بيئة نوج "NooJ" اللسانية، أداة لمعالجة وتعليم اللغات:
تطبيقات صرفية – معجمية على العربية[1]

أ/ د.عزالدين غازي
جامعة القاضي عياض، مراكش

تقدبم :

1. لقد أضحت أنظمة معالجة اللغات الطبيعية (*NLP*) توفر خدماتٍ جليلةً للبحث اللغوي من خلال إعداد برمجيات تطبيقية مندمجة في البرامج الإلكترونية الخاصة بالتعليم المبرمج، بمقارباتٍ هندسية متطورة هدفها العام خدمة الثقافة والمعرفة والتربية والتعليم. وقد أدى هذا التطور، من جانب آخر، إلى العمل في صميم التنمية العلمية والمعرفية والثقافية للإنسان وكل ذلك هدفه تشجيع المحتوى العربي على مستوى الشابكة حتى تنافس المحتويات الأجنبية المختلفة.. وفي هذا الإطار حظيت مسألة معالجة اللغة العربية آلياً، باهتمام العديد من الخبراء والمهتمين بالهندسة اللسانية عربياً وعالمياً، بغرض تمكين الحاسوب/ الآلة من التعامل معها، ليصبح معها قادراً على التعامل مع معطيات نظامها باعتبارها مدخلات لألغورثمات ومخرجات تجيب عن مختلف الإشكالات التطبيقية، لاسيما وأن هذه الأخيرة تعتمد أساساً على التحليلات اللسانية الآلية التي لا تزال تعتمد المقاربات الخطية بصفة إجمالية. نركز هنا على الأدوات اللسانية والأدوات العقلانية المعلوماتية كضرورة منهجية ومعرفية في الإعداد القبلي للمعالجة الآلية

(1) ورقة قدمت ضمن أعمال المؤتمر الدولي الثالث حول "اللسانيات وتطوير تعليم اللغة العربية" يومي 4 و5 مايو 2014، كلية الآداب والعلوم/ جامعة قطر.

بغية بناء أنظمة معلوماتية واقعية لا خطية من حيث أثرها العرفاني، تساعد الإنسان على التعلم والاستهلاك المعرفي والإنتاج الثقافي بشكل ذاتي، تلقيناً وتفاعلا. مستغلين ما صممه مهندسو البرمجيات من مدققات إملائية ونحوية وأسلوبية ومعالجات للنصوص للتلخيص وللتوثيق..، وأنظمة التعرف على الحروف العربية ثم تحليل التراكيب والبنيات اللغوية بداية بالمحلل الصرفي-المعجمي والمحلل التركيبي-الدلالي وصولا إلى المُولّف النصي وتوليد الكلام، هدفها العام الاستجابة للأغراض التطبيقية/ المعرفية، المعتمدة على قواعد البيانات المؤسسة على الشبكة الدلالية وعلى الموسوعات والمعاجم التي أصبحت العماد الرئيس في كل عملية تعليمية.

2. سنجيب عموما في هذه الورقة عن الجوانب التطبيقية الخاصة بالصرف-المعجمي مكتفين ببعض الأمثلة البسيطة في اللغة العربية حتى نبين أهمية البيئة قي الإجابة عن الإشكالات التعلمية والمعرفية من قبيل: تطبيق قواعد النحو المحلي للأمثلة التي سنكتفي بسردها، وذلك بشكل تفاعلي حتى يتمكن المتعلم من توليد صرفي-معجمي تلقائي بناء على الموارد المتاحة، مثل أن يتتبع اشتقاقات فعل "درس" و"كتب" الممكنة مع مختلف الصيغ الصرفية والمسائل الإعرابية المصاحبة بحيث يتم إنشاؤها تلقائياً بالتدريب، مع ضمان جودة البيانات المستخرجة، وإذا لزم الأمر تصحيح كل حالة على حدة مباشرة من قبل المعلم/ المشرف وذلك بتأكيد الصلاحية والتصفية، بالشرح التفاعلي للمدونة والتي تصبح في النهاية مشاريع تعليمية للمتعلمين الذين سيتمكنون بالتالي من التفاعل مع سلسلة من التحويلات اللغوية لتصنيف المبنى، وهذا ما يضفي على هذه الأنشطة المقترحة التقويم التربوي المتساوق مع كل عملية بتطبيق إجراء الموافق الآلي(*conordancer*)، وقد تساهم العملية بدورها لاحقاً في تطوير البيئة من جانب الأنشطة التعليمية للغة العربية.

وقبل البدء في التطبيقات المقترحة في هذا البحث، سأعطي لمحة عن تركيبة "بيئة نوج" وعن مكوناتها البرمجية والخوارزمات التي تتحكم فيها ومجالاتها التطبيقية حتى نقرب البيئة بشكل أفضل للقارئ.

1 – نظرة عن بيئة "نوج" (*NooJ*) وتحليل النصوص المعجمية:

تمكن مخبر المعالجة الآلية والتوثيق اللساني (*LADL*) بباريس7 ومختبر (ELLIADD) بجامعة فرانش كومتي (Franche Comté) من بلورة بيئة إلكترونية لغوية تعمل كمحرك ومعالج لساني، يشتغل بمنظومة معاجم "ديلا" (DELA) الآلية، سمي بنظام أو بيئة "نوج" اللغوية، لهذا السبب أصبحت هذه التقنيات محط اهتمام جل التطبيقات اللسانية للكثير من البرمجيات والأنظمة الآلية التي تستخدم كبيئة للتنمية اللغوية مفتوحة المصدر،مثل (NooJ)[2] لصاحبه ماكس سلبرشتاين[3]، إذ يتأسس في عمله اعتماداً على تكنولوجيا أوتومات الأوضاع النهائية[4] كما هو الحال في بيئة (UNITEX)[5] لصاحبه سبستيان بوميي[6] ومثل

(2) شكل هذا النظام موضوع تطبيقات لسانية على مجموعة من اللغات الطبيعية كاللغات البرتغالية والإسبانية والأرمينية (النظام الإعرابي) والهنغارية والبلغارية والصربية والدنماركية والماسيدونية، وأعمال أخرى خصت لغات أخرى كاستخراج أسماء الأعلام في الإنجليزية وبناء قاموس أسماء الأعلام في الصربية والعربية. ومعالجة الغموض الدلالي ومعالجة المركبات الاسمية في الفرنسية ومعالجة التعبيرات في الفضاء الزمني والفهرسة الآلية للنصوص الطبية وتصنيف البنيات التركيبية لجمل المدونة وتطبيقات نظام (NooJ) على اللغة العربية.

(3) M.Silberztein (93) و(99 M. Silberztein).

(4) نذكر بعض الأعمال التي انصبت على التعرف بالمتواليات الصوتية (الفونولوجية) والخطية (المعجمية التركيبية) وعلى رأسهم Mohri Revuz (1991) و(1997) وRoche & Schabes (1997) على التوالي تمثيلا لا حصراً.

(5) لقد قام كل من بوميي سبستيان وايريك لابورط (المجموعة اللسانية التابعة لمعهد كاسبارد مونج (Gaspard Monge) ببناء نظام (*UNITEX*) البرمجي الصوري الذي يشبه إلى حد بعيد النظام الذي بين أيدينا شكلا ومضمونا ومنهجا وصورة يهدف إلى معالجة اللغات الطبيعية كالفرنسية والإسبانية والعربية والإنجليزية والألمانية وغيرها. للمزيد يرجى زيارة الموقع التالي:
http://www-igm.univ-mlv.fr/~unitex/index.php?page=3

(6) مزيداً انظر :

Sébastien Paumier.(2000). Nouvelles méthodes pour la recherche d'expressions dans de grands corpus. In A. Dister (ed.), Actes des 3èmes Journées INTEX. Revue Informatique et Statistique dans les Sciences Humaines, 36ème année, n° 1 à 4.

خزانة تشغيل العبارات المنتظمة في أوتومات الأوضاع النهائية المستخدمة في نظامي (Xerox 2000)[7] و(AT&T)[8]. هذا ويقع نظام "نوج" (NooJ)[9]، النسخة المطورة لأنتكس، على رأس القائمة، في التمثيل الدقيق للمدونات النصية، بما تتيحه مختلف أنواع هذه الأوتوماتات التي بواسطتها تُرتَّب المفردات وتُصنَّف في جهةٍ تُناسِبُ كل وحدة معجمية على حدة. بمعنى آخر، حتى تُعالجَ بدقةٍ كل حالة نحوية على حدة، لا بد أن يكون هناك تلاؤمٌ للأوتومات مع الأنحاء المحلية، وتستدعي إثر ذلك، كل عملية، أوتوماتاً للقيام بالعمليات الخوارزمية.

(7) انظر (2000.) Mohri et al.

(8) انظر (1997) Karttunen et al.

(9) تعتبر بيئة نوج (NooJ) رديفة لنظام (INTEX)، من ميزتها أنها يجمع المعاجم الآلية الخمس في معجم آلي واحد، قام مصمموها بدمج المفردات البسيطة والمركبة بعضها مع بعض وبالتخلص من المعجم الآلي للمفردات البسيطة المعربة (DELAFs) وكذلك التخلص من المعجم الآلي للمفردات المركبة المعربة (DELACs) وإضافة معلومات أخرى للمعجم الآلي للجذور (DELA) ومعالجتها بنماذج جداول المعجم – التركيبي. ويرجع الفضل في ذلك إلى النظام الإعرابي الذي يعالج المفردات البسيطة والمركبة في نفس الآن بفعل تلك العمليات التصريفية التي يتيحها المحرك الإعرابي المبرمج في البيئة، فهو يعادل كومة من الأتوماتات التي توظف عادة في معالجة اللواحق المتصلة بالمداخل المعجمية وغالباً ما يرمز إليها بالصوغ التالي:(< B > ، < D >،< E >،< C >) حيث يترك الحرية للمستخدم إما بإضافة أو حذف شكل ما. ولكل لغة طبيعية عبارتها الخاصة فمثلا اللغة العبرية عبارتها < F > التي تعني المحرك الإعرابي الخاص بها. ومن الخاصيات الجديدة لهذا النظام أنه يعالج المتواليات المتلازمة والمركبة ويعتبرها في خوارزمياته مدخلا معجمياً مستقلا.كما يفعل مع المتلازمات في اللغة الإنجليزية مثلا: (*bag of hicks*)،(*N+FLX+ACTOFGOD*) فكل صنف *number of the opposition sex =N+FLX=ACTOFGOD* من هذه الأصناف يعالج بمحرك إعرابي واحد حتى تتطابق العمليات بين المفردات المعربة ومعجم المفردات المركبة المعربة في آن واحد وذلك أثناء المعالجة والتحليل. انظر بيئة التطوير NooJ إصدار رقم v.4.1b 0117. وقد شرع في تطبيق هذا النظام على اللغة العربية بهدف تطوير برمجيات عربية أصيلة، ولاسيما فيما يتعلق بالتحليل الصرفي الإعرابي وإنجاز معجم آلي للأفعال والأسماء والصفات (جردت حوالي (40000) وحدة : فعل واسم وصفة) كما وصفت الأصناف الإعرابية، وأما الغرض فهو بناء معجم آلي إعرابي للغة العربية. كما يقوم (NooJ -FLX) بالتصريف الآلي والتحليل المعجمي للنصوص العربية (للمزيد انظر ابتهال فراوي (2004 ,Développement d'un module arabe de NooJ Université http://tln.li.univ-tours.fr/Tln_Colloques/JIntex2004/resumes.htmlLASELDI-de Franche-Comté وكذلك سليم مصفار Electronic dictionaries for the automatic analysis of texts.from INTEX to NooJ)، أشغال الورشة الثانية للمعجم الحاسوبي التفاعلي المنظمة بمدينة الملك عبد العزيز للعلوم والتقنية بالرياض أيام 5 و6 و7 مايو 2008.

