

# المعاجم الإلكترونية للغة العربية

أ. د/ محمد محمد الحناش

E- mail: [contact@al-erfan.com](mailto:contact@al-erfan.com)

ملاحظة: كتبت هذا البحث قبل أكثر من عشرين سنة، ضمن مشروع علمي كبير نفذ منه الكثير وما يزال بعضه ينتظر، وقد ارتأيت نشره على موقع مؤسسة **العرفان** لما لمست فيه من استمرارية معرفية تفيد جيلنا الصاعد من الباحثين الشباب، فهو يتضمن أفكاراً ما زال الكثير منها لم يطبق على لغة الضاد، حيث كثر أديعاء حوسبة المعجم العربي دون تبني أي مرجعية لسانية هندسية، فأفسدوا المعرفة اللسانية بالمعنى الهندسي، وباتوا ينظرون لما هم به أجهل، يخدعون شبابنا الباحث عن المعرفة الحققة بمصطلحات مزيفة لا علاقة لها بأي سبيل علمي يمكنهم من الإسهام في بناء مجتمع المعرفة، ينشرون العبث المعرفي، وهم لا يدرون أنهم يزيدون الجهل "غنى" والمعرفة فقراً؛ فأصبحوا مجرد بياعي بضاعة مزورة تنخرها الهشاشة المعرفية التي لا "تغني" ولا تسمن من جوع.

## تقديم

يعرف العالم اليوم ثورة تقنية كبيرة فرضت تغييراً جذرياً في التعامل مع المعلومات التي تقذف بها مراكز البحوث يومياً ، وأصبحت الأمم تتسابق إلى اقتناء أكبر قدر منها وتوفير سبل الاستفادة منها في وضع خططها العلمية المستقبلية ، ومن نتائج هذا التطور التقني نذكر :

- 1- تيسير التعامل مع المعلومات بمختلف أنواعها ، بهدف استغلالها وجعلها في متناول الجميع ، فكل من يملك حاسوباً شخصياً يستطيع أن يقوم بعمل فريق من الباحثين وفي مكان ضيق يكفي أن يستوعبه هو وحاسوبه .
- 2- تخزين كم هائل من المعلومات في عدد جد محدود من الأقراص لا يتجاوز حجمها كفاً مستعملها ، كما أن الآلة نفسها اختصرت وأصبحت في حجم كتاب يمكن حمله في اليد .
- 3- ولا حاجة للمثقف في عصر ثورة المعلومات إلى السفر إلى البلدان النائية للبحث عن كتاب أو معلومة ، فبفضل وسائل الاتصال يمكنه قراءة ما يوجد مثلاً في مكتبة الكونجرس الأمريكي وهو في بيته في الرياض أو في مكان آخر في العالم ، إذ يكفي أن يمتلك الباحث خطأً هاتفياً وجهاز حاسوب ليلج أضخم بنوك المعلومات .
- 4- سهولة فرز المعلومات والحصول على ما يحتاج إليه الباحث ، فمن ملايين من المعلومات التي تفد علينا يومياً يمكن انتقاء ما يناسب منها وتأجيل استغلال الباقي إلى وقت لاحق .

- 5- سرعة إدخال المعلومات إلى الحاسوب مع توفير إمكانية استرجاعها حسب الطلب وفي أقصر وقت ممكن .
- 6- ضمان سرية نقل المعلومات بحيث لا يلجأ إليها إلا المختصون والذين لديهم الحاجة الملحة في استخدامها .
- 7- ومن أهم ما جاءت به التقنية الحديثة سرعة الاستفادة من المعلومات المخزنة بشكل يضمن سرعة تنفيذ المشاريع العلمية التي كانت تنجز في وقت طويل .
- 8- القدرة على نشر المعلومات في زمن قياسي ، ففي وقت وجيز يمكن تسريب خبر أو معلومة دقيقة إلى الملايين من الناس عبر شبكات الاتصال العالمية ، وذلك دون مراعاة للخصوصيات اللغوية ، إذ يمكن نشر المعلومة بأي لغة .
- وهناك إمكانيات أخرى يضعها هذا التطور التقني الحديث رهن إشارة الباحثين في العالم ، دون مراعاة للفوارق الجغرافية ولا الفوارق اللغوية أو العرقية ، وما علينا إلا الاستفادة منه حتى نلحق بالركب .
- لكل هذه الأسباب أصبح مفهوم الذخيرة اللغوية أساسياً ومركزياً في بناء المعرفة عموماً وفي بناء المعاجم خصوصاً ، فهي تعرض الأساليب القديمة التي أصبحت تصنف بمقياس التكرور التقني من التقنيات الأسطورية .
- وأهم ما تقدمه الذخيرة اللغوية توفير الوقت مع ضمان الإتقان والإنجاز، إذ بفضل القدرة التخزينية التي توفرها التقنية اليوم يمكن إنجاز عمل ثلاث سنوات في سنة واحدة ، هذا مع توفير إمكانية التغيير والتجديد الفوري الذي تتيحه الأدوات المعلوماتية المتكورة التي توظف في بناء قواعد البيانات .
- إن توفير الوقت ينعكس إيجابياً على الجانب الاقتصادي ، مثلاً إن بناء معجم مؤلف من 1000 صفحة قد يتطلب طاقة بشرية هائلة تقوم بمهام مختلفة من بداية العمل إلى نهايته . بينما مع وجود الذخيرة يمكن الاستغناء عن الثلثين من العالمين في نفس المشروع ، ولهذا فوائد اقتصادية كبيرة يوفرها المشرفون على مثل هذه المشاريع العلمية .
- لهذا يمكن اعتبار الذخيرة اللغوية مقدمة أساسية في بناء المشاريع العلمية الضخمة مثل بناء المعاجم في جميع المجالات ، ذلك لأنها تمثل المادة الأساسية الذي يحتاج إليها واضعو المعجم . فإذا ضبقت الذخيرة بشكل جيد سينحصر عمل المعجمين في فرز المعلومات منها وتصنيفها بالشكل الذي يتناسب مع الخطة التي توضع عادة في بداية أي مشروع علمي . لهذا نرى أن إنشاء الذخيرة يعد مشروعاً منفصلاً عن فكرة إنشاء المعجم في حد ذاته ، وذلك لأن الذخيرة لن تكون صالحة فقط لوضع معجم واحد ، بل يمكن أن تنفع منها معاجم كثيرة ، كما أنها ستكون صالحة لإنجاز مشاريع علمية كثيرة غير معجمية .

## 0. مدخل

لن نكون في حاجة إلى التأكيد على دور الأدوات المعلوماتية في وصف أنظمة اللغات الطبيعية ,  
فذلك أصبح من بديهيات الأمور العلمية في عصرنا الراهن , كما أن المتنفس الطبيعي لكل النظريات والمناهج  
اللسانية أصبح هو ما تحققه المعلومات من تقدم في صياغة البرامج القادرة على تقييس ( simulation )  
دماغ الإنسان , وبذلك أصبحت قادرة على صياغة قوانين صورية تقوم بدور مزدوج : من جهة وصف النظام  
اللغوي في جميع مستوياته باستعمال لغة عقلانية ( Rationnel ) , ومن جهة أخرى , توظيف هذه اللغة في  
بناء برامج التوليد الآلي لجميع مفردات وبنيات اللغات الطبيعية وفق قوانين استعمالها العادي . وبهذا التقارب  
استطاعت الدراسات اللسانية حالياً تحقيق تقدم ملموس ظهرت نتائجه على شكل معاجم آلية ( إلكترونية ) قابلة  
للاستعمال العادي من قبل الباحثين والمتعلمين على السواء .

لقد تكونت لدينا القناعة الراسخة أن أي عمل علمي يصب في هذا الاتجاه يجب أن يتوفر له شرطان  
على الأقل : من جهة ضرورة تفر إطار نظري ومنهجي لساني واضح المعالم , وبفضل أن يكون هذا المنهج أو  
ذلك الإطار النظري معتمداً على أساليب صورية ذات طبيعة رياضية في تعامله مع النظام اللغوي . وفي هذا  
الصدد فإننا نرى ضرورة اعتماد الإطار النظري المعروف بالمعجم التركيبي ( Lexique - grammair ) الذي  
سوف نعرض به ملامحه الأساسية في ثنايا هذا البحث , ومن جهة أخرى ضرورة توفر حاسوبيين لهم تكوين أو  
على الأقل اهتمام باللسانيات حتى يتمكنوا من صناعة برامج تعالج بها أنظمة اللغات الطبيعية أي مكننة  
المعلومات أو النماذج التي يقدمها اللسانيون عن لغة من اللغات , إن موقفنا هذا الذي نلح عليه هنا ناتج عما  
لاحظناه على التجارب الحالية المتعلقة بمكننة اللغة العربية . فرغم المجهودات الكبيرة التي بذلت لحد  
الساعة والتي استغلت فيها إمكانيات تقنية معلوماتية كبيرة , فإنها لا تعدو أن تكون أعمالاً فردية , يغلب عليها  
الطابع المعلوماتي , ويغيب عنها الطابع اللساني , وما ذلك إلا لغياب الأساس النظري اللساني الذي نرى أنه  
يجب أن يكون الأساس العملي لكل مكننة . ولحد الساعة لم نطلع على تجربة في هذا الميدان استغلت إطاراً  
نظرياً لسانياً واضحاً , هذا على الرغم من التقدم الكبير الذي حصل في هذه النظريات في العصر الحالي .

