دار المنظومة
DAR ALMANDUMAH
الرواد في قواعد المعلومات العربية

العنوان: قواعد البيانات العربية : معجم التعابير المسكوكة
المصدر: حوليات كلية اللغة العربية بمراكش - المغرب
المؤلف الرئيسي: الحناش، محمد
المجلد/العدد: ع 5
محكمة: نعم
التاريخ الميلادي: 1995
الصفحات: 21 - 41
رقم MD: 221184
نوع المحتوى: بحوث ومقالات
قواعد المعلومات: AraBase
مواضيع: المعاجم اللغوية، اللغة العربية، اللسانيات ، الألفاظ العربية، قواعد البيانات، الأجروميات، الحاسبات الإلكترونية
رابط: http://search.mandumah.com/Record/221184

© 2016 دار المنظومة. جميع الحقوق محفوظة.
هذه المادة متاحة بناء على الإتفاق الموقع مع أصحاب حقوق النشر، علما أن جميع حقوق النشر محفوظة. يمكنك تحميل أو طباعة هذه المادة للاستخدام الشخصي فقط، ويمنع النسخ أو التحويل أو النشر عبر أي وسيلة (مثل مواقع الانترنت أو البريد الالكتروني) دون تصريح خطي من أصحاب حقوق النشر أو دار المنظومة.

قواعد البيانات العربية :
معجم التعابير المسكوكة

د. محمد الحناش
كلية الآداب - فاس

ملخص البحث :

تعرف الدراسات اللسانية الحديثة توجهين علميين في التعامل مع اللغات الطبيعية :

الأول : الاتجاه الاستدلالي الذي ينظر الى الظاهرة اللغوية من خلال الأمثلة النموذجية، وهو اتجاه فرضي يضع المعطيات اللسانية في المقام الثاني أي بعد الافتراض. ومن نتائج هذا الأسلوب الزعم بمطلقية القاعدة اللسانية ومن ثم إغفال الشاذ في النظام اللغوي.

الثاني : وهو الاتجاه التراكمي الذي ينظر الى الظاهرة اللغوية من خلال المعطيات اللغوية التي تتأسس عليها المعالجة اللسانية، وهذا يفضي الى وضع قواعد نسبية في المرحلة الأولى، والى نسقية قواعدية مطلقة في المرحلة الثانية. ولتحقيق هذه الغاية العلمية يتبنى هذا الاتجاه أسلوب بناء قواعد البيانات Bases de données ، وذلك بهدف بناء معاجم آلية للغة العربية تمكن من استغلالها في برامج معلومياتة حديثة تهتم بالعلاج الآلي للغات الطبيعية.

سنتناول في هذا البحث التقنيات الموظفة في بناء قواعد البيانات العربية بهدف حصر هذا الصنف من مداخل اللغة العربية وفق الاتجاه اللساني الثاني المشار إليه أعلاه. وسنركز بالدرجة الأولى على تقنيات بناء قاعدة بيانات التعابير المسكوكة التي تضم أكثر من ثلاثين ألف تعبير مسكوك، جمعت من المعاجم العربية قديمها وحديثها، وقد اعتمدنا في حصرها بعض الأبحاث المنشورة في مجلة حوليات الجامعة التونسية. وهذه التعابير تمثل نسبة 40٪ من الرصيد اللغوي المخزن في الكفاية اللسانية للمتكلمين باللغة العربية. سنرفق البحث بنماذج ورقية ومعلوماتية من قاعدة البيانات الخاصة بهذه التعابير.

كما يهدف البحث الى تقديم أداة إجرائية لسانية حاسوبية للتعرف الآلي على التعبير

المسكوك في اللغة العربية وتمييزه عن التعبير العادي، وذلك من خلال قاعدة بيانات ضخمة تشكل أساس معجم آلي وضع لهذه التعابير العربية. ويتمثل عملنا في ضبط خصائص هذا النوع من التعابير حتى يتمكن البرنامج من التعرف عليها أثناء المعالجة الآلية للنصوص العربية غير المشكولة.

تقديم

سنعالج في هذا البحث طريقة التعرف الآلي على التعابير المسكوكة في قاعدة بيانات التعابير المسكوكة في اللغة العربية، والتي تضم حوالي 30000 تعبير مسكوك. وذلك بهدف بناء معجم آلي لهذه اللغة يتضمن سائر مداخلها مؤطرة في بنياتها التركيبية، والغاية من هذا العمل تكمن أساسا في صناعة برنامج التعرف الآلي على جميع مفردات النصوص العربية (مشكولة أو غير مشكولة)، والى وضع برنامج الترجمة الآلية، والتدقيق الاملائي، وأخيرا امكانية تعليم اللغة العربية بالحاسوب سواء للناطقين بها أو لغيرهم.

وبعد شروعنا في بناء معجم آلي لمفردات اللغة العربية تبين لنا أن الاعتماد على المفردة وحدها غير ممكن الا بصفة نسبية، وذلك نابع من طبيعة اللغة العربية نفسها، فنظامها كلغة طبيعية يقدم لنا معطيات غير متجانسة إطلاقا، فمن وجهة نظر لسانية صرفة لا يمكن اعتبار المفردة - معزولة عن السياق الهيكلي - وحدة دالة، هذا على الرغم من اعتباطية هذا المفهوم من وجهة النظر الحاسوبية التي تعرف المفردة كمتوالية كرافية محاطة ببياض على المستوى الخطي. لكن هذا لا يكفي من جهتين : لسانيا تعتبر المفردة جزءا من وحدة الدلالة الصغرى أي الجملة، هذا إذا كنا نرغب في التعامل مع العربية كلغة استعمالية وفق برنامج مخزن في قدرة الأفراد يمكنهم من التواصل والاحالة بهذه اللغة. وحاسوبيا تنقسم المفردات العربية الى نوعين : بسيطة ومركبة، أما البسيطة فقد كانت وما تزال الميدان المفضل الذي برعت فيه جل البرامج الموضوعة للمعالجة الآلية للغة العربية، والجميع يعرف المجهود الكبير والناجح الى حد كبير الذي قام به باحثونا العرب في هذا الميدان. لكن الذي لم يهتم به أحد لحد الآن هو جانب الأسماء المركبة في اللغات الطبيعية، ونقصد بها تلك المتواليات الاسمية التي تتضمنها التعابير المسكوكة في النظام اللغوي العربي والتي تمثل نسبة 40٪ من مجموع الرصيد اللغوي

