي في يد شخص واحد تسلم له مقاليد قيادة البلاد نحو التقدم المنشود .

---

[1] أنوزلا ، علي ، 19 أبريل 2017

https://www.alaraby.co.uk/opinion/2017/4/18/

شوهد يوم 11 أوت 2017 على 18سا11د

حالة الاستقطاب الداخلي التي أحدثتها هذه التعديلات، وقسمت المجتمع التركي إلى مؤيد ومعارض لأردوغان، تجاوزت الحدود التركية إلى العالمين العربي والإسلامي، واخترقت حتى صفوف الإسلاميين في كل مكان، ممن كانوا ينظرون إلى تجربة أردوغان في الحكم نموذجاً لنجاح الإسلام السياسي في القيادة والحكم والتسيير .

ولا شك في أن أردوغان نجح في قيادة بلاده خلال الخمس عشرة سنة الماضية، ووضعها في مصاف الدول المتقدمة ومن بين الأنظمة الاقتصادية القوية في العالم. وقد بدأ عهده بسياسة "صفر نزاعات"، أي تجاوز كل ميراث النزاعات القديمة التي كانت بين الإمبراطورية العثمانية والدولة التركية التي خَلَفتها ومحيطها الجغرافي الإقليمي. وساهمت تلك السياسة الذكية في النهوض بتركيا قوة إقليمية صاعدة، إلى أن أصبحت اليوم من أقوى الدول المؤثرة في محيطها وفي العالم. وقد تمكّنت تركيا من تحقيق نجاحها الاقتصادي بفضل نظامها الديمقراطي الذي أتاح، لأول مرة، لحزب إسلامي قيادة البلاد، وأدى هذا النجاح إلى إيجاد نموذجٍ لما سمي نظام الحكم الإسلامي الديمقراطي .

اليوم، ومع التعديلات الدستورية التي ستنتقل بتركيا من نظامٍ برلماني تعدّدي إلى نظام رئاسي،

كل السلطات ستصبح فيه مركزة في يد شخص واحد. يخشى مؤيدو أردوغان، قبل معارضيه، من أن يؤدي ذلك إلى انهيار النموذج السياسي والاقتصادي الذي باتت تمثله تركيا في عالم

اليوم، وأن تتحول قصة النجاح التركي إلى كابوس مروّع .

وهناك اليوم من يتساءل إن كانت تركيا بالفعل في حاجةٍ إلى مثل هذه التعديلات؟ فتركيا حققت أهم نجاحاتها السياسية والاقتصادية والاجتماعية في ظل النظام الحالي الذي جاءت التعديلات لتغييره. وحتى أردوغان الذي يُعتقد أن إغراء السلطة دفعه إلى طرح هذه التعديلات كان يتصرف كرئيس له كامل الصلاحيات .

يعتقد من صوتوا أو أيدوا هذه التعديلات أن أردوغان في حاجةٍ إلى صلاحيات واسعة أكثر من الممنوحة حاليا لرئيس الجمهورية لتحقيق النهضة التركية التي وعد بها. أما المعارضون فلا يرون فيها سوى انتقالٍ من نظام برلماني تعدّدي إلى نظام رئاسي فردي سيجعل أردوغان يستمر في الحكم حتى 2029، مسجلا بذلك رقما قياسيا ستضاهي به تركيا أسوأ الدكتاتوريات في تاريخ العالم، عندما يستمر حكم أردوغان 28 عاما بالتمام والكمال .

أول متضرّر اليوم من هذه التعديلات الدستورية هو أردوغان نفسه الذي يُعتقد أنه المستفيد الأول منها أيضا، فالنموذج "الأردوغاني" المبني على الحكم الفردي، والذي طالما جذب معجبين كثيرين به هو اليوم في طريقه إلى التأسيس لدكتاتورية جديدة، "دكتاتورية منتخبة". ومن يعود إلى قراءة التاريخ يكتشف أن أسوأ الدكتاتوريات في العالم وأكثرها ضررا للبشرية كانت هي "المنتخبة ."

عاش أردوغان، أخيرا، تجربتين عصيبتين، المحاولة الانقلابية الفاشلة في الصيف الماضي،

والاستفتاء على التعديلات الدستورية لتوسيع صلاحياته رئيساً، وفي الحالتين خرج أردوغان منتصرا، لكنه انتصارٌ بطعم الخسارة، حتى لا أقول "بطعم الهزيمة" كما وصفت ذلك "بي بي سي"، والخاسر الأكبر في الحالتين هو النموذج الديمقراطي التركي الذي كان أردوغان مؤسسه وعنوانه .

وبهذه التعديلات الدستورية، وضع أردوغان بلاده وشعبه أمام مفترق طرق كبير وحاسم، إما أن يختار أن يكون موحّدا ليعيد ترميم التجربة الديمقراطية التي بدأت تتآكل، حتى قبل المحاولة الانقلابية الفاشلة، أو أن تؤدي نتائج الاستفتاء إلى تعميق حالة الاستقطاب الحاد داخل المجتمع التركي، المنقسم على نفسه، ما قد يمهد لحكم استبدادي طويل، أو قد يُسرع بمحاولة انقلابية جديدة، تعود بالبلاد إلى حكم العسكر الذي لم يودّعه الشعب التركي إلا قبل سنوات قريبة .

فتركيا الحديثة، وعلى الرغم من تجربة حزب العدالة والتنمية التي حاولت أن تصالح بين العلمـــانيـــة والديمقراطية، ما زالت لم تفطم بعد، وتاريخها الاستبدادي الذي رعته، ونظرت له "أصولية علمانية"، فـي سياق تاريخي وثقافي معقد، تحول فيه الجيش فيه إلى حارس للعلمانية ومناهض للديمقراطية. وخلال فترة حكم" العدالة والتنمية " (حزب أردوغان) بدا وكأن تركيـا أصبحت قـادرةً على المصـالحـة بين "الإسلام السياسي" و"العلمانية" و"الديمقراطية". لكن، اليوم هناك مخاوف مشروعة من أن المحاولة الانقلابية الفاشلة وخطر

تكرارها الذي مازال قائما، والاستفتاء على التعديلات الدستورية التي زادت من تقسيم المجتمع التركي، سيعيدان أكبر بلد سني في الشرق الأوسط إلى المربع الأول، ويضعانه أمام خيارين صعبين :العودة إلى "الأصولية العلمانية" أو الخضوع لأصولية جديدة هي "الأصولية الدينية الديمقراطية" التي يخيف شبحها كثيرين في العالم، في منطقة تشهد مخاضا تاريخيا، يصعب التنبؤ بما ستؤول إليه الأمور في أغلب بلدانه. وما بين الخيارين، يوجد خيار ثالث هو خيار الديمقراطية التي تتطلب تنازلاتٍ، وتُبنى على توافقات يفتقدها المجتمع السياسي التركي اليوم .

وأخيرا، من المؤسف أن يتحوّل أردوغان الذي كان يُعتبر، بدون منازع، أول قائد للتجربة الديمقراطية التركية الفتية، ونموذجا للإسلام السياسي الديمقراطي، إلى أكبر خطرٍ يهدّد نموذجه وتجربته، وأكبر خطرٍ على الديمقراطية نفسها المهدّدة اليوم في تركيا!