يتأسس المعجم الآلي للغة العربية على وصف المفردات اللغوية من وجهة نظر تصنيفها  
conjugaison واشتقاقها derivation مع ربط هذا الوصف بالمستوى النحوي أي بالمعجم التركيبي للغة  
العربية . والحاجة ملحة الآن لوضع هذا المعجم لتجاوز المشاكل الكثيرة والعويصة التي تعترض المعالجة الآلية  
للغة العربية والتي من بينها التدقيق الإملائي والتركيبي ( النحوي ) للنصوص المكتوبة , والتعرف على الأشكال  
اللغوية في سائر مستوياتها , ونخص منها التعرف على مفردات النصوص المحررة باللغة العربية غير المشكولة  
وتمييز السليم منها من غيره , ومنها بناء معجم آلي عام للغة العربية يراعي تدرج مستويات المستخدمين في  
سائر أنحاء العالمين العربي والإسلامي , ومنها وضع برامج الترجمة الآلية أو بالحاسوب من وإلى اللغة العربية ,  
ومنها تعليم اللغة العربية بالحاسوب سواء للناطقين بها أو لغيرهم وهم كثيرون في العالمين الإسلامي والعربي ,  
وهناك أهداف أخرى كثيرة تصب كلها في تيسير التعامل مع اللغة العربية آلياً , ونشر قواعدها على أوسع نطاق  
مستفيدين في ذلك من التطور التكنولوجي الذي توصلت إليه البشرية في عصرنا الراهن .

وسيساهم هذا المعجم بشكل عملي في معظم هذه القضايا , وذلك ببناء قاعدة بيانات لمفردات اللغة العربية تتأسس على إطار نظري ومنهجي صوري , مبني على رزنامة من الخوارزميات التي تولد الفرع من الأصل , والأصل من الفرع , ومن الثابت المنهجية لهذا المنهج في المستوى التركيبي نذكر : نهائية الأشكال التركيبية للغة التي حصرت بنياتها الأساسية في خمس , تتفرع كلها عن البنية الأصلية ذات الشكل الصوري : ف س 0 ك , حيث إن : ف = الفعل ( ثابت أول ) , س 0 = الفاعل ( ثابت ثان ) , ك = متوالية المفاعيل في الجملة ( ك=0 , 1 , 2 ) ( متغير )

وبما أن المفردات والجمل المشتقة تخزن في الدماغ البشري بواسطة تحويلات محصورة العدد والشكل بالربط بين المستوى الأساسي والمستوى الفرعي المشتق باعتماد مفهوم الإطناب ( Paraphrase ) حيث تبين أن توظيف هذا المفهوم الإجرائي بشكل صوري أدى إلى نتائج إيجابية في الوصف اللساني لنظام اللغة العربية , فعلى كل مشروع يعمل في هذا الاتجاه أن يعتمد هذه الأدوات باعتبارها خوارزميات تمكن من بناء معجم تركيبى آلي للغة العربية على غرار ما هو موجود الآن في اللغة الفرنسية وفي غيرها من اللغات العالمية , والأوروبية منها بخاصة .

بالإضافة إلى قاعدة بيانات الجذور والمفردات التي يوظفها المعجم الآلي في التعرف على المداخل الصرفية يشمل كذلك بناء قاعدة البيانات لجميع بنيات اللغة العربية بأنواعها الثلاث : البنيات الأساسية العادية , والبنيات ذات الفعل العماد , والبنيات المسكوكة . وأغلب قواعد البيانات التي تم إنجازها لحد الآن للغة العربية لم تفرق بين هذه الأنواع الثلاثة , ومن شأن هذا النقص أن يعرقل وضع وتطبيق علاج آلي حقيقي للغة العربية .

إن أي مشروع للعلاج الآلي للغة العربية يجب أن يتوفر فيه شرطان : أولاً بناء قاعدة بيانات للمفردات العربية بسيطها ومركبها , وثانياً الاستناد إلى نظرية لسانية محددة<sup>(1)</sup> . الأول يجنبنا وضع القواعد اعتماداً على الظن مما يعرقل عملية التطبيق الآلي , والثاني يجنبنا مزلق وضع قواعد جزئية في البرنامج المولد لمفردات وبنيات اللغة . لقد تطورت المعالجة اللسانية للظواهر اللغوية بشكل يدعو إلى الاطمئنان إلى القواعد التي وضعها اللسانيون العرب للغتهم , وقد زاد هذا التطور في السنين الأخيرة مع تطور أدوات المعالجة التي توظف هي أيضاً لغة عقلانية تقترب بها إلى مجال المعلومات . لكل هذه الأسباب أصبح الاعتماد على نظرية لسانية في المعالجة الآلية للغة العربية أمراً حتمياً . وقد دفعنا إلى التركيز على هذه النقطة ما لاحظناه من غموض في القصد على البرامج التي وضعت لعلاج الآلي للغة العربية هنا وهناك تتعلق بمختلف مستوياتها . ومن هنا جاءت دعوتنا إلى التعاون بين اللسانيين والمعلوماتيين لبناء قواعد البيانات الصحيحة للغة العربية والتي ستكون أساس أي برنامج للعلاج الآلي لهذه اللغة , وهذا بالضبط ما نحن بصدد تنفيذه .

تجب الإشارة هنا إلى أن غرضنا من وضع هذا المعجم يتمثل أساساً في تقييس الدماغ البشري لسانياً في مستوى المعجم , وذلك في محاولة منا استكناه الكفاية المعجمية عند الفرد العربي وتعميد إنتاجية هذا

(1) انظر كتابنا : خوارزميات الاشتقاق التركيبي للغة العربية , منشورات التواصل اللساني , 1996 .

المستوى معلوماتياً و/ أو خوارزميةً . ولهذا فإن بناء هذا المعجم يدخل في شكل أساسي في ما أصبح يعرف اليوم باللسانات الحاسوبية العربية .

يتألف المعجم الآلي للغة العربية من ثلاث مستويات متكاملة : الأول وهو مستوى الجذور , وفيه يصاحب كل جذر بالمعلومات النحوية والمورفولوجية التي تستخرج منه , المستوى الثاني يتألف من معجم المفردات البسيطة , حيث أثبتت كل مفردة في قاعدة البيانات بناء على المعلومات النحوية والصرفية المتعلقة بكل واحدة منها . ويتفرع عن هذه القاعدة من المفردات البسيطة قاعدة أخرى تتألف من المفردات البسيطة . أما المستوى الثالث فيتأسس على قاعدة بيانات من المفردات المركبة . وهي بدورها تتفرع إلى مفردات مركبة تشمل العادي منها والمسكوك , ومفردات مركبة معربة . وهذا هو القسم الأول من قاعدة البيانات التي أنجزناها . وهناك قاعدة بيانات أخرى تتعلق بالمعجم التركيبي الآلي للغة العربية . وسيتم الربط بين القاعدتين ببرنامج يسمح بمساءلهما للقيام بالمعالجة الآلية في المستويين معاً , والهدف هو جعل الآلة تقوم بتركيب نصوص عربية سليمة , والتعرف على الجمل العربية السليم منها وغيره , القيام بمراجعة الأخطاء الإملائية والتركيبية في النصوص العربية , والتعرف على الجذور اللغوية وتمييز الزائد من الأصل في المفردات العربية , وتوليد واشتقاق المفردات صرفياً وتركيبياً , الخ .

هناك فرق بين المعجم الآلي للغة العربية كما تقدمه هنا والمعجم العادية اليدوية , ويكمن هذا الفرق في نوع المعلومات المقدمة في كل منهما . فقاعدة البيانات الموجهة لبناء المعجم الآلية تكون أشمل وأوفى في المعلومات التي تقدمها عن المدخل المعجمي الموصوف فيها , أي لا تغفل أي جزئية تتعلق به , حتى تلك التي تعتبر بديهية , وذلك لحاجة البرنامج إليها عند المعالجة . مثلاً نحن نعرف أن الفعل المضارع للمتكلم المفرد على وزن أفعل , وهي حقيقة صرفية وتصريفية بديهية في المعجم العادية التي ربما تتحاشى ذكر هذا الوزن مع كل مدخل فعلي , لكنها أساسية في المعجم الآلية , لأن البرنامج المولد سيحتاج إليها , فإذا لم يعثر عليها في قاعدة البيانات قد يتعثر عمله , دعك من الاعتماد على البرامج الآلية التي وضعت هنا وهناك لتصريف الفعل العربي<sup>(2)</sup> , فهذه في رأينا ما زالت في بداية الطريق , فهي لحد الآن لا تقوى على التعامل مع أبسط القواعد التصريفية , حيث إنها تصرف أي جذر بغض النظر عن مقبوليته أو حتى وجوده في اللغة العربية , ناهيك عن الأخطاء الإملائية التي تسقط فيها , وما ذلك في رأينا إلا لأنها برامج لم تركز في بنائها الخوارزمية الاشتقاقية على قواعد بيانات عربية صرفية - معجمية . ومن هنا جاءت الحاجة إلى الإحاطة بسائر الخصوصيات الصرفية والتصريفية للمدخل المعجمي في قاعدة البيانات . ليس هذا فقط , بل يجب أن تكون المعلومات المعطاة عن كل مدخل دقيقة وشاملة تتميز بالوضوح والاطراد , أي لا يجب أن يكون هناك أي نقص في كمية المعلومات المقدمة عن كل مدخل .

هذا هو الوصف اللساني الموجه لبناء المعجم الآلية للغة العربية . ولا تخفى صعوبة هذا العمل , فهو لا يمكن أن يسند لأي كان , وحدهم الراسخون في علوم اللغة العربية واللسانيات الحاسوبية هم الذين

(2) نقصد بهذا التعليق التجارب التي يقوم بها بعض الباحثين في أوروبا مثل تجربة الباحث يفري الذي وضع محللاً صرفياً للفعل العربي أطلق عليه اسم : تصريف .