الذي تضمه قاعدة بيانات هذه اللغة. وقد تبين لنا أن الطريقة الوحيدة التي يمكن بها التعرف على هذه المفردات هي حصرها في قاعدة بيانات شاملة، وقد لجأنا الى هذا الحل بعد أن تعثرت المحاولات التنظيرية التي قمنا بها في معالجتها ببرامج حاسوبية تعتمد الدخل input والمتمثل في التحليل، والخرج output والمتمثل في التوليد، الأمر الذي أدى الى فشل مطلق في وضع أي خوارزم algorithme لتوليد هذه التعابير في اللغة العربية. وهذا الأمر جعلنا فعلا نشكك في قدرة البرامج المعدة لحد الآن للعلاج الآلي للغة العربية، حيث ثبت لدينا أن هذه البرامج لا يمكن أن تعالج الا نسبة 60٪ من قاعدة بيانات اللغة العربية، بينما تبقى النسبة الخاصة بالتعابير المسكوكة غائبة عن هذه البرامج. ومن هنا نقول بأن الحاجة ملحة اليوم الى وضع برامج إضافية خاصة بهذا النوع من التعابير اللغوية.

فما هو التعبير المسكوك، وما هي خصائص الاسم المركب ؟ كيف نتصور بناء قاعدة بيانات التعابير المسكوكة للغة العربية ؟

1. المعالجة اللسانية :

1- 1 التعبير المسكوك :

يكون التعبير اللغوي مسكوكا إذا تميز بما يلي :

1) ضعف أو استحالة إمكانية استبدال أحد عناصره (المفعول غالبا) بعنصر آخر ولو كان من نفس الفصيلة المعجمية. مثلا (1) :

(1) علت [البسمة]. (وجه س)١ [2]

(2) قضى س. [نحبه] ١

(3) يرقم س. [على الماء] ١

---

1) في إطار الوصف اللساني الصوري الذي ننهجه في بناء قاعدة بيانات التعابير المسكوكة سنوظف الرموز التالية : س. = الفاعل، س ١ = : المفعول، مس = الاسم المسكوك، مظ = : الظرف المسكوك، مص = الصفة المسكوكة، الخ.

2) تمثل العناصر الموضوعة بين معقوفين [...] المنطقة المعتمة في التعبير المسكوك.

حيث لا يمكن استبدال العنصر المشار إليه (بين معقوفين) في هذه التعابير بعناصر أخرى من نفس الصنف دون أن يحدث ذلك خللا في المعنى الذي تعبر عنه بصورتها تلك، لهذا فإننا نلاحظ تغيرات جديدة ستطرأ على التعابير السابقة :

(1*) * علت الضحكة وجه علي

(2*) ٨ قضى زيد حاجته (3)

(3*) ٨ يرقم زيد على الورق

إن أي تغيير يمس العناصر المسكوكة يؤدي إلى فقدان المقبولية نهائيا وهو ما يعني طرد التعبير بأكمله من الكفاية اللغوية للأفراد (1)، أو إلى إضعاف تلك المقبولية أو على الأقل تغيير دلالة التعبير المعبر عنها بالصورة المسكوكة (2)، (3).

2) عدم ارتباط هذه التعابير بالمعجم في حصر دلالتها المعبر عنها بصورتها المسكوكة. مثلا إننا لن نستفيد من المعجم في فهم وتحديد دلالة البنية المسكوكة التالية :

(4) ضرب س.[في الأرض] ١

إذ أن المعروف عن المدخل المعجمي ضرب في لسان العرب أنه فعل علاجي يؤدي القيام به الى إلحاق ضرر مادي بالآخرين. لكنه يفقد تلك الدلالة ليتحول الى معنى السفر برا. وبذلك تنقطع الصلة بينة وبين المعنى المادي لنفس المادة. وعلى الرغم من وجود هذا التعبير كذلك في لسان العرب وفي نفس المادة فإنهما يختلفان في انتماء كل واحد منهما الى تعبير مغاير. وقد كان من المفروض أن يوضع كل مدخل منهما في باب خاص، لكن هذا يحتاج إلى إعادة النظر في بنية وهيكل المعجم اللغوي العربي بشكل جذري، وهو ما تؤسس له هذه الدراسة.

نستخلص من هذا أن هذه التعابير تتميز بالتعتيم الدلالي (Opacité Sémantique) على عكس التعابير العادية التي تتميز بالشفافية transparence ، وهذا ما يجعلنا نراجع ثنائية الدال والمدلول التي صاغتها النظرية البنيوية. وهذا التعتيم هو ما تعبر عنه الدراسة المنجزة حول هذه التعابير بالمظهر اللاتأليفي Non compositionnel ونعني به عدم

3) علامة ٨ تعني انتقال الجملة من التعبير من المسكوك الى التعبير العادي.

التأليف أو التناسق بين مستويي الشكل والمعنى، خلافا للتعابير العادية التي تتميز بالشفافية الدلالية من حيث ارتباطها بشكل مباشر بالرصيد الدلالي المتوفر في المعجم، حيث إن تلك الدلالة تحسب انطلاقا من كلماتها الواردة في المعجم.

إذا توفر هذان الشرطان فإننا نحصل حتما على تعبير مسكوك، وهما شرطان يجمعان بين المستوى الصوري والمستوى الصوتي والتركيبي المحدد للعلاقات المادية بين العناصر اللغوية وبين المستوى المعجمي والدلالي المحدد للعلاقة بين الشكل اللغوي ومرجعه اللساني الإخباري.