وهكذا فكل نحو يتحول إلى أوتومات للأوضاع النهائية[10] (تمثلها محولات الأوضاع النهائية التي تقوم بدور حذف أو تعويض عناصر معينة في المتواليات) تناسب المدونة النصية المطلوبة في المعالجة. ونظراً لأهمية علاقة النظام كمحلل وممثل للمعطيات ومدقق ومفهرس وموثق.نرى، من المفيد، إعطاء فكرة عن تركيبة هذا النظام الآلي وعن بعض الأساسيات المتمثلة في المكونات الفيزيائية والتصورية وشكل منصة البيئة وخوارزمياتها وكذا مجالات تطبيقاتها اللغوية وغيرها، كما في العناصر الآتية:

1-1- منصة ومكونات البيئة اللغوية:

بيئة نوج (*NooJ*) عبارة عن منصة (platforme) تتلاءم مع العديد من أنظمة الاستغلال بإصداراتها المختلفة (Microsoft وMac OSX وLinux وBSD وSunOS) تستخدمها مجموعة هائلة من مراكز ومختبرات البحث والتطوير، كما تعتبر متقدمة على أنظمة عديدة من حيث النتائج المحصل عليها في التحليل والمعالجة اللغويين، لا سيما أنها مُزودة بقالب إحصائي يمد المستعمل بمعلومات عبارة عن معطيات رقمية لمختلف عمليات توزيع المفردات في المدونة المعالجة، إذ يمكن الاستفادة منها في تحليل ظواهر وأنماط خطابية مختلفة. كما يستطيع هذا البرنامج بالأساس، تحليل المدونات المرتكزة على المعاجم الإلكترونية ديلا (*DELA*) وعلى الأنحاء المحلية التي انتشرت كثيراً في اللغات: الفرنسية والصينية والإنجليزية والعربية والألمانية والدانمركية والإيطالية والبرتغالية والإسبانية واليونانية والبلغارية...وخاصة في مراكز البحث المتخصصة حيث يتم تكوين فرق عمل لتطوير معالجة اللغات الطبيعية وبناء الأنحاء الإلكترونية للغات وتطبيقات برمجية – صناعية مختلفة.

أما البيئة فتتألف من خمس طبقات من المعاجم، وهي تشتغل في تناسق تام ومكتمل، كما تحتوي على نحويات إلكترونية محلية، علاوة على كونها مزودة

(10) قمنا بإنشاء مجموعة من الأوتوماتات التي تعالج المتواليات المركبة والمتلازمات في اللغة العربية.

بالطرق الخوارزمية التي تقوم بالصيانة وبخلق موارد معجمية جديدة صرفية وتركيبية تساهم في كل ما يمكن أن يساهم في رفع اللبس وخاصة مع التعبيرات المركبة والمتلازمة. وهذه من بين العمليات الأولى لهذا البرمجي[11] بحيث يقوم برفع اللبس عن بعض المتواليات المعقدة وأسماء الكيانات التي غالباً ما تظهر بين الصفات والظروف في لغات طبيعية كما في البلغارية مثلا[12]. وتتركز الوظيفة الأساسية لهذه البيئة في إدارة المعطيات المعجمية التي زودت بها المعاجم التالية كما هي في اللغة الفرنسية:

أ- يتضمن هذا المعجم المفردات البسيطة (في الفرنسية) ديلاس (*DELAS*)، أكثر من (80000) مدخل معجمي، ويُسند لكل مدخل رقم، هو عبارة عن سنن، ويعتبر جزءاً من الخطاب، كما يُسند إليه سنن رقمي يحدد الإعراب (flexion)، في اللغة العربية تسنن المداخل كما في المثال: [جناح، س21]، حيث يشير الرقم (21) إلى أن ذلك الاسم: مذكر، وفي حالة الجمع تضاف إليه لاحقة (ات) في [جناحات]. ويوظف هذا المعجم حوالي (500) رقم، ويتراوح السنن الرقمي ما بين (1- 500) كما يستعمل (2Mo) لتخزين ملفات النصوص. وتمكن الأرقام في المعجم الآلي للمفردات المركبة ديلاك (DELAC) من توليد حوالي (60000) شكل لغوي مأثور: صيغ تصريف الأفعال،الجمع والإفراد بالنسبة للأسماء ،المذكر والمؤنث،المفرد والجمع بالنسبة للصفات، ويسمى المعجم الآلي للمفردات المعربة ديلاف(DELAF)[13].

ب – تتألف مداخل ديلاف (DELAF) من مفردة وسنن رقمي، ويوظف أقل من سنن رقمي، يتراوح ما بين (1-500)، ويشغل مساحة (2Mo) مخصصة لملفات النصوص، منها فقط (1-Mo) لمعجم (DELAF).

(11) اعتمدنا النسخة الحالية رقم:v.4.1b 0117 وهي قابلة للتحميل من الموقع التالي: www.nooj4nlp.net

(12) انظر المرجع السابق.

(13) أنجزت بلاندين كورتواز معجماً آلياً للمأثورات اللغوية في اللغة الفرنسية زودت به ديلاكف للمزيد ينظر B.Courtois (1984)

ج – يحتوي معجم ديلاك (*DELAC*) على المداخل المكونة من المفردات المركبة [-Si] [Gr]. وتعرف المفردة المركبة بأنها متوالية من المفردات البسيطة، ذات خصائص تركيبية غير صالحة للمفردات البسيطة نفسِها. ومن الناحية الصورية، فإن مداخل ديلاك (DELAC) هي سلسلة من المفردات البسيطة المقترنة بمعلومات إعرابية. وعلى العموم فإن هذه الأرقام الموظفة في ديلاك (DELAC) هي نفسها الموظفة في ديلاف (DELAF). غير أن المعجم المنبثق عن ديلاكف بني بصفة شبه آلية، مما يسمح بعمليات للنشر اليدوي، وهذا ما يسمى ب(NooJ PERL)[14] والتي تعود لجداول المعجم –التركيبي (الموصوفة من قبل موريس كروس (93)) التي تحتوي أسطرها على مداخل معجمية للمركبات أو للمتلازمات، ولكل عمود خصائصه، تتحكم فيه قيم السلب والإيجاب، وهذا ما يمكن إلى حد ما التعرف على المتلازمات والمركبات في النص[15]. وقد كيفت البيئة هذه الجداول مع ما توصل إليه (M.Silberztein (1999 وسمي بالميتارسوم (metagraghes) [16]. وبصفة عامة، تعتبر جداول المعجم– التركيبي العنصر الذي تعود إليه محركات البحث لتنهل منه كل ما شاءت الحصول عليه، مما هو مدرج في المدونة النصية، كما يعمل ضمن مجموعة المعاجم، ويتعلق أساساً بقائمة الأرقام المسندة للجمل، والممثلة انطلاقاً من فعلها،كما تشكل مجموع هذه الأرقام الصيغ النحوية للمفردة.

د – هناك معاجم أخرى لا تقل أهمية عما سبق ومنها: المعجم الآلي الفونيمي ديلاب (*DELAP*)، تناسب مداخله المداخل البسيطة غير المعربة ديلاس (DELAS) مرفقة بأوصاف صوتية ونطقية وخاصة مع الأشكال المعربة[17]. والمفتاح فيه هو المفردة، أما المعلومة المقدمة هي الفونيمية (Phonemization).

(14) انظر سليم مصفار، (2008).

(15) نفس المرجع السابق.

(16) نفس المرجع السابق.

(17) انظر M.Silberztein (1993).p.34.

إن ما يميز هذا النوع من المعاجم هو كونها مركبة ومهيكلة[18]، حيث لكل مفتاح، مدخل تسند إليه معلومات مغايرة.ويهدف معجم ديلاب إلى ربط مفردات لغة ما بمعلومة فونولوجية (صوتية) وتركيبية (صرفية) ولغوية، ويمكن تحديده بدرجة معينة للغة وهو ما يعرف بالمعجم الجزئي[19]. والذي يتناول جانباً نحوياً أو تركيبياً للمفردة.

إن تشغيل نظام معاجم ديلا (*DELA*) يطرح مشاكل كثيرة على مستوى قواعد المعطيات[20]، لأن معايير النجاح تتسم بالتغير والتطوير، وهذا ما لم يجعل من المساحة المخصصة لتمثيل المعلومات، والزمن المستغرق لنجاح استغلال المعطيات من أن يسيرا على قدم وساق، ويرجع ذلك إلى النتائج المحصل عليها، مما فرض على مصممي هذا النظام تطويره ليناسب كل ما استجد على مستوى التطبيقات الحاسوبية واللغوية على حد سواء.

1-2- خوارزميات البيئة اللغوية:

إن الشكل العام للأنحاء المحلية في هذا النظام تلائمه كما تناسبه تماماً تقنية أوتوماتات الأوضاع النهائية (*finite state automata*). لذلك فهو يقوم بوظيفته تلك، مزوداً بخوارزميات صارمة (في حالة استخدام محولات الأوضاع النهائية). وبما أن حزمة الأنحاء المحلية عند وصف كل لغة، بل وحتى مستوى من مستوياتها، في تضاعف مستمر، بسبب البنيات المستحدثة في اللغة نتيجة الاستعمال، فإنها تستدعي لا محالة تطوير تقني جد فعال لإدارة سرعة المعالجة حتى تستجيب لحجم المدونات وتحليلاتها التركيبية. وقد لاحظ الباحثون في

(18) غازي .ع ولهلال.م (2007).

(19) انظر: محمد زايد، دراسة في المعاجم الحاسوبية الفرنسية،أشغال الورشة الثانية حول المعجم الحاسوبي التفاعلي، الرياض أبريل2008.

(20) انظر:

Slim Mesfar, electronic dictionaries for the automatic analysis of texts, from INTEX to NooJ.(2008).Arabic conference proceeding on electronic dictionaries.Riad.April 2008.

المختبرين المذكورين أعلاه، أنه بين تحليل النصوص -عن طريق تقنية الأوتومات المذكورة- وبين حساب تدفق الرسوم توازن نسقي كبير ودقيق، وهذه إمكانية كبيرة لتسريع البحث في نصوص المتن. كما اقترحوا أيضاً، إضافة تتعلق بإدماج النتائج الرياضية الجبرية مع البيانات الإحصائية الأصلية، وهذه العملية طورت بصفة كبيرة بناء النظام، علاوة على المناهج الأخرى التي تقوم بتصفية (*filtrer*) قواعد المعطيات وملاءمة مسارد النص المزمع معالجته مع الأنحاء المحلية قبل الشروع في تحليلها.

1-3- مجالات تطبيقات البيئة اللغوية:

لقد كان طموح واضعي هذه البيئة أن يجعلوها المعيار الصوري العام لكل اللغات الطبيعية ومن ضمنها اللغة العربية. لهذا جهزوها بمحلل معجمي، عبارة عن قالب (*module*) يستطيع معالجة جميع أصناف المتواليات اللسانية البسيطة والمعقدة كالتعبيرات المتلازمة والمفردات المركبة والأساليب.

