مجلة اللغة العربية و آدابها
السنة الاولى-العدد الثاني-صيف و خريف ١٤٢٦ /٢٠٠٥م
ص١١٢-٩٥

مدخل على طريقة كمترة النصوص العلمية في اللغة العربية*

الدكتور سيد محمد محمودي**
الدكتور علي رضا محمدرضايي**

خلاصة:

ترسخت انظمة المعلومات الالكترونية، باتساع تكنولوجيا المعلومات و التبادلات الفكرية(ICT)[1] و بتطوّره، ترسخا شاملا ومسرعا الى جميع النشاطات البشرية و المكانيكية فاثرت تأثيراً عظيماً في ازدياد التدقيق في المعلومات و تأمينها و تسريعلا مبادلتها.

التطبيقات المتنوعة سهّلت تبادل الانسان و الاجهزة المكانيكية و هيأت ارضية فهم اللغات الطبيعية (نطقاً و كتابة) و ادراک الكامبيوتر اياهااي معالجة اللغات اتوماتيكياً.

كي يتمكن الكامبيوتران يفهم لغة طبيعية علينا قبل كل شيء ان نحلّل اللغة تحليلاً دقيقاً و ذلک يقتضي استخدام علوم مختلفة خاصة علوم الكامبيوتر،و علم المعرفة و علم اللغة.

الغرض الرئيس من هذا المقال الذي ظهر اثر الدراسات التطبيقية و العملية، هو تقديم الاساليب و طرق التحليل لاعداد اللغة العربية الكتابية و تنفيذها علي انظمة كامبيوترية كي نتمكن ان ندخل نصوص العربية الى الجهاز ثم نستخرج الكلمات الرئيسة بعد تحليل اتوماتيكي حتى نحصل علي ما يحتويه النص و علي مفهومه.

الكلمات الرئيسة: المعالجة الاتوماتيكية للغة الطبيعي[2]، الذكاء الاصطناعي[3]، وضع قائمة اتوماتيكي4، الانواع الاسمية5 الواصف[6]، محلّل(MS)[7]

1- Information & Communication Technology
2- Automatic Processing Of Natural Language
3- Artificial Intelligence
4-Automatic Indexing
5-Noun Phrases
6-Descriptor
7- Morpho-Syntactic

* تاريخ الوصول ٨٤/٣/٣؛تاريخ القبول:٨٤/٣/٢٩

** استاذان مساعدان في اللغة العربية و آدابها بفرديس قم

مقدمة:

منذ ثلاثين سنة اتسعت ارضية الحاسوب التطبيقية اتساعاً لا يتصور. تسرب اليوم الانظمة الكامبيوترية الى كثير من زوايا الحياة الاجتماعيـــة و الاقتـــصادية و العلمية و تزداد يوماً بعد يومٍ.الجهازات الكامبيوترية و الاســـتطلاعية تـــتمكن ان تلعب دورا رئيساً هاماً في تقدير المعلومات و معالجتها و في العزم على الامـــور و الرسم و الابتكار، والتعليم، و التعرف على المشاكل، و الدراســـات العلميـــة، و النظامات المتخصّصة و الخبيرة حتى فهم اللغات الطبيعية. للكامبيوتر – كما اشرنا آنفاً – تطبيقات متنوعة عديدة في الامورالانسانية العصرية فلنهتمّ منـــها بالمعالجـــة الاتوماتيكية للغات الطبيعيةالتي تعتبر فرعاً من فروع الذكاء الاصطناعي .اتـــسعت تكنولوجيا المعلومات الاتوماتيكي في الحقيقة باتساع اجهزة المعلومات الاتوماتيكية (محمودي، ١٣٨١ هـــ، ص ١٩٤،١٧١).

المعالجة الاتوماتيكية لكل لغة نموذج علمي جديد متطوّر له تطبيقات متنوعة و فروع عديدة يمكن الاشارة منها الى الترجمة الاتوماتيكيـــة،و تحليـــل الـــنص و المحتوى،و تخليصه، و معالجة الكلام و الصورة و وضع قائمة اتوماتيكياً. التطبيـــق الهامّ و الاخير لمثل هذه الاجهزة هو فهم اللغات الطبيعة و ادراكها كالانجليزيـــة و الصينية و الفرنسية و الفارسية و العربية و غيرها بالكامبيوتر.