يمكنهم القيام بهذه المهمة ، إذ إن الأمر لا يتعلق بعمل تقني محض بل يتجاوزه إلى الاجتهاد النظري في إطار مشروع لساني محدد ، وفي إطار المعرفة بالصناعة المعجمية المبنية على قوانين المعلومات . وذلك لأن بناء قاعدة البيانات من هذا النوع تتأسس على عملية إسناد سنن code لكل مدخل معجمي في القاعدة ، ويجب أن يكون هذا السنن صحيحاً أي موضوعاً في المكان المناسب ، خاصة وأن الحاسوب يتعامل مع الأشكال اللغوية لا مع دلالاتها ، ومن ثم يجب إسناد السنن المناسب للمدخل المناسب ، وإلا وقع خلط في المعالجة التي ستأتي فيما بعد.

## 1. المعاجم الآلية : الإطار المنهجي

تدخل عملية بناء المعاجم آلية ( ويطلق عليها أيضاً معاجم إلكترونية ) في معظم البرامج التي تعتمد طريقة من طرق التعرف على الوحدات اللسانية الدالة في الكلام البشري ، فقد عرفتنا هذه البرامج على أن نصاً ما في مرحلة أولى يكون قابلاً مبدئياً لعملية التجزئ إلى وحدات كرافية على شكل رسوم ( Graphes ) ، وهي المفردات / الكلمات . ولذلك فإن المعجم يبنى أساساً لتحديد كل واحدة من هذه الوحدات من وجهات نظر مختلفة : صورية ، دلالية ، صرفية ، صوتية ، الخ ، فإذا لم نعر على مفردة في المعجم فإنه سعيد ناقصاً ، وبالتالي فإن أي تحليل آخر في أي مستوى لغوي لاحق ، ولو كان أعمق من الأول سوف يتوقف ، أو على الأقل سوف يتعثر . وعليه فإن مفردات النص يجب أن تتلاءم مع مفهوم المدخل المعجمي ( Entrée lexicale ) بطريقة دقيقة تعتمد حتماً على أساليب جديدة تراعي فيها التقنيات مستمدة من علوم أخرى . وهذا العمل يحتاج إلى ما يلي :

### أ) المستوى الصرفي

- بناء قاعدة بيانات المفردات العربية تستخلص منها قاعدة معارف صرفية تتضمن جميع قواعد التوليد الصرفي في اللغة العربية ، وهذه الأخيرة تنتج عنها :
  - مولد صرفي
  - محلل صرفي
  - مدقق إملائي
- إجراءات تتعلق بمعالجة الكلمات غير الخوارزمية : أسماء الأعلام ، الكلمات الدخيلة ، المصطلحات العلمية الخ . وهذا يدعونا لمعالجة النقط التالية :

### ب) المستوى التركيبي

- قاعدة بيانات التراكيب الأساسية في اللغة العربية : العادية .
- قاعدة معارف القواعد المولدة للبيانات اللغوية في مستواها التحويلي الاشتقاقي .
- قاعدة بيانات بالبيانات العمادية في اللغة العربية .
- قاعدة بيانات بالتعبير المسكوكة في اللغة العربية .

هذا بالإضافة إلى مختلف التقنيات المعلوماتية التي سنتحدث عنها وهي مجموعة برامج خاصة بكل مستوى على حدة , والتي تعكس بشكل بارز ما تم وضعه لحد الآن وما سوف نصنعه مستقبلاً في إطار المنهج اللساني الذي نتبناه في أعمالنا كلها .

## 1.1 المنهج

منهجنا المتبع في هذا العمل تصنيفي يهدف أساساً إلى بناء نحو صوري خارج السياق , قوامه رصد الخوارزميات التي تتولد بموجبها المتواليات اللسانية , مفردة كانت أم جملاً , هدفه ضبط أساليب توليد البنيات اللغوية من الأصل النظري المفترض في كفاية المتكلم العادي , خلاصته بناء قاعدة معارف تشمل جميع القواعد الصورية المعتمدة في التحليل والتوليد , خطواته بناء قواعد البيانات التي تستخلص منها القواعد الصورية ( قواعد معارف ) , تجنباً للاعتماد على الظن الموهوم بالشمولية .

ولهذا كانت خطوات العمل تتمثل فيما يلي :

1- بناء محلل صرفي مؤسس على قاعدة بيانات للمفردات اللغوية في المعجم , وهو ما سنبني عليه قاعدة معارف القواعد الصورية , وتأسس هذه القاعدة عملياً على معجم للمفردات البسيطة والتي تستنتج بدورها من قاعدة بيانات الجذور العربية التي بنيناها انطلاقاً من المعاجم العربية , قديمها وحديثها .

2- بناء محلل تركيبى يقوم أساساً على قاعدة بيانات الأشكال اللغوية الصحيحة اعتقاداً منا بأن الأشكال اللسانية المؤلفة من متواليات المفردات ( الجمل ) هي الوحدات الدالة في النص اللغوي . أما المفردات فلا تعتبر كذلك إن نظر إليها خارج سياقها التركيبى , وهذه المتواليات تنقسم إلى ثلاثة أنواع :

(1) جمل عادية : يتم فيها توزيع العناصر الاسمية وغيرها مع الفعل بشكل قابل للاستبدال , لكن دلالتها قابلة للحساب واستخلاص النتائج انطلاقاً من المعجم العادي .

(2) جمل مسكوكة : وهي تلك المتواليات اللغوية التي تتضمن مناطق معتمدة على شكل أجزاء ثابتة غير قابلة للاستبدال بعناصر أخرى , ولا للتحريك من مكان إلى مكان داخل الجملة , كما أن دلالتها لا تستخلص من معنى المفردات الواردة في المعجم العادي بل نحتاج معها إلى رصيد من التجارب المتكونة مع كل لغة .

(3) جمل الفعل العماد : وهي تلك التي تتضمن عنصراً تتوفر فيه سائر خصائص الفعل من الناحية المورفولوجية , دون أن يؤدي إدماجه في البنية إلى تغيير دلالة المتواليات التركيبية الأساسية ..

وفي جميع هذه الحالات يكون الفعل الرئيسي (= المبنى للمعوم في البيئة المركزية ) هو العنصر المتحكم في توزيع المواقع الوظيفية على العناصر المؤلفة معه : أسماء , حروف , ظروف , الخ , كما يتحكم

في عددها وطبيعتها الدلالية في التركيب . ومن هذا المنطلق جاز أن نعتبر الفعل جملة بسيطة لأنه حتماً يركب مع عناصر مسيطرة لطبيعته الصورية .

كل هذا يعني أننا في حاجة ليس فقط إلى المعاجم تظهر فيها الكلمات منظمة بشكل من الأشكال مع تحديد الدلالة المناسبة لكل واحدة منها , بل لعلنا في حاجة أكثر إلى معاجم للتركييب اللغوية , فالتركيب هو الذي يعطي للمفرد وظيفتها اللغوية المناسبة في سائر المستويات . لنأخذ هذين المثالين :

1) ضرب زيد علياً

2) ضرب زيد موعداً ل س2

حيث نلاحظ أن الفعل ضرب يأخذ هنا دلالتين : في (1) يعني حدثاً مادياً , وفي (2) يعني قولاً أو واقعاً , وما كان لنا أن نفرق بين هذين المعنيين بالاختصار على معاينة المفردتين خارج التركيب , وهذا هو حال اللغة , وأي معجم يخصص لوصف اللغة يجب أن يدخل في بنائه المستوى التركيبي , وإلا فلا فائدة منه .

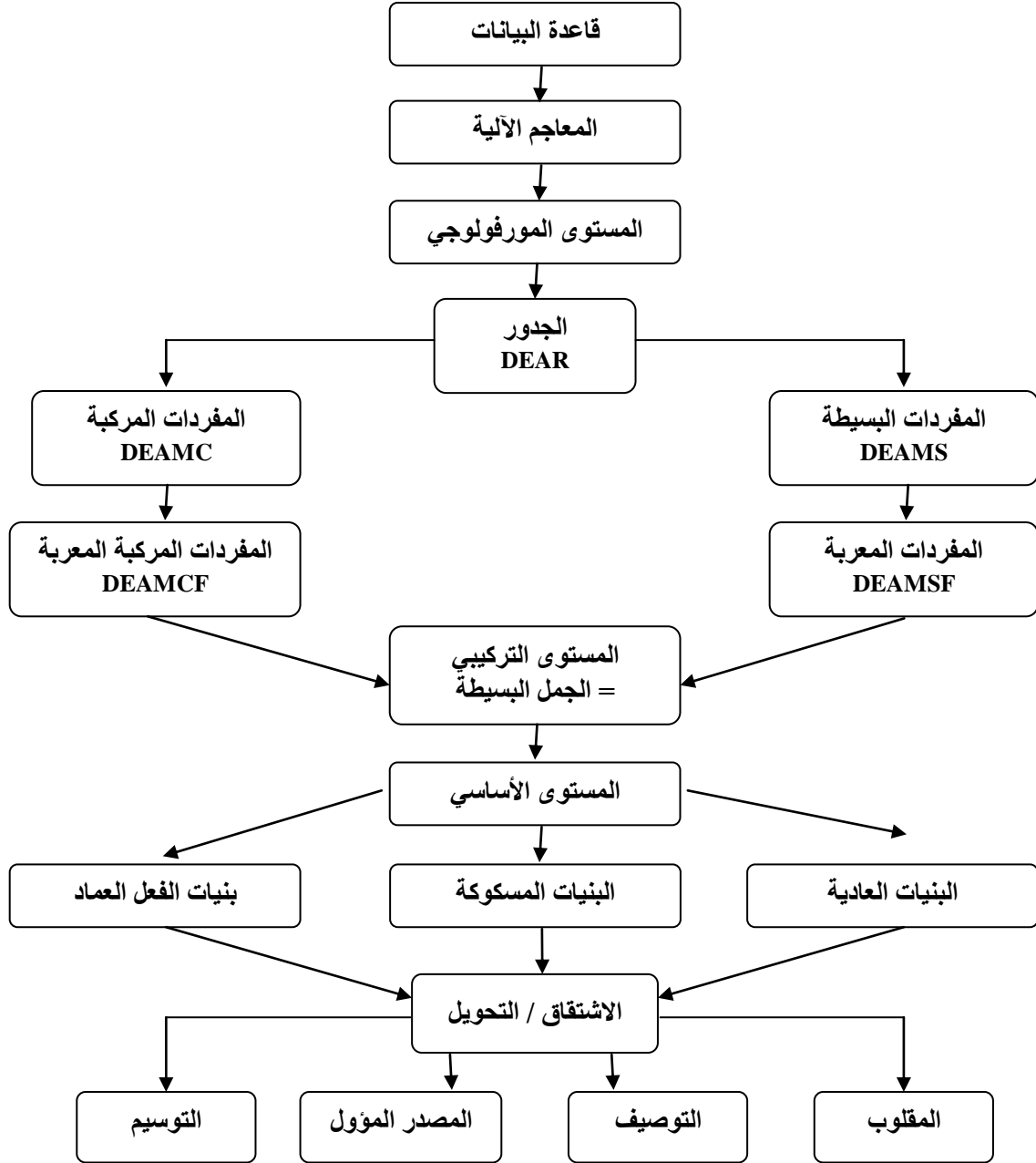
بيد أن هذا العمل يجب أن يسبق بوصف المفردات وضبط خصائصها بمعزل عن سياقها اللساني كمرحلة أولى , وقد قمنا بهذا العمل من الناحية المنهجية حتى نتمكن من التعرف أولاً على الخصائص المورفولوجية للمفردات مع ضبط سائر الإمكانيات الاشتقاقية التي يتوفر عليها نظام العربية . , وثانياً حتى نتمكن من تكوين رصيد من القواعد اللسانية التي تستغلها البرامج المعلوماتية أثناء عملية الممكنة وبناء معاجم آلية للغة العربية . والشكل التالي يلخص طبقات المعجم الآلي كما نتبناه :