إن هذه التعابير الكثيرة العدد والمختلفة البنيات في نظام اللغة العربية، يجب أن تحظى ببرنامج خاص في أي مشروع يخصص للترجمة الآلية أو الفورية التي تهدف الى بنائه في الوقت الحالي جميع اتجاهات اللسانيات الحاسوبية، ذلك لأن هذه التعابير غير قابلة للتفكيك إلى وحدات معجمية تمكن من وضعها على شكل مقابلات للمداخل المعجمية للغات الهدف Cible، والسبب هو أن هذه البنيات تحضر في النظام اللغوي على صورة أشكال صورية صماء وكأنها لا أبعاض لها، حيث تكون في كليتها مداخل معجمية ثابتة، هي بصورتها المتراصة كليا تؤدي دلالة لا تتوضح إلا إذا عبرنا عنها بمدخل معجمي مقابل بسيط وعادي يشرح دلالتها، وهذا يرجع إلى مظهر التعتيم الدلالي الذي تتميز به هذه التعابير. وعليه فكل عملية للترجمة يجب أن تأخذ هذه التعابير على شكل مداخل تشغل مساحة مخصصة للجملة. وترجمتها لن تكون ممكنة إلا بتعابير مسكوكة من اللغة الهدف، وهذا يفرض بناء معجم آلي مقارن للتعابير المسكوكة للغات الخاصة.

نفس التعابير تطرح صعوبات جمة ربما من نفس النوع أمام المعالجة الآلية للغات الطبيعية وهو ما يصطلح عليه التحليل الآلي للنصوص (analyse automatique des textes) وهي العملية التي تتطلب تفكيك السلسلة الكلامية الى مكوناتها الصغرى بهدف تقديمها للآلة لتتعرف عليها، الأمر الذي يجعلها تفقد معناها المسكوك لتؤول إلى دلالة أخرى لن تتمكن أية عملية تركيبية استدراكه بالرجوع إليه ثانية، وما ذلك إلا لأن معاني هذه التراكيب لا تبنى وفق قواعد اختيار مطردة تتمتع بالانتاجية المعروفة في التعابير العادية بل إن نسيجها التركيبي غير التأليفي Non compositionnel يأتيها من تكوينها الأصم الجامد اللا توزيعي الذي لا يقبل التفكيك، وذلك ما يعطيها مسكوكيتها.

وفي هذا الصدد نشير إلى أن التعابير المسكوكة تنقسم من حيث البساطة والتعقيد إلى نوعين : نوع تكون سائر عناصره بسيطة (وهو النوع الذي تحدثنا عنه إلى الآن)، ونوع تكون بعض عناصره مركبة، مثلا :

وقع س.[في حيص بيص] ١

حيث لا يمكن الفصل بين مكوني المركب الاسمي (في حيص بيص) المؤلف من ثلاثة عناصر مسكوكة، بما في ذلك حرف الجر الذي ينسج مع نفس المركب الاسمي. وسنعود الى الطبيعة التركيبية لهذه التعابير فيما بعد.

2. المعالجة الحاسوبية :

2- 1 بناء قاعدة البيانات :

لقد اتبعنا في بناء قاعدة بيانات التعابير المسكوكة طريقتين :

الأولى، وتتمثل في كتابة جميع التعابير وفق نمط من الترقيم الذي نتخذه مرجعا عند بناء برنامج التعرف على هذه التعابير، وتتمثل هذه الطريقة في اسناد رقم مرجعي لكل صنف منها، ويتحدد الصنف في هذا المعجم بنوع الاسم المركب الذي يمثل المنطقة المعتمة في كل تعبير مسكوك، مثلا :

1200، مس ص = [صورة ناطقة]

حيث يتعلق الأمر هنا بمنطقة معتمة مؤلفة من اسم مسكوك وصفة مسكوكة أيضا وقد اتبعنا هذا الأسلوب لاحتضان الأسماء المركبة التي لا تذكر عادة مع الفعل الذي يتحكم في توزيعها في اللغة العربية، علما بأننا لم نستثن وصف المناطق المعتمة التي تنجز مع الفعل الذي يعتبر هو الآخر مسكوكا.

الثانية، وتتمثل في بناء أجروميات عددية قائمة على ثنائية السلب والايجاب، أو القبول والرفض، وهي بدورها تنقسم الى أصناف مبنية على نوع الاسم المركب (المنطقة المعتمة) الذي يتضمنه كل صنف.

والطريقتان معا تتكاملان في نهاية المطاف، ذلك لأن كليهما تقوم على الطريقة التصنيفية عينها، أي تأسيسا على نوع الاسم المركب الذي يتضمنه كل صنف من التعابير المسكوكة.

إن بناء قاعدة بيانات التعابير المسكوكة يتطلب منا حصر هذه التعابير بشكل مضبوط، وهكذا اعتبرنا كل تعبير مسكوك ممثلا لمدخل معجمي مستقل غير قابل للتجزيء وهذا لا يتعارض نهائيا مع الفكرة القائلة بأن بناء قاعدة بيانات لغوية في إطار معجم آلي يتطلب التعرف على سائر مفردات اللغة، فهما معا يتكاملان، ذلك لأن التعرف على المتواليات المسكوكة هو من قبيل التعرف على نظام المفردات اللغوية ما دامت هذه التعابير لا تقبل التجزيء الى مكونات صغرى مع الاحتفاظ بالدلالة المسكوكة (المجازية) غالبا.