تطبيق النهج المدمج لتلخيص النص :

الملخص هو ناتج جمع فقرات تطبيق طريقة الكلمات المفتاحية، وطريقة العنوان، وطريقة الموضع، وطريقة التعبير الإشاري

تطبيق الطريقة الأولى :

الكلمات المفتاحية في النص هي : تركيا ، دكتاتورية منتخبة ، الاستفتاء ، التعديلات الدستورية

نتحصل على الجمل التالية :

1) أدى الاستفتاء على التعديلات الدستورية في تركيا إلى إحداث حالةٍ من الاستقطاب الحاد داخل المجتمع التركي بين مؤيديه ومعارضي نتائجه.

2) اليوم، ومع التعديلات الدستورية التي ستنتقل بتركيا من نظامٍ برلماني تعدّدي إلى نظام رئاسي،

كل السلطات ستصبح فيه مركزة في يد شخص واحد.

تطبيق الطريقة الثانية أي طريقة العنوان : نبحث في النص عن الفقرات التي تحوي العنوان أو جزء منه، فنجد أن أقرب فقرة للعنوان هي :

1) أول متضرّر اليوم من هذه التعديلات الدستورية هو أردوغان نفسه الذي يُعتقد أنه المستفيد الأول منها أيضا، فالنموذج "الأردوغاني" المبني على الحكم الفردي، والذي طالما جذب معجبين كثيرين به هو اليوم في طريقه إلى التأسيس لدكتاتورية جديدة، "دكتاتورية منتخبة".

ثم نطبق الطريقة الثالثة وهي طريقة الموضع والتي –للتذكير– تأخذ الجمل الأولى والأخيرة من كل فقرة فنتحصل على الجمل التالية :

1) وبالنسبة لمعارضي هذه التعديلات، فإنهم يخشون أن تؤدي إلى ترسيخ حكم الفرد الواحد، وتؤسّس منظومة ديكتاتورية فعلية في بلادٍ لم تتعاف بعد من حكم العسكر، ولم تتخلص بعد من خطر الانقلابات العسكرية.

2) حالة الاستقطاب الداخلي التي أحدثتها هذه التعديلات، وقسمت المجتمع التركي إلى مؤيد ومعارض لأردوغان، تجاوزت الحدود التركية إلى العالمين العربي والإسلامي، واخترقت حتى صفوف الإسلاميين في كل مكان، ممن كانوا ينظرون إلى تجربة أردوغان في الحكم نموذجاً لنجاح الإسلام السياسي في القيادة والحكم والتسيير.

3) ولا شك في أن أردوغان نجح في قيادة بلاده خلال الخمس عشرة سنة الماضية، ووضعها في مصاف الدول المتقدمة ومن بين الأنظمة الاقتصادية القوية في العالم.

وأخيرا نطبق الطريقة الرابعة والتي تدعى التعبير الإشاري وتتمثل في البحث في النص الأصلي عن جمل تشمل تعابير مثل : لنلخص ما قيل، وأخيرا، مجمل القول ....

فنتحصل على :

1) وأخيرا، من المؤسف أن يتحوّل أردوغان الذي كان يُعتبر، بدون منازع، أول قائد للتجربة الديمقراطية التركية الفتية، ونموذجا للإسلام السياسي الديمقراطي، إلى أكبر خطرٍ يهدّد نموذجه وتجربته، وأكبر خطرٍ على الديمقراطية نفسها المهدّدة اليوم في تركيا!

وبدمج النتائج المحصل عليها –وبدون تكرار– نتحصل على الملخص التالي :

1) أدى الاستفتاء على التعديلات الدستورية في تركيا إلى إحداث حالةٍ من الاستقطاب الحاد داخل المجتمع التركي بين مؤيديه ومعارضي نتائجه.

2) اليوم، ومع التعديلات الدستورية التي ستنتقل بتركيا من نظامٍ برلماني تعدّدي إلى نظام رئاسي،

كل السلطات ستصبح فيه مركزة في يد شخص واحد.

3) أول متضرّر اليوم من هذه التعديلات الدستورية هو أردوغان نفسه الذي يُعتقد أنه المستفيد الأول منها أيضا، فالنموذج "الأردوغاني" المبني على الحكم الفردي، والذي طالما جذب معجبين كثيرين به هو اليوم في طريقه إلى التأسيس لدكتاتورية جديدة، "دكتاتورية منتخبة".

4) وبالنسبة لمعارضي هذه التعديلات، فإنهم يخشون أن تؤدي إلى ترسيخ حكم الفرد الواحد، وتؤسّس منظومة ديكتاتورية فعلية في بلادٍ لم تتعاف بعد من حكم العسكر، ولم تتخلص بعد من خطر الانقلابات العسكرية.

5) حالة الاستقطاب الداخلي التي أحدثتها هذه التعديلات، وقسمت المجتمع التركي إلى مؤيد ومعارض لأردوغان، تجاوزت الحدود التركية إلى العالمين العربي والإسلامي، واخترقت حتى صفوف الإسلاميين في كل مكان، ممن كانوا ينظرون إلى تجربة أردوغان في الحكم نموذجاً لنجاح الإسلام السياسي في القيادة والحكم والتسيير.

6) ولا شك في أن أردوغان نجح في قيادة بلاده خلال الخمس عشرة سنة الماضية، ووضعها في مصاف الدول المتقدمة ومن بين الأنظمة الاقتصادية القوية في العالم.

7) وأخيرا، من المؤسف أن يتحوّل أردوغان الذي كان يُعتبر، بدون منازع، أول قائد للتجربة الديمقراطية التركية الفتية، ونموذجا للإسلام السياسي الديمقراطي، إلى أكبر خطرٍ يهدّد نموذجه وتجربته، وأكبر خطرٍ على الديمقراطية نفسها المهدّدة اليوم في تركيا!

وبهذا نكون قد تحصلنا على نتيجة جيدة جدا، فقد استطعنا أن نلخص نصا من حوالي أربع صفحات الى نص ذو معنى يشمل سبع جمل فقط.

الخـــاتمـــــة

إن أهم مايميز هذه الدراسة هو ارتكازها على محورين رئيسيين هما :

أ- يعتبر هذا البحث إسهاما علميا في بناء الإطار النظري المتعلق بالتلخيص الآلي للنصوص العربية

ب- محاولة اقتراح وتصميم طرق معالجة آلية للنصوص العربية

إن معالجة اللغة العربية حاسوبياً أصبحت اليوم أمراً لا غنى عنه ولا مفرّ منه، وخاصة أن استثمار الدراسة الحاسوبية والمعلوماتية – بصفة عامة – يحقق نتائج كبيرة للغة العربية، في مجال التعريب، والإحصاء اللغوي، والمعالجة الآلية، وتعلم اللغات، والترجمة الآلية، وفي مجال التربية والتعليم .

في نهاية هذه الدراسة، تبين لنا أن الحاسوب يمكن تطويع آلياته وأنظمته خدمة للدراسات اللغوية والأدبية العربية.

