

بناء معجم التعبيرات المتلازمة والمركبة: تقنيات التعرف الآلي

عز الدين غازي

جامعة محمد بن عبد الله، فاس، المغرب

جمعية اللسانيات الحاسوبية العربية - المغرب

ربط هذا الوصف بالمستوى التركيبي. وفي ظل تقادم المشاكل التي تعترض العلاج الآلي للغة العربية وهي عديدة. كانت الحاجة ملحة لبناء معجم من هذا الحجم، وهذا ما يقتضي شرطين أساسيين هما: بناء قاعدة معطيات شاملة للمفردات العربية، ثم الاستناد إلى نظرية لسانية محددة تستطیع وصف الظواهر اللغوية بدقة متناهية. وهذان الشرطان يجنبان وضع القواعد اعتماداً على الظن الذي يعرف التطبيق الآلي، كما يجنبنا مزلق وضع قواعد جزئية في البرنامج المولد لبيانات اللغة العربية. بناء على هذا التوجيه، فإنني سأبسط القول في شكل المعجم الآلي الذي سيستوعب الكفاية المعجمية الطبيعية، وكذا المحتوى القادر على تمثيل البنيات المتلازمة والمركبة للمداخل والمواد المعجمية، وهذا ما سيعالج الإشكالية المطروحة المتمثلة في تقييس الكفاية المعجمية حاسوبياً واختيار الأدوات العقلانية والمعلوماتية التي تستطیع تمثيل ومعالجة هذه المداخل، شكلاً ومحتوى.

ملخص -- تروم هذه الورقة تقديم الأدوات اللسانية والمعلوماتية المناسبة لبناء المعجم الآلي للغة العربية، باستغلال التقنيات المعلوماتية التي أبانت عن نجاعتها في معالجة لغات عديدة في سائر المستويات المختلفة وخاصة الكفاية المعجمية. وذلك بوضع معجم آلي يركز على وصف المفردات اللغوية بسيطها ومركبها، وبناء عليه ستركز مداخلتي على محورين اثنين هما: شكل ومحتوى المعجم المتمثل في البنيات المتلازمة والمركبة للمداخل والمواد المعجمية وتقنيات التعرف الآلي المتمثلة في خوارزمات التعرف على المنطقة المعتمدة للتعبيرات المتلازمة بتوظيف تقنية أوتومات الأوضاع النهائية والمحولات والرسوم (graphes). مما سيجيب عن كيفية تقييس الكفاية المعجمية حاسوبياً، وعن الأدوات العقلانية والمعلوماتية المناسبة لتمثيل ومعالجة المداخل المعجمية بجميع أصنافها ودرجاتها.

الكلمات المفتاحية-- التعبيرات المتلازمة- المعجم الإلكتروني- أوتومات الأوضاع النهائية- المحولات- الرسوم- المنطقة المعتمدة- التعرف الآلي.

1- تقنيات التعرف الآلي على البنيات المتلازمة:

للتعبيرات المتلازمة والمركبة خصائص تصريفية وتركيبية دلالية، تتحدد طبيعتها، انطلاقاً من المنطقة المعتمدة الكامنة بين الرأس والقاعدة المتلازمة والتي على ضوئها يمكن القيام بعملية وبناء النماذج المعتمدة على الصورة لغايات تصنيفية محضة، هدفها موقعة هذه البنيات اللغوية بسيطة أو مركبة، وجعلها دالة (fonction) رياضية تُحدّد درجة التحجر والتلازم. أما بعد هذه العملية الوصفية سيتضح- عند معالجة هذه التعبيرات- ظهور مشاكل وصعوبات تتعلق

متكلم الطبيعي. وللتعرف الآلي على هذه البنيات المركبة، نرى أولاً ضرورة الاستناد إلى نظرية المعجم- التركيبي التي استنادت عمقاً وتصوراً من نماذج فرضتها أساليب الوصف العلمي في مجالات متعددة وعلى رأسها الهندسة المعلوماتية التي استعارت اللغة العقلانية الرصينة كالتوليد والمعالجة والتحليل والأتمنة بأنواعها. حيث أبانت بجلاء الخصوصية التركيبية والدلالية لهذا النوع من التعبيرات، الذي يفترض بناء نماذج صورية كعمليات أولية في المعالجة النصية المتعلقة أساساً بمدى بناء برنامج آلي يمكننا من التعرف على البنيات اللسانية المركبة والمتلازمة، ومدى التعامل مع هذه البنيات والوحدات المركبة. إذ لا يمكن- مهما كان الأمر- للبرامج الآلية المعدة للتعامل مع المعجم الآلي العربي، أن تتعرف على جميع مكونات النص المستند على عينة قليلة العدد، إلا إذا احتوى هذا المعجم على جميع أنواع وأصناف التعبيرات المركبة في اللغة العربية. وقد بينت التجربة، كذلك، إمكانية تطبيق نفس المنهجية التي تتبع في التعرف الآلي على المفردات البسيطة أن تحدد شكلها الصوري (متوالياتها الحرفية المحاطة ببياضين). وقيل البدء في صياغة نماذجها، لا بد من التذكير بماهيتها الصورية، حيث أن كل

تقديم :

إن الهدف العام من هذه الورقة، هو تقديم الأدوات اللسانية والمعلوماتية المناسبة لبناء المعجم الآلي للمتلازمات في اللغة العربية، وذلك باستغلال التقنيات المعلوماتية الحديثة في سائر مستويات الكفاية المعجمية. أما الهدف الخاص فيرتكز على وصف المفردات البسيطة والمركبة من وجهة نظر تصريفها وصرفها، مع كيفية التعرف الآلي على هذه المتواليات 1 في نظام اللغة العربية التي تختلف قلباً وقالبا عن باقي أنظمة اللغات. ذلك ما أدى بإحاثين أمثال ماكس. سلبرشتاين (1993) إلى اعتماد المبدأ العقلاني الصوري لبناء أنظمة وبيانات مفتوحة المصدر تستغل في معالجة اللغات الطبيعية مثل بينتي (INTEX و NooJ) حيث استطاعتنا بناء نماذج صورية دقيقة تستجيب لخصوصية كل نحو محلي مستخدمة تقنية أوتومات الأوضاع النهائية لمعالجة التعبيرات المتلازمة والمركبة بهدف صورنتها (تصنيفها) وإدخالها في خوارزميات التعرف الآلي التحليلية والتوليدية، سنساعدنا هذه التقنية، لا محالة، في معالجة مثل هذه الظواهر النشيطة والعالية الإنتاجية لدى كل

[1] تطرح التعبيرات المتلازمة إشكالية كبيرة عند معالجتها الآلية، وعلى الرغم من وجود أنظمة آلية مثل ما ابتكره شينك (A.Schenk 1986) في اقتراحه نظاماً لترجمة الآلية لهذه التعبيرات يقضي بوجود تعبيرات في اللغة الهدف. غير أن = هذه المنهجية لم تتوصل إلى وضع اليد على الداء في معالجة آلية مستقلة في نظام مستقل كما فعل إيريك لا بورت (E.Laporte, 1988).

