المحاضرة الخامسة

اللغة العربية وشبكة المعلومات، الإنترنت

الأستاذ الدكتور عبد المجيد بن حمادو
أستاذ العلوم الاقتصادية والتصرف
قسم الإعلامية–جامعة صفاقس

الثلاثاء 29 صفر 1422هـ 22 أيار 2001

1. مقدمة :

الإنسان طموح بطبعه، فهو منذ تواجده ، يحاول تحقيق طموحاته العديدة والمتنوعة حسب العصر الذي يعيش فيه.

في البداية كان يحلم بالصيد الوافر والمحاصيل الزراعية الكثيرة لإشباع نهمه. وعندما تحقق له ذلك بدأ يحلم باجتياز المحيطات والتحليق في الأجواء البعيدة ليلتقي بالقمر والشمس.

ومنذ بداية القرن السادس ق. م. اتسعت دائرة نظره وتعلقت همته بالمعارف العلمية التي مكنته من صنع الآلات وتحقيق البعض من أحلامه. وكلما تقدم في كسب المعارف، كثرت طموحاته وزادت تعقيداً.

وهكذا تمكن الإنسان من اختراق البحر الأبيض المتوسط ثم المحيطات ونجح في إطلاق الصواريخ التي حملته إلى سطح القمر. ومع هذا تبقى بعض أحلامه معلّقة. ولعلّ أسطورة بابل تمثّل أكبر تحدّ تواجه البشرية اليوم؛ في بداية التاريخ كان الناس يتكلمون نفس اللغة، فاتحدوا لبناء برج عظيم يجعلهم متساوين مع الإله. لكن الله، ليثبت لهم غرورهم، هدّم هذا البرج العظيم وشتتهم في كل أنحاء المعمورة وجعلهم يتكلمون لغات عديدة ومختلفة لكي لا يتفاهموا ويتجمعوا من جديد لبناء بابل.

ومنذ ذلك الزمن والإنسان يحلم بالتواصل والتخاطب مع أخيه الإنسان مهما كانت جنسيته ولغته ومكان تواجده على الكرة الأرضية.

ومع عدم التحقيق الكلي لهذا الحلم، فقد خطا الإنسان خطوات هامة لرفع هذا التحدّي تتمثل في اختراع الطباعة، والهاتف، والراديو والتلفزيون والأقمار الصناعية والناسوخ "الفاكس".

ولعله باختراع الحاسوب مع منتصف القرن الماضي والتقنيات الجديدة للمعلومات والاتصال بما فيها خاصة شبكة الإنترنت **(Internet)**،التي تكتسح كل

أرجاء المعمورة، قد اكتسب آليات ذكيّة يمكن تطويرها لتجاوز ، في المدى القريب ، العقبة اللغوية في معاملاته اليومية مع أخيه الإنسان حيثما كان.

هذه التقنيات الجديدة للمعلومات والاتصال ، تعتبر اليوم نوعاً من المنعطف الثقافي والتكنولوجي الذي تحياه أغلب المجتمعات الحديثة. بل إنها تشكل تحدياً يواجه مجموع اللغات المكتوبة وبالتالي كل الثقافات. بحيث يكون رفع هذا التحدي هو الاختيار الوحيد أمام أية لغة وأية ثقافة تريد أن يُكتب لها البقاء والازدهار.

ومع دخول هذه التقنيات وبالخصوص شبكة الإنترنت بسرعة إلى فضاءاتنا التربوية ومنشآتنا الصناعية بل في بيوتنا أيضاً، لكي تصبح عن قريب مثل سائر الوسائط كالتلفاز والهاتف مثلاً، يبقى استعمالها في الوطن العربي محتشماً نسبياً. لكن أغلب المؤشرات تدل على نمو سريع سيطال كل جوانب مجتمعنا (اقتصاد، تربية، تعليم، ترفيه، بحث علمي،الخ...).

والسؤال الذي يطرح نفسه هنا يتمثل في معرفة ما هي اللغة التي سيتعامل بها المواطن العربي مع هذه التقنيات ؟ وهل وفرنا له الأرضية المعلوماتية الملائمة لتشجيعه على استعمال لغته؟

2. علاقة إنترنت باللغات :

إن شبكة المعلومات لها علاقة وثيقة باللغات. فاللغات هي المادة الأولى التي نمثل بواسطتها المعلومات المتوفرة في صفحات "الواب" المتبادلة بين الأشخاص والمؤسسات، والتي تُجرى عليها المعالجات الآلية المتعددة مثل البحث بالاعتماد على المضمون، التصفية قصد انتقاء المعلومات المفيدة والمركزة داخل مواقع الواب، فهرسة الوثائق للتعرف على المفاهيم المستعملة داخلها، إلى غير ذلك.

وهي أيضاً ركيزة التخاطب (**communication support**) مع الشبكة لطلب الخدمات ولصياغة الأسئلة الموجهة للنظم المتوفرة في الشبكة .

وانطلاقاً من هذه الاعتبارات نستنتج أن اللغة المستعملة بطريقة مكثفة في شبكة الإنترنت، هي الإنجليزية وذلك لعدة عوامل لعل أهمَها يرجع إلى تاريخ هذه

الشبكة التي انطلقت من الولايات المتحدة واستمرّت هناك سنوات عديدة قبل أن تنتشر في بقية أيجاء العالم. فكانت جلّ الصفحات والمواقع إنجليزية والشبكة لا تتعامل إلاّ باللغة الإنجليزية عند التخاطب مع مستعمليها.

وقد تغيرت هذه الوضعية اليوم وأصبحت جلّ لغات العالم المكتوبة ممثلة بطريقة متفاوتة وذلك بتوفر الصفحات والمواقع المكتوبة بهذه اللغات، والأنظمة المتعددة اللغات مثل محركات البحث، وأنظمة البريد الإلكتروني.