## 2. طبقات المعجم الآلي

عندما فكرنا في بناء قاعدة البيانات التي سيبنى معجم آلي متكامل للغة العربية وضعنا الخطاطة التالية

:



يمثل هذا الشكل النتائج التي توصلت إليها الدراسات اللسانية في تعاملها مع أنظمة اللغات المختلفة .  
ونلاحظ على طول نزول هذا المخطط ثلاث محطات تتجمع فيها الفروع بشكل تراتبي :

- قاعدة البيانات
- المستوى المورفولوجي
- المستوى التركيبي
- المستوى الاشتقاقي التحويلي (لن نتعرض له في هذه المحاضرة نظراً لضيق المجال) .

يتعلق الأمر بمجموعة من المعاجم المتكاملة التي تؤلف في مجموعها المعجم الآلي للغة العربية ,  
وهي تتألف بشكل تدريجي من الجزء إلى الكل : من البسيط ( المفردات ) إلى المركب ( التراكيب اللغوية ) .

## 2 المستوى المورفولوجي

لقد بنينا برنامجاً متكاملًا لتوليد المفردات من الجذور العربية بجميع أنواعها : الثلاثي والرباعي ،  
الصحيح والمعتل ، الخ . في الأفعال والمشتقات والمصادر . وفيما يلي وصف مختصر لهذا البرنامج :

### 1.2 المستوى الصرفي

يهدف هذا المشروع إلى بناء قاعدة معارف لقواعد تصريف مختلف مكونات نظام العربية الفصحى ،  
وتهدف هذه القاعدة إلى :

(1) بناء قواعد بيانات المكونات التالية :

- الجذور
- الزوائد والسوابق واللواحق : في الأسماء والأفعال والصفات والمصادر
- الأدوات
- الأوزان والصيغ الصرفية والاسمية والفعالية والمشتقات

(2) تحديد خوارزميات الاشتقاق في مستويي التحليل والتوليد تقوم بتحديد مواقع الزوائد في الكلمة  
من جميع المستويات مع حصر شامل للاستثناءات

(3) الواجهة التي تسمح بإدخال وتحديث البيانات :

- من الجذر والوزن إلى الكلمة
- من الكلمة إلى الجذر والوزن

(4) برنامج يساعد في بناء مدقق إملائي لمفردات اللغة العربية

(5) برنامج يساعد في ضغط النصوص

## 1.0.2 المحاصيل الرئيسية

\* برنامج حاسوبي متكامل يحيط بجميع مفردات المعجم العربي :

- الأسماء البسيطة والمعربة : في الأفراد التثنية والجمع والتذكير والتأنيث
- تحديد أبنية المصادر وكيفية اشتقاقها ( أكثر من مائة بناء غير مطرد أو غير قياسي )
- الأفعال في مختلف الأزمنة
- المشتقات بجميع أنواعها وطرق اشتقاقها
- المصادر بجميع أنواعها
- الأدوات بمختلف أنواعها

\* قاعدة بيانات تشمل جميع القواعد الاشتقاقية انطلاقاً من الجذر بجميع أنواعه :

(1) في الأفعال :

- الثلاثي الصحيح والمعتل
- الرباعي ومزيده , صحيحه ومعتله
- الأجوف والمثال والناقص ( بالواو والياء )
- إسناد الفعل إلى مختلف الضمائر المتصلة

(2) في الأسماء :

- البسيط الذي لم يخضع بعد لأي عملية اشتقاقية
- المعرب الذي يخضع لمختلف عمليات الاشتقاق : التثنية والجمع والنسبة والتصغير والتذكير والتأنيث , الخ .

(3) في المشتقات : تم حصر قواعد اشتقاق :

- اسم الفاعل
- اسم المفعول
- الصفة المشبهة
- اسم التفضيل
- صيغة المبالغة
- اسم الزمان
- اسم المكان
- اسم الآلة

(4) في المصادر :

- حصر جميع صيغ المصدر غير المطردة ووضعها في أصناف ذات خصائص مشتركة , مع تحديد طريقة توليدها آلياً , وهي كما يلي :

- المصدر العادي وهو غير قياسي
- المصدر الميمي
- المصدر الصناعي
- مصدر المرة
- \* تحديد قواعد الإعلال والإبدال والقلب في الأسماء البسيطة والمعربة .
- \* برنامج التوليد الصرفي الآلي يستجيب لمختلف برامج التعرف على مفردات النص العربي غير المشكل .
- \* برنامج تطبيقي لاختبار صحة المفردة المولدة آلياً .
- \* برنامج تطبيقي لتصريف الأفعال آلياً يقدم الجداول التصريفية الفعلية في مختلف الأزمنة .
- \* معالجة قاعدة بيانات للأفعال البسيطة تتكون من 1300 مدخل معجمي فعلي .
- \* معالجة قاعدة بيانات للأفعال المعربة تتكون من 70.000 مدخل فعلي .
- \* معالجة قاعدة بيانات الأسماء تتكون من 23000 مدخل اسمي بسيط .
- \* معالجة قاعدة بيانات الأسماء المعربة تتكون من 600.000 مدخل اسمي .
- \* معالجة قاعدة بيانات جميع الأدوات في اللغة العربية الفصحى .
- \* معجم آلي لتوليد جميع الصيغ الصرفية في اللغة العربية : الأسماء والأفعال والأدوات والمشتقات والمصادر .
- \* معجم بالقواعد المولدة للمفردات الشاذة في اللغة العربية .

## 2.0.2 المقاييس الفنية :

- التعرف على الحرف الأصلي والحرف الزائد الذي يتم بموجبه توليد المفردة بنسبة 100 % .
- التفرقة بين جذر وساق الكلمة بنسبة 100 % .
- خوارزميات التعرف على موطن الحرف المولد للمفرد 100 % .
- تحديد الحركة المعجمية في بنية الكلمة المولدة صرفياً 100 % .
- قواعد النقل والقلب والإعلال والإبدال في المفردات المعتلة بإحدى الحروف : أ و ي 100 % .
- التعرف الآلي على نوع المفردة هل هي اسم أو فعل أو من صنف المشتقات أو المصادر ؟ 100 % .
- التعريف الآلي على المفردة هل هي معربة أو بسيطة , مشتقة أو غير مشتقة 100 % .
- تحديد مسار المفردات المشتقة , هل هي مشتقة من الجذر عبر مسار الفعل أم عبر مسار الاسم 100 % .
- تعيين المشتق الذي له جذر والمشتق الذي ليس له جذر معروف 100 % .
- إسناد جذور لبعض المفردات الحديثة التي دخلت إلى اللغة العربية حديثاً 100 % .

- تحديد القواعد الفونولوجية التي تتحكم في بناء الصيغ الصرفية 100 % .

ويتكون هذا المستوى من ثلاثة أنواع من المعاجم الآلية :

## 1.2 معجم الجذور ( DEAR )<sup>(3)</sup>

لقد تمثلت الخطوة الأولى من هذا العمل في بناء معجم للجذور العربية , وذلك بتصنيفها بالشكل الذي سيجعلها تستغل من قبل المعاجم اللاحقة في هرمية هذا البناء التصنيفي للمعاجم الآلية العربية . هذا على الرغم من الصعوبات التي تنجم عن بناء هذا المعجم نظراً لتعددية التوليد انطلاقاً من جذر واحد . ونظراً لمحدودية عدد الجذور في اللغة العربية ( 9464 جذراً ) , بما في ذلك الجذر الثلاثي والرباعي . كما أن جذراً واحداً يشتق منه الفعل والاسم والصفة والمصدر الخ . مما يعقد عملية بناء هذا المعجم وربما جعله صعب الاستعمال من قبل البرامج المعلوماتية اللاحقة .