Zaid a cassé le verre / Zaid a cassé le jeune

مفردة مركبة هي متوالية مولفة على الأقل من مفردتين بسيطتين كما في الصوغ التالي:2

مفردة مركبة = (م ب + 1 م ب 2) + ح³

سيتم اختيار، إما فعل "فَطَرَ" (تناول الفطور) أو "كَسَرَ"، في هذه الحالة يتدخل المستوى الصرفي هنا لأمن اللبس، ولن نستعين بالمستوى التركيبي إلا إذا اضطر الأمر لذلك. وهكذا سيكون دور المحلل التركيبي المساعدة على تحديد إمكانية بروز هذا الفعل أو ذلك. وتوظف نفس المنهجية بالنسبة للمفردات البسيطة التي يمكن تعريفها باعتبارها متوالية من الحروف مكونة من أكثر من بياض واحد. ويتبع هذا التصنيف كطريقة تراعي خصوصيات مقولة المفردة التي تتألف منها المفردة المركبة¹⁰.

11-2- المتواليات المترابطة والمتواليات غير المترابطة:

تعتبر المتواليات المتميزة بمنطقة مترابطة بنية صماء لا تتغير، فإن قبلت على المستوى التركيبي، دمج إحدى العناصر فيما بين مكوناتها، اعتبرت غير مترابطة (zone non connexe) أما إذا لم تقبل ذلك اعتبرت مترابطة (zone connexe) نحو: حينئذ، مازال، ما فتئ/ لكن: * ما زال، ... فالتعرف الآلي على هذه البنيات يقتضي، اعتماد المستوى الصرفي كمرحلة أولية. لأنه رغم تكرار بروز بعضها في النصوص المعالجة، كما سنرى، سيتمكن البرنامج من التعرف عليها، بدون أن يعود عند التحليل للمستوى التركيبي مثلاً: **حيص بيص**- **شذر مذر**، كما هي في المتواليات التاليتين: **وقع زيد في حيص بيص**، و**تفرق العرب شذر مذر**.

إن التحليل المعجمي هنا، من شأنه أن يقدم جميع الفرضيات المعجمية الممكنة، بحيث لا يهمل في هذه المرحلة الحسم في رفع اللبس إلا في مرحلتى التحليل التركيبي أو الدلالي. وبعد العودة إلى قواعد البيانات الموسعة المصنفة ضمن المعجم الآلي للمفردات، بفحص جميع الخصائص المورفولوجية والتركيبية المحتوية عليها. نقوم أولاً بفحص المتواليات البسيطة وتحليلها ثم توزيعها بين عناصر أخرى تختارها الكفاية المعجمية للمدخل المركب، بمعنى آخر، يحل العنصر في الاستعمال العادي ثم بعده الاستعمال المركب¹²، كما في استعمال لفعل "أكل" التالي:

1- **أكلت هند التفاحة (ف س 0 س 1)**، بنية حرة

2- **أكل عليه الدهر وشرب (ف س 0 مت س 1)**، بنية متلازمة

إن منطقة التلازم في المثال (2) تقع بين الفعل والمركب "عليه الدهر" (ح + ضمير + س 1). في هذا المثال يبرز العنصر الحر الذي يمكن أن نستغني عنه في تحديد المنطقة المعتمدة، بحيث نترك الفعل "أكل" والتعبير "عليه الدهر" منفصلين لتظهر العلاقة الدلالية بينهما، فإن وزعنا "عليه الدهر" مع الفعل المذكور سنحصل على

[¹⁰] لن نتم هنا إلا بالتصنيف المعتمد على القوانين التركيبية متجاهلين إجرائياً الجانب الصرفي من جهة، ومن جهة أخرى نلاحظ أن الأفعال تغيب في هذا المعجم نظراً للخصوصية التركيبية المعجمية لهذه المقولة.

[¹¹] تحتوي المتواليات المتلازمة على منطقة مترابطة وصماء (zone connexe) لا تقبل أي تغيير أو استبدال أو دمج عنصر غريب ضمن عناصرها الأصلية، وهذه هي التعبيرات المتلازمة كلياً، أما المتواليات المتلازمة المحتوية على منطقة غير مترابطة (zone non connexe) فهي معرضة لأي استبدال لعناصرها أو دمج عنصر ما بين بياضها، وهذه هي التعبيرات المتلازمة جزئياً. للمزيد انظر ماكس سلبيرشتاين (1993) صص. 144 - 147.

[¹²] نفس المرجع والصفحة.

اقترح ماكس سلبيرشتاين (1993) هذا التعريف باعتباره تعريفاً مقيداً للمفردة المركبة، خاصة وأن مصطلح متوالية هو بالمعنى التقني منطقة مترابطة⁴ (zone connexe) لا يمكن النيل منها لكونها صامدة أمام أي تغيير أو استبدال أيا كان نوعه، وليس فقط بالمعنى اللساني. لنأمل الأمثلة التالية: **صباح مساء-علامة المرور**.

يمكن هنا، أن نجري بعض التعديلات بإدماج عنصر أو عناصر ضمن هذه المتواليات نحو: **العلامة الخاصة بالمرور- علامة خاصة بالمرور لكن: *الصباح المساء- * صباح ومساء - * علامة X الساعة - * الصباح X المساء**.