فما هي إذن وضعية اللغة العربية داخل هذه الشبكة؟

إنّ مكانة اللغة العربية في شبكة الإنترنت يمكن تقييمها بالاعتماد على الجوانب التالية:

أولاً : من حيث عدد الصفحات ومواقع الواب المكتوبة بالعربية.

ثانياً : من حيث وجود محركات البحث التي تتعامل بسهولة وبطريقة ناجعة مع اللغة العربية على مستوى صياغة الأسئلة وعرض نتائج البحث. هذه المحركات هامة جداً عندما تكون متخصصة في اللغة العربية، حيث تساعد أكثر من غيرها من المحركات المتعددة اللغات على كشف المعلومة العربية **(Information visibility)** والوصول إليها بسرعة.

ثالثاً : من جهة توفر آليات و خدمات متعددة ، تعتمد العربية كلغة التخاطب، وكمادة أولى تجرى عليها المعالجة على المستوى الصرفي، والنحوي و الدلالي. ومن بين هذه الخدمات نذكر مثلاً:

- الترجمة الآلية أو شبة الآلية من العربية وإليها والتي تمكّن المستعمل العربي من الاستفادة من المعلومات المتوفرة في الشبكة والمكتوبة بلغات أخرى يجهلها مثل الإعلانات المتعددة، والدراسات المختلفة.

- تصفية المعلومات (**Information Filtering**) التي تمكّن من انتقاء ما يحتاجه المستعمل العربي من بين زحمة المعلومات المتوفرة في الشبكة.

- التراسل الإلكتروني (**e-mail**) لتمكين المستعمل العربي من كتابة رسائله بلغته الطبيعية وعدم اللجوء إلى لغات أخرى مثل الإنجليزية والفرنسية.

- المناقشة الآنية (**chat**) ومنابر *الحوار* (***forums***) لتشجيع المستعمل العربي على الاتصال بأخيه العربي وعلى استعمال لغته والتقرب إليه.

- تعليم اللغة عن بعد (**e-Education**) وذلك بوضع دروس في اللغة العربية على "الواب" مدعومة بآليات تسهّل التعلّم. هنا تصبح الشبكة وسيلة ناجعة وفعّالة لنشر اللغة العربية كلغة ثانية في العالم وخاصة في بلدان لها علاقات متميزة مع البلدان العربية.

3. تطوّر نموّ الشبكة :

أ. نبذة تاريخية

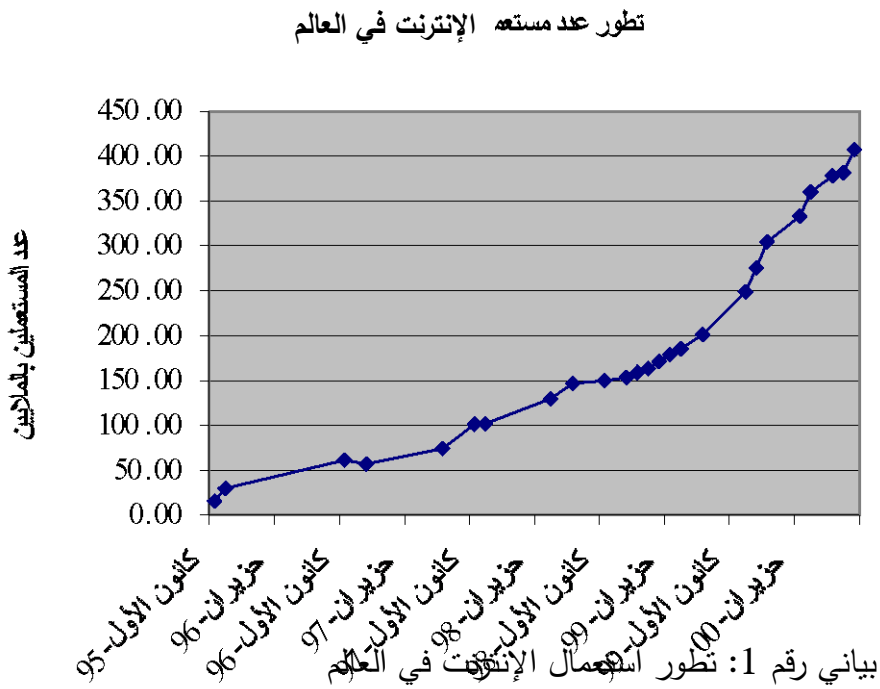
شبكة الإنترنت هي اختراع أمريكي ولد سنة 1962 في مؤسّسات عسكرية ثم انتشرت في الأوساط الجامعية ، وبالتحديد لربط مراكز البحث العلمي بعضها ببعض. وكانت آنذاك تسمّى "أربانات" (**ARPANET**). الفكرة الرئيسية التي ترتكز عليها الشبكة هي استمرارية الاتصال في حالة تعرض الولايات المتحدة إلى هجوم نووي، أي هنالك أكثر من ممر واحد لربط مكان بآخر.

الاستعمال العمومي لهذه الشبكة انطلق منذ سنة 1958، داخل الولايات المتحدة، حيث لاقت نجاحاً كبيراً لدى المستعملين خاصّة من خلال خدمات البريد الإلكتروني وتبادل الوثائق.

ب. تطور الشبكة في العالم :

لقد تطورت شبكة الإنترنت في العالم في السنوات الأخيرة بصفة مذهلة. هذه الشبكة العنكبوتية تمكنت من تغيير نمط التواصل بين الناس بالبريد الإلكتروني ومنابر الحوار (**forums**) وطريقة العمل بالمؤسسات الاقتصادية بالتجارة الإلكترونية (**e-commerce**) خاصة.