ومع هذه الصعوبات لا نشك أننا سنتمكن من بناء هذا المعجم الذي يرفق كل جذر لغوي عربي بجميع الاحتمالات الاشتقاقية , مما سيجعل منه معجماً شاملاً لكل المداخل تمكن من استغلاله بنوع من السهولة التي تناسب حجمه وطريقة بنائه . إلا أنه يبقى مع ذلك معجماً صورياً يمثل رافداً معلوماتياً أكثر منه رافداً لغوياً . وهذه هي الطريقة التي اتبعناها في بنائه . مثلاً :

ض. ر.ب: < ف = 0+ > , < مص = 0+ > , < فا = 1+ > , < فو = 2+ > , < س = +0 > , الخ

ق. و.ل: < ف = 0+ > , < مص = 0+ > , < فا = 1+ > , < فو = 2+ > , < س = +0 > , الخ

حيث إن + = الحروف الزائدة عن الجذر

- = الحروف الناقصة عن الجذر

0 = ليست هناك زيادة عن حروف الجذر

1 = أو أي عدد آخر يعني عدد الحروف الزائدة عن الجذور لتوليد الصيغة المشار إليها

بالحرف قبل العلامة =

وهذه الطريقة تعتبر أنجح وسيلة لتعرف سائر المداخل المتولدة عن كل جذر من جذور اللغة العربية المحصورة العدد . وللإشارة فإن المقولات المعلمة في هذا المعجم كلها في المفرد أي بالنسبة للاسم والمصدر والصفات والظروف ، وفي الزمن الماضي للغائب المفرد بالنسبة للفعل . ولا نريد إيقال كاهل هذه المعاجم بالرموز المحيلة على العمليات التي سيتم التعرف عليها في المعاجم الأخرى اللاحقة : معجم المفردات البسيطة العادية والمعربة ، معجم المفردات المركبة العادية والمسكوكة .

(3) DEAR المعجم الآلي العربي للجذور .

ومن الصعوبات التي واجهتنا في هذا الجزء من المعجم إعطاء سنن للأدوات والحروف التي لا تعرف جذوراً مطردة وعادية كما هو الحال مع بقية أقسام الكلم الأخرى ، لذلك اضطررنا إلى معاملتها بشكلها الاستعمالي لا بصورتها الأصلية غير الموجودة أصلاً .

لقد اتبعنا هذا الأسلوب لاعتقادنا بأنه هو الأسلم ، فاللغة العربية ( خلافاً للغات الأوروبية ) لا تستجيب للنظام الخطي التسلسلي في الزيادة ، فهذه الأخيرة لا تعرف لها موضعاً محدداً في بنية الكلمة كما هو الشأن في اللغة الفرنسية مثلاً حيث تتمركز جميع العلامات التصريفية إما في نهاية الكلمة أو في بدايتها ، ولا يتموضع إلا القليل منها في وسط الكلمة ، فموضع حروف الزيادة في اللغة العربية يختلف في الكلمة حسب المقولة المولدة ، من الزائدة إلى الواسطة إلى اللاحقة . وهذا ما يدعو إلى طريقة تعامل خاصة تراعي فيها أولاً وقبل كل شيء خصوصيات هذه اللغة ، ولا نجد بديلاً عن طريقتنا هذه في ضبط الجذور العربية وربطها بالمستوى المورفولوجي الذي سيؤدي إلى المستوى التركيبي أخيراً .

وتجدر الإشارة إلى أننا نعتبر الحركات جزءاً من خوارزميات التوليد التي تطبق على الجذر ، شأنها في ذلك شأن بقية حروف الزيادة التي تنتظم في قواعد صورية محددة الموقع مع كل مقولة لغوية على حدة . فالتوليد يتم بموجب إقحام أحد المتغيرات أو أكثر في الجذر ، ولكل مقولة لغوية ( مفردة ) قانونها الخاص بها في التوليد انطلاقاً من الجذر . فالفعل له خوارزمياته الخاصة به والتي لا يمكن إسنادها للاسم ، والعكس صحيح . وكذلك الصفة وسائر المشتقات في النظام اللغوي العربي . فهذه الزيادات التي تتم وفق قواعد مضبوطة والتي تلحق الجذر هي التي نسميها خوارزميات التوليد . وهي التي يتعامل معها البرنامج المعلوماتي الطبيعي المخزن في كفاية المتكلمين لإنتاج بنيات لغوية طبيعية .

## 2.2 معجم المفردات البسيطة ( DEAMS )<sup>(4)</sup>

وفيه وصف صوري لكل مفردة من مفردات اللغة العربية موضوعة خارج سياقها التركيبي . والمفردة بالنسبة إلينا متوالية من الحروف محاطة بياض أو بعلامة أخرى متعارف عليها في استعمال اللغة العربية : كحروف الجر اللاصقة ، حروف العطف ، الخ . ولم نهتم إلا بالمفردات التي تستعمل في النصوص الحديثة ، أي تلك التي تذكرها المعاجم الحديثة وما زالت في الاستعمال العادي للغة العربية سواء بمعناها القديم أو بمعناها الحديث .

وقد نظمنا هذه المفردات بشكل يسمح بربطها آلياً بمكونات النصوص العربية عند تطبيق الإجراءات المتعلقة بالتعرف ( Reconnaissance ) أو تلك التي تتعلق بتوليد الجمل آلياً . والحقيقة أن شيئاً من هذا سبق للنحو القديم أن قام به ، حيث تم حصر أقسام الكلم في أوجهها المختلفة . لكننا لم نقف عند تلك التصنيفات الأولية والتي استفدنا منها فعلاً ، بل أضفنا إليها أقساماً أخرى كالصفة ، والمصدر ، ومختلف أنواع الظروف ، وقد قمنا بهذا تسهياً للعمل الذي ستقوم به البرامج المعلوماتية حيث سيسهل عليها ملاءمة هذه البيانات مع طريقة تكوين النصوص العربية . أما الطريقة التي اتبعناها في وصف هذه المدخل فتمثل في إسناد

(4) DEAMS المعجم الآي العربي للمفردات البسيطة .

رقم عددي لكل خصيصة مورفولوجية يعد مرجعاً نهائياً للصنف الذي ينتمي إليه كل مدخل معجمي في قاعدة البيانات .

أما تصنيفنا هذا فيختلف جذرياً عن جميع التصنيفات السابقة من حيث الهدف ومن حيث الطريقة . نريد إقامة تصنيف يقوم على الكلم وبراغي شيئين: الخصوصية النحوية للمدخل المعجمي ( المفردة ) , وربطها بالمستوى المورفولوجي المناسب لها ( الصيغة والاشتقاق ) . وسنسير على النمط التالي :

ف : الفعل في الزمن الماضي للمفرد الغائب المذكور

س : الاسم في المفرد للمذكر أو المؤنث

ص : الصفة وسائر المشتقات بجميع أنواعها

ظ : الظرف بجميع أنواعها

ح : الحروف والأدوات بجميع أنواعها

### 3.2 معجم المفردات البسيطة المعربة ( DEAMSF )<sup>(5)</sup>

ويتعلق الأمر هنا بمفردات اللغة العربية من خلال احتمالات ظهورها الممكنة في النصوص العربية ، حيث تأخذ الكلمة أشكالاً مورفولوجية مختلفة . فالاسم الجامد يعرف التغيرات الإعرابية التالية : مفرد ، مثنى ، جمع ( = : سالم ، تكسير ، قلة ، وتكسير كثرة ) ، مؤنث ، مذكر ، تصغير ، نسبة ، تعريف ، تنكير ، علم ، عام ، الخ . أما الفعل فهو أيضاً قد يظهر في نص واحد بأزمته الثلاثة المعروفة في اللغة العربية ، والاسم بجميع علامات الشخص والعدد والنوع ، والأمر نفسه بالنسبة للمشتقات والظروف والصفات . ومن المفروض في كل مفردة أن تأخذ الشكل المورفولوجي الذي يحدده السياق التركيبي داخل النص .

بهذه الطريقة تتميز مداخل المعجم الآلي ، حيث يتفرع كل مدخل من مداخل DEAMS إلى أكثر من مدخل معرب، وقد استطعنا لحد الآن حصر سائر الخصائص الإعرابية للمداخل ويمكن لأي حاسوب مزود بقاعدة البيانات هذه أن يكون بشكل آلي معاجم آلية صغرى للمفردات المعربة .

وهنا نشير إلى نقطتين :

1- إن DEAMSF يستند بشكل مباشر إلى قاعدة بيانات DEAMS لكنه يفرع كل مداخله إلى مداخل صغرى تراعي فيها الإمكانيات التأليفية لكل مدخل . وفي هذا الصدد فإننا اضطررنا للاحتفاظ ببعض العلامات التصنيفية من DEAMS رغم ما يطرحه هذا من حشو يثقل كاهل البرنامج المولد لمفردات النصوص انطلاقاً من DEAMSF .

2- إن DEAMSF يصلح كذلك كمدقق إملائي للغة العربية ، إذ إنه يتوفر على سائر المعلومات المتعلقة بكل مفردة يفترض أنها مركبة داخل النص العربي ، كما أن DEAMSF يعتبر مقدمة لبناء معجم تركيبية كما سنرى بعده .

(5) DEAMSF المعجم الآلي للمفردات البسيطة المعربة .

بقي أن نشير إلى أن عدد مداخل هذا المعجم الآلي DEAMSF يضاعف ست مرات عدد مداخل المعجم الآلي DEAMS الذي لا يتضمن إلا معلومات ذات طبيعة تصنيفية أولية.