وأما الغرض الأساسي من هذه البيئة فهو بصفة إجمالية:

1 - تعليم اللغات الطبيعية ومعالجتها آلياً في جميع مستوياتها.

2 - التدقيق الإملائي والنحوي والأسلوبي.

3 - *رصد المفردات البسيطة والمركبة والتعبيرات المتلازمة انطلاقاً من نصوص مدونة معينة.*

4 - المساعدة على بناء نظام للترجمة الآلية يتمكن من ترجمة نصوص وتعبيرات اللغات الطبيعية.

5 - استخراج المصطلحات ومتغيراتها الممكنة، بدراسة الأدوات المعلوماتية واللسانية.

6 - البحث عن المعلومات في مدونات الوثائق الإلكترونية عن طريق سؤال-جواب *QA*.

كما أن هذه البيئة تجمع كومة ضخمة من الأنحاء المحلية تقوم بوصف وتحليل الجزئيات التركيبية في السياق (فنحو الأسماء الأعلام هو عبارة عقلانية)، ومجهز بالتوافق الآلي (*concordance*) وبمحرك للأنحاء المحلية.

إن وظيفة أوتومات الأوضاع النهائية هي وظيفة محض خوارزمية تصف "المفردات" اللغوية، فهي عبارة عن (أبجديات)، تقوم بوصف "جمل" هي (متواليات) معروفة سلفاً لأنها من المفترض أن تكون موصوفة نحوياً وإلا رفضها برنامج التوافق الآلي وهو كاشف سياقي لكل المعطيات المعالجة. ولأوتومات الأوضاع النهائية خصوصيات خوارزمية هامة تتمثل في المعالجة الخطية للظواهر اللغوية معالجةً آنية، مثلا:

(كتب، بحث) كتب " ك ت ب " * نص

تستغل محولات الأوضاع النهائية هنا (*finite state transducers*) في جميع العمليات الخوارزمية التي تقوم بتحويل مجموعة من المتواليات في الدخل، تضاف إليها متواليات أخرى يراها الخوارزم ضرورية في الخرج. ويقوم المحول بتعديل هذه المتواليات في الدخل مسحاً وتعويضاً فنحصل على :

3 ج/ كتب / كاتب / اكتتب / ت * نص

كما أن هذا النظام يحيل إلى الأنحاء المحلية (local grammar) متى رأى ذلك ضرورياً، فالنحو المحلي محدد وواصف لمجموعة من المتواليات اللغوية، عبر القيام بمجموعة من العمليات التي تناط به كالفصل والوصل والنفي الذي يلحق الأوتومات ومحولات الأوضاع النهائية، كما تستغل الأنحاء المحلية في التحليل التركيبي الجزئي للغة، وهذا يعني أن مهمته تنحصر في وصف ومعالجة أجزاء أقسام الكلم كنحو الاسم ونحو الفعل ونحو الحرف..إلخ، كما يبين الشكل التالي :

[أ- ب] [أ- ب] +

ولبيئة نوج أربعة تطبيقات معلوماتية رئيسة تساعد على التدقيق والتوافق والفهرسة في كل عملية وهي:

أ-المدقق الإملائي الآلي:

يعمل هذا الأخير في التدقيق في المتواليات المعجمية خارج السياق النصي، حيث يميز بين المفردات المصنفة في المعجم الآلي للمفردات البسيطة وتلك المصنفة في معجم المفردات البسيطة المعربة وذلك بهدف أمن اللبس. ويعمل المدقق الإملائي كبرنامج للتحليل المعجمي استجابة لأهداف محددة، وهي التعرف على أقل عدد ممكن من الأخطاء الواردة أثناء القراءة الآلية، وتجدر الإشارة إلى أن هناك إمكانية لتدخل المستعمل في تصحيح الخطأ حينما يكون التصحيح الآلي غير ممكن. أما البرنامج فيتأكد أولا وقبل كل شيء من صحة الكلمة من خلال قاعدة المعطيات المصنفة في النظام.

ب - المدقق النحوي:

يشتغل هذا المدقق بصفة أساسية، بمحولات الأوضاع النهائية التي تقوم بدور رئيس ألا وهو رفع اللبس وأمنه. وبهذا الأسلوب يمكن وصف الأنحاء المحلية التي تنتسب إليها المتوالية، فعلى الآلة أن تتعرف على طرفي المحول، فإن تعرفت حصراً على الجزء الأول من المتوالية أرسل البرنامج خطاب الخطأ (*error message*) وهو:

< V : 3s >، لأن القراءة غير مكتملة، فمثلا عند تحليل الجملة المأخوذة من اللغة الفرنسية[21] التالية: (*il ne mangeras)، يرسل إليك البرنامج رسالة تخبرك بالخطأ الذي طرأ عند قراءة المتوالية المذكورة.

ج- برنامج التوافق الآلي:

يوظف معجم المفردات (البسيطة والمركبة) برنامج التوافق الآلي، بهدف إيجاد البنيات الصحيحة مورفو-تركيبياً في سياقاتها النصية الأصلية، وذلك

(21) انظر: M.Silberztein (1993). P.169.

للمطابقة والتحقق. وإذا كانت هذه التقنية قد استخدمت في النظرية الأسلوبية بالعودة، إن أردت أن تبحث عن أصل لفظة ما، إلى نصوص قديمة مثل الشعر أو الحديث والقرآن الكريم لتستشهد بها، فإنك على مستوى النظام الآلي ترتكز مبدئياً على برمجي يحتوي على جميع المفردات الموجودة في النص، فيقوم بالبحث في نصوص المدونة ليطابق النتائج التي توصل إليها بالتحليل والمعالجة وما هو موجود في المدونة الأصلية مرفقة بحزمة الأنحاء المحلية التي تساعد على إقامة هذا التطابق. كما يقوم هذا البرمجي بمعالجة نصوص ثابتة وأخرى دينامية، وهذا ما يميز بيئة نوج كونها مفتوحة المصدر، تفاعلية من حيث مقارنتها بالمعطيات الجديدة، ولذلك تطلب الأمر تطبيق مفاهيم لغوية وكيانات اسمية وعبارات عقلانية لتنظيم وفرز وتصدير هذه النتائج وفي النهاية يقوم البرنامج ببناء تقرير إحصائي اعتماداً على أرقام دقيقة عن البنيات المعالجة.

هذا وتشغل بيئة التطوير "نوج" ما يسمى بالموافق الآلي (*concordancier*)[22] الذي يعتبر أداة معلوماتية تسمح بالكشف السياقي عن المتن النصي باستخدام بعض الأبنية الخاصة ومقاماتها حيث تبرز فيها هذه الأخيرة محددة لسانياً وفي موقع ضمن أقسام الكلم، فيخصص لها نحو معين يستطيع وصفها بدقة كما هو الشأن مع نحو المفردات المركبة مثلا. علاوة على ذلك تجرى تطبيقات بتشغيل الموافق الآلي باختيار مقولات معينة للمعالجة حيث يقوم النظام باستخراج المفردات المناسبة من المدونة تصنيفاً وتشخيصاً. سواء كانت مصطلحات أو مفردات أو أسماء أعلام أو كيانات جغرافية أو وحدات اسمية مختلفة[23]. وإذا ما أردنا رصد ألفاظ تتعلق بحقل "الحب" في نص روائي مثلا، فإننا سنستدعي الموثق الآلي في جرد جميع المصطلحات والمركبات والألفاظ المتعلقة بالمجال المبحوث عنه وبتفعيل العبارات العقلانية، يعمل الأوتومات على رصد كل ما له انتماء أو علاقات دلالية بالكلمة – المفتاحية.

(22) فضلنا هذا المقابل في اللغة العربية اعتماداً على قاموس مصطلحات المعلوماتية، فرنسي– عربي لصاحبه إ.و.حداد، مكتبة لبنان حرف c.

(23) غازي.ع، معالجة الوحدات الاسمية في اللغة العربية: الأعداد المركبة نموذجاً، أشغال الندوة الدولية*CITALA*، معهد الأبحاث والدراسات للتعريب، 2012، الرباط/ المغرب .

د – الفهرسة الآلية:

إن عمل الفهرسة الآلية التي يستخدمها النظام للتعامل مع معجم المفردات البسيطة والمركبة، يقوم بفهرسة جميع المداخل المصنفة في المدونة[24] سواء كانت معربة أم غير معربة، وهذه الكلمات المفتاحية هي عبارة عن مصطلحات تقنية حاملة للدلالة يستخدمها هذا المفهرس لأغراض التوثيق الآلي، ومع ذلك فالمفردات التي يستخدمها هذا المفهرس يجب ألا تكون غامضة[25]، فلا يمكن أن يقوم هذا الأخير بالتلخيص الآلي انطلاقاً من العناصر المكونة للمتواليات خاصة منها المركبة، لأنها لا تختزل بالضرورة دلالتها في عنصر واحد يمكن الاقتصار عليه كما يحصل الأمر مع التعبير التالي:

تفرق العرب شذر مذر

لأنك لا يمكن أن تكتفي بتفرق العرب فقط، مما يعرض ذلك إلى إجراءات تقنية للتصحيح والتدقيق في بنية يعالجها محول الأوضاع النهائية.

ويسمح هذا النظام بإمكانية عالية للوصول إلى البيانات عن طريق برمجيات حاسوبية أخرى ملحقة بالبيئة[26] والاستفادة منها في معالجة قضايا كثيرة. ولما كانت هذه البيئة اللغوية متصفة بهيكلة بياناتها المتراصة من حيث العلاقات والروابط بينها، فإن ذلك يجعلها تُمَثِّل بياناتها النهائية بشكل سريع وسهل المنال بالنسبة لأي مستخدم، بكل المقاييس المعيارية المعروفة والتي تستجيب لأساليب العرض بلغة "أكس إم إل" (XML) أو حتى "ميتاموديل" (Meta Model) للهيكلة "تي إم إل" (TML).

---

(24) لا يمكن للمفهرس أن يعثر على البنيات المركبة والمتلازمة في النص إذا لم تكن مدرجة ومصنفة في المدونه.

(25) انظر: M.Silberztein (1993). P.179.،

(26) تم تزويد نظام أنتكس(INTEX) (قبل نوج) ببرمجيات صغيرة وظيفية وأساسية لاستكمال عمله الشامل، وفي هذا المضمار، نجد ما توصلت إليه الأبحاث الإحصائية–لمختبر تابع لجامعة فرانش كومتي ببيسانكون بفرنسا. إذ تمكنت من تزويد المعجم الآلي للغة العربية بحوالي (340000) مدخل قاموسي وتولدي إلى حوالي (500) ألف شكل فرعي.

وللتحقق من توافق البيانات المتوصل إليها في التحليل، فإن البيئة تقوم بوصف قواعد التصريف والإعراب والتحليل الصرفي، كما تتعرف بشكل آلي على أسماء الكيانات، بالتدقيق الإملائي لتصحيح بعض الأخطاء مع التشكيل الآلي بالنسبة للغة العربية والمفهرس الآلي مع الاختبار والتثبيت بالتوافق الآلي مع محتوى الشابكة في كل ما يتعلق بالسياق.