ففي هذا المقال نشرح بصورة موجزة **المراحل المختلفـــة لابـــداع محلـــل** (MS) [1] الذي يستخرج محتوى النص اتوماتيكاً ضمن تقديم **الطرق لفهم نصوص العربية بالكامبيوتر:**

1-Morphosyntactic.

طرق ادراک النص و تحليله اتوماتيكاً:

كما اشرنا آنفاً - يعتبر وضع الفهارس الاتوماتيكي احدى التطبيقات الهامة لـ "PNL". وضع الفهارس الاتوماتيكي الذي يبتني على تحليل علمي للغة يطلق على تيار يتمكن الكامبيوتر به ان يحلّل النص بعد وروده اليه و يستخرج محتواه للمستخدمين.

الغرض من وضع الفهارس الاتوماتيكي نفس ما يكون في الاعتيادي من التعرف و تفريق عناصر المعلومات و عواملها التي يحتوي عليها النصّ ثم تفصيلها و استخراجها ثمّ تنسيقها و تصنيفها في بنک للمعلومات حتى يمكّن الكامبيوتر ان يدرک النصّ و يصفها (لو گرن[1]،١٩٩١،ص٢٣)

احدى التطبيقات الهامة لوضع الفهارس اتوماتيكاً هو البحث السريع و السهل بالكامبيوتر عن المعلومات المطلوبة في النصوص الكبيرة و العبارات الطويلة.

ينفذ وضع الفهارس الاتوماتيكي بتحليل النص تحليلاً اسنادياً لا يمكن الوصول اليه الا بعد ابداع محلّل اتوماتيكي و تشغليها على الجهاز حتى يتمكن ان يحقق تشكيل النص و ترسيمه ثم استخراج المعنى النهايي لكل وثيقة او نصٍ.

لوضع الفهارس اتوماتيكياً اساليب فلنقف لحظة إلى كلّ منها:

١-١- الطريقة الاحصائية:

جدير بالذكر ان الغرض الرئيس من وضع الفهارس الاتوماتيكي هو استخراج معلومات ذات معنى في كل نص لفهمه و ادراكه او بعبارة اخرى هو البحث

1- Le Guern

الاتوماتيكي عن المعلومات في كل نص. احدى الطرق القديمة لوضع الفهارس هيالتحليل الاحصائي لانتخاب العناصر التي تكون ذات معنى في مجموعة المعطيات. يعالج نظام وضع الفهارس في هذه الطريقة الاحصاء الصرف لتراوح المفردات و تقديرها التي تظهر في النص فتمكن بقياس تراوحاتها ان نتلقى تلقيات شتى من النص. و ان كان هذا الاسلوب سهلا و لكن ليس له تاثير وافر و شامل لان نصوص العربية مليئة باستعارات وكنايات و تشبيهات لا تظهر في النص فقط بل يتلقى منها تلقيات مختلفة فلهذا اذا افترضنا ان يستخرج المفردات من النص فليس هذا مبيناً لمعنى النص الحقيقي بل الانسان هو القادر على ادراكه فقط.مثال ذلك في اللغة العربية اذا استخدمنا عبارة: «هو ربيب ابي الهول» فنحن نقصد منها شدة الكتمان للسرّ و لكن الحاسوب لا يتمكن ان يتلقى هذا المعنى و إن استخرج كلماتها. هذا و هناک كلمات لاتظهر في النص الا قليلاً ووراءها معان عميقة فكيف يمكن عبر احصاء تراوحاتها ان يبلغ الى معانيها الحقيقية (بوشه[1]، ١٩٨٨م،ص٢) على ذلک أنّ هذا الاسلوب الاحصايي يفيد للمعطيات الاحصائية و الكمية و لا يفيد لكيفية افادة مطلوبة.

٢-١- طريقة الاختطاط الارجاعي القياسي[2]

لنافي هذه الطريقة نموذج قياسي من النصوص المختلفة يسمى «بعالم المرجع» يمكننا أن نقيس نصنا بنص المرجع لنبلغ الى تحليل ما و نتمكن أن نتعرف على معنى كل نص قانوني او طبّي او ...و نتمكن ان نصل الى نتائج رائعة تبلغ معنى

1- Bouché 2- Projection

النص الى حدّ ما.هذا المنهج صعب جدّا و يحتاج الى تقنيات الذكاء الاصطناعي التي يضيق المجال لتحدّث عنها.