انطلاقاً من هذه الأمثلة يتبين أن بعضها يحتل إدماج عنصر ما ضمن متواليته البسيطة، لأنها وحدات معجمية تقبل التواجد ضمن وحدات تؤلف التعبير المركب والتي تختلف نوعاً ما من حيث قابليتها لإدماج عناصر غريبة عنها مقارنة مع التعبيرات المتلازمة، لأنها تقبل إدماج بعض المتواليات بصفة أكثر حرية من التعبيرات المتلازمة كلياً، لنأمل التعبيرات المركبة التالية: **صباح مساء - علامة الساعة**، إذ سنجد مداخل غير قابلة للتجزئ، لأنها تشكل في أغلبها مداخل متحركة، وهذا ما يستدعي القيام بوسمها⁵ (étiquetage) في المعجم⁶ وتحليلها تركيبياً وإعرابياً، مما يتطلب وصفها صرفياً وتركيبياً ودلالياً بصفة خاصة. ولأجل رفع اللبس الحاصل في جميع هذه المراحل وجب القيام بتحليل تصاعدي للجملة المركبة⁷. وهذا يستدعي الجمع بين نوعين من الوصف، الاهتمام أولاً بالوضع المورفولوجي للعناصر التي يتألف منها الاسم المركب، ثم ثانياً وصف وضعه الجديد في التاليفة التي أنشأها منه نظام العربية. وعملياً فيتم تحويل النص (inversion du texte) من البسيط إلى المركب⁸، خاصة وأن هذه المركبات تجمع في الغالب بين فصليتين من الأسماء الموجودة في معجم المفردات البسيطة بين الاسم والصفة، أو بين الاسم والحرف أو بين الاسم والظرف. وهذا الجمع يعني أن الاسم المركب يتألف من فصيلة صرفية كذلك، وهو ما يتطلب معالجة وافية. وقد تم تحديد الأصناف (انظر الجدول (1) أسفله) التي نعتمدها في طرق التعرف الآلي وأمن اللبس المتواليات المعقدة، لنأمل مثلاً البنية المأخوذة من اللغة الفرنسية التالية⁹:

[²] Max Silberstein , (1993)

[³] م.ب= مفردة بسيطة ، ح= حرف

[⁴] نفس المرجع. ص. 144.

[⁵] يسمح الوسم المعجمي للمفردات بإسناد الملامح الإعرابية المتمثلة في الجنس والعدد والشخص والزمن واستخراج أشكالها الأصلية (الفعل في صيغة الماضي المفرد الغائب) وفي حالة غموض المتوالية مصرفة في النص المعجمي، فإن الوسم يعطي مجموعة من الإمكانيات المتمثلة في الوصف الدلالي والتركيبي للمفردة المعالجة.

[⁶] E.Laporte, 1988.p.117.

[⁷] نفس المرجع والصفحة.

[⁸] من بين التقنيات التي اقترحها ماكس سلبيرشتاين في التعرف الآلي على التعبيرات

المركبة تقنية تحويل النص (inversion du texte) انظر نفس المرجع

السابق.ص.118.

[⁹] يمكن أن أضيف تعبيراً أكثر وضوحاً في اللغة الفرنسية نحو: (Amr a casse

sa pipe) التي تعني مات.

أكثر من فصيلة صرفية أيضاً. ومن المفروض أن تكون هذه المفردات كلها موجودة في معجم المفردات البسيطة. كما يعني أن المركب يتكون ليس فقط من متواليات من المداخل بل وأيضاً من متواليات من البياضات الفاصلة بين كل مدخل بسيط، تأخذ بعين الاعتبار هذا المظهر المعقد من وجهة نظر تقنية، أي وضع برامج تستجيب لهذه الحالة، فلكل معجم طريقته في التصنيف. فإذا كان المعجم المحوسب يقوم على طريقة التصنيف التي تجمع بين الترقيم الصرفي، والإحالة النحوية، فإنه سيقوم كذلك على أسلوب آخر في ترتيب المعلومات. ولأن استدعاء المستوى التركيبي في عملية الوصف المورفولوجي يمكننا فعلاً من إقامة تصنيف مورفو- تركيبى لهذه المركبات. وبفضله سنتمكن فعلاً من إنشاء هذه الأصناف التي تسهل عملية ضبط هذا المعجم وبالتالي بناؤه بالشكل الذي يستجيب به للعلاج الآلي المتوخى من هذا العمل. ولذلك تم تحديد هذه الأصناف، إلى أصناف صورية كبرى وصنفيات صغرى مما سيسهل عملية المعالجة الآلية، ويبقى علينا البحث عن المتواليات المناسبة في الاستعمال العادي للغة العربية مع تجنب ما أمكن المهمل منها. وهذا أيضاً ما لم ينتبه إليه أغلب المعجميين العرب في وضع معجم خاص بالمفردات المركبة على كثرتها وإنتاجية بنيتها في النظام العربي. نظراً لهذا الخط التصوري في تطبيق الإجراءات التصريفية والتركيبية، فأيهما سيكون الأولى أم هما معاً وكيف؟ هذا مما يفقد الصيغ التي تشملها المعجم العربية خاصيتها أي مفردات بسيطة أم مركبة؟ وكأن المفردات المركبة والتعبيرات المتلازمة غير موجودة. مع أن بداية هذا الاهتمام برز مع اللسانيات الحاسوبية. وحسب جل المعجميين الحاسوبيين فإن المفردة المركبة هي متواليات لسانية اسمية مؤلفة من أكثر من مدخل معجمي بسيط من مداخل المعجم. وهي تتألف تارة من عنصر مشتق وآخر جامد، وتارة أخرى من عنصرين جامدين، علماً بأن اسمين مشتقين نادرين في اللغة العربية. ومن المعجميين من يزعم أيضاً، أن المفردة المركبة عبارة عن جملة مختزلة خضعت عناصرها لتغيرات الصرفية وهذا ما يعقب عليه **محمد الحناش** 15 بعدم جدواه، ومرد ذلك إلى أنهم يتخذون المحمولات الاسمية المولدة بظاهرة التوسيم كنموذج يعممه على المفردات المركبة ذلك أن بعض المتواليات المركبة لا تستجيب لهذا التحليل نحو: **شذر مذر- حيص بيص - مثنى مثنى ...**

وحتى في حالة التوسيم المشار إليه أعلاه، يصعب العمل به كإجراء مورفو- تركيبى مولد للأسماء المركبة في اللغة العربية، لما ينتج عن هذا الاختزال من ضياع في الخبر، لأنه يصعب الرجوع بها دائماً إلى الجملة التي تولد عنها المركب الاسمي نحو: **مقود السيارة - حائط المبكى..** لسبب واحد هو فقدان العلاقة الدلالية بين هذه المركبات الاسمية وما يحتمل أن يكون أصلاً جملياً لها.