وإذا كان من السهل التعرف على مفردات التعابير العادية بفضل الطريقة المتبعة في الحناش 1992(*)، فذلك يصعب مع التعابير المسكوكة، وذلك لتضمنها مناطق جامدة Fixe لا تتغير، كما أن هذه المناطق تختلف طبيعتها من صنف الى صنف، وهو ما يدعو الى التريث في وضع قواعد تصنيفية جديدة تغاير تلك التي تعودنا عليها مع التعابير العادية، مثلا جرت العادة اعتبار المفاعيل فضلات في الجمل، لكن هذا المفهوم سيتغير نهائيا مع هذه التعابير التي يمكن أن نعتبر عدم وجود الفضلات فيها إحدى مميزاتها الأساسية، وهذا من شأنه أن يعقد ظاهريا عملية التعرف على جميع مفردات اللغة العربية بأي برنامج معلومياتي لا يتضمن قواعد خاصة بهذه التعابير تمكننا من إدخالها في هرم المعجم الآلي الذي نتوق الى بنائه للغة العربية، انسجاما مع ما توصل اليه هذا المشروع في اللغات العالمية والاوروبية على الخصوص.

وبالنسبة الينا فإننا قمنا بتصنيفها من خلال المناطق الجامدة والتي تكون غالبا مناطق اسمية مختلفة الشكل والنوع. أما الفعل فنظرا لفقدانه القدرة التوزيعية التي كانت تمكنه من اختيار العناصر الاسمية المناسبة في التعابير العادية، فإنه لم يعد ممكنا اعتماده كعنصر أساسي في التصنيف، هذا على الرغم من حضوره الاجباري في كل تعبير مسكوك إما ظاهرا أو مضمرا. نخلص من هنا الى أن التعرف المطرد على التعابير المسكوكة لا يمكن أن يتيسر الا ببناء قاعدة بيانات شاملة لهذه التعابير.

(*) يرجع في هذه الاحالة ومثيلاتها إلى لائحة المصادر والمراجع.

ومن الصعوبات التي واجهتنا في بناء المعجم الآلي للغة العربية نذكر اضطراب مفهوم المفردة في التعابير المسكوكة، فإذا كانت المفردة في التعابير العادية تعرف بأنها متوالية لسانية دالة محاطة ببياض من الجهتين، فإنها في التعبيرالمسكوك قد تتضمن أكثر من متوالية بسيطة متضمنة لأكثر من منطقتين فارغتين لكنها غير دالة على أكثر من مفردة واحدة، مثلا :

يضرب س[(أخماسا) ١ (في أسداس) ٢] س

حيث تتأسس المفردة/المتوالية : (أخماسا في أسداس) على أكثر من بياضين، ولا يمكن لأي برنامج معد لتوليد المفردات اللغوية أن يدخل في الاعتبار أكثر من بياضين، بينما تبقى البياضات الأخرى بدون أي دور في عملية التعرف على هذه التعابير. ذلك لأن هذا النوع من المركبات الاسمية لا يقبل التجزيء إلى مكونات صغرى دالة تنفرد كل واحدة منها ببياضيها في الاستعمال العادي للغة العربية، بل ولا تسمح حتى باستبدالها بمتوالية لسانية.بسيطة معادلة لها دلاليا، فهذا النوع من المركبات الاسمية يبقى هكذا منطقة سوداء معتمة لا يمكن النفاذ اليها الا بالاحصاء والتصنيف حتى تقدم الى المعجم الآلي بشكل يسمح له بتوظيفها في العمليات التي ننتظر منه أن ينجزها.

غير أن هذا لا يجب أن يفهم منه أن كل التعابير المسكوكة قائمة على المتواليات الاسمية المركبة، فهناك العديد من هذه التعابير تقوم على عناصر اسمية بسيطة، مثلا :

ضرب س. [السكة] ١

وقع س. [في الفخ] ١

شد س. [(الرحال) ١ (الى مظ)].

أرهق س.[كاهل س] ١.

وفي هذه الحالة نعتبر الاسم جزءا من الفعل حيث يقع بينهما انصهار دلالي كلي يستحيل معه القيام بنفس الحساب الدلالي خارج هذه المنظومة الفعلية، وهذا ما دعانا الى تنويع أسلوب بناء قاعدة بيانات هذه التعابير في اللغة العربية، مع اعتقادنا بأن الأسلوب المناسب لها هو طريقة الأجروميات وليس الأسلوب الترقيمي، وهي الطريقة المفضلة في نظرية المعجم التركيبي المطبقة هنا.

2- 2 وصف المناطق المعتمة :

لم يفكر المعجميون العرب في وضع معجم خاص بالمفردات المركبة على كثرتها وإنتاجية بنياتها في نظام اللغة العربية. فكل الصيغ التي تشملها المعاجم الحالية وكتب النحو خاصة بالمفردات البسيطة، وكأن الأسماء المركبة غير موجودة.

الاسم المركب هو المتوالية اللسانية الاسمية التي تتألف من أكثر من مدخل معجمي بسيط من مداخل معجم المفردات البسيطة الموسوم ب DEAMS (4) ، وهي تتألف تارة من عنصر مشتق وآخر جامد، وتارة أخرى من عنصرين جامدين، ولم نصادف في بحثنا أي اسم مركب مؤلف من عنصرين اسمين مشتقين.

والأسماء المركبة على نوعين :

1- تلك التي تكون مسكوكيتها جزئية، وخاصة تلك التي تأتي على شكل مركب إضاف مثل :

[علامة المرور]

[غرفة العمليات]

حيث يتكرر اسم علامة وغرفة مع مجموعة من المركبات الاسمية : علامة النجاح، غرفة النوم، الخ. في حين ان المضاف اليه يتغير، وهو مع كل مدخل جديد يؤدي الى خلق اسم مركب جديد وبمعنى جديد.

2- ثم المسكوكة كليا، ويكون الاسم المركب مسكوكا إذا استحال ربطه بجملة، وفي هذه الحالة فإننا نعتبره مدخلا معجميا موحدا رغم فاصل البياض الموجود بين مكونيه جامدين كانا ام مشتقين. ويتميز الاسم المركب المسكوك كليا بظاهرة اللاتاليف Non compositionnalité ، أي عدم الانسجام بين شكله ومضمونه : المفردة في المعجم العادي تعني شيئا وفي الاسم المركب تعني شيئا آخر. كما يتميز هذا النوع من المركبات الاسمية برفضه

4) نعني بهذا معجما آليا للمفردات العربية البسيطة (انظر الحناش 1990).