٣-١- طريقة تحليل صدي الاعماق

هذه الطريقة التي تكون اسهل من السابق تقوم على تمثيل جمل النص لتجد بعض الكلمات الرئيسة التي تظهر في النهاية في جدول ما. فلهذا لا يرى اكثر الكلمات لاسيما الافعال و القيود حتى الصفات. تفيد هذه الطريقة للنصوص الكلاسيكية و المعينة. لكنها لا تضمن ان تستخرج ما تحتويه النصوص الطويلة من المعاني الكلية.(تاون سند[1]،١٩٨٨م،ص٤).يستخدم هذا الاسلوب كثيرا لفحص البريدات الالكترونية: فاذا اردنا ان نتعرف على البريدات التي تحتوي على كلمات مثل: الاسلحة، الارهاب، المحترقات، يعتبر هذه كلمات رئيسة. فيتمكن الحاسوب بنظامه الاتوماتيكي ان يدّخر آلاف آلاف بريدٍ الكتروني يتضمن هذه الكلمات. حتى يمكن تحليلها تحليلاً ذهنياً.

٤-١- طريقة فقه اللغة

اذا اراد الحاسوب ان يفهم لغة الانسان فعليه ان يدرک قبل کل شيء القواعد و المنطق الحاکم على اللغة. فنعرف من هنا أن دور علم اللغة لمعالجة اللغات اتوماتيكية دور حياتي. لا يمكن الاجتناب عنه(کَرِه[2]،٢٠٠٤م،ص٥٩).

هذه الطريقة تبتني على التحليل المتزامن للتراكيب الشكلية و المعنوية: بعبارة اخرى علينا في هذا التحليل الذي يقف على فقه اللغة ان ندرک و نعرف عناصراللغة: کالحروف و المفردات والجمل التي لکل واحدة منها هوية صرفية او

1-Townsend 2-Carré

نحوية مستقلّة او مرتبطة تستنبط منها معان مختلفة في ظروف او مقامات مختلفة. على أنّ طريقة تحليل الكلام بفقه اللغة تعني طريقة تحليل اللغة الصرفي و النحوي الذي يمكّننا أن نحدّد محتوى النص بتحديد الاسماء في كلّ وثيقة.

كما ذكرنا سابقاً ان الطرق التراثية كانت تقوم على استخراج الكلمات المكرّرة بل الكلمات الرئيسة و احصاءها و التعرف عليها احياناً.لكنها ضعيفة جداً. و اما بهذه الطريقة الجديدة لا تعالج الكلماته الجذلة او الهامّة فحسب بل تعرف و تسخرج جميع الواصفات بتقنية ذكيةالانواع الاسمية : الاسماء المنعوتةو الصفات و الموصوفات و المضاف و المضاف اليه و الظروف التي على ما يعتقده ارسطو، تتداعى لنا جوهر النص المعنوي.

الواصفات التي تطلق عليها الانواع الاسمية لا تعبر كلماتٍ للنص فقط بل مجموعة ذات معنى في كل حوار او مبحث يمكن بها أن ندرک ما يحتويه النصّ. (لو گرن[1]،١٩٩١م،ص٤٣) يقصد من الانواع الاسمية مجموعة من الكلمات المركبة (وصفياً او اضافياً) تظهر في النص كواحدة ذا معنى نحو: كلمة «كتاب» لا يعتبر من الواصفات لانّه ليس مركباً بل مفرداً و لكنه اذا اضيف إلى كلمة «زيد» (كتاب زيد) ظهر المقصود اكثر دلالة و دقّة و معنى (روبرت[2] ،٢٠٠٥م،ص٤٣) يذهب «سوسو» في كتابه الشهير: «علم اللغة العام» الى ان الانواع الاسمية تتألف على الاقل من كلمتين. فنحن نرى الانواع الاسمية الكثيرة واصفاً لانّ دلالتها عامة لكننا اذا ادخلنا عليه «ال» و اصبحت معرفة (الجميل) فانها ايضاً تعتبر من الواصفات اي الانواع الاسمية.