ولنا أن ننظر في الرسم البياني رقم 1



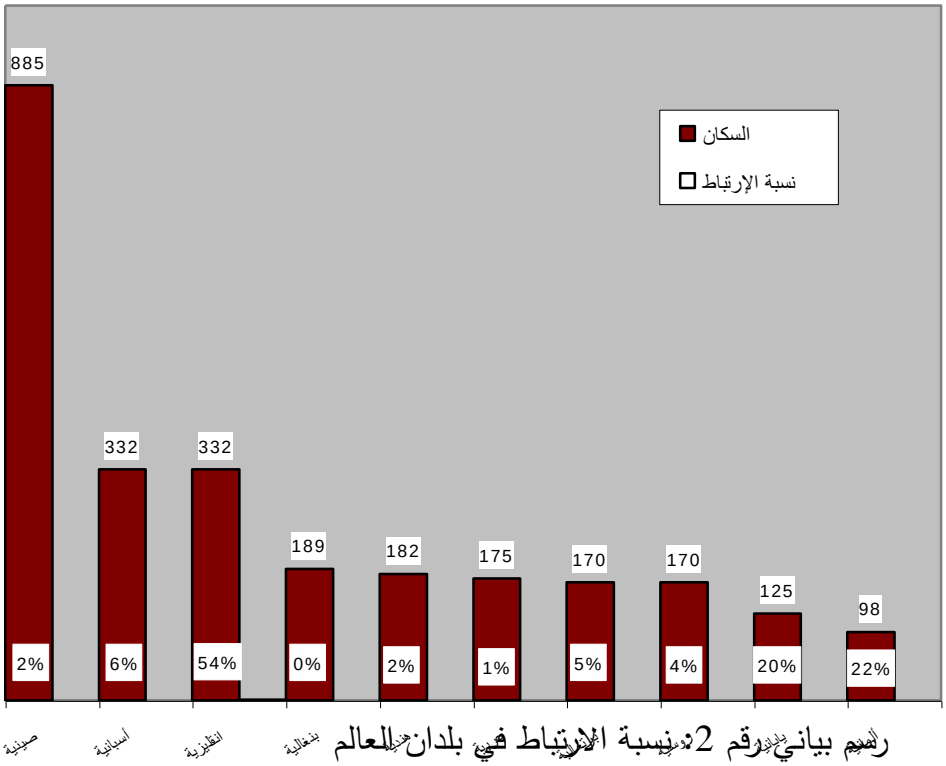
رسم بياني رقم 1: تطور أعداد مستعملي الإنترنت في العالم

وتجدر الملاحظة هنا أن هذا التطور لم يتم بصفة متساوية بين جميع البلدان وبين البلدان من نفس الجهة.

ج. تطور الشبكة في الوطن العربي :

بالرغم من العدد الكبير لسكان الوطن العربي نلاحظ أن نسبة الارتباط بشبكة الإنترنت تبقى من بين النسب الضعيفة في العالم (ما يقارب من 1/100). وسنحاول تفسير هذه الحالة في الفقرة الآتية:

الرسم البياني رقم 2 يعطي فكرة شاملة على نسبة الارتباط في بلدان العالم مجمعة حسب اللغة المعتمدة.

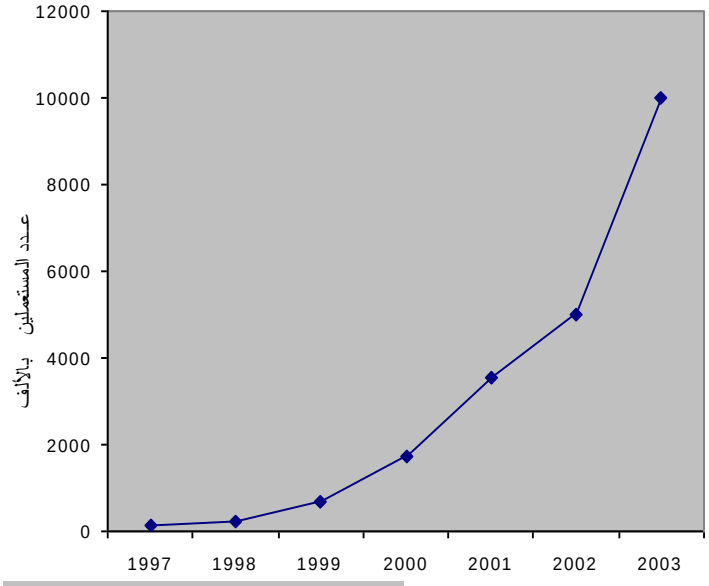


رسم بياني رقم 2: نسبة الارتباط في بلدان العالم

Source: http://www.ditnet.co.ae

أما التطور الكمي لعدد مستعملي شبكة الإنترنت في العالم العربي فلم يخرج عن ما هو عليه في سائر بلدان العالم. غير أن ما نلاحظه أن انطلاق هذه الظاهرة بدأ متأخراً مع بداية سنة 1997 إذا استثنينا بعض البلدان مثل تونس (1991). كما أن استعمال الشبكة يقتصر في أغلب الحالات على الفئات المثقفة من باحثين وجامعيين ومهندسين.