الاستجابة للاستبدال، حيث تبقى عناصره صامدة في وجه أي تغير تركيبي (الحناش 1991).

ومن خصائص الاسم المركب أنه يقع في الغالب في موقع المفعول، وهو بذلك يبتعد عن منطقة التطابق مع الفعل، ولعل هذا هو السر في عدم اهتمام المعجميين العرب به، اذ غالبا ما كانوا يهتمون بالمداخل التي تلعب دورا صرفيا/ إعرابيا في التركيب.

من خلال استقصائنا لهذا النوع من التعابير تبين لنا أنها تتضمن من الأسماء المسكوكة مع الفعل : أسماء بسيطة، وأخرى مركبة. وقد تم تحديد الأصناف الصورية للمفردات المركبة التي تعتبر مناطق ثابتة في بعض أصناف التعابير المسكوكة على الشكل التالي :

1- المركبات المأخوذة من اللغات الأجنبية
2- ح س
3- الاعداد المركبة
4- المصطلحات العلمية المركبة
5- ص س
6- س ص
7- ح س ح
8- س ح س
9- س س
10- اشكال أخرى

هذه كلها بنيات صورية تصنف فيها الأسماء المركبة في اللغة العربية، ويمكن تقسيم كل صنف كبير من هذه الأصناف الى أصناف صغرى مما يسهل عملية المعالجة الآلية. يبقى علينا البحث عن المتواليات المناسبة في الاستعمال العادي للغة العربية، مع تجنب ما أمكن المهمل منها.

ونظرا للتكوين المعقد للاسم المركب كما قدمناه، فإنه يتطلب أسلوبا خاصا في المعالجة يختلف بها عن المعاجم السابقة. فلن نتمكن من إسناد أرقام متسلسلة للمداخل الاسمية نظرا لأن كل مدخل مركب يتضمن أكثر من مدخل بسيط، كما لن نتمكن من الاكتفاء برمز نحوي تصنيفي موحد، لان الاسم المركب يتضمن في الغالب أكثر من قسم من أقسام الكلم العربي. وأخيرا لن نتمكن من سلوك نفس الطريق التصريفي في بناء المعجم الاعرابي لهذه المفردات. كل ذلك فرض علينا اتباع الاسلوب التالي :

تقسم الاسماء المركبة الى أصناف توزيعية تقوم على متوالية الرموز التي يتألف منها

كل صنف مع مراعاة الفاصل الموجود بين مكونات كل صنف، وذلك بوضع علامة خاصة مكان البياض على الشكل التالي :

صنف الصفة ا	صنف الصفة ب
تع / س / تع / ص	ص / تع / س
ال / قيمة / ال / مضافة	بصيص / ال / امل
ال / قرض / ال / فلاحي	طويل / ال / نجاد
عملية / تنوين / جراحية	كثيرا / ال / رماد

حيث تعني العلامة (/) الفاصل الموجود بين المفردات البسيطة التي يتألف منها الاسم المركب. وهكذا يقسم معجم المفردات المركبة الى ملفات تختلف في شكلها باعتمادها الرموز المؤلفة لكل صنف. ويمكن تقسيم كل صنف الى أصناف صغرى، وبذلك يشمل الوصف كل الجزئيات المتعلقة بمداخل هذا المعجم.

2، 3 الطريقة الثانية (الأجروميات) :

وتتمثل في تقسيم التعابير المسكوكة إلى أصناف تركيبية، كل صنف منها يمثل مدخلا معجميا مستقلا، وقد قمنا بتجزيء كل متوالية الى مكوناتها الأساسية النحوية من ناحية الشكل فقط. وذلك تيسيرا لوصفها بهدف إدماجها في المعجم الآلي العام للغة العربية. وقد كونا بهذا قاعدة بيانات على شكل معجم تركيبي مصنف بحسب نوع البنيات الصورية التي يتأسس عليها منهجنا والتي تلخصها البنية ف س. ك. بهذه الطريقة توصلنا الى الآن الى تصنيف أكثر من 30000 تعبير مسكوك. وقد صنفنا ذلك كله وفق الأشكال الصورية التالية :

I - ف [مس] :

(البنية التي منطقتها الثابتة هي الفاعل).

II- ف س. [س] ١

(بنية منطقتها الثابتة المفعول له المباشر البسيط)

III- ف س. [حرف مس ١] :

(بنية منطقتها الثابتة المفعول المجرور)

IV- ف س. مس حرف مس] ١ :

(بنية منطقتها الثابتة مفعولها المركب من عنصرين : أولهما مباشر والثاني مجرور).

V- ف س. [حرف مس حرف مس] ١ :

(بنية منطقتها الثابتة مفعولها المباشر المركب من عنصرين غير مباشرين)

VI- ف س. [مس مس] ١ :

(بنية منطقتها الثابتة مفعولها المباشر المركب من عنصرين اسميين مباشرين)

مس = : العنصر الثابت في البنية المسكوكة، فقد يكون اسما بسيطا أو مركبا أو صفة أو ظرفا، الخ، كما أن حرف = : أي نوع من أنواع حروف الجر في اللغة العربية، لكن وجود أي من هذه المكونات في البنية تحت أي شكل كان يؤدي الى سك الفعل معه ومن ثم إعطاء الجملة دلالة خاصة وجعلها رسما (Graphe) غير قابل للتجزيء.