1- LeGuern, 2-Robert

جدير بذكر أن هناک انواع فعلية يعتبرها المتخصصون بمعالجة اللغة الاتوماتيكية عبارات مكمّلة تظهر في جملة، نحو:"النهيرات تصنع الانهار الكبيرة "ففي هذه الجملة نرى «النهيرات» و« الانهار الكبيرة» من الواصفات و فعل «تصنع» التحليل و تنظر اليها بصورة مستقلة و منفصلة عن الاخرى. بناء على هذا نتمكن ان نضع اول قاعدة لمحلّلنا الشهير (M.S):

P=NP+VP	Phrase=P
VP=NP+V	Noun phrase=NP
NP=n+a	Verb Phrase=VP
	Verb=V
	Adjective=a
	Name=n

جدير بذكر ايضاً ان كل محلّل (M.S) يحتاج الى نظام من القواعد نتعرض لها في المستقبل.بناء على ما قدّمنا من اسلوب جديد فاذا ما تمكنا ان نرسم محلّلاً اتوماتيكياً يستخرج الانواع الاسمية في كل نصٍ، بالقواعد المحدّدة من قبل، و ينظمها و يدخرها في بنک من المعلومات، فقد تقدمنا الخطوة الاولى للتعرف على ما يحتويه النص من المعنى، الخطوات التالية تحتاج الى ترسيم نظامات كامبيوترية ذكية جدّاً تستلزم دراسات و بحوث علياء نتحدث عنها في فرصة اخرى.

ما أنس لا أنس أنّ هناک اساليب اخرى معقدة و مؤثرة جداً تكمل طريقة التحليل بعلم اللغة بل اعمق غورامنها في مستوى التحليل، فعلى سبيل المثال؛ لنا

التحليل المنطقي– الدلالي[1] (اي التحليل بعلم المنطق و علم دلالات الالفاظ و تطوّرها) يجرى، بعد اسلوب التحليل بعلم اللغة، بالبحث عن الظروف المنطقية الدلالية لكل نص ثم يقرب المحتوى المستخرج، بتحليل التشبيهات و الاستعارات و الكنايات و الكلمات الغامضة، من المحتوى الواقعي القابل للادراک.

كما ذكرنا مرّات عديدة ان اول خطوة لتحليل النص بعلم اللغة في اللغة العربيةهو ترسيم محلّل M.S الاتوماتيكي.

٢ ــ المراحل المختلفة لترسيم محلّل(MS)

الغرض من ترسيم هذا المحلّل (M.S) هو اعداد برنامج كامبيوتري يتمكن ان يتلقى النص ثم يستخرج، بعد التحليل، الانواع الاسمية و يذخرها في بنک من المعلومات. الترسيم يشمل تعريف الامكانيات الضرورية و الاهداف و تعيينها ثم توصيف النظام و تحليله تحليلاً كاملاً باعداد قواعد خاصة و تدوينها حتى نتمكن ان نفسّر و ندرک عناصر اللغة. لترسيم كل محلّل نحتاج إلى إلمام كامل بعلم اللغة و قواعد الصرف و النحو ثم التعرف على لغة وضع البرامج الكامبيوترية خاصة لغة المعالجة الاتوماتيكية يعني Prolog .على أنّ علم اللغة و علم الحاسوب من الضروريات لتحليل كل لغة بالحاسوب.فلنقف لحظة الى كل من مراحل الترسيم:

المرحلة الاولي: تصنيف[2] **الاجزاء التي تركّب اللغة.**

الغرض من التصنيف هو التعرف على الوحدات المشكّلة و تحليلها: مثل: الاسماء و الصفات و الافعال و غيرها حتى نتمكن أن نخصّص بها قيماً و معالم ظاهرية. علينا ان ننتبه بان قواعد التسمية في الصرف و النحو معينة ثابتة لكننا عند المعالجة

1-Logic Semantics 2-Categorization

الاتوماتيكية لابدّ لنا ان نضع علامات مختصرة ليدركها الحاسوب عند التحليل فالاسماء هي الاسماء و الافعال هي الافعال لكننا نضع للاسم حرف «n» و للصفة «a» و للفعل «V» فنحن نقترح لاجزاء الجمل في العربية العلامات و الرموز التالية:

١- الاسماء المشتركة = nco

٢- الاسماء الخاصة = npr

٣- المصدر = inf

٤-الحروف الجارة = e

٥-الفعل = V

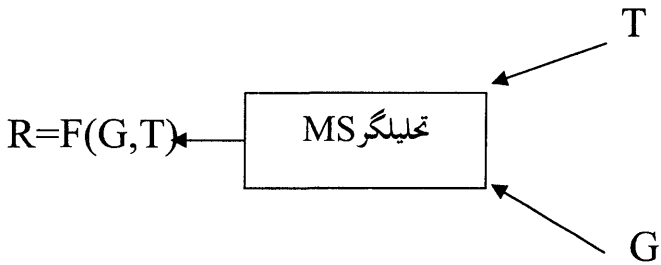
٦-القيود = w

المرحلة الثانية: استخراج القواعد

لكلّ لغة قواعد خاصة لتركيب الكلمات و تنسيقها في الجمل. لا يمكن الفصل بين المتغيرين: النصّ و القواعد فليس لاحدٍ دون الاخر معنى.

يتلقّى كلّ محلّل M.S نصوصاً مختلفة فيحلّل النص بقواعد خاصة ترتبط بتركيب الجمل و بالانواع الاسمية و الفعلية و غيرها حتى تتمكن أن تستخرج نتائج تسمّي اصطلاحاً «R» (Result) فاذا واجهنا نصاً (T) و قواعد خاصة(G) يمكننا ان نقدّم الصياغ التالي على ما بحثه (براندونه[1]،١٩٨٣م،ص٢٥)

1- Berrendonner



فعلى المحلّل أن يستخرج القواعد من النصوص الكلاسيكية و يدخرها في بنک من المعلومات أن يعمل على النموذج السابق فعلينا بعد التعرف على الانواع المختلفة أن نعدّ صياغاً سهلاً، في النصّ يبين جميع قواعد اللغة

نحو:

-عوى كلب زيد الاسود
-جاء زيد من المدرسة
-ارجو ارجاع الكتب إلى المكتبة

1-p = v+n+n+a
2-p = v+n+e+n
3-P = v+I+n+e+n

المرحلة الثالثة: استخراج قواعد لتجديد كتابة الانواع الاسمية:

كما ذكر مرات عدّة كان غرضنا تأسيس نظام كامبيوتري لادراک ما يحتوي النصوص العلمية في اللغة العربية. قدّمت طرق مختلفة. يكون التعرف الاتوماتيكي على الانواع الاسمية اكثر حسناً من الاخرى فعلينا ان نبحث عن طريق نتمكن به أن نحدّد و نستخرج الانواع الاسمية من النص. ثم يقيم معالم القواعد العامة للغة و يدخره في محطة القواعد. ففي هذه المرحلة علينا ان ندرس النصوص و نحدّد القواعد المرتبطة بكيفية تركيب الانواع الاسمية و نضع لكل منها رسماً خاصاً مثلا

نتمكن ان نضع للعبارة الاولى(كلب زيد الاسود)، وهي نوع اسمي،نظرا الى تركيبه مع الفعل، قياسا جديداً:

P=NP+v

المرحلة الرابعة: تنظيم البرامج لمحلّل M.S

لايقدر الحاسوب على اقلّ عملٍ او احصاء و حساب. فعلينا أن نضع برامج ليتمكن ان يقوم باعمال ذكية:

١-برنامج «المستخدم»[1]الذي يختص بتلقي النصوص الواردة ثم ادخارها

٢-برنامج لفصل الكلمات[2] و ادخارها في بنک من المعلومات منفردةً

٣-برنامج لانشاء بنک للمعلومات لمفردات اللغة[3].

٤-برنامج لايجاد محطة القواعد النحوية[4] تدخر بها انواع القواعد النحوية المرتبطة بالجملة و الانواع الاسمية.

المرحلة الخامسة: الاستقرار و الاختبار و تنفيذ البرامج:

بعد ان وضع برنامج، يجب ان يعدّ على الحاسوب ثم يختبر بالتشغيل حتى يتبين ما له من سقم او صحة.فاذا عرفت صحته فيتمكن المحلّل (M.S) لانه يتمكن، بالترقيم و وضع العلامات كالنقطة و الفارزة والفاصلة المنقوطة و النقتطين و علامة السؤال و التعجب، أن يميز بين الجمل؛ لأنه يكفي للمحلّل أن يقدّر النص. و عندما بلغ الى علامة من هذه العلامات يعرف بين العلامة السابقة و اللاحقة ثم يدخرها في البنک على نظام واحد.