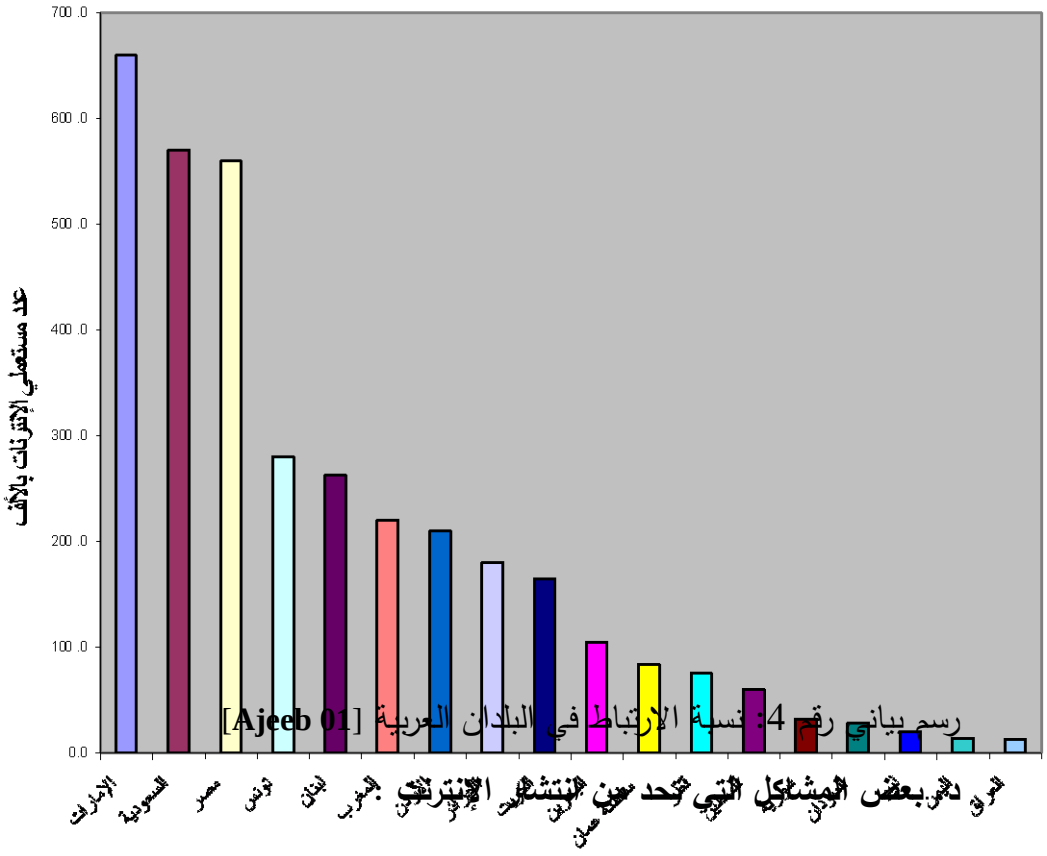
الرسم البياني رقم 3 يوضح تطور عدد مستعملي الإنترنت في العالم العربي ويعطينا فكرة عن ما سيكون عليه في نهاية سنة 2003.



Sources: http\\w w w .nua.org et

لرسم البياني رقم 3: تطور عدد مستعملي إنترنت في العالم العربي

كذلك نلاحظ من خلال الدراسات التي أنجزت أن هذا التطور تم بين البلدان العربية بصفة متفاوتة لكن هنالك بلدان مثل الإمارات العربية قد وصلت فيها نسبة الارتباط إلى درجة 24 % تفوق بذلك بعض البلدان الأوروبية مثل فرنسا حيث تبلغ فيها نسبة الارتباط 19 % فقط



رسم بياني رقم 4: نسبة الارتباط في البلدان العربية [**Ajeeb 01**]

المشكل الرئيسي الذي يحد من انتشار استعمال شبكة الإنترنت هو الحالة التي عليها البنية الأساسية للمواصلات. فإذا أخذنا مثلا البلدان العربية نلاحظ أنّ ثلث العائلات فقط يتمتع بخط هاتفي، الشرط الأدنى للارتباط بالشبكة **[UIT00]**.

كذلك تطوير شبكة الاتصال الموجودة لتصبح رقمية قادرة على التعامل مع تقنيات الإنترنت بصفة مجدية، يتطلب اعتمادات كبيرة وكبيرة جدّاً.

كما يمكن أن نضيف فقدان المنافسة بين مزوّدي خدمات الإنترنت، الشيء الذي يحد من جودة المنتوج المقدم للمستعمل. لذلك تبقى تكاليف الربط باهظة بالنسبة للبلدان العربية، إذا قارنّاها بما هي عليه في سائر بلدن العالم. غير أن هذا لا يمنع أن بعض البلدان تقوم بمجهودات جبّارة لتخفيض هذه التسعيرة وكذلك تخفيض تكلفة اقتناء الحاسوب لأنّه لا يكفي توفير الخط الهاتفي للارتباط بالشبكة.

وأخيراً يجب أن لا ننسى مكانة التكوين في نشر ثقافة الإنترنت عبر سائر أفراد المجتمع ، وتشجيع استعماله في هذه الأوساط بتوفير ما يسمى بمقاهي الإنترنت (**cyber-cafés**) كما هو موجود في المملكة الأردنية الهاشمية والجمهورية التونسية (**Publinet**).

4. قابلية اللغة العربية للمعالجة الآلية

أ. مقدمة :

إن الاهتمام بالمعالجة الآلية للغات الطبيعية بدأ منذ نهاية الأربعينيات. وكانت التطبيقات تقتصر آنذاك على إنشاء القواميس الإلكترونية. ومع اشتعال فتيل الحرب الباردة اتجهت الأنظار إلى نظم الترجمة الآلية لاستغلالها لأغراض عسكرية بحتة. لكن النتائج المتحصل عليها كانت دون الطموحات والاعتمادات الضخمة المرصودة لهذه الأبحاث. إذ تبين للجميع أنّ موضوع معالجة اللغات الطبيعية آلياً شديد التعقيد ويتطلب دراسة نظرية معمّقة حول كيفية تمثيل المعارف اللغوية من قواعد وقواميس واستنباط طرق ناجعة لتحليل الجمل على المستوى الصرفي والنحوي والدلالي.