أما طريقة العمل فقد تم تحديثها على الشكل التالي :

أ - الأسماء الجامدة (أسماء الذوات)

لقد تبين لنا من خلال تعاملنا الطويل مع مكونات النظام اللغوي العربي أن أقسام الكلم على نوعين من الناحية الحاسوبية: قسم يعرف نوعاً من الاطراد التصريفي يسهل التعامل معه حاسوبياً ، ويدخل في هذا القسم: الأفعال والمشتقات (=: الصفات) ، وقسم يرفض الاطراد التصريفي يصعب التعامل معه حاسوبياً ، وهو يضم الأسماء الجامدة والمصادر . وهذا ما يفسر تركيز أغلب الدراسات الصرفية والبرامج الحاسوبية على الأفعال والمشتقات نظراً ليسر التعامل معها حاسوبياً .

ونظراً لرغبتنا في بناء برنامج حاسوبي قوي بدأنا من الجانب الصعب، أي من الأسماء والمصادر، وفيما يلي وصف للبرنامج الذي وضعناه لمعالجة الأسماء الجامدة في اللغة العربية .

وفي البداية كان علينا أن نضع تصوراً تصريفياً حاسوبياً متكاملًا نقوم فيه بتشفير سائر المداخل الاسمية الجامدة ، يتجاوز ما قدمه الصرفيون العرب عن الأسماء ، ولذلك صممنا الجدول التصنيفي التالي:

تجب الإشارة هنا إلى ما يلي :

- تم تشريع في وضع البرنامج الحاسوبي إلا بعد أن اكتمل الوصف اللساني لجميع المداخل اللغوية . وبعد أن اكتمل بناء قواعد البيانات بشكل سليم .

- لأن البرامج المقصود هنا يلخص العمل اللساني ويترجم الخوارزميات اللغوية على خوارزميات حاسوبية .

- العمل اللساني هو الأصل في كل هذا ؟ .

العلامة	الرقم	أسماء	المج
ون/ ين صيغ صيغ	1 5 - 2 28 - 6	المذكر الذي يجمع جمع مذكر سالماً المذكر الذي يجمع جمع تكسير (قلة) المذكر الذي يجمع جمع تكسير (كثرة)	I
أت صيغ صيغ	29 33 - 3 32 - 34	المؤنث الذي يجمع جمع مؤنث سالماً المؤنث الذي يجمع جمع مؤنث تكسير (قلة) المؤنث الذي يجمع جمع مؤنث تكسير (كثرة)	II
X	85 - 56	أسماء للمذكر والمؤنث المذكر الذي لا مؤنث له من لفظه المؤنث الذي لا مذكر له من لفظه	III



X	82	أسماء مذكرة لا جمع لها من لفظها	IV
X	87	أسماء مؤنثة لا جمع لها من لفظها	
X	88	أسماء مذكرة في الجمع لا مفرد لها	V
X	89	أسماء مؤنثة في الجمع لا مفرد لها	
X	105 – 90	الضمائر	VI
X	106	أنواع أخرى	VII
	106	عدد المجموعات الاسمية	

يشمل هذا الجدول جميع المداخل الاسمية البسيطة ، العادية والمعربة ، وكل مدخل اسمي يجب أن يحمل رقماً هو عبارة عن سنن code يعتبر بمثابة المرجع له في المعجم . وسيتم استغلال هذا الجدول على الشكل التالي :

معلم : س56 ، رجل تربية وتلقين المعرفة للأطفال .

وردة : س85 ، نبات برائحة زكية ، وفيه أنواع ... إلخ .

حيث يحيل الرقم إلى النوع والعدد والصنف التصريفي بالإضافة إلى التعريف الدلالي ، أي تحديد المرجع العام والاستعمالي للمدخل الاسمي .

أما الأسماء المركبة فسيكون لها تنظيم آخر في المعجم ، ويمكن الرجوع إليه في أبحاثنا التي نشرناها في مجلة التواصل اللساني . غير أن الذي نشير إليه هنا باستعمال هو أن أغلب الأسماء المركبة تنتمي إلى صنف التعابير المسكوكة ، ويمكن الرجوع إليها في معجم هذه التعابير الذي بنيناه في إطار مشاريع مجلة التواصل اللساني ، وهو جاهز الآن للاستعمال ، وهو يضم حوالي 30000 تعبير مسكوك في اللغة العربية بما في ذلك التعابير الفعلية والإسمية وجميع الأسماء المركبة .

لقد بنينا برنامجاً للإدخال والتوليد بلغة البرمجة Visual basic وهو الآن في طور الإنجاز في إطار المشاريع العلمية التي تتبناها مجلة التواصل اللساني في تعاونها مع جهات علمية وأكاديمية متعددة .

يتأسس البرنامج على جميع الأوزان الاسمية العربية في سائر مستوياتها ، وهو مزود بخوارزميات للتوليد التي تؤدي مهمة التوليد ، حيث يقوم البرنامج بالتنسيق بين الجذر ومختلف المشتقات الاسمية الجامدة المتولدة عنه ، مع التعرف على قضايا الإغلال والإبدال والقلب الضرورية لكل مدخل والبرنامج مزود بقاعدة بيانات واحدة فقط هي معجم الجذور العربية ، وهو يتولى بنفسه وباعتماد الخوارزميات توليد جميع الاسما البسيطة والمعربة .

## ب - الأفعال

لن يكون من السهل ضبط الأفعال في أرقام محددة نظراً لتشعب علاقاتها بسائر أقسام الكلم وخاصة منها المصدر وسائر أنواع الصفة . وقد تمكنا مع ذلك من عزل الفعل عن سائر أقسام الكلم باعتباره وحدة مستقلة عن التركيب . هذا على الرغم من صعوبة الفصل بين الفعل وفاعله نظراً لغياب صيغة infinitive في اللغة العربية ، حيث يتهيأ للمصنف أن إعطاء رقم معين للفعل هو في نفس الوقت رقم للفاعل ، وهذا ما قد يؤخر إنجاز عملية التقييم قليلاً .

إن اعتبارنا للفعل كمدخل مستقل عن تركيبه بفرض اختيار الصيغة الأقرب إلى التمثيل الصوري ، وقد جاء اقتراحنا لصيغة الفعل الماضي كأساس لخلوها من الزوائد واللواحق التي تحمل علامات الزمن والشخص والعدد ، وبذلك تمكنا من اتباع الأسلوب التالي :

قسمنا الأفعال إلى مجموعة تصنيفية يتمثل فيها أقصى حد ممكن من الاطراد في الانتماء إلى الصنف المناسب . وهذه الأصناف تحدد من خلال الحركات التي يأخذها الفعل في المضارع ، حسب الصحة والإعلال ، وحسب الزيادة أو عدمها . وهذه كلها يتضمنها النموذج التصنيفي التالي :

### الثلاثي<sup>(6)</sup>

الثلاثي الصحيح	{	فعل (1) ← يفعل (= صنع)
		فعل (2) ← بفعل (= ضرب)
		فعل (3) ← بفعل (= نصر)
المضعف	{	فعل (4) ← يفعل (= سر)
		فعل (5) ← بفعل (= فر)
المعتل الفاء بالواد	{	فعل (6) ← بفعل (= وعد)
		فعل (7) ← يفعل (= وضع)
		فعل (8) ← يفعل (= وجل)

وقد سرنا على هذا النموذج الذي تصل عدد أصنافه إلى 55 صنفاً تصنيفياً توصف بموجبه جميع المدخل الفعلية في اللغة العربية. ونملك الآن قاعدة البيانات الفعلية الموصوفة بهذه الطريقة على (ACCESS) على الحاسوب من فصيلة IBM . وهي جاهزة للاستعمال من قبل برامج التوليد الآلي للغة العربية في مستويها الصرفي والتركيب .

(6) هذا مجرد نموذج من طريقة التصنيف المتبعة في بناء قاعدة بيانات الأفعال في اللغة العربية .

### 3 المعجم التركيبي الآلي (LEXIQUE-GRAMMAIRE)

يتألف المستوى التركيبي من عدة أنواع من المعاجم الصغرى يمثل كل واحد منها مستوى لغوياً متميزاً ، أولها مستوى الأسماء المركبة ، يليها مستوى الجمل العادية ثم يأتي بعدها مستوى العبارات المسكوكة فتركيب الفعل العماد التي تؤطرها ظاهرة التوسيم ، وهي ظاهرة عالية الإنتاجية التركيبية في اللغة العربية (الحناش 1988) .

#### 1.3 معجم المفردات المركبة (DEAMC) <sup>(7)</sup>

تعد الأسماء المركبة أول خطوة في مجال الوصف التركيبي للغة العربية ، فنحن لا نتعامل مع الاسم المركب إلا بعد أن يكون قد تجاوز مرحلة البناء الصرفي ، هذا على الرغم من الصعوبات العملية التي تواجهها في إدراج المستوى الصرفي في هذه المركبات التي جاء معظمها على شكل تعابير مسكوكة أو أجزاء من التعابير المسكوكة .

تتضمن اللغة العربية نوعين من المفردات : البسيطة وهي تلك التي تعرضنا لها في (2.2) ومفردات مركبة وهي على نوعين كذلك: عادية ومسكوكة. الأولى هي تلك التي تقوم بوظائف مختلفة داخل الجمل العربية مثل:

(1) [منارة المسجد] عالية.

(2) شاهد س [منارة المسجد] .

حيث يقوم المركب الاسمي منارة المسجد بوظائف مختلفة يفرضها السياق التركيبي. والثانية هي تلك التي نطلق عليها الأسماء المركبة. وهي تتخذ الأشكال المختلفة التالية:

س ح س : بناء على هذا

س س : عجلة السيارة

ح س ح س : بمنأى عن هذا

وهي مركبات تتلازم فيها العناصر البسيطة لتكون بذلك مركباً اسماً .