ولقد وصفنا نهجنا (نظام أوجز) لإنتاج التلخيص الآلي للنصوص باللغة العربية، استنادا إلى مرحلة أولية نحلل من خلالها النص ثم نطبق أساليب الاستخراج المدمج (الجمع بين عدة أساليب للاستخراج لإعطاء نتائج أفضل).

تستند قوة هذا النهج على النقاط التالية :

• تحليل النصوص العربية بطرق علمية.

• يتم الحصول على التلخيص الآلي من خلال تطبيق طريقة مدمجة تتكون من مزيج من عدة طرق للاستخراج

• بساطته من حيث التنفيذ

❖ **النتائج العامة للدراسة**

تم عرض نظام أوجـز الذي يهدف للتلخيص الآلي للنصوص العربية، ولقد تمت دراسة هذه اللغة آليا(حاسوبيا) واعتمدت الخصائص المشتركة مع اللغات اللاتينية الأكثر تطبيقا في مجال المعالجة الآلية للغة مع الأخذ بعين الاعتبار خصائص اللغة العربية.

أعطت طرق الاستخلاص نتائج جيدة وذلك عند تطبيقها على مدونة نصوص علمية وتقنية عربية.

كما تم التحقق من أهمية بعض العوامل حيث أثبتت الكلمات المفتاحية دورا مميزا في نظامنا للتلخيص الآلي.

❖ **وكتوصيات ومقترحات و آفاق لهذا البحث** يمكننا أن نقول أن نظام أوجـــز يمكن أن يصبح أكثر فعالا لو تتم إضافة وحدة حذف تراكيب الحوارات غير المباشرة مع معالجة وتصحيح وجود الإشارات دون سوابق.

محور آخر يمكن أن يكون هاما في الدراسات المستقبلية ألا وهو استغلال واستثمار خصائص اللغة العربية لإنتاج نصوص جديدة مختصرة آليا تعتمد على فهم النصوص المدخلة.

محور آخر هو إنتاج ملخصات آلية مع ارتباطات تشعبية للنصوص الأصلية

كما يمكن إعتباره من التطبيقات المستقبلية، تطوير برامج حاسوبية تعتمد أساسا على نظامنا للتلخيص الآلي من أجل قياس فاعليته ونجاعته.

كما يمكن إبداء جملة من المقترحات، أحسب أنها تسهم في توطين توجه المعالجة الآلية للغة العربية في الدراسات اللغوية المعاصرة، وتدفع به نحو الأمثل والأفضل كما يرى ذلك أيضا العارف عبدالرحمن بن حسن من جامعة أم القرى بالسعودية، وهي على النحو الآتي :

أولاً : أن تتضافر الأعمال في مجال اللسانيات الحاسوبية العربية وتتآزر بين اللغويين والحاسوبيين في أي مشروع علمي يهدف إلى برمجة الأنظمة اللغوية للعربية، وتحليلها، ومعالجتها آلياً. وأرى في هذا المقام أن أي عمل منفرد، أو مستقل عن الطرف الآخر من المعادلة، أو غير منسق بين هذه الفئات العلمية، إنما هو بمثابة جهد ضائع لا طائل من ورائه.

إن التعاون والتنسيق في هذا الميدان – إن تَمَّ – ستكون نتائجه غاية في الأهمية، وستدفع به خطوات واثقة إلى الأمام، وسوف تتغير قناعات ومسلمات كانت أشبه بالحقائق التي لا تقبل الجدل والنقاش حولها، كما ستتأكد قضايا في اللغة كانت أقرب ما تكون إلى الاحتمال والرجحان والظن، وهذا ما توفره اللسانيات الحاسوبية للعلوم الإنسانية بعامة.

والواقع أن هذا الأمر، أعني التقاء علماء اللغة وعلماء الحاسوب، كان – ومازال – مطلباً ملحاً، وضرورة قصوى، لأي عمل ناجع في هذا الميدان على وجه الخصوص، وهو ما أكدت عليه جميع المؤتمرات التي عقدت لمناقشة قضية الحاسوب واللغة العربية.

ثانياً : ترجمة جميع الأعمال العلمية في مجال اللسانيات الحاسوبية العربية، التي كتبت باللغات الإنجليزية، والفرنسية، والألمانية، وغيرها من اللغات، التي وضعها باحثون عرب وأجانب، ونقلها إلى العربية.

ثالثاً : عدم ترك أمر البرمجيات الحاسوبية العربية بيد الشركات ومراكز البحوث الغربية، بل ينبغي أن يصممها أبناؤها، فهم أقرب الناس رحماً بهذا المجال.

رابعاً : صناعة معجم موحد لمصطلحات اللسانيات الحاسوبية، بالعربية والإنجليزية، وفق المتعارف عليه علمياً في هذا النوع من المعاجم الاصطلاحية أو المصطلحية .

خامساً : أن يكون علم اللغة الحاسوبي مقرراً دراسياً معتمداً في أقسام اللغة العربية بكليات الآداب، والتربية، كما هي الحال في كلية الآداب بجامعة الكويت.

سادساً : إنشاء قسم خاص للغويات الحاسوبية في الكليات والجامعات العربية يمنح درجة الليسانس في هذا التخصص، على غرار ما هو موجود بجامعة الأمير سلطان الأهلية بالرياض.

سابعاً : نشر تلك الرسائل العلمية التي كتبت أصلاً باللغة العربية عن قضايا استخدام اللغة العربية في الحاسوب، وهي تُشَكل فيما اطلعت عليه إسهاماً فاعلاً في ترسيخ هذا الاتجاه، ومعالجةً موضوعيةً وعلميةً لكثير من مشكلات تعامل العربية مع الحاسوب.

الملاحق

ANNEXE A

Overview of Text Summarization History

(from Udo Hahn's presentation at Automatic Summarization Workshop, Seattle, 2000)

Early Extraction 1955 – 1973

- Luhn, 1955, lexical occurrence stats
- Edmundson, 1969, positional indicators, lexical cues, cuewords, cue phrases
- Mathais, 1973, cohesion streamlining

Linguistic Approaches 1961 – 1979

Artificial Intelligence Approaches 1980's

- 1982. DeJong. FRUMP, using scripts
- 1985. SUSY, logic and production rules
- 1988. Reiner and Hahn.TOPIC. frames and semantic networks
- 1989. Rau et al. Hybrid Representations

"Renaissance" – 1990's and begin of 2000's

- Recurrence of statistical techniques
- hybrid approaches

ANNEXE B

A List of Summarization Projects

Research : Sentence Extraction

Surrey University: Summ-It applet

This summarization system works by extracting sentences using Lexical Cohesion.

Royal Institute of Technology (Sweden): SweSum

SweSum extracts sentences to product an extract type summary. It is closely related to the work at ISI. Summaries are created from Swedish or English texts in the either the newspaper or academic domains. Sentences are extracted by ranking sentences according to weighted word level features and was trained on a tagged Swedish news corpus. The summarization tool can be hooked up to search engine results.

University of Ottawa: The Text Summarization Project

Not much is available about this research project except their project proposal. In it they proposed to use machine learning techniques to identify keywords. Keyword identification can then be used to select sentences for extraction. They planned to use surface level statistics such as frequency analysis and surface level linguistic featurs such as sentence position.