واعتماداً على ما ذكرناه، فإن المفردات المركبة على نوعين:

أ- **المتلازمة جزئياً**، وخاصة تلك التي تأتي على شكل مركب إضافي نحو: **علامة المرور- غرفة النوم** حيث يتكرر اسم علامة وغرفة مع مجموعة من المركبات الاسمية: **علامة النجاح، غرفة العمليات الخ،** في حين أن المضاف إليه يتغير، وهو مع كل مدخل جديد يؤدي إلى خلق اسم مركب جديد وبمعنى جديد أيضاً.

ب- **ثم المتلازمة كلياً**، ويكون الاسم المركب متلازماً إذا استحال ربطه بجملة. وفي هذه الحالة فإننا نعتبره مدخلاً معجمياً موحداً رغم فاصل البياض الموجود بين مكونين جامدين كانا أم مشتقين. وتتميز المفردة المركبة المتلازمة كلياً بظاهرة اللاتاليف، أي عدم الانسجام بين شكلها ومضمونها: فكلما " ألقى " في المداخل البسيطة تعني شيئاً وفي المداخل المركبة يعني شيئاً آخر، ويتميز هذا النوع من المركبات الاسمية كذلك، برفضه الاستجابة للاستبدال، حيث تبقى عناصره صامدة في وجه أي تغيير تركيبى. هذا ما يجعل المفردات

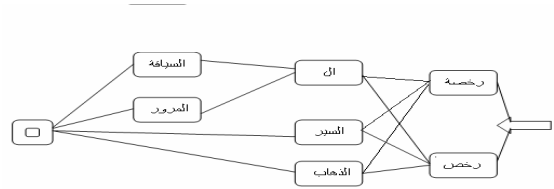
المتلازمة المقبولة في الكفاية المعجمية، فالمنطقة المعتمدة هي التي تسمح بالعدد المحدد للعناصر البسيطة لولوجها بحسب تغييراتها الصرفية 13 التي تلحقها في التعبير المتلازم نحو:

3- **أكل عليهم الدهر وشرب / أكل عليها الدهر وشرب / أكل عليهم الدهر وشرب**

على الرغم من التحويل الذي حصل في المثال (3)، فإننا نلاحظ تثبيت حرف الجر بالضمير، وهذا ما يؤكد العلاقة المتلازمة بين الحرف والضمير من جهة كتعبير مركب، وبين الفعل والتعبير المركب من جهة أخرى.

1-2-3- تمثيل بيانات المفردات المركبة والمتلازمات:

عند عملية التعرف على الوحدات البسيطة التي تشبه المفردات المركبة الموجودة في المعجم الآلي، نعتد بتقنية العبارات العقلانية وبالرسوم للتمثيل، بهدف إبراز نتائج التعرف الآلي للمفردات المركبة، ونورد هنا للتمثيل مثالا من اللغة العربية 14 المفردة المركبة التالية " **رخصة السياقة** ". التي يمكن تمثيلها بالرسم (graphe) التالي:



الرسم (1): أوتومات التعرف على المفردات المركبة

بحيث لا يمكن للبرامج المعدة للتعامل مع المعجم الآلي أن تتعرف على جميع مكونات النص المكتوب، إذا لم يكن هذا المعجم متضمناً لجميع أنواع الأسماء المركبة في اللغة العربية، ذلك أن عدداً كبيراً من المداخل المعجمية جاءت مركبة بشكل يمنع تناولها مجزأة مثل: **علامة المرور- بصيص الأمل- المثنى الأخير- أحد عشر، وسائر الأرقام المركبة- ذهاباً وإياباً- مثنى مثنى .. الخ.** وهي مداخل غير قابلة للتجزئة، لأنها تشكل في أغلبها مداخل متلازمة، وبالتالي تتطلب وصفاً آلياً صرفياً من نوع خاص. هذا الوصف الذي قد لا يفيد معه الأسلوب العلاجي الصرفي المتبع مع المفردات البسيطة، فهي لا تدخل في أي من التصنيفات الخاصة بأقسام الكلم كما حددناها: الاسم والصفة والظرف والحروف والأدوات، بل تتأسس في الغالب في أكثر من قسم من أقسام الكلم كما أشرنا. ولهذا كان من الطبيعي أن يخلق هذا الوضع صعوبات أمام المعالجة التقليدية التي عملت بها الدراسات اللسانية الحاسوبية لحد الآن، فكان من الضروري إدخال مفاهيم وصفية إجرائية جديدة تتلاءم مع هذه الوضعية الجديدة في المعالجة. إن الأمر يدعو إلى الجمع بين نوعين من الوصف: الاهتمام أولاً بالوضع المورفولوجي للعناصر التي تتألف منها المفردة المركبة، ثم ثانياً وصف الوضعية الجديدة التي أنشأها منه نظام اللغة، خاصة أن الاسم المركب يجمع في الغالب بين فصليتين من الأسماء وهي: إما بين الاسم والصفة أو بين الاسم والحرف، أو بين الاسم والظرف، الخ. هذا الجمع يعني أن المفردة المركبة تتألف من

[13] نفس المرجع. ص. 119.

[14] للمزيد حول التعرف الآلي على التعبيرات المركبة والمتلازمة وعلى التمثيلات

بالطريقة الأقواسية وبالرسوم انظر، M.Silberztein (1993) p.138.

[15] محمد الحناش، (1985) صص 35-83.