وإذا تمعنّا في الأبحاث المنجزة وجدنا أن اللغة الإنجليزية قد حظيت بنصيب الأسد وتأتي بعدها لغات أوروبا كالفرنسية والإسبانية. أما بالنسبة إلى اللغة العربية فنلاحظ أنّ الأبحاث في هذا المجال انطلقت بصفة تكاد تكون منفردة في بداية السبعينات في مخابر بلدان أجنبية وبقيت محتشمة نسبياً إلى يومنا هذا. والسؤال الذي نطرحه هنا يتعلق بمدى قابلية اللغة العربية للتحليل الآلي مقارنة ببقية اللغات كالإنجليزية والفرنسية؟

فيما يأتي سنحاول الإجابة عن هذا السؤال من خلال تجربتنا الخاّصة وبعض تجارب أخرى.

ب. القابلية على المستوى التحليل الصرفي :

إن التحليل الصرفي يهتم بالتعرف على مكونات الكلمات وخصائصها الصرفية النحوية (الصفة النحوية، العدد، الجنس، ...). والغاية من هذا التحليل هو توفير المعلومات اللازمة لمرحلة التحليل النحوي.

وللقيام بهذا العمل يجب تصميم نموذج صرفي يفسر كل إمكانيات تكوين مفردات اللغة انطلاقاً من المورفيمات المكونة للكلمة ثم يأخذ بعين الاعتبار كل الظواهر التعاملية بين الأصوات والتي تظهر على مستوى كتابة الكلمة من إدغام وقلب، وتباين.

وما يمكن أن نفيد به هنا أن الصرف العربي غني بأوزانه وقواعده التي يمكن تمثيلها بسهولة في ذاكرة الحاسوب للرجوع إليها في عملية التفكيك الآلي. لكن عدم توفر الشكل في الكلمات يولد أشكالاً هاماً خاصة في مرحلة توليد الخصائص الصرفية النحوية بصفة مدققة (فكلمة ذهب يمكن أن تفسر كاسم أو كفعل) .

كذلك يبقى تمثيل الظواهر التعاملية بصفة ناجعة مشكلاً يصعب حلّه بدون دراسة معمقة حول إمكانية تقنين هذه الظواهر ووضعها في قواعد حاسوبية عامة تطبّق في كل الحالات واستعمال تقنيات معلوماتية تمكن من الاستغلال الأمثل لهذه القواعد.

وحسب إحصائيّاتنا تمسّ هذه الظواهر نسبة مهمّة من مفردات القاموس (ما يقارب عن 40%).

وإذا تمعنّا في المبادرات التي أنجزت في هذا المجال وجدنا أنه أجمالاً وعلى المستوى التطبيقي تم التغلّب على مشاكل التحليل الصرفي العربي. لكن الأنظمة المقترحة لا تستند أغلبها إلى نماذج نظرية متطورة، الشيء الذي يحد من شموليتها وإمكانية تطورها بسهولة واستعمالها في تطبيقات لغوية متعددة كتعليم الصرف العربي بالاستعانة بالحاسوب والتدقيق الإملائي والترجمة الآلية.

وفيما يأتي نقترح عليكم نموذجا للصرف العربي يمكن اعتماده للتحليل الزوائدي للكلمات. هذا النموذج يحترم قواعد الصرف العربي ويأخذ بعين الاعتبار جل الظواهر التعاملية التي تطرأ داخل الكلمة [**BEN93**].

النموذج الصرفي المقترح يرتب مفردات القاموس العربي إلى المجموعات الخمس:

مجموعة المفردات القياسية ونرمز إليها ب **Vd** وهي التي تحتوي على الكلمات التي تم تكوينها من سابقة ولاحقة وجذر وحشوه. وهي التي لم يطرأ على بنيتها الأساسية أي تغير.

كلمة ما نرمز إليها ب **w** من هذه المجموعة يمكن صياغتها:

w = D (R,P,I,S)

حيث إنّ :

- **D** تمثل محمول الاشتقاق (**Derivation Predicate**).
- **R** يمثل جذراً من مجموعة جذور اللغة العربية **Vr**.
- **P** تمثل سابقة من مجموعة سوابق اللغة العربية **Vp**.
- **I** يمثل حشوة من مجموعة حشوات اللغة العربية **Vi**.

- **S** لاحقة من مجموعة لواحق اللغة العربية **Vs**.

وعلى هذا لأساس نتحصل على :

$$Vd = \{w \mid R \in Vr, P \in Vp, I \in Vi, S \in Vs, (R,P,I,S)\}$$

- **مجموعة المفردات القياسية المقيدة نرمز إليها ب Vds**

إنّ داخل مجموعة **Vd** نجد كلمات لا تنتمي للقاموس بالرغم من أنها متكونة من جذر وزوائد ولهذا فرضنا على هذه المجموعة قيداً دلاليّاً نرمز إليه ب :

Sm (R, P, I, S). هذا القيد الدلالي يمكننا من تجنب الكلمات غير المقبولة في اللغة أو غير الموجودة في قاموس مفردات اللغة (مثل انوقف).

وبهذا تصبح المجموعة **Vd** الممثلة داخل الحاسوب تقتصر على المجموعة المقيدة دلالية (أو المحدودة) **Vds** حيث :

$$Vds = \{ w| (R \in Vr, P \in Vp, I \in Vi, S \in Vs) \text{ مع } D(R,P,I,S) \text{ و} Sm(R,P,I,S)\}$$

- **مجموعة المفردات المتغيّرة (التي طرأت عليها تغيّرات صرفية صوتية)**. نرمز إلى هذه المجموعة **ب Vts**.