Columbia University: FociSum

The FociSum system takes a question and answer approach to summarization. Sentences that answer key questions regarding participants, organisations and other wh-questions are extracted. The result is a concatenation of sentence fragments and clauses found in the original document. The system first uses a named entity extractor to find the foci of the document. A question generator is used to suggest relationships between these entities. The document is parsed to find candidate answers for these question on the basis of syntactic form. Sentence fragments and clauses are pulled out of the selected sentences.

University of Southern California: ISI Summarist

Summarist is produces summaries of web documents. It has been hooked up to the Systran translation system to provide a gisting tool for news articles in any language. Summarist first identifies the main topics of the document using statistical techniques on features such as position, and word counts. Current reseach is underway to use cue phrases and discourse structure. These concepts must be interpreted so that of a chain of lexically connected sentences, the sentence with the most general concept is selected and extracted. Subsequent work will take these extracted sentences to construct a more coherent summary.

Deep Understanding

The Sheffield University TRESTLE

This project produces summaries in the news domain. It uses MUC to extract the main concepts of the text which then presumably is used to generated summaries. Unfortunately, not much information is available on the official website regarding the system architecture.

Columbia University: SUMMONS

Summons is a multi-document summary system in the news domain. It begins with the results of a MUC-style information extraction process, namely a template with instantiated slots of pre-defined semantics. From this, it can generate a summary by using a sophisticated natural language generation stage. This stage was previously developed under other projects and includes a content selection substage, a sentence planning substage and a surface generation stage. Because the templates have well-defined semantics, the type of summary produced approaches that of human abstracts. That is they are more coherent and readable. However, this approach is domain specific, relying on the layout of news articles for the information extraction stage.

Hybrid Approaches (These combine extraction techniques with more traditional NLP techniques)

Columbia University: MultiGen

MultiGen is a multi-document system in the news domain. It extracts sentence fragments that represent key pieces of information in the set of related documents. This is done by using machine learning to group together paragraph sized chunks of text into clusters of related topics. Sentences from these clusters are parsed and the resulting trees are merged together to form, building logical representations of propositions containing the commonly occuring concepts. This logical representation is turned into a sentence using the FUF/SURGE grammar. Matching concepts uses linguistic knowledge such as stemming, part-of-speech, synonymity and verb classes. Merging trees makes use of identified paraphrase rules.

Copy and Paste

The Copy and Paste system is a single document summariser that is domain independent. It is designed to take the results of a sentence extraction summariser and extract key concepts from these sentences. These concepts are then combined to form new sentences. The system thus, copies the surface form of these key concepts and pastes them into the new sentences. This is done by first reducing the sentence removing any extraneous information. This step uses probabilities

learnt from a training corpus, and lexical links. The reduces sentences are merged by using rules such as adding extra information about speakers, adding conjunctives and merging common elements.

Commercial

(These are mostly extraction based)
note: The following descriptions are derived from information found on the products' official websites

- **Datahammer by Glucose**

  Datahammer is a product to designed to summarise online texts and works in conjunction the user's web browser. It extracts sentences by using an algorithm called 'Microword Tree Trimming' which they created. A demo version is available from their website.

- **Text Analyst by Megaputer**

  Text Analyst extracts sentences of documents on the user's computer. The official website of Text Analyst describes the summarization process they use. A semantic network is constructed from the source document using a neural network. They state that the construction of the semantic network is not dependent on prior domain specific knowledge. A graphical

representation of concepts and relationships in the source document is shown to the user for selection. Sentences with matching concepts and relationships are extracted.

- **IBM summarization products**

  IBM Japan incorporates summarization tools in two of its products: Internet King of Translation (Japanese) and Lotus Word Pro (Japanese Version). The type of summaries produced are sentence extracts, selected using rhetorical relations and position within the document. Extraction is done statistically and can utilise genre specific features.

  IBM also has a toolset called Text Analysis, which has as one of its components, a summarization tool. This toolset is part of the Intelligent Text Miner product. It produces summaries by extracting sentences. As with most commercial products, this is done by ranking the sentences by a measure of importance and then selecting the topmost ranks. Ranking is achieved by word level features and the user can select the extract length.

- **Websumm by Mitre**

  Mitre's WebSumm performs sentence extraction over single or multiple documents in conjunction with a search engine. The resulting summary is an extract of sentences based on a users

query. This is done by representing the source document(s) as a network of sentences. Using the query terms to select nodes which are related, the sentences are extracted. The summarisation tool is able to handle both similar and contrasting sentences across multiple documents.

- **InText Search Engine**

  This search engine development kit is used to make a search engine for a particular website. In doing so, it is supposed to allow dynamic creation of summaries of documents in the website when requested by users. There is no information as to how this is done on their website.

- **InText by Island Soft**

  InText extracts key sentences by using key words, although the exact technique is is not mentioned on their website. Their description mentions that the user may choose one of several extraction techniques. InText is a product that the user installs and uses on documents already residing on the computer.

- **British Telecom (ProSum) or NetSum**

  It is difficult to get to official site, so this paragraph is based on second-hand descriptions. British Telecom produced a summarization tool for both offline and online texts which works by selecting key sentences and extracting them as a summary. However, since ProSum and NetSum are commercially available products, the internal mechanisms behind the extraction process. Alternatively, instead of extracting the sentences, the tool highlights them in the original document. The user is able to change the length of the summary produced. According to researchers at the University of Ottawa, ProSum works best with factual documents of a single theme. These include genres such as news, paper articles and technical journals. It doesn't work as well with lists and narrative works.

- **inXight (LinguistX)**

  inXight's Summary Server is an application that creates extraction-based summaries offline. Users view summaries when they move their mouse over a hypertext link to a document that has been previously summarised. It uses statistical extraction techniques based on features such as sentence position, sentence length and keywords. The application allows the user to specify length and salience of certain keywords. It

also provides the ability for further training on structured documents of other genres.

- **Tetranet Software (Extractor)**

Extractor picks out keywords from a document. From my home page, extractor produced the following list:

- Natural Language
- Communication Sciences
- Intelligent Interactive Technology
- CMIS
- Technology Group
- Macquarie University
- Honours student.

Extractor simply finds key phrases and the sentences that use them. The extractor algorithm uses a set of parameters (such as stem length) which are tuned by a genetic algorithm (GenEx). The extracted phrases are then matched with their occurences in the document and the corresponding sentence is extracted. There is simple stemming and morphological analysis of extracted key words to score non-noun phrases lower, however, there is no synonym detection. The list of extracted sentences is filter based on heuristics regarding presentation.

- **IBM Japan Summarization Project**

  "The importance of a sentence is determined by some surface clues such as the number of important keywords, the type of sentence (fact, conjecture, opinion, etc.), rhetorical relations in the context, and the location in which a sentence exists in a document."