2 – بيئة "نوج" *(NooJ)*، أداة للمعالجة الآلية وللتعلم اللغوي:

إن الهدف الأساسي من بناء بيئة "نوج" لم يكن ترفاً تقانيا (تكنولوجياً)، بل ضرورة لتحليل المدونات اللغوية إلكترونياً وبجميع اللغات الطبيعية، وإذا كان الأمر يهدف إلى معالجة اللغات والترجمة الآلية والتوثيق الآلي والاستخراج المصطلحي.. فإن الغاية الكبرى هي بيداغوجية تمكن المستخدم/المتعلم مستكشفاً المعلومات اللسانية الدقيقة وخاصة منها الأنحاء المحلية أوحتى كذلك الوظائف الدلالية والتركيبية، وهذا يحصل بصفة تفاعلية وبمعالجة قبلية للمدونة. فعند دراسة المستوى الصرفي–المعجمي (كما سنرى في بعض الأمثلة)، يتعرف النظام على البنيات المحولة والمشتقة ويقترح تصنيفا لها حسب نماذج البناء. لكن مع ذلك هناك صعوبة في تعلم اللغة واللسانيات عموماً، وهذا يرتبط بالإنجازية المثلى للغة لدى الأفراد، ولذلك وجب الحذر في التطبيقات البيداغوجية، مع أن نوج استطاع بتطبيقات إجرائية صارمة أن يتخطى هذه الصعوبة، وسنقتصر في مقالنا هنا على بعض التطبيقات المبسطة من قبيل الوسم الصرفي–التركيبي لما له من نتائج مرضية على المستوى التعليمي، لكونها ليست بحال من الأحوال نماذج للأخطاء التي تشوش نوعاً ما على العملية، ولهذا السبب آثرنا اعتماد معالجة أولية وأخرى مراقبة من قبل أدوات المعالجة الآلية. هذا إضافة إلى أن الموارد اللسانية المعتمدة في المدونة نسبياً مركبة وهي في طور الحلول لدى المختصين لكنها مع ذلك واعدة بالنسبة للتطبيقات البيداغوجية وعند المدرسين بصفة خاصة.

سيساهم هذا المقال، أيضا، في تقديم نظام نوج باعتباره بيئة لتعليم اللغة واللسانيات منذ نشأته إلى الآن[27] تنقيحاً وتعديلا، وسنركز على الأنشطة التعليمية ولاسيما المستوى الصرفي – المعجمي في سياق تعلم اللغة العربية في المراحل المتقدمة أوكلغة أجنبية.

يستطيع "نوج" أن يبني مدوناته النصية ويختبرها ويديرها ويصفها صورياً على نطاق واسع وعلى شكل أنحاء إلكترونية ومعاجم آلية، كما يستطيع أن يرضي جمهور اللسانيين والمعلمين في وصف المستوى الصرفي والتركيبي والمعالجة الآلية للوثائق، مثل تحليل المدونة وخاصة حينما يتعلق الأمر باستخراج المعلومات والبحث عنها كلما تعلق الشأن بالاشتقاق أو بالتحويل الصرفي أو حتى شرح مداخل معجم ما، وهذا ما يفضله المدرسون أكثر من اللسانيين لأن غاياتهم تعليمية خالصة.

وبما أن مصادر البيئة مفتوحة، فإن وظائفها رهن إشارة المستخدمين ونتائج معالجاتها النصية قابلة للفهم والمعالجة في آن. مما يجعلها نظاماً تفاعلياً بامتياز، يكون قادراً على أن يبني أنحاءه المحلية في بضع دقائق، ويجعل بالتالي معلمي اللغة، غير المعلوماتيين منهم، يتكيفون بسهولة مع القاموس وأنحاء البيئة، لاسيما عند تطبيقها مع متعلمين متقدمين وعلى مدونة نصية من اختيارهم. وعلى هذا الأساس جهزت البيئة بالأدوات التالية:

1. جهز "نوج" بلغات خاصة، ليصبح قادراً على معالجة المدونة والوثائق بكافة أنواعها (وثائق *XML* و*Microsoft Word* وصفحات الويب...إلخ).

2. كما جهز بأدوات آنية شبه – تطبيقية مثل مختبر (*TP* أو *Labs*..) لإعطاء الفرصة للمستعملين حتى يتمكنوا من التدرب والتمرن (في مرات متكررة) على معالجة بعض المشاكل اللغوية والمعلوماتية الخاصة (تسنين النصوص وأبجدية وكتابة كل لغة وأنواع العبارات العقلانية التي تتعلق بالصرف الإعرابي مثلا).

(27) M.Silberstein 99.

3. وقد يقوم المستخدم متى شاء بإضافة وتخزين تمارينه على شكل مشاريع كماً وكيفاً، يمكن أن يعود إليها وقتما شاء، فمثلا يمكن للمدرس أن يحضر مدونةً ومعجماً وبعض الأنحاء الضرورية (لبناء الرسوم) وموارد معجمية وتركيبية، وحتى كذلك تحرير وثائق أو جملة من التمارين وتخزينها مما يتيح للمتعلمين ملاءمة المشاريع مع أهدافهم التعليمية والتي تجيب عنها الوثائق المحتفظ بها كمشاريع.

4. تدبير ونشر الموافق الآلي الخاص بالكشف السياقي وتطابق المعلومات وذلك بناء على مجموعة من الطلبيات في الآن نفسه وتخلص العملية بفرز النتائج غير الصحيحة ثم حفظ الصحيحة بناء على ما توصل إليه الموافق الآلي وهذا ما يتيح معالجة سلسة وتدريجية للمشاكل اللسانية الموجودة.

نطلب من المتعلم التمكن من معرفة ألفاظ معينة موجودة في المدونة، مثل البحث عن المصادر أو الظروف أو الأعداد للتعرف على التواريخ بحسب ما يحصل عليه الموافق الآلي، إذ يصبح المتعلم بعد ذلك ذا قدرة معرفيةٍ على القيام بتمارين وعلى مواجهة ظواهر لغوية أخرى وذلك بواسطة طلبيات جديدة متقدمة، آخذا في عين الاعتبار، سياق الكلمة المعالجة وكل مرحلة تكون موضوع التمرين، حيث يصبح الموافق الآلي- هنا- بمثابة مراقب ومشرف يساعد المدرس على نشر وانتقاء وحفظ المعلومات الصحيحة الموجودة والكامنة في الكفاية اللغوية لدى المتحدثين بالعربية.

2-1- تطبيقات صرفية – معجمية على اللغة العربية:

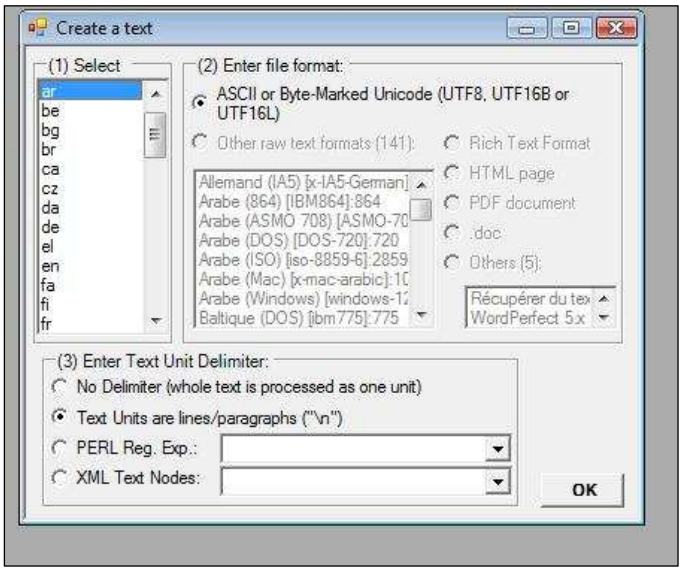
سنقدم تطبيقاً صرفياً على العربية مشيرين إلى أن نظام "نوج" كما هو يحتوي على معلومات تقوم بدور المحلل اللساني، وهو ضمن أبجديات التحليل النصي اللساني (سلبرشتاين 91). وقبل البدء، لا ضير من ذكر أوليات عمل البيئة المتمثل في التحميل والتثبيت واختيار اللغة[28] كما يلي:

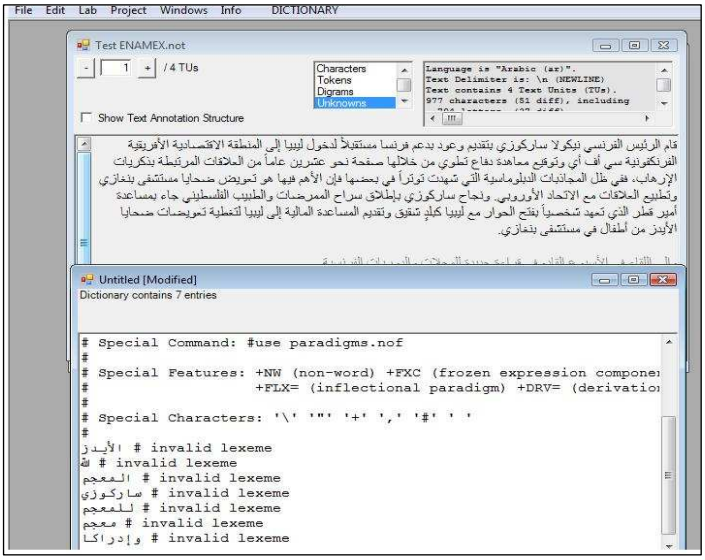
(28) نظام نوج ليس بعدُ مجهزاً براصد لغوي وهذا لا يزال من بين الطموحات العلمية للمهندسين الباحثين، لأن ذلك يقتضي مدخلات لغوية متعددة، فتم الاقتصار في البيئة على الاختيار اليدوي للغة وعلى اعتماد اللغة الإنجليزية كلغة دخل تلقائية فقط.

لاختيار اللغة العربية والمدونة يجب اعتماد الخطوات التالية:

File → new → file → text

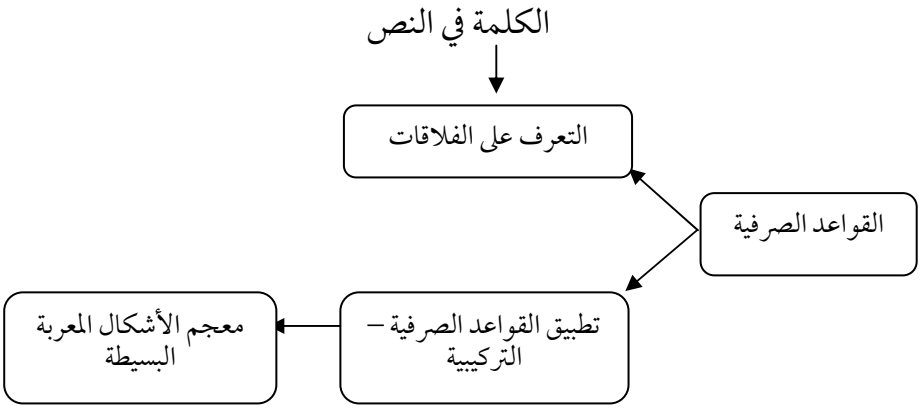
File → open → text





نوافذ لاختيار اللغة والمدونة والقاموس في نوج

بفضل جودة المعلومات ومناهج التمثيل والتخزين يسهل على المستعمل توظيف المعطيات واسترجاعها، وبما أن العربية إعرابية وتتطلب تشكيل حروفها فإن التطبيقات الحاسوبية تفترض تحليلا مدققاً لمتوالياتها، وخاصة ما يتعلق بالتشكيل الآلي[29] والبحث عن المعلومات، فيجب عدم الاكتفاء بالتأكد من علاقة الكلمة بالمعجم، بل كذلك بتقسيم وبتفليق الكلمة إلى مورفيمات بحيث يصبح كل جزء منها في علاقة مع المعلومات الصرفية-التركيبية المناسبة. الأشكال المعربة في اللغة العربية، ويمكن توضيح عملية التعرف على الكلمة كما يبين الشكل التالي:

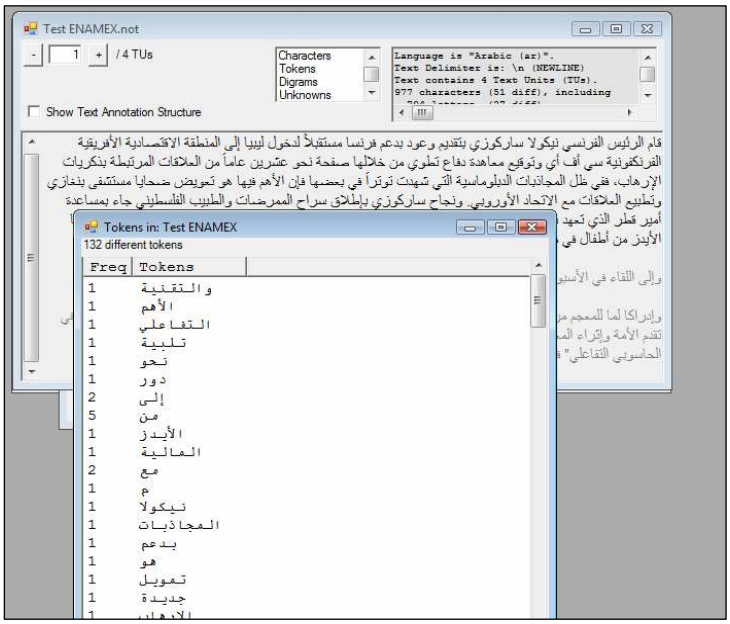


شكل 1: المعلومات اللسانية المتعلقة بالشكل المتعرف عليه (*Tokenisation*)

معمارية المحلل الصرفي: سلسلة أجزاء الكلمة

(29) عملية التشكيل الآلي من بين صعوبات التحليل الآلي للنص العربي لأنه على مستوى النتائج يطرح إشكالية غموض المتواليات، ويحاول الباحثون هنا اعتماد سلسلة من الإجراءات المتمثلة في: وضع قواعد ورسوم (graphes) للشدة والهمزة والتنوين أولا (لأنها تصاحب التحليل الصرفي) ثم فرز الصوائت عن المصوتات مع الاحتفاظ بكل حالة قصد التحليل المعجمي، مع ذلك لم تف هذه الإجراءات بشيء، وأضاف واضعو نوج تطبيقاً جديداً يتمثل في البدء بالعقدة الأولية لمحولات الأوضاع النهائية بقراءة السلسلة المعالجة مع وسمها والتعرف عليها بحيث يمكننا رصد الإدغام في السلسلة، فإذا تعلق الأمر بالمصوت، نستمر في اتجاه عقد المحول الموالي حتى نعثر على الصائت، واذا تعلق الأمر بالتنوين نتوقع صائتاً مائعاً (ي) في وسط السلسلة، أما اذا كان الرمز المقروء من قبل المحول هو صامت قنستمر إلى غاية العقدة النهائية، وعلى العموم إن تمكن المحول من قراءة السلسلة برمتها نقول إنها معروفة ويمكننا إضافة باقي المعلومات، وإن لم يحصل كذلك قلنا إن المتوالي غير معروفة، للمزيد انظر : سليم مصفار (2008) ص.111.

إن فوائد عملية إحصاء تدفق الأشكال اللغوية (*Tokenisation*) تكمن في إعطاء فكرة للمستخدم بأهمية أجزاء الكلم في النص دون تصنيفها إلى مقولات اسمية أو فعلية أو حتى حرفية كي لا تتعرض إلى غموض مما يسهل على البرمجي ولوج التحليل الصرفي – المعجمي المبتغى من العملية كما تظهر نافذة نوج التالية:



شكل 2 يبين إحصاء تدفق الأشكال: (أسماء/ حروف/ أفعال) (Tokenisation)

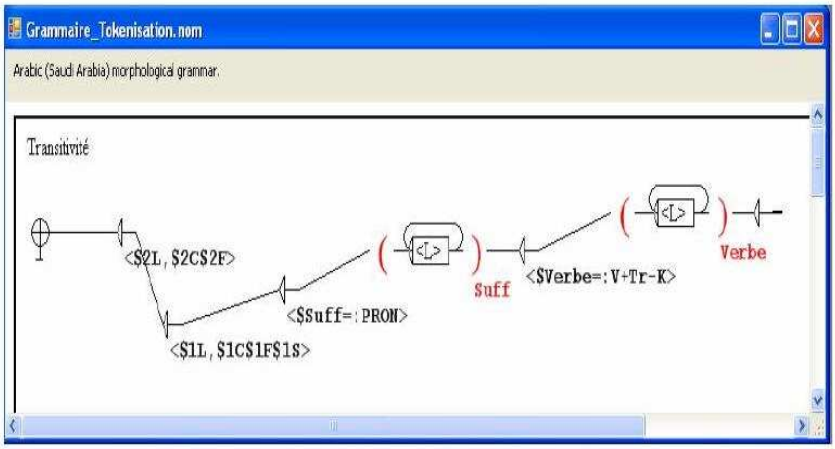
2-2- التشكيل الإعرابي واطراد الفعل:

من بين ما واجه الباحثين في عملية التعرف على الوحدات الفعلية في نص المدونة المعجمية العربية مشكلة الاطراد الفعلي في نظام اللغة العربية، فكلما تعمق المرء في وصف الحالة الإلصاقية للفعل ازدادت الصعوبة، وهذا ما يقتضي وصفاً تركيبياً أولياً عند بناء المداخل المعجمية[30] مما يسمح بفهم سمة الفعل أهو متعدٍ

(30) ينظر سليم مصفار 2008.

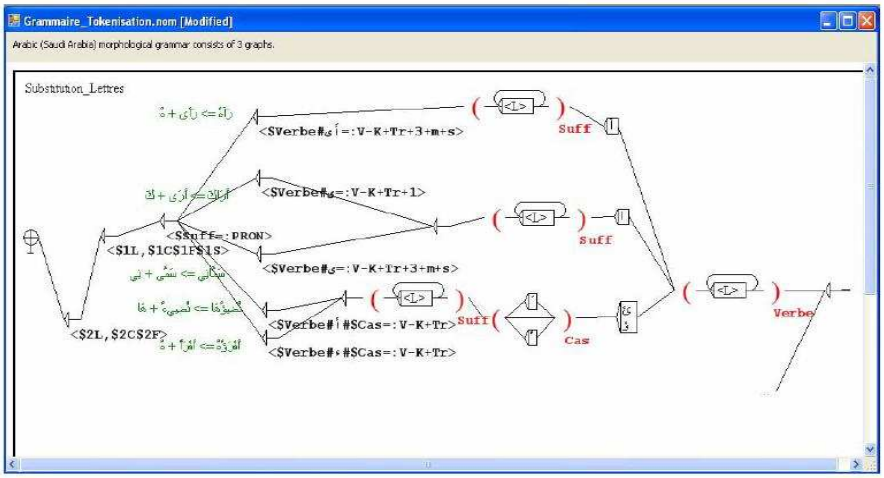
أم لازم وكذا طريقة تصريفه مبنيا للمجهول أم للمعلوم مثلا الفعل "كتب" (متعدٍ) يقبل إلحاق الضمير "ه" في كتبه (kataba+hu)، في حين لا يُقبل هذا في حالة الفعل مصرفاً للمبني للمجهول في كتب+ ه (kutiba + hu) نفس الأمر في الحالة الأخيرة يحصل مع فعل "مات" (لازم) الذي لا يقبل زيادة ضمير " ه " في مات +ه (maata+hu).

وهذا الرسم يمثل كيفية معالجة الأتومات لمسألة تعدي الأفعال كما يلي:



رسم1 يبين اطراد الفعل المتعدي في اللغة العربية

انطلاقاً من هذا الرسم يتبين أن الشكل المتعرف عليه كلاصقة فعلية "ه" ينبني على سلسلتين اثنتين تشكلان متوالية حرفية يمكن تمثيلها في البرنامج بالعبارة التالية: *<L> ومحفوظتان في المتغيرين الفعليين : Verbe $ و $Suff بحيث يجب أن يعثرا على الاطراد المعجمي الموجود بين" < " " > " وهذا ما يعتبر ضماناً لأمن لبس هذه البنية كما هو مبين في الرسم الموالي التالي:



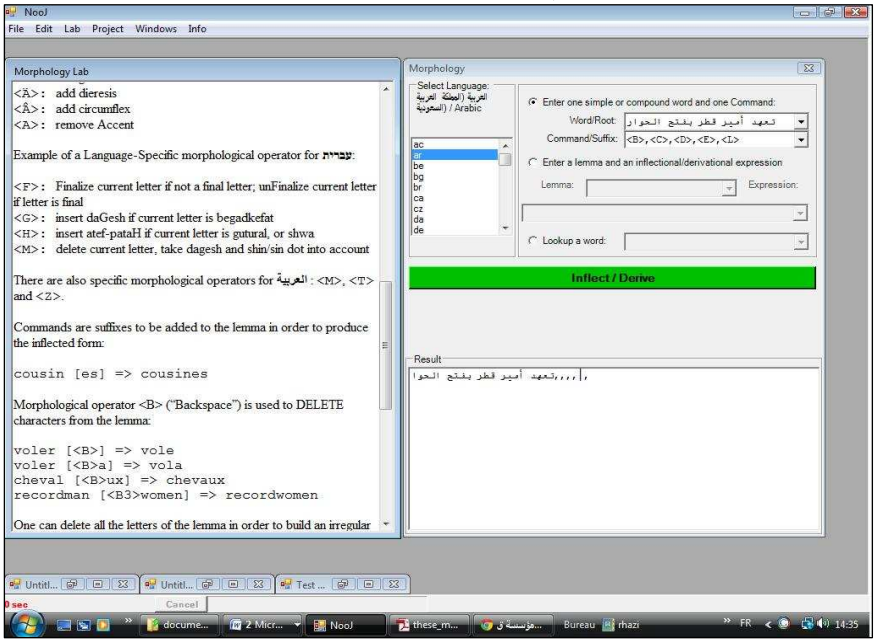
رسم 2 يبين القواعد الصرفية للأشكال : اطرادات صرفية بتعويض الحروف

وقد تكون مسائل الاطرادات الصرفية- المعجمية غير متعلقة بالتشكيل الإعرابي فقط، بل بالإملاء حيث تتشابه بعض المتواليات والاطراد الفونولوجي مما يتطلب وضع ألغوريثم جديد للتحليل المعجمي كمراحل لإعادة التشكل الآلي بهدف إضعاف الغموض المعجمي لبعض الأشكال الفعلية، وخاصة تلك المتلوة باسم موصول أو بحرف وهذا ما يفرض كذلك تصحيح بعض الأخطاء في الكتابة : المد والهمزة مثالين لمعالجة العلة (ي) وإعادة كتابة (ة) في آخر الكلمة، ثم تصحيح التنوين.