ننطلق من العمليتين الأساسيتين الآتيتين :

- إبدال حرف بحرف آخر مثل (ن+ت+ث ⟵ ن+د+ث).
- حذف حرف مثل (ن+ت+ث ⟵ ن +ث).

إذن نسمّي تغييراً صرفياً صوتياً والذي نرمز إليه ب **T** تسلسل عمليات مثل التي ذكرنا (حذف أو إبدال) :**T= t1,t2,…t3**

- **مجموعة المفردات التي طرأت عليها تغيّرات صرفية صوتية هي :**

$$Vts = \{w|\exists X \in Vds \text{ مع } X^{ti} \Longrightarrow w, w \in (v+\backslash vd)\}$$

- **مجموعة المفردات القياسية المتغيّرة :**

نرمز أيضا **ب Vdst** إلى المجموعة التي نتحصل عليها بتطبيق التغيّرات العكسية T^{-1} على المجموعة **Vts** داخل مجموعة **Vds**.

$$Vdst = \{X \in Vds \mid \exists Ti \in T \text{ مع } X^{ti} \Rightarrow w, w \in (v+\backslash vd)\}$$

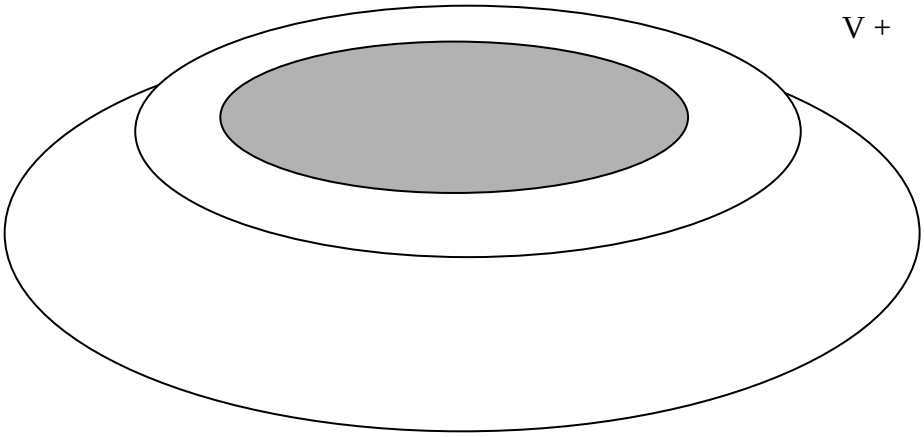
- **مجموعة المفردات غير القياسية :**

هذه المجموعة التي لا يمكن صياغتها بالاعتماد على **D (R,P,I,S)** .

نرمز إلى هذه المجموعة ب **Ve**.

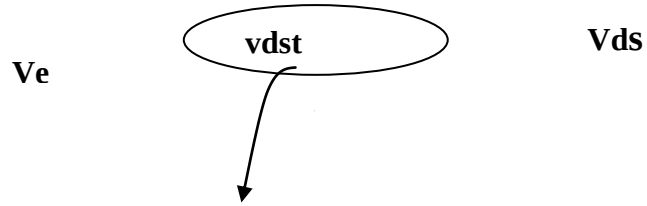
إذن تتكون مفردات اللغة العربية **LN** حسب النموذج المقترح من اتحاد مجموعة المفردات القياسية المقيدة بعد استثناء مجموعة **Vdst** مع المجموعة **Vts** (المساحة السوداء في الشكل 5).

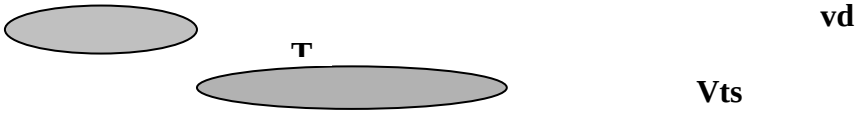
$$LN = (Vds\backslash Vdst) \cup Vts \cup Ve$$



V^+ يمثل كل السلسلات (الكلمات) التي يمكن أن نتحصل عليها من خلال إلصاق الحروف العربية (الألف ياء).

شكل5 :ترتيب المفردات العربية حسب نظام تكوينها.





ج. القابلية على مستوى التحليل النحوي:

التحليل النحوي (أو التركيبي) يمكّن من التعرف على بنية الجملة. هذه البنية يمكن أن تمثّل في صور مختلفة نذكر مثلاً الشجرة التركيبية **(syntactic tree).**

إن الأعمال المنجزة في مجال التحليل النحوي للجمل العربية لم تصل إلى مستويات مرموقة. بل هنالك محاولات عديدة لتطبيق أنماط وآليات تحليل جربت على اللغات اللاتينية مثل نحو السباق الحر **(Context free Grammar)** والنحو التسلسلي **(String Grammar)** والنحو التوحيدي **(Unification Grammar).**

هذه الأنماط والآليات التي نستعملها لا تتماشى دائماً مع خصوصيات اللغة العربية ونذكر هنا بالتحديد عدم احتواء هذه اللغات على شكل آخر حروف كلمات الجملة (الأعراب).

التحليل التركيبي للجمل غير المشكولة يفضي غالباً إلى عدة بن يات تركيبية (حالة غموض تركيبي). لفك هذا الغموض وتخفيض عدد الهياكل (أو البنى) التركيبية المتحصل عليها يجب استعمال آليات تحليل متين مثل ما يسمّى بشبكات الانتقال المزيدة :

(ATN : Augmented Ttransition Networks [WOO70])

وتجدر الملاحظة أن التحليل التركيبي يجب أن يعالج أيضاً الظواهر البلاغية مثل الحذف. ولمعالجة الحذف يجب تقنين ودراسة هذه الظاهرة دراسة نظرية تبرز أشكالها وتصنفها لتصميم خوارزميات ملائمة للمعالجة الآلية.