- **Copernic Summarizer**
- **Sinope Summarizer**
- **Pertinence Summarizer**
- **Microsoft (MS Word Auto Summarization)**
- **Apple**
- **Discontinued Summarisation Products**
- **Webcompass by Quarterdeck**
- **Sakh Arabic Summariser**

المصادر والمراجع

قائمة الكتب

1- الإبراهيمي، خولة طالب ، مبادئ في اللسانيات، دار القصبة للنشر 2000

2- أمان، محمد محمد ، خدمات المعلومات والإحاطة الجارية، الرياض، دار المريخ، 1985

3- بن عمر، محمد صالح ، الثورة التكنولوجية واللغة، ط1، دار الشؤون الثقافية العامة – وزارة الثقافة والإعلام، بغداد – العراق، 1986م

4- ابن منظور، محمد بن مكرم، لسان العرب، الناشر: دار صادر – بيروت، 2010

5- جرجيس، جاسم محمد، وعبد الجبار عبد الرحمان، المراجع والخدمات المرجعية، مركز التوثيق الإعلامي (1985) لدول الخليج العربي

6- حشمت، قاسم، خدمات المعلومات : مقوماتها وأشكالها، القاهرة : مكتبة غريب، (1984)

7- الحليمي، رفيق حسن، فن كتابة التلخيص والمختصرات، مركز المخطوطات والتراث والوثائق، 2002 الكويت ط1

8- الخويسكي، زين كامل، قواعد النحو والصرف، الوفاء لدنيا الطباعة والنشر مصر 2002

9- خياط، محمد غزالي، تمثيل الدلالة الصرفية في النظم الآلية لفهم اللغة العربية، دار تعريب، 2001

10- زغيبي، علي نبيل، اللغة العربية والحاسوب، دراسة بحثية، شركة العريض (2002)

11- العارف، عبدالرحمن بن حسن، جامعة أم القرى، توظيف اللسانيات الحاسوبية في خدمة الدراسات اللغوية العربية " جهودٌ ونتائج "، مجمع اللغة العربية الاردني، 2016

12- العناتي ، وليد أحمد ، اللسانيات الحاسوبية العربية – رؤية ثقافية، المعرفة، 2015

13- عز الدين، محمد، الناقل العربي : تصميم برنامج للترجمة الآلية، دار الهدى، تونس، 2008

14- القزويني، جلال الدين محمد بن عبد الرحمن، كتاب التلخيص في وجوه البلاغة ضبط وشرح عبد الرحمان البرقوقي دار الفكر العربي 1999

15- الموسى، نهاد ؛ العربية – نحو توصيف جديد في ضوء اللسانيات الحاسوبية، دار الفكر، 2000م

16- نبيل، علي، "اللغة العربية والحاسوب"، دار تعريب، 1998

17- الوعر، مازن، قضايا أساسية في علم اللسانيات الحديث، الوفاء لدنيا الطباعة والنشر مصر 2009

18- Beau, Michel, L'art de la thèse : comment préparer et rédiger une thèse de doctorat, de magister ou un mémoire de fin de licence, Guide approches, Casbah éditions 1999.

19- Benbrahim, Ahmad, Text Summarisation : the Role of Lexical Cohesion Analysis. The New Review of Document & Text Management, p.321-335, 1995.

20- Bloomfield, Leonard. .Language. Revised from 1914 edition. New York: Holt. 564 pages. (Fischer-Jorgensen 1975.) Location: Dallas SIL Library 410 B655l. Interest level: academic, 1933

21- Bonneval, Pière et André, Le résumé, Agnès, edition Eska 1997

22- Bouchard, Monique et Lespingal, Brigittre Reaute, Le résumé de texte, Edition Hachette, 1993

23- Chomsky, Noam, Syntactic Structures, Mouton France 1957

24- Chomsky, Naom, Aspects of the theory of syntax. MIT Press USA 1965

25- Conte, Alain, La révolution d'informations : l'utilisation des ordinateurs pour le stockage et la récupération d'informations,

traduction Hechmat Kacem et Chawki Salem 2ème édition, Kuwait : l'agence des publications, p172, 1979.

26- Coullomb, Renaud, Méthodologie du résumé de texte en anglais, Edition Ellipses 2007

27- Geneviere, Clerk, 50 modeles de résumés de textes, edition Marabout 1992

28- Gomes, Claudia de Olivira, Réeussir la note de synthèse, Groupe Studyrama-Vocatis, 2010

29- Guenitcha, Cleber et Manou, Michel, Introduction générale des sciences et techniques d'informations et d'archivage, UNISCO, p171, Tunis, 1987

30- Joshi, A. K. An Introduction to Tree Adjoining Grammars. In: A. Manaster-Ramer, Ed., Mathematics of Language. John Benjamins, Amsterdam. Holand 1987.

31- Minel, J. Luc, Filtrage sémantique, Edition Lavoisier, 2001.

32- Moreno, Juan Manuel Torrres, Resume automatique de documents, Lavoisier 2011, Paris, Hermes

33- OWENS, Jonathan, The foundations of Grammar: An Introduction to Medieval Arabic Grammatical Theory, éd. J. Benjamins (SiHoLS, 45), Amsterdam, (1988).

34- Owlley, Jennifer, Abstacting and indexing, London : Clive Bingley, p12-20, 1982

35- Paice, D. The Automatic Generation of Literary Abtracts : An Approach based on Identication of Self-indicating Phrases. In Norman, O., Robertson, S., van Rijsbergen, C., and Williams, P., editors, Information Retrieval Research, Butterworth, London 1981.

36- Régine, Prime, La methode pas à pas de la contraction de texte, Edition Ellipses, 2015

37- Sabah, Gerard, Le TALN & les modèles de représentation des connaissances, Hachette, 1988

38- Salton, G., Singhal, A., Mitra, M., and Buckley, C., Automatic Text Structuring and summarization. Information Processing & Management, 33(2) :193-207, 1997, USA.

39- Yvon, François, Introduction au traitement automatique des langues naturelles, Hachette, Paris, 2000

الرسائل الجامعية

40- حلمي، علي موسى ، دراسة إحصائية لجذور معجم الصحاح باستخدام الكمبيوتر، جامعة الكويت 1973

41- A. Chen and F. Gey : Building an Arabic Stemmer for Information Retrieval. Proceedings of the Eleventh Text Retrieval Conference (TREC 2002). National Institute of Standards and Technology, Nov 18-22, 2002, pp631-640.

42- DUBEAU, Michel, L'activité resumante, le resume de texte, aspects didactiques, collection didactique des textes, université de Metz, 1992

43- El Kassas, Dina, thèse de doctorat 'Une étude contrastive de l'arabe et du français dans une perspective de génération multilingue', Université Paris 7, 2005

44- Farzindar, Atefeh, Développement d'un système de résumé automatique de textes juridiques, Université de Montréal 2001.

45- Ferfera, Soufiane, Proposition d'une architecture multi-agents pour l'analyse syntaxique de la langue arabe, mémoire de magister, CRSTDLA, Alger, 2003

46- F. S. Douzidia, Résumé automatique de texte arabe
Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de M.Sc en informatique
Septembre, 2004, Université de Montréal, Canada, 2004.

47- Hadj Salah, Abderrahmane, Linguistique arabe et linguistique générale : Essai de méthodologie et d'épistémologie du "ilm al-arabiyya", Sorbonne, Paris, 1979.

48- Horacio, Saggion, Génération automatique de résumé par analyse sélective, Thèse de doctorat, Université de Montréal 2000.