وقد قمنا ببناء أجروميات تضم المداخل المسكوكة مع خصائصها التوزيعية. وقد اتبعنا في ذلك النموذج أسفله، وهو عبارة عن جدول يقوم على خطين : أفقي وقد كتبت عليه التعابير المسكوكة بنصها الكامل. وعمودي وقد كتبت عليه خصائص هذه التعابير، وعند نقطة التقاء الخطين وضعت علامة (+) وتعني وجود هذه الخاصية لذلك التعبير، و (-) وتعني العكس أي عدم توفر هذا التعبير على تلك الخاصية التوزيعية، نموذج ذلك :

الخصائص التوزيعية

بنية التعبير المسكوك	الفاعل	المنطقة المعتمة (المفعول)	

الشكل

السطر يمثل جميع الاحتمالات التركيبية (التوزيعية والاشتقاقية) التي تتعلق بالمدخل المعجمي (التعبير المسكوك)، وقد قسمنا قاعدة البيانات الى مجموعة من الجداول، لكل صنف منها أجرومه الخاص به، بحيث لا نضع في الصنف الواحد الا الخصائص التي تتعلق بذلك الصنف. وهكذا فقد خصصنا للصنف الأول أجرومه الذي يغيب عنه عمود المفعول به لأن المنطقة الثابتة فيه تتبع الفعل مباشرة، هذا على الرغم من عدم توفر قاعدة البيانات على تعابير مسكوكة الفاعل في بنيتها الأصلية، فهي كلها مبنية المفعول المحول الى فاعل بموجب إجراء تركيبي أدى بها الى هذه الوضعية التركيبية. وهي الإجراءات التي حددناها في الحناش 1988. أضف الى هذا أن سك الفاعل يجعل التعبير مسكوكا كليا، وهو ما قلنا عنه بأنه يمتنع نظرا لانقطاع الصلة بينه وبين التعبير العادي. لكل هذه الأسباب كنا نشك دائما في وجود تعابير مسكوكة الفاعل أي لازمة. وقد أدرجناها هنا بهذه الصورة فقط لأننا نعتبرها مداخل مستقلة إذ يصعب ربطها ببنياتها الأصلية التي اشتقت منها. فنحن نقوم بوصف التعبير من خلال منطقته الثابتة المحققة بصرف النظر عن كون هذه المنطقة المعتمة مشتقة من منطقة عادية كانت تحتل موقعا آخر في بنية لم تعد قابلة للإنجاز بشكلها المسكوك. وفي هذا الصدد نشير الى أن بقية أصناف التعابير المسكوكة تطرح نفس المشكلة من حيث إن أغلبها يتضمن مناطق ثابتة مشتقة، ومن ثم أمكن الشك في كونها تحتل موقعها الصحيح والأصلي في التعبير، فنحن نرجح أن يكون أغلبها منقولا الى منطقته التي وصلنا فيها ولا نستطيع إرجاعه الى مكانه بدون التضحية بالمسكوكية، ذلك لأن أي محاولة لتطبيق الاجراءات المناسبة للعودة

بمكونات التعبير الى أماكنها الأصلية يؤدي غالبا الى نسف مسكوكية التعبير، وهذا ما دعانا كذلك الى التشبث بالبنيات كما هي في أجروملياتها المناسبة دون محاولة الرجوع بها الى الأصل الذي انطلقت منه، هذا على الرغم من توفر قاعدة البيانات على نسبة كبيرة من البنيات التي جاءت أصلية دون أن يكون قد طرأ عليها أي تحويل اختياري أو إجباري، مفرد أو معمم. وهذا هو السبب في وضعنا في جداولنا تعابير مسكوكة تتضمن ظروفا وصفات لم تكن تعتبر في تصنيفنا للتعابير العادية مكونات أساسية في البنيات اللسانية التي تؤطرها، بل كانت تعتبر فضلات وامتدادات ما صدقية، وقد أدرجت هنا باعتبارها عناصر أساسية نظرا لغياب مفهوم الفضلة من هذه التعابير كما سبق الذكر، فكل العناصر التي يأتي بها التعبير المسكوك أساسية فيه ولا مجال لحذفها أو اعتبارها عناصر مقحمة. ليس هذا فقط بل إن التعبير المسكوك لا يقبل التغيير حتى في أداة التعريف التي تعتبر مكونا أساسيا في المجموعة الاسمية في اللغة العربية، حيث لا نتصور وجود اسم بدون تعريف. وهذا يعني أن المسكوكية تطال ليس فقط المكون الاسمي في عمومه بل تطال سائر جزئياته أيضا.

إلى جانب التعابير المسكوكة التي لا تقبل الاشتقاق من أي بنية أخرى وبأي عملية تركيبية كانت، تتضمن قاعدة البيانات تعابير مسكوكة تقبل وبشكل لا يقبل الجدل الخضوع لمختلف أو لبعض العمليات التركيبية سواء بشكل جزئي أو كلي (الحناش 1991 ب).

إن تصنيفنا للتعابير المسكوكة من خلال مناطقها الثابتة التي يختلف محتواها من صنف الى آخر جعل عدد أصناف بنياتها الصورية الموصوفة يتجاوز عددها في التعابير العادية الأساسية في اللغة العربية، ومرد ذلك الى أننا لم نتمكن من تقسيم التعابير المسكوكة الى بنيات أصلية وأخرى محولة، أي مولدة بخوارزمات صورية كما هو الشأن مع التعابير العادية التي لا نثبت منها الا البنيات الأصلية الخمس (الحناش 1988) في حين يتكفل الخوارزم اللساني بتوليد سائر البنيات المشتقة. فنحن ندرج هنا كل بنية مسكوكة في صنفها المناسب، وهذا الصنف المناسب قد يكون متكونا من عناصر أساسية فقط، مثل :

أثقل س.[كاهل س] ١

كبح س.[جماح س] ١

أو من عناصر أساسية مضافا اليها بعض المكونات التي نصفها بالفضلة لكنها لم تعد

كذلك كما كانت في التعبير العادي، مثل :

شد س.[(الرحال) ١ الى مظ]

حيث لا اختيار لنا في ذكر الظرف المسكوك (مظ) والذي يؤدي حذفه الى نسف التعبير نهائيا :

* شد زيد الرحال

هذا على الرغم من اعتبارنا هذا التعبير أصليا، أي غير محول بأي عملية اشتقاقية تركيبية معروفة في النظريات اللسانية الحالية.