٢-بعد استخراج الجمل يبلغ الدور إلى الفصل بين الكلمات في كل جملة.

1-User **2-Segmentation** **3-Lexicon** **4-Syntactic**

وضع برنامج المحلّل كي يتمكن ان يفصل المفردات عن الجمل. فهذه العملية سهلة جداً ؛اذ ان المحلّل بمعالجة الجمل عند ما بلغ الى فراغ يتعرف على سلسلة ثم يدخرها في بنک.

٣-بعد الفصل بين الجمل و المفردات يبلغ الدور إلى عملية ذكية تسمّى **التحليل الصرفي و النحوي**[1]. في كل تحليل موروفولجي وسينتاكتيكي تعرف قيمة المفردات الاعرابية و دورها في الجمل ثم تدخّر في بنک ما.

اذا عرف ماهية الكلمات تقاس الجمل المدخرة بالقواعد النحوية في محطتها.

٤-اما المرحلة الاخيرة فهي تجزئة الجمل و تحليلها ثم استخراج الانواع الاسمية و ادخارها في بنک المعلومات الخاص يمكّن المستخدمين ان يتلقوا ما يحتويه النص.

نتيجة البحث:

المراحل التي قدّمت لوضع محلّل M.S ما كانت غير مقدمة لاستنباط محتوى كلّ نص في اللغة العربية بالحاسوب. فاذا اردنا ان نحلل نصاً باللغة العربية بالحاسوب بصورة عميقة و مؤثرة ثم ننال الى تلقيه فعلينا الانتباه بما يلي:

١-لنحاول أن نختار نصاً كلاسيكياً للتحليل؛ لان الحاسوب لا يتمكن ان يدرک الشعر و الادب و ظرائف ... جدير بالذكر ان لا يوجد هناک بعد نظام كامبيوتري، على رغم التقدمات في الذكاء الاصطناعي، يقدر على أن يدرک اكثر من ثلاثين او خمسين بالمأئة من نص كلاسيكي.

1- Morpho – Syntactic

٢-كثير من المشاكل اللغوية التي لا يقدر الانسان ان يفهمها لن يتمكن الحاسوب ان يعالجها ابدًا لذا يجب ان يجتنب عنها شديداً فهي:

Homonymy يتعلق بالاشتراک اللفظي:الهوجل، الساعة.

Homotaxy يتعلق بجمل نتمكن ان نتلقى منها تلقيات مختلفة الا ان تستخدم علامات الوقف والترقيم او علامات اخرى لحصرها :كيف لا تدرس؟!

كما انتم تلاحظون ان الانسان يتمكن بالفارزة و الوقف و... أن يعرب عن غرضه و لكن الحاسوب، و لو حلّل بدقة كاملة، كيف يتمكن ان يفسّر جملة تفسيرين مختلفين؟

Synonymy يعني قد يوجد مفردات لها معانٍ مشابهة:عذل=لوم

Allotaxy ان تعبّر جمل مختلفة غرضاً واحداً:ما لهذا الرسول يأكل الطعام و يمشي في الاسواق؟!كيف تكفرون بالله و أنتم تعلمون؟!

Definition ان تتداعى كلمة، نفس معنى تتداعاه جملةاخرى:الاحسان!الاحسان!

Contexi يرتبط بما تقتضيه المفردات من المعاني في الظروف و الحالات و المقامات المختلفة: المعاني الضمنية و التجنيس و ان كانت في النصوص الادبية اكثر ظهوراً لكنّها في النصوص العلمية التي تودّي الى غموض في الادراک لان يؤثر المعالجة الاتوماتيكة علينا ان ننتبّه بالحالات و المقتضيات عند المعرفة اللفظيةالجديدة للنص.

(اليويه و ونيفلد[1]،٢٠٠٠م،ص٢٢١)

٣-لا يكفي التحليل الصرفي و النحوي للغة بل يجب أن يتمكن الحاسوب ان يحلّل النصوص تحليلاً منطقياً معنوية.(متزگر[2]،١٩٩٦م،ص٤١٣)

1- olivier & weinfeld. 2-Metzger

٤-يجب الفصل بين الشكل و المحتوى.(اشميد[١] ،١٩٩٢م،ص٤٥) علينا أن ننتبه بان الحاسوب يدرک الشكل فقط دون المحتوى؛فاذا تمكنّا ان نقارن التحليل النحوى بالمنطقي، الذي ليس سهلاً، نتمكن ان نتوقع من كمترة اللغة.