ولمزيد من التعرف على المشاكل التي تطرحها المعالجة الآلية للحذف في اللغة العربية والطريقة المتوخاة لمعالجتها يمكن الاطلاع على أطروحة الدكتور قيس الهدار[**HADD 98**].

5. مكانة اللغة العربية في شبكة الإنترنت :

كل من يحاول لأول مرة الولوج إلى موقع معرب يكون غالباً مآله الفشل. فبدلاً من أن يتحصّل على حروف عربية يتحصل على حروف لا معنى لها كما توضحه هذه الصورة. والأسباب عديدة ولعل أهمها أن البيئة المعلوماتية المستعملة في الحاسوب لا تتعامل بصفة طبيعية مع خصوصيات اللغة العربية.

فما هي البيئات التي يجب توفيرها للشبكة كي تتعامل مع لغتنا العربية ؟ هذا ما سنحاول الإجابة عنه في الفقرات الآتية ونبدأ بعملية تشفير الحرف العربي.

أ – تشفير الحرف العربي في الشبكة :

تشفير الحروف يمكّن من تمثيلها داخل ذاكرة الحاسوب ، فهي إذن عملية أساسية ومهمة جداً.

الشفرة العربية مرت بمراحل عديدة نذكر أهمها :

- سنة 1981 وقع اقتراح ما يسمّى ب **Codar-u** الذي يستعمل 7 "بيتات" **(7 bits)**
- سنة 1982 وقع تطوير هذه الشفرة لتصبح **ASMO.44** من طرف المنظمة العربية للتنميط والتق يس. هذه الشفرة متكونة أيضاً من 7 "بيتات" **(7bits)**.
- سنة 1986 ظهرت شفرة **ASMO-708** المتكونة من 8 بيتات **(8bits)**

ومع انتشار نظام تشغيل النوافد لشركة "مايكروسوفت"ورغبة هذه الشركة في التعامل المتعدد اللغات وحل مشاكل تنوع الشفرات وقع تصميم ما يسمّى **UNICODE** الذي اعتمد قياساً **(ISO 10646)**. هذه الشفرة تمكن من تمثيل كل الحروف العربية بكل أشكالها داخل الكلمة.

وتجدر الملاحظة هنا أن الحلول الأولى التي استخدمت بالنسبة إلى شبكة الإنترنت تتمثل في تخزين الصفحات أو النصوص كصور **(GIF)**. هذه الحلول سرعان ما اختفت نظراً لكلفتها بالنسبة إلى مساحة التخزين والوقت الذي يتطلبه نقل هذه الصفحات عبر الشبكة .

أما اليوم فقد حلت الحروف محل الصور، لكنها مشفرة بصور مختلفة حسب المواقع التي توجد فيها، وهذا التنوع يضع مشكل التقييس **Normalization** الذي يحد من سهولة انتقال الصفحات العربية من موقع إلى آخر.

ب.التصفح أو الإبحار باللغة العربية

التصفح أو الإبحار باللغة العربية ممكن عن طريق نظم الإبحار مثل نظام **Internet Explorer 5.0** تحت نظام التشغيل ويندوز معرب أو غير معرب مع إضافة برنامج خاص باللغة العربية. كذلك يمكن استعمال نظم أخرى مثل **Netscape** مزود ببرنامج سندباد.

ج. البريد الإلكتروني باللغة العربية e-mail

هو نظام لإرسال الملفات والرسائل الإلكترونية بس رعة فائقة لأي كان في العالم له عنوان إلكتروني وحاسوب متصل بالشبكة آنذاك. ومن أمثلة العناوين الإلكترونية نذكر العنوان التالي:

abdelmajid.benhamadou@fsegs.rnu.tn

نظم التراسل مثل **eudora** الموضوعة تحت نظام تشغيل معرب تمكن من التراسل بالعربية شريطة أن يكون المتلقي مزوداً بنظام تشغيل معرب.

وهناك كثير من المحركات التي تمكن من التراسل باللغة العربية مثل : "نسيج" **naseej.com** و"مكتوب" **maktoob.com** و"عرب فيستا" **arab vista** التي تشتغل تحت نظام تشغيل غير معرب وذلك بتمكين المستعمل من طرف عربي افتراضي كما نرى في الصورة الموالية.

د. البحث عن المعلومات بالعربية

البحث عن المعلومة المفيدة من بين رصيد كبير جداً من المعلومات المخزنة داخل الشبكة يتم بوساطة محركات البحث (**Search engine**).

هنالك محركات البحث وهنالك أيضاً ما يسمى بالبوابة أو الدليل **(portal / annuary)**.

كل منهما يمكّن من البحث عن المعلومات داخل شبكة الإنترنت لكن الفرق بينهما يتجلى في كيفية البحث. فبالنسبة للدليل يتم البحث حسب مواضيع معينة ومقترحة على المستعمل، أما بالنسبة إلى محركات البحث فهي مفتوحة لكل المواضيع التي يقترحها المستعمل بواسطة كلمات مفاتيح **(Key words)**.

وهنالك عدد كبير من البوابات العالمية التي تتعامل مع اللغة العربية مثل : **Arab vista** وغالباً باستعمال طرف افتراضي.