وقد اتبعنا في إحصائها ووصفها نفس الأسلوب المتبع في DEAMS و DEAMSF . فالاسم المركب كذلك قد يكون مؤنثاً أو مذكراً ، مثنى أو جمعاً ، مكبراً أو مصغراً ، الخ .. باختصار إن الاسم المركب لا يختلف عن الاسم البسيط في الخصائص ، لكنه يختلف عنه في صورته بما يؤدي إلى إضافة رموز أخرى إضافية في وصف هذا النوع من المفردات العربية ، وقد وصلنا لحد الساعة إلى حصر أكبر عدد من هذه المفردات يتلازم فيها عنصران فأكثر في الاستعمال العادي للغة العربية .

#### 2.3 معجم المفردات العادية

(7) DEAMC المعجم الآلي العربي للمفردات المركبة .

انطلاقاً من المبدأ العام الذي يقوم عليه منهجنا اللساني ، والذي مفاده أن المستويات اللسانية لا تتمتع إلا باستقلال نسبي ، (المورفولوجيا مثلاً في خدمة المستوى التركيبي) ، انطلاقاً من هذا المبدأ فإننا نعتبر أن النوع الأول من المعاجم الآلية الذي تعرفنا عليه سابقاً هو مجرد مستوى من مستويات المعجم التركيبي للغة العربية . فالمعلومات التي تتضمنها المعاجم السابقة يجب حتماً أن توظف في هذا المعجم بشكل مباشر ، ذلك أن توزيع العناصر داخل البنية يكون دائماً بخصائصها المورفولوجية التي حددناها في المعاجم السابقة ، وتشكل هذه الخصائص المستوى التوزيعي الذي يسمح أولاً بظهور العنصر في البنية . مثلاً :

كتب محمد المدرس

لكن \* كتب هند المدرس

وقد نتجت عدم المقبولية عن غياب المستوى المورفولوجي المتمثل في المطابقة بين الفعل وفاعله . ومن شأن المعجم التركيبي أن يهتم بهذا التلاؤم ليستقيم بذلك التركيب محدداً الخطأ ونوعه ومقترحاً حلولاً مناسبة لكل بنية ساقطة تركيبياً . فقد أدخلنا هند إلى معجم المفردات البسيطة المعربة على أنه اسم علم مؤنث . ومن المفروض في كل اسم من هذا النوع إذا وقع فاعلاً أن يسبق بعلامة مناسبة . وإذا كان معجم المفردات البسيطة يتضمن قواعد مورفو - دلالية ثابتة فإن المعجم التركيبي من شأنه أن ينسق بين الوحدات داخل البنية ، فهو مكون أساسي من قواعد سياقية تهتم بمحيط كل عنصر داخل الجملة .

لقد كانت الوحدة الصغرى في النوع الأول من المعاجم الآلية هي المفردة ، أما الوحدة الصغرى في هذا النوع من المعاجم فهي الجملة . ولكي يستجيب هذا المعجم للغاية المرجوة منه فقد قسمنا الجمل إلى نوعين: جمل أساسية وجمل محولة. وضعنا كمنطلق للجملة الأساس القاعدة الرتبوية التالية: ف س ك ، وهي بنية صورية قادرة على احتواء جميع أنواع الجمل البسيطة في اللغة العربية. وقد ألحقتنا بهذه البنية الأصلية رموزاً صورية ذات طبيعة سياقية مثل : س = + - حي ، + - مقيد. وإذا كانت الثنائية الأولى لا تطرح أي إشكال بالنسبة لا غلب الأسماء في اللغة العربية حيث يسهل تعرف نوع الاسم هل هو + حي أو - حي ، فإن الأمر مختلف تماماً مع سمة: + - مقيد ، ونعني بهذه السمة العلاقة غير الإرادية بين الفعل وفاعله كما نعني المكانة التوزيعية المفتوحة ، فالاسم غير المقيد غالباً ما يكون قابلاً للاستبدال بأي عنصر اسمي بدون قيد توزيعي ، مثلاً :

أقلق { هذا  
زيد  
سقوط المطر  
أن يسافر محمد } علياً

حيث نلاحظ انعدام وجود قيود توزيعية من أية طبيعة كانت على نوع الفاعل ، ومن الأكيد أن ذلك يخضع لنوعية الفعل ، لذلك قمنا بحصر الأصناف الفعلية التي تقبل هذا النوع من الأسماء ، أو بالأحرى نتحكم في هذا النوع من المواقع التركيبية . وبهذه الطريقة تمكنا من حل إشكال ظل يتردد بين البلاغة وأصول الفقه والدراسات النحوية ، ونعني به مشكل التعابير المجازية في اللغة العربية ، إذ تبين لنا أثناء بنائنا للمعجم

التركيبية الآلي ، أن هذا النوع من التعابير لا يعدو كونه عملية توزيعية قابلة للدراسة والحصر ، بل وإعادة الإنتاج ، أي إحداث استعارات جديدة بتطبيق قوانين هذا المعجم .

وقد جاء معجمنا هذا شاملاً لوصف التعابير الحقيقية والمجازية ، فهو يشمل قوانين إجرائية قادرة على توليد ووصف الاستعارات العربية . وقد تأتي لنا ذلك باعتماد مختلف مستويات المعجم الآلي للغة العربية : DEAMS DEAMSF و DEAMC فجاءت مداخلة متجاوزة 24000 بنية لغوية أساسية ، وهي مقسمة إلى الأصناف التركيبية التالية :

(1) ف س 0

(2) ف س 0 س 1

(3) ف س 0 ح س 1

(4) ف س 0 س 1 ح س 2

(5) ف س 0 ح س 1 س 2

وهي تمثل البنيات التركيبية الأساسية في اللغة العربية ( انظر الحناش 1988 ) .

وقد تم - في حصر هذه البنيات - اعتماد الفعل كمكون أساسي أي جملة بسيطة ، أما بقية العناصر فامتدادات له تستمد قانونها التوزيعي ووظائفها النحوية من ذاكرة الفعل . فالفعل في تصورنا دالة وبقية العناصر هي متغيراتها يُوَظَرها الخوارزم Log N2 الذي سرنا على هديه في جميع عمليات الحصر والتصنيف .

بالإضافة إلى هذه الخطوة قمنا بتصنيف الجمل البسيطة العادية وحصر خصائصها التحويلية . فبعد أن فرغنا من حصر الجمل البسيطة حددنا التحويلات التي يقلبها كل صنف والتي تبين لنا بوضوح تام أنها لا تطبق بشكل اعتباطي على سائر جمل اللغة ، بل إنها تطبق وفق القوانين التوزيعية المحددة لكل صنف ، مثلاً : البناء المقلوب (Passif) لا يمكن استخلاصه إلا من بعض البنيات الأصلية ، وهي ذات الشكل : ف س 0 س 1 ويشترط أن يكون مفعولها اسماً بسيطاً ، مثلاً :

سر هذا الكلام عليا

سر علي بهذا الكلام

فإذا كان هذا المفعول اسماً مركباً أو مصدرًا مؤولاً فإن هذا يؤدي إلى تعثر البناء المقلوب :

يكره زيد ( أن يفعل هذا )

\* يكره ( أن يفعل علي هذا )

وللإشارة فإن ظاهرة المقبولية عندنا ترادف تجربة الفرد مع لغته ، وليست هي ما ورد في المتون العربية . فقد تكون هذه الجملة مقبولة في متن عربي قديم ، لكننا لا نعيد إنتاجها في عربية العصر الحاضر ، ولذلك فإننا لا نقبلها . يؤكد هذا ما نحن بصدد التنظير له أي بناء معاجم آلية للغة العربية تنظر إلى اللغة خارج المقام التواصلية . إننا لا نعبأ في هذا الصدد كذلك بالقواعد القياسية التي طبقها النحاة ، فقد قالوا بأن البناء المجهول ينتج عن بنية متعددة مباشرة ، وهذا قياسي ، لكننا لن نتمكن من توليد هذه الظاهرة من جميع الجمل العربية التي يتوفر فيها شرط التعدية المباشرة ، مثلاً :

يهم هذا الأمر عليا

لكن : \* يهم علي

وقد دفعنا هذا إلى اقتراح مجموعة من الحلول اللسانية، ساعدنا على ذلك المعجم الآلي للمفردات البسيطة المعربة . ولما كانت هذه الظاهرة لا تطبق بشكل قياسي حتى داخل الصنف الواحد من التراكيب اللغوية العربية فإن حصرها يدعو إلى مكننة جمل اللغة ، فهذه الطريقة وحدها يمكنها خدمة الأهداف المتوخاة من بناء معاجم آلية شاملة للغة العربية .

وقد اتبعنا نفس الطريقة في معالجة ظواهر تركيبية كثيرة توطر الجمل الأساس البسيطة في اللغة العربية . كظاهرة التوصيف ، والبناء الموسم ، والتهيكال .