وبعمليات رياضية حسابية تبين أن القواعد المعتمدة في البحث المتوالياتي لا تناسب نسب تواتر بعض المفردات المركبة في نصوص اللغة الألمانية، لذلك اقترح سيمون أولمان استراتيجيات خوارزمية أخرى للتحليل، تتمثل في عدم اعتماد التسلسل الخلفي في البحث، وذلك بالاقصصار على عدد محدد من الدوال الرياضية لحساب تقطيع العُجَيّمات البديلة أو المشتقة في المفردة المركبة المعالجة، فيختار صَوغاً رياضياً يعين الرمز الذي يمثل العُجَيّمات المؤلفة للكلمة بداية من حرف i الذي يتضمن عدد معين من الحروف j بحيث تكون المعادلة على الشكل التالي: $0 \leq j < n - i$

ترتكز هذه العملية الرياضية في التحليل على معطيات أساسية شاملة للطول الأقصى لبعض الكلمات الممكنة التي وقع عليها الحساب. وفكرة اعتماد الكلمة الأقصى طولاً أعطت نتائج هامة لأنها تمكنت بالملحوس من إحصاء الكلمات المحتملة في التحليل، وبالتالي يسهل التوصل بهذا إلى تحسين أداء الخوارزم المذكور، وذلك باعتماد مجموعة من الإجراءات في كتابة. وتتلخص في اعتماد لائحة معينة من المفردات المركبة للتحليل ثم بداية وهيكل ونهاية البرنامج [21]، حيث الحصول على الحجم أو العدد النهائي من العُجَيّمات المكونة للمفردة المركبة موضوع المعالجة.

ومع الصعوبات التي تطرحها البنيات المتلازمة عند معالجتها والتعرف على المنطقة المعتمدة فيها، صرفياً وتركيبياً- معجمياً ودلالياً لأهداف تطبيقية متعددة، رغم وجود أنظمة معلوماتية، تستطيع معالجة هذه المداخل. فإنه على مستوى ظاهرة التعبيرات المتلازمة بات ضرورياً البحث عن مناهج عامة واقعية في الوصف اللساني باعتماد متن موسع وشامل لجميع البنيات الممثلة للكفاية المعجمية للفرد. لأن معالجة هذه الظواهر قد تطورت بفضل أدوات الوصف التي توظف اللغة العقلانية وأساليب المعلوماتية، فأصبح لزاماً أن تعتبر المفردة خارج السياق لا معنى لها أو هي عنصر خارج لساني إذ أنها على مستوى التعبير المتلازم لا تستقل بدلالاتها الخاصة بل تظل منصهرة ومتلاحمة أو حتى صامدة في وجه أي تجزئ أو تفكيك مما يستدعي بناء قاعدة بيانات للتعبيرات المتلازمة محصورة في أشكال صورية مضبوطة، خاصة وأن المدخل المعجمي مستقل بذاته غير قابل للتجزئ، بسبب الخاصية الدلالية اللاتألفية التي تميزه عن غيره من الحروف والألفاظ البسيطة لأنه يتميز باستحالة إمكانية استبدال أحد عناصره (المفعول غالباً) بعنصر آخر ولو كان من نفس الفصيلة المعجمية. وبسبب الخاصية الدلالية اللاتألفية لبعض الجمل البسيطة فإنها كذلك تصبح تعبيرات متلازمة كما يذهب جاك لابيل: 22 في أن التعبير المتلازم لا يمكن تأويله دلالياً بفعل بسيط، لأنه وحدة دلالية مستقلة يجوز تعويضه بلفظة واحدة، إذ تحسب عناصره بمدخل معجمي واحد قد يقوم مقام التعريف بالبنية المعجمية- المتعددة (polylexical figée) وهذا له دور رئيسي في بناء أنظمة الآلية لترجمة والتوثيق والتلخيص الآليين... فإذا نكون إزاء بنية صماء لا مقابل لها في اللغة الهدف يسهل علينا إيجاد حل لترجمة بعض التعبيرات المعقدة، نحو: شُد زيد الرحال: [سافر]- نادى زيد ربه: [مات]- ضرب زيد في الأرض: [تاه].

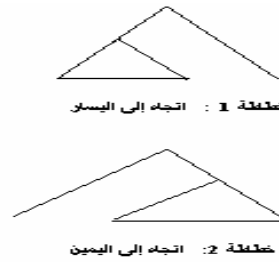
وإذا كان من السهل التعرف على مفردات التعبيرات الحرة والعادية، بفضل طرق المعالجة الآلية المتقدمة، فذلك صعب مع التعبيرات المتلازمة لتضمنها مناطق ثابتة و صماء لا تتغير، وهذه المناطق تختلف طبيعتها من صنف إلى صنف: من المتلازمة كلياً إلى المتلازمة نسبياً، مما يدعو إلى المرونة في الوصف واعتبار المفاعيل مناطق قارة و صماء وليست مجرد فضلات كما يذهب

المركبة الاسمية المتلازمة تختلف عن غيرها بانعدام إمكانية إدماج أي عنصر بين مكوناتها نحو: **أخماسا في أسداس / لكن: * أخماسا X في أسداس** خلافاً للمفردات المركبة غير المتلازمة أو المتلازمة جزئياً والتي تسمح بهذه إمكانية.

2- خوارزميات التعرف الآلي على المفردات المركبة والمتلازمة:

تتخصر هذه الخوارزميات في التعرف الآلي على المتواليات المركبة وذلك بالارتكاز على البحث المتوالياتي (recherche séquentielle) 16، انتهج هذا الأسلوب إيريك ثرنهر (E.Thurnherr 1993) الذي حاول تحليل المركبات انطلاقاً من اللغة الألمانية، على الرغم مما لاقاه برنامج الآلي من صعوبات- لاسيما مع تطور أنظمة الاستغلال وبنية المعطيات النصية- تتمثل في سرعة معالجة المعلومات وفي الطريقة التي يبحث بها البرنامج، كان يعتمد البحث الخلفي لتحديد رأس المركبات أو ذيلها. منطلقاً بتحليل المفردة المركبة وتجزئتها إلى وحدات صغيرة بسيطة مثل ما يحصل في تعبيرات اللغة الألمانية التالية: (ingänge و haus) حيث يتبين أن عدد مرات الولوج المباشر لتحليل هذه المتواليات، حرفاً بحرف، حوالي تسعة. ولكن مارتان سيمون أولمان (Martin Simon Ulmann) 18 حاول تطوير عملية التعرف المباشر للنصوص المتوفرة على هذه المركبات بدقة وسرعة تتمثل في إدراج حوالي 640 و لوج مباشر لتحليل معطيات المركبات في هذه اللغة (بمقدار 6.6 حرفاً بالثانية). و تعود هذه الصعوبة إلى طول المفردة المركبة في اللغة الألمانية من جهة، وإلى الولوج المفرداتي المتعدد للبرنامج من جهة ثانية، لذلك وجب بناء خوارزم يتناسب مع هذه المعطيات لرفع اللبس الحاصل أثناء المعالجة الآلية 19. ولتحديد عناصر المتواليات المركبة بشكل دقيق ومقبول في الكفاية المعجمية، اعتمد إيريك ثرنهر تحليلاً آخر يتجه من اليمين إلى اليسار (الشكل 1) و آخر يتجه من اليسار إلى اليمين (الشكل 2) حيث تبين أن نسبة 70% من المتواليات التي تتكون من ثلاثة عناصر تتجه إلى اليسار بينما حوالي 30% تتجه إلى اليمين 20، علماً بأن هناك متواليات تتوفر على أكثر من ثلاثة عناصر كما في اللغة العربية نحو:

البطاقة الرمادية ← [ال + بطاقة + ال + رمادية]



الشكلان (1) و (2) لتحليل المركبات باتجاهين معاكسين

[16] Martin Simon Ulmann. (1994) p.1-16

[17] عن مارتان سيمون انظر، نفس المرجع السابق .

[18] نفس المرجع والصفحة.

[19] توجد في اللغة الألمانية بعض المركبات من نوع (bauchtanzschule) التي

تتفرع عنها مركبات أخرى مثل bauchtanz- و bauch-tanzschule و bauchtanzschule و bauch-tanz-schule) مما يستدعي برنامجاً يقلص الغموض الحاصل

فيهما.

[20] نفس المرجع والصفحة .

[21] نفس المرجع السابق. صص. 7- 10.

[22] J. Labelle (1988)

تقوم بالدور المحوري في تصنيف وضبط البنيات المذكورة تأسيساً لقاعدة البيانات، خصوصاً وأن عدداً كبيراً من المداخل المعجمية تأتي مركبة وليست مجزأة لأنها مداخل غير قابلة للتجزئة، ونظراً لتكوينها المعقد، فإن ذلك يستدعي هذا الأسلوب المميز في المعالجة سواء أكانت البنيات متلازمة كلياً أو جزئياً،²⁷ فلن نكتفي بإسناد رمز نحوي تصنيفي موحد لأن المفردة المركبة تتضمن أكثر من قسم من أقسام الكلم العربي. كل ذلك يجعل إتباع أساليب وصفية تصنف التعبيرات المتلازمة وفق بيئاتها الصرفية-التركيبية والمعجمية-التركيبية كذلك، وتتمثل في كتابة جميع التعبيرات وفق نمط من الترتيب الذي نتخذه مرجعاً عند بناء برنامج ألي للتعرف عليها، بإسناد رقم لكل صنف من أصنافها التوزيعية، مع مراعاة الفاصل الموجود بين مكونات كل صنف على حدة بوضع علامة خاصة مكان البياض على الشكل التالي:

صنف : تع / س / تع / ص / ال / قيمة / ال / مضافة
عملية / انتحارية / ال / بنك / ال / شعبي
صنف : ص / تع / س / بصيص / ال / أمل

إن الرمز الوحيد الذي يرسم البياض بين العناصر البسيطة في النص هي العلامة (/)، يساعد على تقسيم التعبيرات المركبة إلى ملفات، وهي عبارة عن أصناف كبرى ثم أصناف صغرى وأجزاء صغيرة، تتمثل في العمليات الإعرابية الناتجة عن التغيرات التصريفية التي تتحكم فيها الأدوار التركيبية. وهذا ما يعبر عنه ماكس سلبزرتاين بالبنيات الصغرى الغامضة التي تحتاج التعامل الدقيق²⁸، بالتعامل إلا مع رأس المتلازمة، لأنه هو الذي يتحكم في العملية التوزيعية. فيجوز مثلاً، إعراب الأسماء المركبة ويتعامل صرفياً إلا مع الجزء الأول نحو: جواز السفر - جوازي السفر - جوازات السفر وهكذا.

غير أن دور المستوى التصريفي يقلص تماماً هنا نظراً لرفض هذه البنيات أي تحرك صرفي نحو: حيص بيص/ لكن *حيص (جمع/مثنى/تصغير) بيص

إن الأشكال الصورية المتبعة في الوصف هنا، هي نفسها التي اعتمدت مع التعبيرات الحرة، فالهدف واحد هو بناء نظام ألي يستطيع حوسبة الأشكال اللغوية بسيطها ومركبها. ولما كانت التعبيرات متميزة بالثبات الكلي في التعنيم الدلالي فإنه يصعب معها استدراج أي برنامج للتعرف الألي. وهذا ما بقي قيد التجريب والاختبار لحد الساعة ببناء أنظمة وبيانات للمعالجة اللغوية المتطورة. وأما فكرة تصنيف البيانات اللغوية لمعالجتها فينبني أساساً على تصور خوارزمي مضبوط²⁹ ينسب كل جملة أو تعبير إلى الصنف المناسب لها دون أي تردد، وذلك لتمكين البرامج الخاصة بالمعجم الآلية أن تتعرف على هذه التعبيرات بواسطة جداول معجمية - تركيبية بنيت لهذا الغرض. وهذه الأصناف محصورة صورياً في البنيات³⁰ الست التالية:

جدول(1) يلخص الأصناف الصورية للبنيات المتلازمة

القضاء من الناحية²³، لأننا لا يمكننا أن نستبدل التعبير المتلازم الآتي: **انكسر أنفه** بتعبير آخر هو **انكسر يده**، فالمراد من التعبير الأول هو **انهزم**. لأن العلاقة بين المفعول والفعل هنا في علاقة تلازم محضة. ومن صعوبات التعرف الألي نجد أيضاً أن بعض التعبيرات تتضمن أكثر من منطقتين فارغتين لكنها غير دالة على أكثر من مفردة واحدة كما في المثال الآتي ذكره: **يضرِب (أخماساً) في (أسداس)** يتأسس هذا التعبير على أكثر من بياضين، وهو لا يقبل التفكيك إلى مكونات صغرى لأنه لا يسمح باستبدال عنصر من عناصره بمتواليه لسانية بسيطة²⁴. وإذا كان الفعل في التعبير العادي هو الذي يتحكم في اختيار عناصره المؤلفة معه، فإنه مع التعبير المتلازم يفقد قدرته التوزيعية وأن علاقته تظل مستقرة وثابتة مع المفعول مثال: **ألقي زيد الحبل على الغارب/ *أزاح زيد الحبل على الغارب**. بينما يمكن لعنصر الفاعل أن يستبدل بوحدة معجمية أخرى، كما يجب في بعض الحالات الاعتماد على السياق لفهم المعنى المراد من التعبير مثلاً: **دبت بينهم العقارب = وقعت بينهم العداوة**.

1-2- وصف المنطقة المعتمدة (zone fixe):

إن المناطق الثابتة في التعبيرات المتلازمة هي مناطق اسمية مختلفة الشكل، فهي تحتوي على عدد ثابت من المفردات البسيطة رغم قابلية هذه الأخيرة للتحويل الصرفي²⁵ إلا أنها تتصهر دلاليًا مع العناصر الأخرى، فيستحيل معها القيام بنفس الحساب الدلالي خارج هذه المنظومة الفعلية، ولذلك فإن المفردة المركبة كمتواليه هي التي تمثل المنطقة المعتمدة في التعبير، لأنها تتألف من أكثر من مدخل معجمي بسيط، كما أشرنا، المتلازمة جزئياً والمتلازمة كلياً. كما تختلف المركبات الاسمية المتلازمة عن غيرها بانعدام إمكانية إدماج أي عنصر بين مكوناتها (كما أسلفنا الذكر أعلاه)، نحو: **أخماساً في أسداس / لكن *أخماساً X في أسداس** خلافاً للأسماء المتلازمة جزئياً نحو: **علامة المرور، العلامة الخاصة بالمرور**.

يميز العالم الفرنسي ماكس سلبزرتاين، في هذا الصدد، بين المنطقة التي تقبل إدماج عنصر ما بين مكوناتها باستخدام تقنية أوتومات الإدماج (Automate d'insertion)، وأخرى لا تقبل إدماج أي نوع من الأوتومات، بمعنى لا تترك الحرية للتقليل بين العقد قصد اختيار عناصر صورية أخرى، وهذا له دور أساسي في التمثيل الدلالي والمعرفي الألي الذي يروم تيسير التوليد التركيبي- المعجمي والدلالي²⁶. وللاشارة هنا فإن دور المحولات (Transducteurs) ونوع أوتومات الأوضاع النهائية (Automate à état fini). كما سبق مع التجارب المخبرية في مختبرات (LADL و Franche Compte) لمعالجة اللغة الفرنسية وبناء المعاجم الإلكترونية باستخدام بينتي (INTEX) و (NooJ) قد حلت معهما مجموعة من القيود التركيبية للتعبيرات المتلازمة المتمثلة بصفة خاصة، في مسألة إحالة التعريف وخاصة مع البنيات الصورية المصنفة، ويحصل هذا مع البنيات الصورية للتعبيرات المتلازمة في اللغة العربية، وذلك بمساعدة الأتداء الصورية المحلية التي تستطيع تمثيل المعطيات وتوليدها في خوارزمات التحويل نحو: **ضرب زيد أخماساً في أسداس**.

(ي) ضرب زيد ال (أخماس) في ال (أسداس)

إن إتباع المنهجية العقلانية في الوصف من شأنه أن يساعد على ضبط الخصائص التوزيعية للتعبيرات المتلازمة ومعالجتها، لأنها

[23] محمد الحناش، (1996) ص. 81.

[24] نفس المرجع السابق والصفحة.

[25] Eric. Laporte (1988)

[26] M. Silberztein 88 pp : 144..148

[27] محمد الحناش (1992) ص. 102.

[28] M.Silberztien,(1993) , p.141.

[29] محمد الحناش، (2002)، ص. 20.

[30] قمنا بمحصر البنيات الصورية للمتلازمات في اللغة العربية في خمس أصناف لكننا نضيف هنا صنفاً آخر هي البنية رقم ستة والتي تتداخل فيها منطقتين معتمتين مما يجعل =الصنف يتألف من فعل واسم متلازمين كما في البنية (ف س 0 [مت مت])، ولم نعمد هذه الأخيرة في بناء قاعدة البيانات المعجمية للمتلازمات العربية نظراً للغموض الذي قد تطرحه عند استخراجها من النصوص مما يتطلب إغناء قاعدة البيانات بشكل دائم وتفاعلي.

* تتميز التعبيرات المتلازمة بالدرجة في التعنيم ويفسر ذلك درجة البساطة والتعقيد الذي تتحكم في بناء المتواليات المتلازمة فهي مؤلفة من مفردتين بسيطتين بها منطقة مترابطة، فالتى تتوفر على منطقة مترابطة قارة أسهل تعرّفاً من غيرها من التعبيرات، مما يسهل وضعها ضمن قاعدة البيانات الخاصة بالمعجم الآلي للمفردات المركبة .

* ينشأ الغموض من أول عملية للتعرف الآلي المطرد وهو من أهم المشاكل المطروحة على المتواليات المتلازمة والمركبة، مما يستدعي بناء برمجيات أكثر تطوراً و ملائمة لنظام اللغة الطبيعي. ولن يتأتى ذلك، إلا ببناء قاعدة بيانات شاملة تتضمن جميع الأصناف وليس الاعتماد فقط على أمثلة قليلة العدد.

* كلما كانت التعبيرات المتلازمة تتمتع بحرية استبدال عناصرها أو تقبلها لتغيير ما، كلما كان التعرف بالآوتومات في النص أقل دقة.