وكما تتضمن قاعدة بيانات هذه التعابير بنيات لسانية أصلية محملة بمكونات إضافية نقلتها ظاهرة المسكوكية الى عناصر أصلية، نجدها تتضمن كذلك بنيات محولة يصعب الرجوع بها الى أصولها التركيبية، مثلا :

باء س.[بالفشل] ١

حيث يعتبر الفعل باء فعلا عماديا Verbe support مؤطرا بإجراء لساني أطلقنا عليه إجراء التوسيم Nominalisation ، وهذا الإجراء من أخصب العمليات اللسانية التي تعرفها بنيات التعابير المسكوكة في اللغة العربية (انظر الحناش 1991 ب). لكننا في تصنيفنا لهذه التعابير لم نتوقف عند هذه القضايا كلها، فقد اكتفينا بإدراج هذه التعابير في البنية الفعلية دون التفريق بين أنواع الأفعال، وذلك في انتظار صناعة برنامج معلومياتي قادر على مساءلة قاعدة البيانات هذه لاستخلاص أنواع التعابير من وجهة نظر لسانية - حاسوبية في المستقبل القريب إن شاء الله.

3 منهجية التعرف الآلي :

يتم التعرف على التعبير المسكوك من خلال نظام الترميز المتبع في بناء قاعدة البيانات الخاصة بهذه التعابير. بالاضافة الى أن كل تعبير يحمل رقما تسلسليا من ١ الى .....٣، وقد حصرت المنطقة المعتمة في كل تعبير بعلامتين كدليل على أن ما بينهما يشكل مدخلا اسميا مركبا غير قابل للتفكيك. كما صنفت هذه التعابير بحسب بنياتها التركيبية التي

صممت لها خصيصا في جدول الهياكل والأطر التركيبية التي توصلنا اليها في إطار البحث الذي نجريه على نظام المعجم التركيبي للغة العربية. وهكذا خصصنا كل تعبير بأحد الهياكل التالية :

ف (مس.)(5)

ف س. (مس ١)

ف س. (ح مس ١)

ف س. (مس ١ ح مس ٢)

ف س. (ح مس ١) (ح مس ٢)

وهذه هي البنيات الأساس في اللغة العربية، كل واحدة منها قابلة للتفريع الى بنيات صغرى، وذلك بحسب نوع حرف الجر الذي يتركب مع أحد مكونات البنية الأساس المسكوكة وبحسب الطبيعة التوزيعية للعناصر الاسمية. هذا بالاضافة الى تنوع البنيات المسكوكة الى وصفية وظرفية ومطاوعة، الخ. وقد حدد كل نوع بالخاصية التركيبية التي جاء عليها. ولا تتضمن قاعدة البيانات هذه الا التعابير غير المحولة، ونحن بصدد بناء خوارزمات خاصة بتطبيق مختلف العمليات التركيبية بهدف توليد مختلف الجمل المحولة، مثل البناء التوصيفي والمتهيكل والموسم، الخ. وبتطبيق هذه العمليات سيتجاوز عدد التعابير المسكوكة المائة ألف تعبير.

ومن خلال التجربة التي قمنا بها على نصوص عربية بتطبيق برنامج FIG الذي بنيناه ليتعرف على التعبير المسكوك وتمييزه من التعبير العادي تبين لنا أنه تمكن من التعرف آليا على جميع التعابير المسكوكة مع نسبتها الى صنفها الخاص بها في قاعدة البيانات، وذلك بفضل الطريقة التي تم بها ضبط هذه التعابير في هذه القاعدة، لنأخذ المثال التالي :

" قطع الرجل على نفسه عهدا أن يدافع على مصلحة الجميع إذا تم التصويت عليه في الانتخابات، لكن وبعد مضي أقل من شهر على تقلده منصبه الجديد شرع في تلطيخ سمعة

---

5) يرمز حرف (مس) الى العنصر المعتم (Opaque) في التعبير المسكوك، والرقم الذي يلحق به يشير الى وظيفة العنصر في التعبير.

بعض الذين لا يساندونه في مواقفه، وكان ذلك نابعا من إرث عدائي ظل يكتمه في نفسه عبر السنين الى أن سنحت له الفرصة بتفجير مكبوتاته، فاستغل منصبه ليصب جام غضبه على أعداء الامس".

وقد تعرف البرنامج على جميع التعابير المسكوكة ورتبها بالشكل التالي :

قطع الرجل على نفسه عهدا : ف س. (مس ١) ح مس ٢ ض.)

تقلد منصبه الجديد : ف س. (مس ١ض.) (ص ١)

تلطيخ السمعة : (مس مس)

إرث عدائي : (مس ص)

سنحت له الفرصة : ف (مس.) (ح ض ١)

استغل منصبه : ف س. (مس ١ ض.)

يصب جام غضبه : ف س. (ص ١ مس ١)

أعداء الأمس : (مس ظ)

أما التعابير العادية فقد أهملها البرنامج لعدم تضمنه لخصائصها، ونطمح أن يتم ربط برنامج FIG بالبرامج الخاصة بالمعالجة الآلية للنظامين التركيبي والصرفي للغة العربية.

ومن شأن هذه الطريقة أن تسهم بشكل كبير في وضع برنامج فعال خاص بالتدقيق الاملائي والتشكيل الالي، لكن أهميته ستتجلى أكثر في بناء برامج خاصة بالترجمة الآلية أو الترجمة بالحاسوب.

وقد شرعنا في التنسيق مع باحثين أجانب بنوا قواعد بيانات خاصة بهذه التعابير في اللغات الأجنبية، وخاصة الفرنسية والاسبانية والانجليزية. ويتمثل الهدف الأساسي من هذا التنسيق في وضع برامج موحدة تؤهل لبناء برنامج متكامل للترجمة الآلية على المستوى الدولي.