49- Loukam, Mourad, une plate-forme d'analyse basée sur le formalisme HPSG pour l'Arabe standard,
Université de Chlef 2008

50- Mammeri, Mahmoud Fawzi, Une approche pour l'analyse syntaxique de l'arabiyya basé sur la TNK et sur la grammaire syntaxique guidée par les têtes,
Mémoire de magister CRSTDLA, Alger, 2003

51- Nacer, Hassina, Conception d'un système expert pour le diagnostique de l'aphasie selon le modele linguistique neo-khalilien, mémoire de magister, CRSTDLA, Alger, 2001

52- OUERSIGHNI, Riadh, Modélisation des expressions figées en arabe en vue de la constitution d'une base de données lexicale. Thèse de doct., Univ. Lyon 2. 2001

53- Shariful Islam Bhuyan, and Reaz Ahmed, AN HPSG ANALYSIS OF ARABIC VERB, Department of Computer Science and Engineering, 2009 Bangladesh University of Engineering and Technology, Dhaka-1000, Bangladesh

54- Y. Kadri, A. Benyamina, Système d'analyse syntaxico-semantique du langage arabe, mémoire d'ingénieur, université d'Oran Es-sénia, 1992.

المقالات

55- حاج صالح، عبد الرحمان، منطق النحو العربي والعلاج الحاسوبي للغات، مجلة المبرز، مجلة دورية أكاديمية، المدرسة العليا للآداب والعلوم الانسانية، العدد الثالث 1993

56- حاج صالح، عبد الرحمان، الجملة الاسمية في كتاب سيبويه، مجلة المبرز، مجلة دورية أكاديمية، المدرسة العليا للآداب والعلوم الانسانية، العدد الثاني 1993

57- خضر، محمد زكي، الحروف العربية والحاسوب، مجلة مجمع اللغة العربية، ع78 عمان (2001)

58- الشاعر، شفيق سالم . معلم بوكالة الغوث . فن التلخيص ، مجلة المعلم تربوية ثقافية جامعة، ع134، 2006

59- عمشوش، مسعود، التاريخ : 2013. مجلة المعرفة, عن دار جامعة عدن للطباعة والنشر، ع102 التصنيف : دروس, مقالات, مهارات

60- د. محمد الحناش، مشروع نظرية حاسوب لسانية في بناء معاجم آلية للغة العربية، مجلة التواصل اللساني، المجلد2، العدد2، 1990م

61- محمد صاري، المفاهيم الأساسية للنظرية الخليلية الحديثة، مجلة اللسانيات، مجلة في علوم اللسان وتكنولوجياته، العدد العاشر، مركز البحوث العلمية والتقنية لترقية اللغة العربية ، الجزائر 2005

62- محمد فتحي عبد الهادي، العمليات الفنية في المكتبات ومراكز البحث :
(1983) المجلة العربية للمعلومات، مج4، ع2

63- د.محمود إسماعيل صيني، بنوك المصطلحات الآلية، مجلة اللسان العربي، العدد 48، 1999م

64- النوايسة، غالب عوض، الاستخلاص لأغراض استرجاع المعلومات، مجلة رسالة المكتبة، ع65، (1999) الأردن

65- اليوسفي، أحمد شعيب، كلية الآداب، تطوان، نحو نظام جديد لإدارة المعلومات، مجلة آراء، ع8، 2015

66- A.M. Mutawa , Salah Alnajem , Fadi Alzhouri

An HPSG approach to Arabic nominal sentences

Journal of the American Society for Information Science and Technology, Volume 59 , Issue 3 Pages: 422-434 ISSN:1532-2882 (February 2008)

67- Cours « Information scientifique et technique » 1ère année magister Journalisme scientifique, faculté des sciences, Université de Blida, 2006/2007

68- Edmundson, H. P. New methods in automatic abstracting, *Journal of the Association for Computing Machinery (ACM), vol. 16 N°2,* pp. 264-285, April 1969.

69- GECSEG, ZSUZSANNA, L'adjectif relationnel et le complément du nom, Revue d'Études Françaises No 2 France 1997

70- Liddy, E. The discourse-level structure of empirical abstracts : an exploratory study. Information processing & management, 27(1) : 55 81, USA, 1991.

71- Luhn, P. H. The Automatic Creation of Literature Abstracts, IBM Journal, pp. 159-165, April 1958.

72- Monod, Morad, Le résumé automatique, un petit état de l'art, revue des arts, n66, 2004

73- Mourad, Amine, Glossaire de linguistique computationnelle, Lavoisier, n43, Paris, 1995.

74- Pollard, Carl et Ivan Sag, Head-driven Phrase Structure Grammar, The University of Chicago Press USA, Published August 1994

75- Y. BAHOU L. HADRICH BELGUITH C. ALOULOU A. BEN HAMADOU, Adaptation et implémentation des grammaires HPSG pour l'analyse de textes arabes non voyellés, 2009, Laboratoire de recherche LARIS, Faculté des Sciences Economiques et de Gestion de Sfax – TUNISIE

أعمال الملتقيات والندوات

76- استخدام اللغة العربية في تقنية المعلومات، مطبوعات مكتبة الملك عبد العزيز العامة، الرياض

77- عبد الرحمان حاج صالح، المدرسة الخليلية الحديثة ومشاكل علاج العربية بالحاسوب، مؤتمر اللغويات الحاسوبية بالكويت، 1989

78- وثائق ملتقى فعالية اللسانيات الحاسوبية العربية جامعة ورقلة 2014

79- Barzilay and Elhadad, Using Lexical Chains for Text Summarization. In Proceedings of the ACL/EACL'97 Workshop on Intelligent Scalable Text Summarization, pages 10-17, Madrid, Spain 1997.

80- Blache P. “Introduction à HPSG” in 2LC-CNRS, France 1995

81- Ishikawa, K., Ando, S., Okumura, A., Hybrid Text Summarization Method based on the TF Method and the Lead Method. *Proceedings of the Second NTCIR Workshop Meeting on Evaluation of Chinese & Japanese Text Retrieval and Text Summarization. Tokyo.Japan. pp.5-219-5-224, March 2001.*

82- J. Baxendale, Man-made Index for Technical Litterature - an experiment. IBM J. Res. Dev., 2(4) :354-361, 1985.

83- K. Darwish, Building a Shallow Arabic Morphological Analyzer in One Day. *Proceedings of the workshop on Computational Approaches to Semitic Languages in the 40th Annual Meeting of the Association for Computational Linguistics, (ACL-02), Philadelphia, PA, USA. pp. 47-54.*

84- Larkey, L. S., Ballesteros, L. and Connell M., improving Stemming for Arabia Information Retrieval: Light Stemming and co-occurrence Analysis, In proceeding of the 25th annual International conference on Research and development in information Retrieval (SIGIR 2002), Tampere, Finland, August 2002.

85- Lehmam, A. Automatic Summarization on the WEB. A System for Summarizing using Indicating Fragments : RAFI. In Proceedings of Computer-Assisted, Information Searching on Internet Conference. RIAO'97, pages 112-122, McGill University, Quebec, Canada, 1997.

86- Lin and Hovy, Identifying Topics by Position. In Fifth Conference on Applied Natural Language Processing, pages 283-290. Association for Computational, Linguistics 1997.