-3- تصفية وفرز المعلومات من خلال البحث عن المصادر الفعلية في النص: أمثلة تطبيقية معجمية-صرفية في العربية:

نفترض مجموعة من المتعلمين في مراحل متقدمة في تعلم العربية ولذلك نقترح القيام باشتقاق الاسم من الأفعال مثل "كتب" كما سنرى، فبالإضافة إلى القدرة التواصلية المفترضة للمتعلمين نمكنهم من معالجة المعجم لتعلم اللغة الثانية بشكل فردي لكنها مرتبطة بتمارين الإنتاج والفهم. وقد كان هذا المستوى محط دراسة العديد من المختبرات التي ركزت على استيعاب المعجم وما قد يرتبط به من صعوبات مثل المعجم السياقي الذي يطرح مشكلة المتلازمات

والمسكوكات مثلا. وقد تبين أن جعل المتعلم يقوم بعملية البحث واستخراج الألفاظ المناسبة بصفة بنيوية في السياق[31].



نافذة التحليل الصرفي الاشتقاقي والإعرابي في نوج

الأنشطة التي نفضلها هنا هي ملاحظة الصيغ الصرفية للأفعال أو كذلك صيغة الاسم/ المصدر في النص مثلا بداية من الفرضية الوصفية اللسانية أولا، بغية جعل المتعلم يتعرف على الأدوات المعجمية بصفة دقيقة ولتمكينه ثانياً من معرفة القواعد الصرفية التي تكون الكلمة والقواعد المعجمية المصاحبة وهذا يظهر من خلال المحترف التعليمي كما في الأمثلة الآتية:

مثال **دَرسَ** :

(درس : darrasa) FLXDRV:V+Tr+FLX=V_darrasa+DRV=D_darrasa، درس

هذا الفعل "درس" في حالة المفرد المذكر الغائب مع الوصف الإعرابي : V_darrasa، تولد أشكال تصريفية جديدة تسمى "بالبراديكم"، بحيث أن

(31) Treville 96

عمليات التصريف المتمثلة في الجهة والصيغة والعدد والنوع تنتح عنها نحو 122 شكلا إعرابياً لكل مدخل معجمي. ويمكن أن نذكر، على سبيل المثال، شكلا إعرابياً واحداً من ضمن العدد المذكور كما في التالي:

<LW>ي<R4><S>,<R><S>'/A+B+3+m+s(<LW>yu<R4><S>i<R><S>u /A+B+3+m+s)

هذا التحويل الصرفي في نوج يتضمن الأشكال التالية:

- <LW> : وتعني وضع العلامة (|) في آخر الكلمة (درس|)، عند رأس الوحدة المعجمية نحو اليسار (|درس)
- ادمج (ي) في رأس الشكل (ي درس)
- <R4> : اقفز على أربعة أخرف اتجاه اليمين (يدرر|س)
- <S> : امح الحرف الموالي في (يدر|س)
- ادمج المصوت (i) في (يدرس)
- <R>: اقفز على حرف واحد نحو اليمين (يدرس)
- <S> : امح الحرف التالي في (يدرس)
- ادمج المصوت الأخير في (يدرسُ)

هذه الاجراءات تمكن من توليد الشكل التالى:(يُدَرِسُ) التى تتألف مع الأشكال الإعرابية التالية:

V+Tr+A+P+3+m+s

فعل (درس) كفعل متعدٍّ (V+Tr) في المضارع (P) مسند إلى المذكر (m) المفرد الغائب (3) مبني للمعلوم (A)

في "نوج" يمكن تطبيق إجراء "البراديكم" الاشتقاقي المتعلق بالمثال المذكور ألا وهو (D_darrasa).

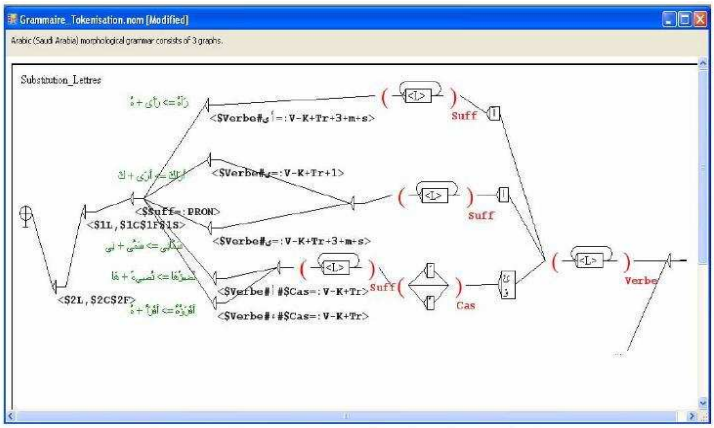
كما يمكن توليد مختلف الاشتقاقات الممكنة الأخرى نحو: (مُدرس) بتطبيق الإجراء التالي:

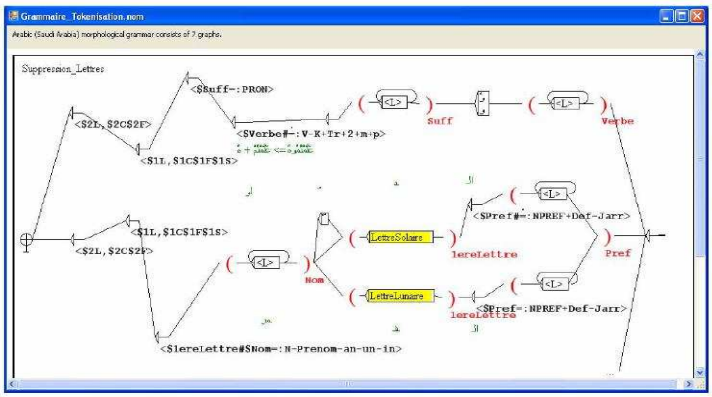
<LW> مُ <R4><S>,<R><S>/N+PA

مُدرِس:

<LW>mu<R4><S>i<R><S>/N+PA

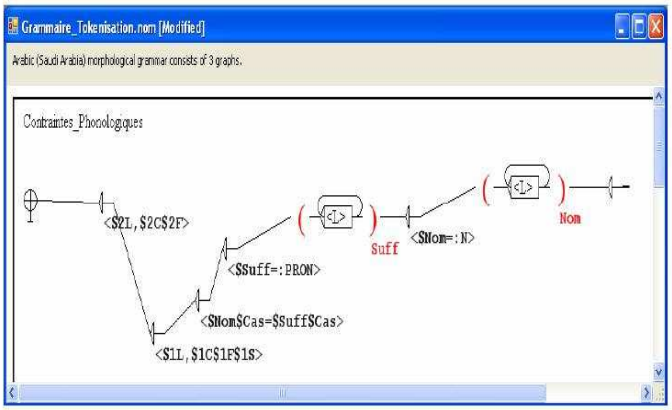
وهكذا مع صيغة المبني للمجهول حيث نحصل على : مُدرَس كما مع الإجراءات التالية:





رسم 3 و4 يوضحان تحويلات فعل "درس" بالتعويض والحذف

ولا مجال لتحليل هذه الأفعال في النظام الصرفي للعربية، كما تذهب نسرين عبد الله شنوف العلواني، إلا باعتبار أصولها من ناحية ووزنها، فعن طريق تصور الحروف الأصول لعزل المعنى البسيط وليس هو نهاية الأمر سوى الحدث بمعنى المصدر، وعن طريق اعتبار الوزن نتمكن من تشخيص المعنى الإضافي المستوعب ذلك أن هذا المعنى الإضافي ليس وليد زيادة حرف أو التصرف في حركة أو حذفها وإنما هو وليد العناصر الحرفية والحركية في تعاقبها وتفاعلها[32]، ففكرة المورفيم الزائد هي فكرة توزيعية تجعلنا ندرك تماماً أن مختلف التصريفات أو الأشكال المستخرجة هي متصلة بعنصر مشترك هو الجذر (د ر س) كما نستطيع أن نميز عناصر أخرى تحدد إذا ما كانت الكلمة اسماً أو فعلاً، وكذلك الطبقة اللغوية التي تنتمي إليها من حيث النوع، مذكر أو مؤنث أو من حيث العدد، مفرد أو مثنى أو جمع ومن حيث الشخص، متكلم أو مخاطب أو غائب. وهذه العناصر عبارة عن مورفيمات وقد تطرح في قلبها قضايا مورفو-فونولوجية من شأنها أن تحدث غموضاً في تحويلات الأشكال المعربة وهذا ما يوصف بالاطرادات الصواتية حيث ينشأ له رسم يعالج هذا الأمر كما يلي:



رسم 5 يمثل الاطراد الصواتي الناتج عن القلب أثناء التحويلات الصرفية

(32) نسرين عبد الله شنوف العلواني، البحث الصرفي في الدراسات اللغوية العربية الحديثة، أطروحة جامعية قدمت إلى مجلس كلية التربية – ابن رشد – جامعة بغداد لنيل درجة دكتوراه فلسفة في اللغة العربية/ لغة، بإشراف الأستاذ الدكتور هاشم طه شلاش، سنة 2003. ص. 330.

ففى المثال (درس) نجد المورفيم عبارة عن مقطع يقع في أول الكلمة، أي سابقة (ي) وهو يحدد أيضاً أن الفعل المسند إلى المفرد الغائب كما يدل على زمن وقوع الفعل في الحال أو الاستقبال، وذلك في مقابل (ندرس) أو(أدرس)، أو (تدرس) . كما نجد في كلمة (يدرسون) المورفيم (ون)، أن الدراسة واقعة من جماعة الذكور، وهو لاحقة، كما أن النون مورفيم دال على علاقة هذا الفعل بغيره من العناصر الداخلة في التركيب والواو هو مورفيم آخر يدل على الفاعلين.

نفس الأمر يمكن أن نقيسه على أمثلة أخرى مثل (كتب) لكن ثمة اطراد صرفي كما ذكرنا ذلك سابقا في حالة الفعل المتعدي، حيث يتحقق النظام من حالة التعدي من أجل رفع اللبس في إلحاق مورفيم الضمير (ه) التأكد من أن الفعل في حالة (+Tr) وغير مبني للمجهول (-K)

نافذة تمثل عمل برنامج التوافق الآلي في البحث عن المصدر من "ك.ت.ب"

يتوفر "نوج" على محلل صرفي يقوم بالبحث في المدونة وبالمعالجة النصية بناء على عبارات عقلانية صورية موحدة تحكم الأشكال والبنيات والصيغ والمقولات التركيبية والمعلومات الصرفية مثلا : العبارة التالية تمكن من البحث عن الكلمات التي تنتهي باللاحقة (ة) مثلا:

درس، مدرسة، وبمدرستهم

نطلب من المتعلم استخراج الألفاظ المتعلقة بالمصادر مثلا مستعينا ببراديكم عبارة عن متواليات من الفلاقات وأو مورفيمات في مدونة صغيرة الحجم (مقال صحفي) ومن اختيار المعلم، وللقيام بذلك وجب على المتعلم إعداد طلبيات متلائمة مع النص واستخراج الكاشف السياقي ثم أخيرا فرز النتائج الممكنة ويمكن أن تساعده شروحات (*annotations*) ليعرف مدى اختيار المداخل التي عثر عليها الموافق الآلي والتي تناسب في الآن نفسه الأسماء الظروف وتحتفظ بها ومحو الألفاظ التي سيقت عرضا بالكاشف : انظر الموافق الآلي التالي:

يستطيع نوج رصد مجموعة من المدونات النصية في نفس الآن ليبحث عن الفعل الأصلي للظروف المشتقة مثال:

يمكن تطبيق هذه الحالة على كل الألفاظ التي تشبه البنيات المحولة في أشكال أخرى مشابهة.