### 3.3 التعبيرات المسكوكة

ونعني بها تلك المتواليات التي تتضمن عنصراً مسكوكاً مع الفعل ، وفي الغالب تقوم المسكوكية غالباً بين الفعل وأحد المفاعيل في الجملة . وهي تعني ثبوت أو جمود أحد العناصر في مكانه حيث يستحيل استبداله بعنصر آخر مع الاحتفاظ بنفس دلالة الفعل . ومن خصائص هذا النوع من الجمل أنه لا يمكن استخلاص دلالاته من العناصر المؤلفة لها ، بل يعول في فهمها على تجربة الفرد مع لغته وطريقة توظيف المجتمع لهذه التعابير في مقاماتها المناسبة . ذلك لأنها تعابير تخزن في الدماغ على شكل رسوم مثبتة في كفاية الأفراد لا تولد بالقوانين العادية الموجودة في كفاية الأفراد ، وكل فرد يملك خزاناً خاصاً به من هذه التعابير ، ومن الصعب العثور على نفس العدد من التعابير المسكوكة عند نفس الأفراد في مجموعة لسانية واحدة . كل هذا لأنها تعابير غير تأليفية ، أي لا تستمد قوانينها التركيبية من DEAMS بشكل مباشر كما هو الشأن بالنسبة للتعابير العادية ، وكل محاولة لربطها بهذا المستوى ستتطلب منا تفكيك مكوناتها إلى وحدات معجمية . وهذا يؤدي إلى ضياع المعنى المسكوك لصالح معنى جديد عاد ، مثلاً :

(1) قضى زيد نجه = : مات

(2) يرقم زيد على الماء = : يضيع وقته

للحفاظ على المعنى المعبر عنه في هاتين البنيتين معاً يجب الإبقاء على كل عنصر في مكانه داخل البنية بما في ذلك الضمير (1) وكذا حرف الجر (2) ، فإذا حدث أي تغيير على هذه الشبكة ضاع المعنى المعبر عنه وانتقل إلى شيء آخر :

(3) قضى زيد يوماً كاملاً في العمل

(4) يرقم زيد على الورق

هذا ما دفعنا إلى اعتبار هذه التعابير مداخل معجمية مستقلة ، أي أنها تقوم مقام المفردات العادية في المعجم العادي ، لكنها مفردات من نوع خاص تحتاج منا إلى معالجة خاصة مختلفة عن تلك التي وضعنا بها المعاجم الآلية السابقة .

ونظراً لأن هذه التعابير تعد رسوماً بيانية محدودة العدد مثبتة في كفاية الأفراد ، فإن هذا يضع موضع شك ما ذهب إليه بعض النظريات اللسانية ، والقائلة بمبدأ الكليات اللغوية ، ولا نهائية اللغة ومبدأ التوليد الطبيعي لنظام اللغة في قدرة الأفراد ، فهذه البنيات لا تركيب وفق القوانين التوزيعية العادية كما أنها لا تعرف ذلك الاطراد التحويلي للتركيب العادية . ولا تخلو لغة من هذا النوع من التعابير ، إنها تكثر أو تقل في لغة من اللغات بدرجة شيوعها ومستوى استعمالها ، ولا نعرف لحد الآن قانوناً لسانياً يولدها بشكل مطرد في لغة من اللغات ، كل ما في الأمر أننا نكتفي بوصفها وضبط خصائصها التركيبية الصورية التي تبين لنا أنها لا تخرج عن قوانين البنيات العادية أي أن الفاعل مرفوع فيهما معاً ، والمفعول منصوب أو مجرور بحرف جر ، الخ ...

وقد كونا بهذا معجماً تركيبياً بهذه التعابير مصنفة وفق الأشكال الصورية التالية :

ف س مس1

ف س حرف مس1

ف س س حرف مس1

ف س مس1 حرف س

مس =: العنصر الثابت في البنية المسكوكة وقد يكون اسماً بسيطاً أو مركباً أو صفة أو ظرفاً ، الخ . لكن وجوده في البنية تحت أي شكل كان يؤدي إلى سك دلالة الفعل معه ومن ثم إعطاء دلالة خاصة للجملة وجعلها رسماً غير قابل للتجزئ .

ونحن الآن بصدد بناء قاعدة بيانات خاصة بهذه التعابير ، ونتوقع أن يتجاوز عدد مداخلها المعجمية 30000 تعبير مسكوك ، وهي تمثل نسبة 40% من الرصيد اللغوي المخزن في كفاية المتكلم العربي ذي التكوين الثقافي المتقدم . وقد قمنا بجمعه من مختلف المعاجم العربية قديمها وحديثها . وستضم هذه القاعدة سائر الخصائص التوزيعية المحددة لكل صنف من هذه التعابير ( انظر البحث الخاص في الحناش 1996 ) .

#### 4.3 معجم تراكيب الفعل العماد

ما عرضناه سابقاً كان يتعلق بالبنيات اللغوية العادية والمسكوسة منظوراً إليها من خلال ذاكرة الأفعال العادية في اللغة العربية ، ونعني بها تلك المداخل التي تستقل مع عناصرها بنيته ودلالاتها . لكن مسيرتنا البحثية اضطررتنا للتوقف عند صنف خاص من البنيات اللغوية وهي التي تتضمن الفعل العماد والتي تُوَظَر ظاهرة تركيبية أطلقنا عليها ظاهرة التوسيم ( انظر الحناش 1996 ) ، ونمثل لذلك بما يلي :

غم هذا الأمر علياً

=: (1) أدخل هذا الأمر الغم على علي

أقلق هذا الأمر علياً

=: (2) يشعر علي بقلق ( من + إزاء ) هذا الأمر

حيث إن الفعلين أدخل و يشعر لا يحدثان أي تغيير دلالي بالنسبة لما هو معبر عنه في الجمل الأساس . وهذا يبين بوضوح أن هذا النوع من الأفعال تحضر في البنيات اللغوية التي يتحول فيها الفعل

الأساسي إلى مصدر على شكل مورفيمات تعبر عن الزمن والشخص والعدد ، إنها لا تتحكم في توزيع العناصر الاسمية في البنية . ومن ثم جاز أن نعتبرها مورفيمات غير دالة / فارغة . وذلك خلافاً للأفعال العادية التي قلنا عنها بأنها تملك ذاكرة سجلت فيها سائر البيانات التوزيعية والوظيفية للعناصر المتوزعة معها . هذا على الرغم من توفرها على سائر العناصر الشكلية للأفعال العادية: الجذر، الصيغة، الصرف، الخ ...

ومن الأكد أن هذه الأفعال العمادية لا تظهر بشكل اعتباطي مع أي فعل كان، فقد وضعنا اليد على مجموعة من الضوابط الصورية التي يتم بموجها إظهار هذا الفعل العمادي في هذه البنية المحولة دون أن يحدث خلل في التركيب . وقد تبين لنا - بفضل بناء المعجم التركيبي - أن هذه الأفعال مختصة فعلاً ، أي أن كل مجموعة منها لا تظهر إلا مع التركيب الفعلي المناسب لها ، ووفق الظاهرة التركيبية المناسبة ، وقد أحصينا أكثر من 50 فعلاً عمادياً يختارها صنف الأفعال الدالة على إحساس أو شعور في اللغة العربية . وأكثر من 30 فعلاً عمادياً بالنسبة لصنف أفعال الحركة ، الخ ... ولا يمكن إطلاقاً استعمال الفعل العمادي الخاص بالنصف الأول في توسيم أفعال الصنف الآخر . أكثر من هذا فإننا قسمنا ظاهرة التوسيم نفسها إلى ثلاثة مستويات اشتقاقية : التوسيم المعلوم ، التوسيم المقلوب ، والتوسيم الموصف ، ولا يمكن استعمال الأفعال العمادية الخاصة بالصنف الواحد إلا في نوع واحد من التوسيم . مثلاً :

أقلق هذا الأمر علياً

[ توسيم المعلوم ] : أثار هذا الأمر القلق في علي

قلق علي من هذا الأمر

[ توسيم مقلوب ] : حصل لعلي قلق من هذا الأمر

ولا يمكن استعمال الفعل العماد أثار الخاص بالتوسيم المعلوم في التوسيم المقلوب بالنسبة للفعل

نفسه :

\* أثار لعلي قلق من هذا الأمر

\* حصل هذا الأمر القلق في علي

ومن خصائص هذه الأفعال أن تتركب مع مصدر ، ومع ذلك فقد توصلنا إلى وجود بعض أسماء

الأفعال العمادية التي تتركب مع أسماء غير مشتقة ، مثلاً :

أثار هذا الأمر هذا الأمر حفيظة علي

هذا الاطراد في فصل الأصناف اللغوية في سائر مستوياتها يعتبر الوسيلة الإجرائية الأسلم في سبيل

بناء معاجم آلية للغة العربية ، كما أن هذا يبين بوضوح تام حاجتنا إلى اكتمال الوصف اللساني لمكونات النظام

اللغوي قبل الشروع في الخطوة الموالية ، أي بناء معجم آلي للغة العربية .



#### 4 خلاصة

من خلال ما سبق يتضح تصورنا المنهجي والنظري الذي نلح عليه في بناء معجم آلي للغة العربية ، وعلى الرغم من أن ما قدمناه يتميز بنوع من الاطراد والتكامل بين سائر مستوياته : معجم آلي لخوارزميات للمفردات ، معجم آلي لخوارزميات للتراكيب ، على الرغم من هذا فإننا ما زلنا نعاني من بعض الصعوبات في وصف النظام اللغوي الذي يفاجئنا كل مرة بقضايا جديدة . وإذا كنا قد تمكنا من تجاوز - ولو جزئياً - مشاكل التعابير المسكوكة ، ومشاكل الفعل العماد ، فإننا نتوقع صعوبات أخرى سيكشف عنها الوصف اللساني لسائر مكونات النظام اللغوي .

إننا نؤكد مرة أخرى على أن أي محاولة لبناء معجم آلي للغة العربية يجب أن تبدأ من الوصف اللساني لنظام هذه اللغة ، ولا وصف إذا لم يتمكن من تمثيل إطار نظري ومنهجي لساني واضح ، وقد طبقنا هنا نظرية المعجم التركيبي التي لا نرى بديلاً لها في الوقت الراهن من البحث العلمي اللساني . وبفضل تقنيات هذا المنهج استطعنا إدخال التعابير المجازية في المعجم التركيبي للغة العربية على أنها تعابير عادية ، كما تمكنا من تجاوز صعوبات أخرى كبيرة .