87- Mammeri, Mahmoud Fawzi & Bouhassain, Nacer Eddine

Université de Blida, 3rd International Conference on Arabic Language Processing, CITALA'2009, Implémentation d'un fragment de grammaire HPSG de l'arabe

sur la plate-forme LKB

88- Pardo, T.A.S. L.H.M. Rino, M.G.V. Nunes,

Extractive summarization: how to identify the gist of a text,

International.Bibliographie 63 *Information Technology Symposium - I2TS 2002, Florianópolis-SC, Brazil, pp.245-260, 01-05 October 2002.*

89- Strzalkowski, T., Wang, J. and Wise, B., Summarization-based Query Expansion in Information Retrieval, *Proceedings of 36th Annual Meeting of the ACL, V. 2, pp. 1258-1264, Montreal 1998.*

90- Udo, Hahn, l'atelier de résumé automatique, Seattle, USA 2000

91- ZALILA, Inès & HADDAR, Kais, Étude et analyse des relatives de la langue arabe en HPSG et LKB, 3rd International Conference on Arabic Language Processing - CITALA 2009 – May 4th – 5th 2009, Rabat, Morocco

المواقع الالكترونية

92- أبو العزم، عبد الغني
http://www.aljabriabed.net/n31_04abualazm.%282%29.htm

93- أنوزلا، علي
https://www.alaraby.co.uk/opinion/2017/4/18/

94- مجلة (الجريدة) الالكترونية
http://www.aljarida.com/articles/14622721035883502 00
www.mahaarat.com

95- منتصر، عبد العاطي عبد المجيد، دروس في الترجمة التلخيصية،
http://faculty.ksu.edu.sa/montaser/DocLib8

96- عويس، دينا

http://mawdoo3.com

97- Abeillé, Anne, Cours DEA, www.llf.cnrs.fr/Gens/Abeille/hpsg-04.doc ,

98- Agence universitaire de la francophonie

Traitement automatique des langues

http://www.infotheque.info

99-ALRAHABI, Motasem et Ghassan, MOURAD et DJIOUA, Brahim,

LaLICC (Langage, Logique, Informatique, Cognition et

communication), UMR 8139, Université Paris –Sorbonne,CNRS-

France, Filtrage sémantique de textes en arabe en vue d'un

prototype de résumé automatique

www.lpl.univ-aix.fr/jep-taln04 2009

100- BOUKEDI, Sirine et LOUKIL, Noureddine & HADDAR, Kais

Génération d'un analyseur syntaxique basé sur HPSG avec le

LKB, www.emi.ac.ma/citala2009/docs

101- Bourbia, riad, Analyse Syntaxique du langage naturel

www.badri86.byethost2.com

102- Delafosse, Lionel, Présentation du TAL

http://www.cavi.univ-

paris3.fr/ilpga/ilpga/tal/cours/parcours/introtal.htm

103- Institut de technologie royal (Suède)

SweSum : un systeme de résumé automatique

www.swesum.se

104- Le site Atala Le traitement automatique des langues

http://www.atala.org/AtalaPedie/index.php

105- Le site http://faculty.ksu.edu.sa/montaser/DocLib8

106- Le site ics

Liste de projets de résumé automatique

www.ics.mq.edu.au/~swan/summarization/projects_full.htm

107- Le site apprendrearabe

http://www.apprendrearabe.com/frarabic/grammar.asp

108- Le site http://www.shatharat.net/vb/showthread.php?t=12168

109- Le site de l'université d'ottawa

http://www.site.uottawa.ca/tanka/ts.html

110- Le site de l'université de la Californie de sud

https://www.isi.edu/natural-language/people/hovy/papers/

111- Le site de l' université de Sheffield

http://nlp.shef.ac.uk/trestle/

112- Le site de l' université de Colombie

http://www.cs.columbia.edu/~hjing/sumDemo/

113- Le site :

http://www.mactech.com/content/md1-glucose-releases-data-hammer

114- Le site :

http://www.islandsoft.com/

115- Le site :

https://www.lesechos.fr/british-telecom.htm

116- L'encyclopédie wikipedia, http://fr.wikipedia.org

La grammaire syntagmatique guidée par les têtes HPSG

117- L'histoire de résumé automatique

www.ics.mq.edu.au/~swan/summarization/history.htm

118- Loukil, Noureddine et Haddar, Kais

et Ben Hamadou, Abdelmajid

Laboratoire de Recherche en Informatique et Multimédia, Sfax,

Tunisie, Normalisation de la représentation des lexiques

syntaxiques arabes pour les formalismes d'unification

www.miracl.rnu.tn

الفهـــــرس

الفهـــرس

شكـر

إهداء

مقدمة ..... أ

الفصل الأول

إستخدام علم الإعلام الآلـي في خدمة اللغة العربية

1ـ نبذة تاريخية

على المستوى الغربي 04

على المستوى العربي 07

2ـ ميادين تطبيقات الاعلام الآلي

الميدان الأول : الإحصاء اللغوي 07

الميدان الثاني : علم اللغة الحسابي 12

الميدان الثالث : التعريب 18

الميدان الرابع : العلاج الآلي للغة العربية 20

الميدان الخامس : الترجمة الآلية 32

الميدان السادس : الكتابة العربية ومعالجتها آليا 34

الميدان السابع : تعليم اللغة 36
الميدان الثامن : التلخيص 37
3ـ أعلام واتجاهات 40
4ـ أمثلة من تطبيقات حوسبة اللغة 42
خاتمة ومقترحات 43

الفصل الثاني

Les résumés et les résumés automatiques

A/ Les résumés

1. Les définitions ……..... 46
2. Apparition des résumés ……..... 48
3. L'importance des résumés ……..... 49
4. Les méthodes des résumés ……... 51
5. Les types de résumés ……..... 54
6. Les documents concernés par le résumé ……..... 57

II/ Les résumés automatiques

1. Le résumé automatique …… 65

2. L'histoire de résumé automatique des textes …… 67

3. Les approches de résumé automatique ……..... 68

3.1 Les méthodes d'extraction de phrases ……... 68

3.1.1 Méthodes à bases de mots clés ……... 68

3.1.1.1 Mots-clés prédéfinis …… 69

3.1.1.2 Titres ……..... 69

3.1.1.3 Méthodes de distribution de termes 70

3.1.2 La méthode de la position …….. 72

3.1.3 Expressions indicatives …….. 74

3.1.4 Cohésion lexicale ……..... 75

3.1.5 Classification des éléments ……. 78

3.2 Approches hybrides …… 79

3.3 Méthodes de compréhension et génération ……. 81

3.3.1 Scénarios ……..... 81

3.3.2 Instanciation de patrons et génération ……... 82

4. Les domaines d'applications des résumés automatiques..... 84

5. Avantages et inconvénients ……... 85

Conclusion ……..... 86

الفصل الثالث

Traitement automatique des langues et grammaires

1. Présentation du Traitement Automatique des Langues 89

1.1 Préliminaires ……..... 89

1.2 Les niveaux de traitement ……. 91

1.2.1 Analyse morphologique et lexicographique … 91

1.2.2 Analyse syntaxique ……..... 93

1.2.3 Analyse sémantique ……..... 94

1.2.4 Analyse pragmatique …… 95

1.2.5 Exemples de difficultés …….. 96

1.3 Les domaines du TAL ……..... 99

2. Une liste de projets de résumé automatique …… 100

2.1 Projets et outils du domaine de recherche …… 100

2.1.1 Extraction de phrase …….. 100

2.1.2 Arrangement profond …….. 103

2.1.3 L'approche hybride ……..... 104

2.2 Projets et outils du domaine commercial …… 106

3. Grammaires et formalismes grammaticaux …… 112

3.1 Grammaires classiques ……..... 112

3.2 Grammaire d'unification ……. 117

3.2.1 La grammaire lexicale fonctionnelle …… 117

3.2.2 La grammaire syntagmatique généralisée 118
3.2.3 Les grammaires syntagmatiques guidées par têtes 121
3.2.4 Les grammaires d'arbres adjoints …….. 122
Conclusion ……..... 123