أول تطبيق بسيط لنوج هو استخراج المعلومات الصرفية من "النص الخام" والتأكد من عدم وجود علامات معينة. تعيق البحث، وبما أن البيئة مجهزة بالمحلل الصرفي الذي يسمح بإجراء الاستخراج والعلاج بناء على النصوص المختارة، يتم دمج الأشكال (lemmas)، وموافقتها بالقواعد النحوية أو المعلومات المعجمية. في المثال التالي يمكن البحث عن جميع الأسماء المصادر المشتقة من فعل (درس) كما في الطلبية التالية:

$"> || درس <N+MP="

طلبية بسيطة للعثور على كافة أسماء (N) نهاية ($) لكل فعل، درس

قد يطلب من المتعلم استخراج تكرارات أسماء تنتهي بسلسلة صرفية مختارة على سبيل المثال. للقيام بذلك فإن الطالب عليه صياغة الطلبيات لاستخراج الفهارس، وتصفية نتائج البحث غير المناسبة. وهذا من شأنه المساعدة على نقل النص إلى لغة (PERL) وهذه اللغة موجودة كعلامات الاقتباس في الطلبيات كما في الشكل التالي الذي يبين وجود علاقة مستمدة من طلبيات المتعلم السابقة وهي الاختيار بين المداخلات التي تتطابق مع تلك التي تم الحصول عليها في التوافق السابق كما يحصل مع مصادر الأفعال المشتقة من جذر (ك ت ب) هذا مع حذف أسماء وأبنية مشوشة على بحث الطلبية كما يلي:

نافذة توضح الطلبية الخاصة بالبحث عن المصادر

3 – ولوج المدونة وتصفيتها: البحث عن الاشتقاقات بالشرح التفاعلي للمدونة:

يمكن الاشتغال بالنحو المُمَثل بواسطة الرسوم التي تم بناؤها وحفظها في البيئة لتطبيقها على مدونة تحتوي على الوحدات الاسمية العددية أو الزمنية، كما تستغل النتائج المحصل عليها لشرح النص (annotation)،كما يسمح بتطبيق قواعد النحو (بيانيا). على سبيل المثال، يمكننا تقديم هذا النحو لمتعلمي العربية، ونطلب منهم تطبيقها على نص من النصوص الصحفية...إلخ. ينطبق هذا النحو على التسلسل الزمني لكل كلمة . يمكن للمستخدم اختبار إدخال المعلومات الجديدة والنتيجة يمكن استخدامها للتصحيحات التعليمية مثل الشرح الذي تقوم به البيئة التي يستخدمها الطلاب من جهة والتكيف مع نتائج المعالجة الآلية وتنفيذها تلقائياً، ومن جهة ثانية، التحقق من المعلومات، أي تلك الاشتقاقات أهي صحيحة أم غير صحيحة بناء على المدخلات القاموسية والقواعد الصرفية والنحوية الصحيحة في الكفاية المعجمية.

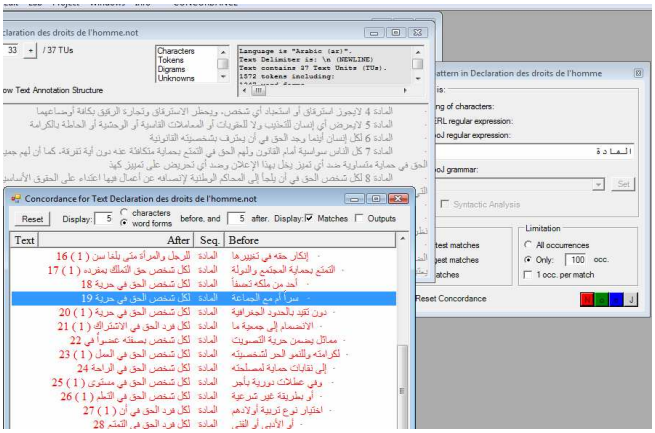
3-1– برنامج التوافق الآلي كمرحلة للتقييم:

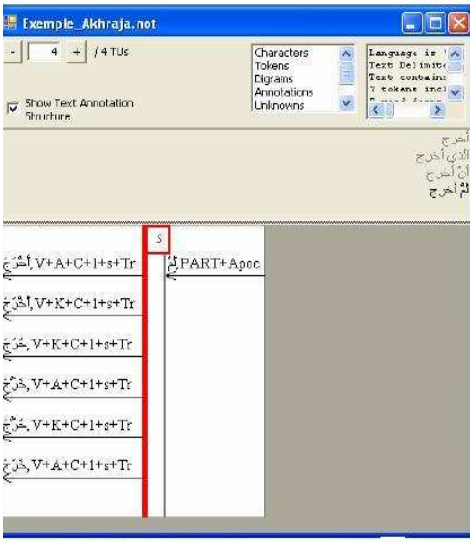
إن الشرح الآلي للنصوص والتحليل الآلي والبحث اللغوي والكشف السياقي، كلها مسميات تؤكد صلاحية وجهوزية المعلومات من حيث الجوانب المعرفية والتعليمية في استيعاب مفردات اللغة في سياق النص وليس خارجه، وهذا ما يضفي على المرحلة طابع التقييم وهو موضوع للتغذية الراجعة التفاعلية بين المستخدم / الطالب والنظام المفتوح المصدر، نلاحظ العديد من الشروح التي يمكن تطبيقها على النص مثلا إذا كان الاسم أو المصدر غير مشتق من الفعل (درس). بطبيعة الحال، فإن برنامج التوافق الآلي سيجيب بعدم تلاؤم المعلومات وتطابقها مما يفرض تعديلات جديدة للطلبيات (بشأن المشتقات اللفظية : من الاختيار إلى النص الرئيس، ثم أخيرا الفعل في صيغة المصدر)، وتعاملنا هنا لا يؤدي إلا إلى التركيز المعرفي تزامنا مع المعطيات السياقية ومع

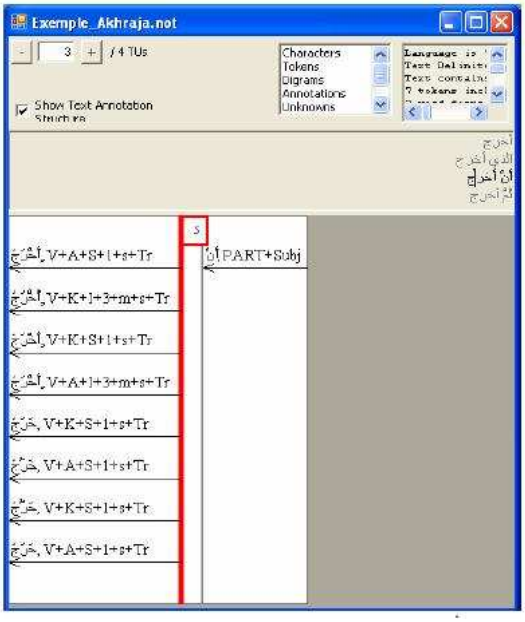
الطلبيات المعروضة على البيئة، تسمح الطلبيات باستغلال هذه الشروح التي سوف تعطي للمتعلم وتمكنه من وضع هذا الشرح في شكل (XML).

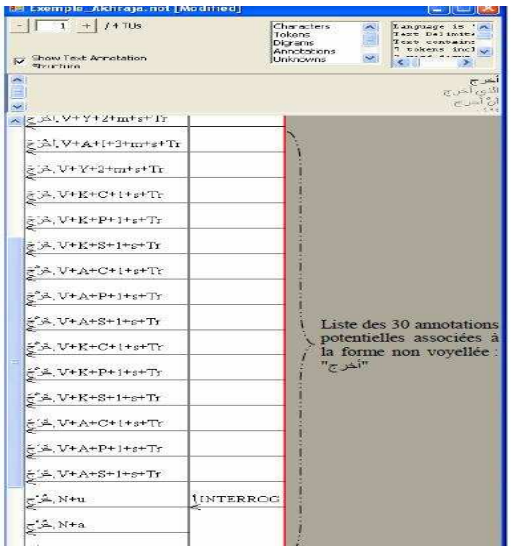
يستطيع الطالب بعد ذلك إجراء الفحوص الدقيقة على المسارد والقوائم، كما هي في نافذة التوافق الآلي مشروحة، استجابة للطلبية التي يمكن دمج نتيجتها وتصفيتها وحفظها في مشاريع لاحقة كأن تطبق على حالة البحث عن الظروف مثلا:

(+ <E> <A>) <DEVERBAL>وفي أكثر <V+INF>









نوافذ تظهرأجوبة على طلبيات الطلاب لفعل "خرج" مع الشروح **(annotations)**

اشتغال البيئة التي بين أيدينا، من حيث كيفية الاشتغال مع منظومة معاجم ديلا وخاصة التعامُل مع ديلاك وديلاكف، وكذا مكوناته وخوارزمياته ومجالات تطبيقاته المتمثلة خاصة في المدقق الإملائي الآلي والفهرسة الآلية، الإحصائية وبواسطة الكلمة المفتاحية و/ أو يدوياً كذلك، ثم تطبيقات ما يسمى بشبكة التوافق[33]. هذا، علاوة على إدراج بعض النماذج من المتواليات المتلازمة والمركبة هي عبارة عن رسوم تمثيلية لأتومات الأوضاع النهائية.

أولا – تطبيق قواعد النحو المحلي لتحديد ظواهر محددة وجديدة. ويمكن إجراء هذه الاستكشافات بشكل تفاعلي بواسطة المتعلمين أو إجراء معالجة مسبقة.

ثانيا – تطوير أدوات لسانية وذلك بالعمل على تخزين نصوص كبيرة ومتنوعة، لكونها أدوات قابلة للتكيف مع أي نوع من النصوص، أو حتى لغات عديدة. ويمكن للطالب، إذن، أن يقترح أنشطة توليد معجمي تلقائي وفقا للموارد المعجمية المتاحة مع تمارين قليلة.

ثالثا – تعلم اللغة العربية، على سبيل المثال، لدراسة آليات التشكيل المعجمي، وهي تتبع الكلمات مثل الفعل واشتقاقاته الممكنة وكذا تصنيف هذه المشتقات في أبنية معينة يحتاجها المستخدم عند استغلال البيئة المذكورة.

رابعا – تمكين المشرفين واستعلام الطلبة (على سبيل المثال)، على تمييز أنواع مختلفة من المصادر، بحيث يتم إنشاؤها تلقائيا بالتدريب، من خلال أخذ عينات من قاعدة البيئة، مع ضمان جودة البيانات المستخرجة، وإذا لزم الأمر تصحيح كل حالة مباشرة من قبل المعلم/ المشرف وذلك بدراسة المشتقات

(33) انظر سليم مصفار،

Electronic dictionaries for the automatic analysis: from INTEX to NooJ.(Université de Franche Comte,Besancon, France).in conference proceeding on Arabic electronic dictionary.Riad.April 2008.

الاسمية وتأكيد الصلاحية والتصفية، بالشرح التفاعلي للمدونة والتي تصبح في النهاية مشاريع تعليمية للمتعلمين.