* تجاوز محنة فقدان كل استقلال معجمي ودلالي عند إدماج الاسم في البنية المتلازمة، ببناء أنحاء محلية جزئية لا تختلف عن النظام اللغوي العام، مما سيساعد برامج التعرف الآلي القائمة على أوتوماتات الأوضاع النهائية بوصف أدق لعناصر التعبير المتلازم، لأنه تركيب خاص في ذاته لا يخرج إطلاقاً عن نظام الكفاية المعجمية.

المراجع :

- [1] سليم م، زابد محمد، دراسة في المعجم الفرنسية الحاسوبية، أشغال الاجتماع الثاني لخبراء المعجم الحاسوبي للغة العربية، مدينة الملك عبد العزيز للعلوم والتقنية ، الرياض- أبريل 2008.
- [2] غازي عزا لدين، بناء المعجم الآلي للغة العربية: معالجة التعبيرات المركبة والمتلازمة أطروحة لنيل الدكتوراه بإشراف أ / د عبد الغني أبو العزم و أ / د محمد لهلال كلية الآداب عين الشق، الدار البيضاء 2009.
- [3] الحناش محمد، برنامج لساني – حاسوبي للتعرف الآلي على التعبيرات المسكوكة في اللغة العربية، مجلة التواصل اللساني م. 3 سنة 1996.
- [4] الحناش محمد، النحو التأليفي (مدخل نظري)، مجلة دراسات أدبية ولسانية. (1985) صص 35-83.
- [6] الحناش محمد، التعبيرات المسكوكة في اللغة العربية، تقديم نظري ومنهجي، منشورات جامعة عجمان (2002)، ص.
- [7] الحناش محمد، المعجم الآلي للغة العربية، بناء قاعدة معطيات، مجلة التواصل اللساني، م. 2 ع. 1 سنة 1992.
- [8] E. Laporte, la reconnaissance des expressions figées lors de l'analyse automatique, in langage n°90, 1988. pp. 117-126.
- [9] Labelle. J, lexique - grammaires comparées, formes verbales figées en français de Québec. in langage 90.1988. Larousse.
- [10] Max Silberstein Dictionnaires électroniques et analyse automatique du textes Masson. 1993. p 137.
- [11] Martin Simon Ulmann, l' analyse des mots composés Allemands, in Notes techniques n°94/8. LATL, Université de Genève p.1-16.
- [12] Merella. Conenna., les expressions figées en Français et en Italien: problèmes lexico-syntaxiques de traduction. in langage n°90.88.

رقم	اصناف البنيات
1	ف [مت 0]، (البنية التي منطقتها الثابتة هي الفاعل)
2	ف [س 0 س 1]، (البنية التي منطقتها الثابتة هي المفعول له المباشر البسيط)
3	ف س 0 [ح مت 1]، (البنية التي منطقتها الثابتة هي المفعول المجزور).
4	ف س 0 [مت ح مت 1] البنية التي منطقتها الثابتة هي المركب من عنصرين: ³¹ أولهما مباشر والثاني مجزور.
5	ف س 0 [ح مت ح مت 1] (البنية التي منطقتها الثابتة هي مفعولها المركب من عنصرين غير مباشرين)
6	ف س 0 [مت مت 1]، (البنية التي منطقتها الثابتة هي مفعولها المباشر المركب من عنصرين اسميين مباشرين)

ثمة عناصر تكون البنية المتلازمة فتمنح الدلالة الخاصة للمتواليات ومن ثم جعلها رسماً (graphe) لا يقبل التجزئ بصفة نهائية³². فبالصنيف نستطيع بنوع من الحرية بناء أنظمة آلية تستجيب مع خصوصيات النحو المحلي للتعبيرات المتلازمة لأنها بمثابة أساليب عملية أصيلة في لغة مخصوصة، وهذا لا يعني نفس القوة والدرجة الموجودة في جميع اللغات الأخرى، ما عبرت عنه الإيطالية ميريليا كونييا،³³ بالترجمة الداخلية للتعبيرات المتلازمة في الثقافة، ففي غياب لدراسات لسانية مقارنة في ما بين اللغات الطبيعية. لا بد من تصنيف هذه التعبيرات وفق قواعد بيانات في كل لغة نريدها أن تلج الأنظمة الآلية المتعددة اللغات، بإخضاعها لترجمة لسانية داخلية في ذات اللغة، شرحاً وترجمة من داخل اللغة المصدر، كما يجب وضع نفس البرنامج في لغات أخرى حتى تتمكن من المقارنة والتوسيط بين المستويات اللغوية، وهذا لن يتأتى إلا ببناء معجم آلية لهذه التعبيرات. بتصنيف المدخل وجعلها في جداول بغية إعداد قواعد بيانات شاملة لها، وذلك لتمكين الكفاية المعجمية من ترجمتها آلياً، كما يظهر ذلك جلياً حينما نكون إزاء وضع لغوي مختلف وخاصة عند الترجمة من العربية إلى الفرنسية أو العكس³⁴.

خلاصة :

وأخيراً أجمل ما خلصت إليه من نتائج كما يلي:

* تتميز التعبيرات المتلازمة بضعف واستحالة استبدال أحد العناصر التي غالباً ما تكون أكثر تحجراً و تكلساً، فهي بهذا الاعتبار ليست شفافة بل معتمة مما يجعلها لا تأليفية شكلاً ومعنى.

* تعتبر التعبيرات المتلازمة حاصل حسابي بين المستوى الصوري والمستوى التركيبي والصوتي، وهو المحدد للعلاقات المادية بين المتواليات اللسانية والمستوى التمثيلي الكامن في الكفاية المعجمية للأفراد.

* استحالة تجزئ هذه البنيات إلى أبعاض وعناصر وعُجُمات، وهذا ما يعتبر حجرة عثراء أمام الترجمة الآلية حيث يجب أن تشغل هذه التعبيرات حيزاً جلياً يلائم المدخل المعجمي باعتباره أدنى وحدة للدلالة مما يستدعي بناء معجم مقارن للتعبيرات المتلازمة في كل اللغات³⁵.

[32] محمد الحناش. (2002)، ص. 21.

[33] Merella. Conenna. (1988). p : 130

[34] تعتبر عملية الترجمة كباقي التطبيقات الحاسوبية: التلخيص الآلي والنشر الإلكتروني والبحث عن المعلومات..

[35] محمد الحناش ، (1996). ص. 81.