الفصل الرابع

نهج مدمج لتلخيص النصوص العربية

1. تلخيص نص 125

2. خصائص التلخيص 129

3. فوائد التلخيص 129

4. كيفية تلخيص النصوص 130

5. قواعد لتلخيص النصوص 132

1.5. قاعدة الحذف 132

2.5. قاعدة الدمج 132

3.5. قاعدة البناء 132

4.5. قاعدة التعميم 133

6. خطوات التلخيص 134

7. الأسس الفنية للتلخيص 138

8. عرض وتلخيص 141

أ. طرق استخراج الجمل 142

أ.1. طرق الكلمات المفتاحية 142

أ.1.1. كلمات مفتاحية معرفة مسبقا 142

أ.1.2. العناوين 142

أ.1.3. طرق توزيع الكلمات 143

أ.2. طريقة الموضع 143

أ.3. التعابير الإرشادية 144

أ.4. التماسك المعجمي 144

أ.5. سلاسل معجمية 145

أ.6. استخراج الفقرات 145

أ.7. تصنيف العناصر 145

ب. النهج المدمج 146

جـ. طرق الفهم والتوليد 158

9. مزايا طرق الاستخراج 146

10. وصف الوحدات الرئيسة المكونة لأوجز 150

1.10. التجزئة للكشف عن الجذر 150

2.10. تحليل صرفي– معجمي 152

3.10. تحليل نحوي 152

4.10. تطبيق طريقة مدمجة للتلخيص (الاستخراج)..... 153

5.10. إستخراج الجمل 155

خاتمة 156

الفصل الخامس

تطبيقات النهج المدمج لتلخيص النصوص العربية

النص الأول 158

النص الثاني 169

النص الثالث 178

الخاتمة 189

الملاحق 195

قائمة المصادر والمراجع 207

الفهرس 234

قائمـــة الجـــداول

Liste des tableaux

Tab. 01 : Résumé des étapes de la méthode de distribution de termes 71

Tab. 02 : Résumé des étapes de la méthode de la position 73

Tab. 03 : Résumé des étapes de la méthode des expressions indicatives 75

Tab. 04 : Étapes de la méthode de cohésion …… 77

Tab. 05 : Étapes de la méthode des chaînes …… 77

Tab. 06 : Étapes de la méthode d'extraction de paragraphes …… 78

Tab. 07 : Résumé des étapes de la méthode de classification …… 79

Tab. 08 : Résumé des étapes de la méthode des scénarios …… 82

Tab. 09 : Résumé des étapes de la méthode des patrons …… 83

Tab. 10 : Avantages et inconvénients des méthodes de résumés …… 85

جدول 11: قوائم معظم السوابق المشتركة واللواحق 151

قائمة الصـــــور

صورة 1 لتصميم نظام أوجز 148

صورة 2 لتصميم نظام أوجز 149

ملخــص

أبرزت الزيادة الحادة في مقدار النص المتاح في الشكل الرقمي، الحاجة لتصميم وتطوير أدوات وبرامج تلخيصية فعالة لتحديد واستخراج المعلومات المهمة في شكل مختصر. إن توافر النصوص العربية ليست استثناء، ولكن هناك نقص كبير في المعالجة الآلية وخاصة الاستخراج. يتمثل بحثنا في اقتراح طريقة لانتاج ملخصات للنصوص العربية.

إن الأسلوب المنهجي المتبع هو دراسة الأساليب و التقنيات القائمة لاستخراج الملخص الآلي وكذلك تحديد و دراسة النماذج التصميمية لتمثيل النصوص.

ثانيا – وهو الهدف الرئيس من هذه الدراسة – سندرس خصائص اللغة العربية بغرض التكيف مع هذه الأساليب التصميمية لتوليد ملخصات آلية للنصوص العربية.

أوجز ، نظام التلخيص الآلي للنصوص العربية الذي صممناه يعتمد على تصاميم مناسبة للغة العربية و تقنيات الاستخراج التي أثبتت فعاليتها في لغات أخرى مثل الإنجليزية .

كلمات البحث: التلخيص التلقائي ، استخراج المعلومات ، أدوات مجردة تلقائية ، أوجز

Résumé

La forte augmentation du volume de texte disponible en format numérique a fait ressortir la nécessité de concevoir et de développer des outils de résumé performants dans le but de repérer et extraire l'information pertinente sous une forme abrégée. Les textes arabes ne font pas exception quant à leur disponibilité mais pour ces derniers, il y a un grand manque d'outils de traitements automatiques spécialement pour l'extraction.

Dans notre travail, nous nous proposons une méthode de production de résumés pour les textes arabes. Notre démarche méthodologique consiste en premier lieu à étudier les méthodes et les techniques existantes en terme d'extraction automatique de résumé ainsi qu'à recenser et étudier les formalismes conçus pour représenter les textes.

En second lieu, -l'objectif principal du travail- nous étudions les spécificités de la langue arabe en vu d'instrumenter et d'adapter ces méthodes et ces formalismes pour générer les résumés automatiques des textes arabes.

AWDJIZ, le système de résumé automatique de textes arabes que nous avons conçu est basé sur une formalisme adapté pour la langue arabe et les techniques d'extraction -qui ont déjà fait leurs preuves pour d'autres langues comme l'anglais- .

Mots clés : résumé automatique, extraction d'information, les outils de résumé automatique, AWDJIZ

Abstract

The increase in availability of text in digital format accentuates the need for design and development efficient summarizer tools to track and extract relevant information in a shortened form. arabic texts are becoming widely available but miss tools for its automatic processing. In this work, we propose to design a method for producing summaries for arabic texts.

Our approach to do this, consist on two steps, first, we study both the existents methods and techniques which are used to extract automatically summaries and the formalisms used to express and represent texts.

In second steep, wish is the main objective of this thesis, we study specificities of the Arabic language in seen to instrument and adapt these methods and these formalisms to generate the automatic summarization of the arabic texts.

AWDJIZ, the system of automatic summarization of arabic texts that we have conceived is based on a formalism adapted for the arabic language and the techniques of extraction which turn out to be effective for other languages such as English.

Keywords : text summarization, information extraction, summarizer tools, AWDJIZ.