4 نتائج البحث :

بالجمع بين الطرق المختلفة المحددة في هذا البحث تمكنا من بناء قاعدة بيانات تتضمن عددا كبيرا من المداخل المعجمية المسكوكة تتجاوز الثلاثين ألف مدخل، وهي موزعة حسب مناطقها المعتمة التي يتضمنها كل صنف. وقد وضعنا لكل صنف أجروما خاصا به يرصد شكل ونوع التعتيم الذي جاء عليه في التعبير المسكوك. وتتضمن قاعدة البيانات هذه أغلب - ولا نقول كل - التعابير المسكوكة التي وردت مفرقة في المعاجم العربية قديمها وحديثها. وقد صنعنا بها معجما جديدا يضاف الى المعاجم العربية المتخصصة.

وبناء قاعدة البيانات بهذه الطريقة هو الذي يسر وضع برنامج خاص بالتعرف الآلي على التعابير المسكوكة خلال عملية العلاج الآلي للغة العربية، وقد أعددنا قاعدة البيانات هذه بشكل يساهم في وضع هذه البرامج حيث أعطينا رقما لكل تعبير مسكوك في العربية مع الاشارة الى نوع التعتيم والجمود اللذين أصابا كل تعبير، دون إغفال الوصف اللساني لكل تعبير، الأمر الذي يسهل على البرنامج المرتقب مهمة التعرف على التعبير المسكوك وتمييزه عن التعبير العادي.

كما أن وضع قاعدة البيانات بهذه الطريقة من شأنه تيسير وضع برامج الترجمة الآلية، وقد بنيناها بشكل يستجيب لكل المتطلبات التي تسمح بالتعامل مع قواعد بيانات سبق وضعها في اللغات الأجنبية، وخاصة تلك التي بنيت في اللغة الفرنسية في مختبر العلاج الآلي والتوثيق اللساني LADL في باريس، وهي نفس القاعدة التي يستغلها برنامج الترجمة الآلية للسوق الأوروبية المشتركة EUROTRA . فقد وظفنا نفس التقنيات المعتمدة في هذين المشروعين، وذلك تيسيرا لإدخال اللغة العربية الى هذا المشروع الدولي الكبير، وقد ساعدنا عملنا المشترك مع هذين المشروعين على تحسين وضع قاعدة البيانات العربية عامة، والتعابير المسكوكة خاصة، حيث تبين أننا نلاقي جميعا نفس الصعوبات في الترجمة الآلية، وخاصة تلك التي تعلق بترجمة التعابير المسكوكة في اللغتين، ومن هنا اضطررنا الى تبني سنن CODE موحد وطريقة موحدة.

من نتائج هذا العمل كذلك أنه سيمكننا من صناعة برنامج حاسوبي يتعلق بالتدقيق الاملائي للنصوص المحررة بالعربية غير المشكولة، ويتعلق الأمر ببرنامج يتعرف على سائر

مفردات النصوص العربية، وخاصة الأسماء المركبة التي تقف غالبا في وجه أي اشتقاق صرفي أو تصريفي من نوع :

(أخماسا في أسداس)

(شدر مدر)

(أهلا وسهلا)

حيث يمكن للبرنامج تصحيحها بكل يسر ان أصابها خطأ في الطبع. هذا بالاضافة الى إمكانية تصحيح كل المفردات العربية البسيطة (انظر الحناش 1992).

المصادر والمراجع :

ذ. أحمد أبو سعد (1987) : معجم التراكيب والعبارات الاصطلاحية العربية : القديم منها والمولد. دار العلم للملاين، بيروت - لبنان.

- د. الحناش محمد، 1980) : البنيوية في اللسانيات، دار الرشاد الحديثة، 420 ص

- د. الحناش محمد، (1985) : النحو التاليفي. مجلة دراسات ادبية لسانية ع.1.

- د. الحناش محمد (1989) : المعجم الالكتروني للغة العربية، مؤتمر الكويت الأول للحاسوب، مارس 1989.

- د. الحناش محمد (1989) الاشتقاق التركيبي في اللغة العربية، المؤتمر الثاني حول اللغويات الحسابية العربية، الكويت نوفمبر 1989.

- د. الحناش محمد، (1990) مشروع نظرية حاسوب - لسانية في بناء معاجم آلية للغة العربية، مجلة التواصل اللساني، المجلد الثاني ع. الثاني.

- د الحناش محمد (1991 أ- ب) : ملاحظات حول التعابير المسكوكة في اللغة العربية. مجلة التواصل اللساني، المجلد الثالث، العددان الأول والثاني.

- د. الحناش محمد (1992) : المعجم الآلي للغة العربية : قواعد البيانات : المؤتمر الأول حول استخدام اللغة العربية في تقنية المعلومات، الرياض 10-14 ماي.

- د. الحناش محمد (1992) : قاعدة بيانات التعابير المسكوكة في اللغة العربية : المؤتمر الأول حول اللغويات الحاسوبية العربية، القاهرة 20-22 يونيه.

- د. الحناش محمد (1992) : المعجم التركيبي للغة العربية : مقدمات في المعالجة الآلية للغة العربية (تحت الطبع) في جزأين، منشورات التواصل اللساني.

- د. الحناش محمد (1993) : قاعدة بيانات التعابير المسكوكة تحتوي على 30000 تعبير، مسجلة على حاسوب صخر. مشروع عمل علمي في طور الانجاز بدعم من العالمية (الكويت).

- الزجاجي بن اسحق : كتاب حروف المعاني، تحقيق : د. على توفيق الحمد، ط. مؤسسة

الرسالة / دار الأمل، الأردن.

- عباس أبو السعود (1971) : الفيصل في ألوان الجموع، ط. دار المعارف بمصر.

- د. مراياتي محمد وآخرون، (1989) : نظام اشتقاق الكلمة العربية بالحاسوب، وقائع ندوة اللسانيات العربية والاعلامية، تونس 1987.

- الميداني احمد بن محمد : نزهة الطرف في علم الصرف، تحقيق لجنة احياء التراث في دار الآفاق الجديدة بيروت.

- د. نبيل على، (1988) : اللغة العربية والحاسوب (دراسة بحثية)، الكويت 591 ص.

- د. يحيى هلال (1990) : العلاج الآلي للعربية، مجلة التواصل اللساني المجلد الثاني، العدد الثاني.