أما عدد محركات البحث التي تستعمل اللغة العربية (محركات متعددة اللغات) فهي قليلة جداً نذكر عن سبيل المثال: **alltheweb (fast)**

طريقة عمل هذه المحركات بالعربية يتم بالنسبة للبحث حسب المراحل الآتية :

(انظر: الشكل عدد6)

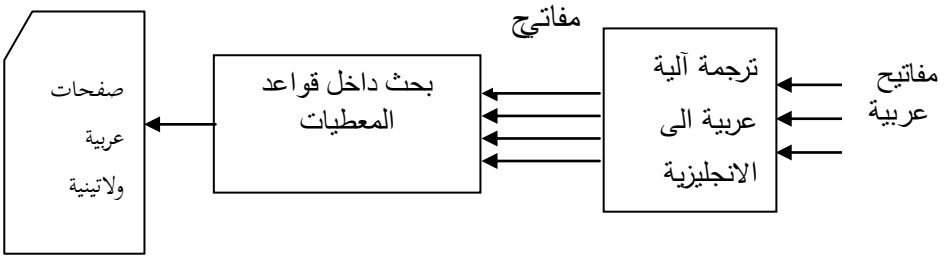
- ترجمة المفاتيح العربية إلى الإنجليزية.
- المفاتيح الإنجليزية المتحصل عليها تستعمل للبحث عن الوثائق التي تحتوي على هذه المفاتيح وجلبها.

وهذه الطريقة تجلب وثائق متعددة اللغات منها ما هو مكتوب بالعربية ومنها ما هو مكتوب باللاتينية.

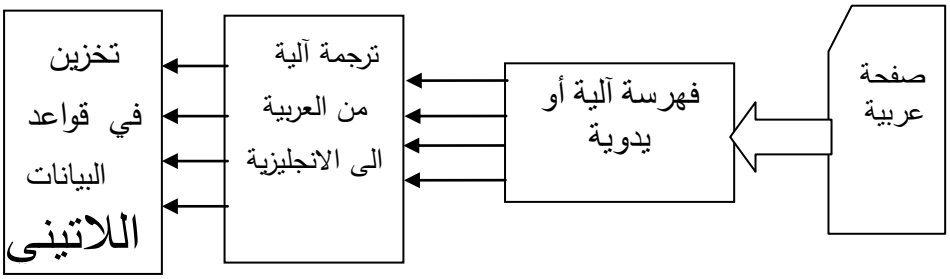
وبالنسبة إلى عملية الفهرسة فتتم حسب المراحل التالية: (انظر الشكل عدد7)

- فهرسة الوثيقة العربية (الصفحة مثلاً) للتحصل على قائمة مفاتيح عربية.
- ترجمة هذه المفاتيح إلى الإنجليزية.
- تخزين هذه المفاتيح الإنجليزية في قاعدة بيانات المحرك.

والجدير بالملاحظة هنا أن الإنجليزية تلعب دور اللغة المحورية بالنسبة إلى لغات أخرى كالعربية.



شكل رقم 6: عملية البحث عن صفحات



شكل رقم 7: عملية فهرسة صفحات

الو اب و تخز ين المفاتيح

الخاتمة :

في هذا المقال حاولنا أن نعطي فكرة عن التطور الذي شهدته شبكة الإنترنت في العالم وفي الوطن العربي.

ورأينا أيضاً كيف أنّ هذه التقنية لها ارتباط وثيق باللغات ويمكن أن تلعب دوراً مهماً في بقاء وازدهار لغة على حساب أخرى.

وبالنسبة إلى اللغة العربية رأينا أنّ استعمالها داخل الشبكة يبقى ضئيلاً جداً

واللغة العربية (المرتبة 18) إذا ما قارناها بلغات أخرى مثل الإنجليزية، بالرغم من أنها تأتي في المرتبة السادسة حسب عدد الناطقين بها.

هذا التأخر ناتج عن عدة عوامل قد شرحناها ولعل أهمها ضحالة المحتوى العربي داخل الشبكة إذ أن أغلب المواقع التي تنتجها الجهات العربية مكتوبة باللسان الإنجليزي ولعل هذا ما يمكن أن نفسره بتجاهل المواطن العربي لهويته الثقافية مجسدة في اللسان العربي. هذا إلى جانب عدم توافر آليات معالجة متنوعة تتعامل مع اللغة العربية ويمكن للمستعمل العربي استغلالها بسهولة.

لذلك تظل دعوتنا ملحة إلى ضرورة الوعي بأهمية العمل الجاد على تغيير هذه الوضعية وإعطاء اللغة العربية المكانة التي تستحقها.

المراجع

[BEN 93] BEN HAMADOU A.

Vérification et correction automatiques par analyse affixale des textes écrits en langage naturel: le cas de l'Arabe non voyellé. Thèse ès-Sciences Informatiques. Faculté des Sciences de Tunis, mars 1993.

[HAD 98] HADDAR K., BEN HAMADOU A.

An Approach to Ellipsis Detection and Classification for the Arabic Language. In the proceedings of the eleventh International Conference on Industrial and Engeneering Applications of Artificial Intelligence and Expert Systems June 1-4, 1998, Castellán, Spain. Lecture Notes in Artificial Intelligence, Springer Verlag N°1415.

[WOO 70] WOOODS W. A.

Transition Network Grammar forNatural Language Analysis. Communication of the ACM, Vol. 13, N° 10, October, 1970

[UIT00] Union Internationale des Télécommunications. Réunion préparatoire de la région arabe à la conférence mondiale de

développement des Télécommunications. Alexandrie (Egypte), 17-19 octobre 2000.

[DITnet99] DITnet "Internet Reaches Layman in middle East", 5 August 1999. http://www.ditnet.co.ae/itnews/me_internet/ecomprofiles.html

[GLReach01] Global Internet Statistics(by Language) http://www.glreach.com/globstats, site visitée en Avril 2001.

[ATI01] Site de l'Agence Tunisienne d'Internet. http://www.ati.nt, site visitée en Avril 2001.

[BABel01] http://babel.alis.com, site visitée en Avril 2001.

[AJEE 01] http://www.ajeeb.com/

[NUA 01] http://www.nua.com/