خامسا: تطوير النشاط التعليمي باستخدام بيئة "نوج" *(NooJ)*، بشكل تفاعلي، بحيث لا نكون في حاجة لاستخدام سلسلة من البرامج المعقدة أو حتى تجهيز أو استخدام الحاسوب من قبل متخصصين فقط في المعالجة الآلية للغات. وهذا ما يضمن الاتساق بين المعالجة اللغوية والبرامج التعليمية التي يتم إجراؤها. بتطوير العديد من الميزات التطبيقية التربوية لبيئة "نوج" *(NooJ)*. إضافة على قدرته على الترجمة واستخراج المصطلحات... وعلاوة على ذلك، وغيرها من الميزات الأكثر تقدما، يمكن للطالب أن يتفاعل مع سلسلة من التحويلات اللغوية لتصنيف المبنى وما إلى ذلك من المسائل التي يتم التعامل معها في البيئة، وهذا ما يضفي على جل الأنشطة المقترحة التقويم التربوي المتساوق مع كل عملية، قد تساهم بدورها كما تحدد لاحقا تطوير البيئة.

خاتمة:

إن بيئة "نوج" تبدو واعدة لتدريس اللغات واللسانيات، ميزتها الرئيسة بساطتها لأنها تسمح لكل من المعلم غير المتخصص في الهندسة الحاسوبية ببناء الموارد اللغوية وباستخدام واجهات بسيطة فتصبح عبارة عن مشاريع تعليمية للمتعلمين وكذا تطوير النشاط التعليمي باستخدام البيئة محل التطبيق، فإنه ليس من الضروري استخدام سلسلة معقدة من المعالجة بعد تطوير العديد من الميزات للتطبيقات التعليمية، وهذا يضمن التنسيق بين المعالجة اللغوية والبحث عن المعلومات وتوليدها بصفة تلقائية في السياق من خلال استغلال نظام "نوج". الذي يمكن المدرسين من بناء الطلبيات الصحيحة التي يجب أن يتبعها المتعلم، وذلك بجعله قادرا على التمييز بين أنواع مختلفة من التحويلات والاشتقاقات، إذ يتم إنشاؤها تلقائيا بواسطة التمارين مع أخذ العينات عشوائياً. مع ضمان

جودة البيانات المستخرجة، حيث يمكن التحقق منها وتصحيحها إذا لزم الأمر في كل حالة على حدة، وغيرها من الميزات الأكثر تقدما في "نوج". كما يمكن أيضا أن تستخدم التحويلات الصرفية-المعجمية على سبيل المثال، في حالة البحث عن المصادر الفعلية، تكوينا للمتعلم بأن يلعب دور المساعد وذلك بالقيام بسلسلة من التحويلات الجديدة لتصنيف المبنى الذي يتم التعامل معه في برنامج التوافق الآلي، وبناء على مرحلة التقييم يمكن اقتراح الأنشطة التي سيتم تدريسها فيما بعد والتي بدورها تحدد التطورات اللاحقة للبيئة.

من خلال التطبيقات السالفة حاولنا أن نقارب بعض المتواليات البسيطة في العربية، وتبين مدى أهمية إجرائها في التحليل والمعالجة، ولأن ذلك يحتاج إلى بحث مستفيض لا يتسع له المقام الآن، فإننا اكتفينا بتوضيح قوة وإنجازية وجودة واتساع الاستغلال المبنين لبيئة نوج التطويرية لما لها من أهمية في:

1 – التنمية اللغوية بإتاحتها بناء واختبار وإبقاء التعريفات الشكلية بغطاء أوسع للغة، في شكل معاجم حاسوبية أو أنحاء إلكترونية يتمم بعضها بعضاً، نظراً لطبيعة عناصر اللغة المشكلة للمادة المعجمية.

2 – تطوير التطبيقات لمعالجة اللغة صرفياً وتركيبياً وتحليل المدونات بالنسبة للمختصين في التوثيق والتطبيقات الحاسوبية.

3 – وضع النماذج المفهوماتية والمنطقية والمادية للمفردات والروابط التركيبية والصرفية والمعجمية، ووضع حلول معلوماتية وبرمجية حاسوبية مع وضع الآليات والأدوات الكفيلة بالنفاذ إلى المعاجم والقيام بعمليات البحث والمعالجة بالتفاعل البيداغوجي المباشر.

4 – كما تتضمن هذه البيئة المجانية المبنية بأدوات عقلانية وهندسية، أدوات للمعالجة الآلية للغة مثل نظام إدارة البيانات ولغة البرمجة مفتوحة المصدر، مما يمنح إمكانية معالجة المدونات المعجمية بطرق البحث والاختبار

(مثل الصيغ الصرفية والتركيبية والدلالية) وهذا ما يتطلب صورنة مكثفة للغة العربية ووضع منصة جيدة للتحليل.

5 – ومن جانب آخر، تعتبر البيئة محللا آلياً لمنظومة المعاجم: (ديلاس وديلاسف وديلاك وديلاكف)، مما يثقل كاهله أثناء عملياته ولاسيما التحليل الصرفي، لذلك تم الرفع من مستوى أداء هذه البيئة لتوحيد المداخل المعجمية (*super-lemmas*) من جهة، ومن جهة أخرى، لتدبير منظومة المعاجم جنبا إلى جنب مع جداول المعجم- التركيبي .وقد أصبح بإمكان المستخدم ولوج معجمه سواء عن طريق الجداول أو عن طريق القوائم،أو حتى ولوج الخاصيات الصرفية والتركيبية في الآن نفسه.

أما الخوارزم الجديد للنظام فقد تقلص وفق الخطوات المنهجية الملخصة في الخاصية التالية[34]:

TRIE construction + minimization

6 – وفيما يتعلق بمحولات الأوضاع النهائية فقد استطاعت التقنية الجديدة أن تشفر المورفيمات وجميع اللواصق في وحدة معجمية ذات شكل موحد. مطبقة على لغات عديدة من بينها اللغة العربية، وتفهرس كل لغة حرفها الأول اللاتيني مثل (A) بالنسبة للعربية.وهذا جدول يلخص المعطيات الإحصائية للغة العربية في بيئة نوج كما في الجدول التالي [35]:

(34) انظر، سليم مصفار (2008) .

(35) نفس المرجع أعلاه.

المعجم الآلي	عدد الوحدات المعجمية	عدد الأشكال المعربة
الأفعال	10500	2734122 شكل منها 1290795 فعلا، و1443327 مصادر وغيرها..
الأسماء	15000	280267 شكلا
الصفات	4600	163866 شكلا
أسماء الأعلام	12400	
الكيانات	5300	

المراجع

- زايد محمد وسليم مصفار، نظرة في التركيبة الحاسوبية للمعجم الفرنسي، أشغال الاجتماع الثاني لخبراء المعجم الحاسوبي للغة العربية، مدينة الملك عبد العزيز للعلوم والتقنية، الرياض، أبريل2008.
- زايد محمد، دراسة في المعاجم الحاسوبية الفرنسية،أشغال الورشة الثانية حول المعجم الحاسوبي التفاعلي، مدينة الملك عبد العزيز للعلوم والتقنية، الرياض، أبريل 2008.
- شنوف العلواني نسرين عبد الله، البحث الصرفي في الدراسات اللغوية العربية الحديثة، أطروحة دكتوراه فلسفة في اللغة العربية، كلية التربية – ابن رشد – جامعة بغداد، بإشراف الأستاذ الدكتور هاشم طه شلاش، العراق، 2003.
- غازي عزالدين ولهلال محمد، معمارية المعجم العربي الإلكتروني: رؤية جديدة لمعجم آلي مُبنين، مجلة الدراسات المعجمية، المغرب، ع.8 2009.
- غازي عز الدين ولهلال محمد، استخراج ومعالجة المصطلحات:تجربة البيئة المجانية المفتوحة المصدر نوج (NooJ)، المنتدى المصطلحي، سوسة، تونس، أيام 2009.
- غازي عز الدين، قراءة في تركيبة المعاجم الحاسوبية التفاعلية بَيْئَةُ التطوير اللغوية (NooJ) نموذجاً، ورقة قدمت في الندوة العلمية حول "المعجمية والقاموسية والمصطلحية والمقاربات اللسانية الحديثة بجامعة منوبة"، تونس، أيام 3 و4 و5 يونيو 2010 .

- غازي عزالدين، معالجة الوحدات الاسمية في اللغة العربية: الأعداد المركبة نموذجاً، أشغال الندوة الدولية (CITALA)، معهد الأبحاث والدراسات للتعريب، الرباط، 2012.

- Benmoumen Elhaj, rapport syntagmatique en langues de spécialité (LSP) in linguistica communicatio.v6n 1et 2. 1994.

- Blandine Courtois et Max Silberztien, Les dictionnaires électroniques DELAS et DELAC, in Lingüística Comunicatio, Vol.1 N°.2.1989. PP. 64-80.

- Blandine Courtois, M.Silberztein, les dictionnaires électroniques DELAS et DELAC. In LinguisticaCommunicatioV1.N°1.1989.41-47.

- Bogaards, p. (1994). Le vocabulaire dans l'apprentissage des langues étrangères. Paris : Credif-Hatier.

- Chanier, T.& Selva,Th.(2000).«Génération automatique d'activités lexicales dans le système ALEXIA ».Sciences et techniques éducatives (STE, Vol.7,2. pp385-412.

- Corbin, D. Morphologie dérivationnelle et structuration du lexique. Lille : Presses Universitaires de Lille,1987.

- Cuq, J.-P. (2004). "Le lexique en situation d'apprentissage guidé : pour une méthodologie d'enseignement interventionniste dans l'enseignement du français langue étrangère". In langues", Linguisticae Investigationes, Tome XXII (1998-1999). pp. 59-70.

- Previuos work of Silberstein (2007) S. Mesfar (2007), S. Mesfar, A. Ben Hamadou (2009) and Kais Haddar (2009).

- Elhannach M, lexique grammaire de l arabe: classe des verbes qualitatifs, (suite) in Linguistica communicatio,V.1,N.2.1989.pp.31-41.

- MESFAR. Slim, Analyse morpho-syntaxique automatique et reconnaissance des entités nommés en arabe standard, Thèse en vue de l'obtention du titre de docteur en Informatique, Université franche comté, France 2008.

- Pierre-Andre Buvet, représentations métalinguistiques de phrases a partir de Transducteurs, Centre Lucien Tesniere, Universite de Franche-Comte (2002)

- Michael Zock & John Carroll, les dictionnaires électroniques, in revue TAL n° 44, 2003.

-Rhazi Azeddine, Mohammed Lahlal , NooJ Application for Arabic terminological extraction, Conference 2011 , 13-15 June 2011, Dubrovnik, Croatia.

- Kraif, O. (2003). "Propositions pour l'intégration d'outils TAL aux dispositifs informatiques d'apprentissage des langues". Lidil, n° 28. pp. 53-165.

- Max Silberstein, NooJ v2 guid, coyright 2002-2008.

- Silberztein. Max 1993) Locale grammar implemented in INTEX linguistic tool to extract multi-wods units.

- Silberztein. Max, Dictionnaire électronique et analyse automatique de textes,INTEX, Ed. Masson Paris 1993.

-Silberztein, M. (1999). "Indexing large corpora with INTEX". Computer and the Humanities, n° 33-3. pp. 265-280.

- Max Silberztein et Agnès Tutin, NooJ, un outil TAL pour l'enseignement des langues. Application pour l'étude de la morphologie lexicale en FLE ; revues alsic, Vol. 8, n° 2 . 2005a

- Silberztein, M. (2004). "NooJ : an oriented object approach". In Royauté, J. & Silberztein,

- Tréville, M.-C. & Duquette, L. (1996). Enseigner le vocabulaire en classe de langue. Paris : Hachette.

Logiciels et sites :

www.nooj4nlp.net (2011).disponible et gratuitement téléchargeable.