١- Schmid

المصادر و المراجع

1- Blache (Philippe), 2001. *Les grammaires de propriétés* : *des contraintes pour le traitement automatique des langues naturelles*, Hermes science, Paris.

2- Bonnet (Alain). 2001. *L'intelligence artificielle* : promesse et réalité. InterEditions Paris, 2001.

3- Bouché (Richard), 1988. *Valeur référentielle et langage d'indexation dans les systèmes d'information documentaires,* communication faite le 28 Novembre 1988 au Colloque "Archives et Temps Réel", organisé à Lille par le CREDO (Université Lille III), l'ADBS Nord, et les Archives du Nord.

4- Carré (R), Dégremont (J.F), Gross (M), Pierrel (J.M), Sabah (G), 1991. *Langage Humain et Machine*, Presses du CNRS, Paris.

5- CROCHEMORE M., HANCART C., LECROQ T. 2001*Algorithmique du texte*, Paris, editions Vuibert

6- GAUDINAT A., BOYER C. 2002 *Automatic Extraction of MeSH terms from Medlinet*, Abstracts. Atelier de NLPBA2002 sur le Traitement du Langage Naturel dans les applications Biomédicales.

7- Grevisse (Maurice), 2001. *Le Bon Usage*, onzième Édition, Grammaire française avec des remarques sur la langue française d'aujourd'hui, Duculot .

8- LE DUFF F., THIRION B., DARMONI SJ. 2003 UMLF *: Construction d'un lexique medical francophone unifié.* Actes de Journées Francophones d'Informatique Médicale, sous presse.

9- Le Guern (Michel), 1991. *Un analyseur morpho-syntaxique pour l'indexation automatique*, in revue "Le français Moderne", N° 1 Juin 1991.

10- Le Guern (Michel), 1994. *Parties du discours et catégories morphologiques en analyse automatique*, publié chez PUL.

11- Mahmoudi (Seyed Mohammad), 1994. *Contribution au traitement automatique de la langue persane : l'analyse et reconnaissance automatique des SN*, thèse de doctorat nouveau régime, présentée devant l'Université Lumière Lyon 2

12- Mahmoudi (Seyed Mohammad), 2002 : *The Role of The Information and Communication Technology in The Reengineering of The Systems , in Persian* .In Revue " Journal of Qom Higher Education Complex " N° 11, pp. 171-194

13- Metzger (Jean Paul), 1988. *Syntagmes nominaux et information textuelle : reconnaissance automatique et représentation*, thèse d'État ès Sciences, Université Claude Bernard, Lyon1

14- Metzger, J. P. & Mahmoudi S. M. : *Proposition pour une reconnaissance automatique des syntagmes nominaux du Persan*, in Revue Internationale de Lingvisticæ Investigations, Tome XX, 1996.

15- Ollivier (Daniel), Weinfeld (Michel), 2000. *Utilisation de rétroaction et de classifieurs adaptatifs pour améliorer les performances d'un système de lecture de montants littéraux de chèques bancaires*, in CIFED'2000 (Colloque International Francophone sur l'Ecrit et le Document), sous la direction de Hubert Emptoz et Nicole Vincent, INSA Lyon, Juillet 2000, pp. 221-229.

16- POULIQUEN B. 2002 *Indexation de textes médicaux par indexation de concepts, et ses utilisations*. Thèse de Doctorat, Université Rennes 1, 2002.

17- Robert : *Dictionnaire Petit Robert*, Paris, 1986

18- RUCH P., BAUD R., Geissbühler A. 2003 *Learning-Free Text Categorization*. Actes de Artificial Intelligence in Medicine in Europe, 199-204.

19-Schmid(Anne-Marie),1992,in *Bulag*(n° 18):Conférence Faite du département de Linguistique de L universite de Besanson,le 17 Mai 1992.

20- SOUALMIA LF., BARRY-GRÉBOVAL C., ABDULRAB H., DARMONI SJ. 2002 *Modélisation et représentation des connaissances dans un catalogue de santé. Actes de Ingénierie des Connaissances* 2002, 139-149.

21- Townsend (Carl), 1988. *Turbo Prolog : applications, Sybex*. Paris