دراسة مسحية للتعرف الآلي على جمع التكسير في النص العربي

الطالبة: للا عائشة قاديري

المشرف: أ.د/عبد الفتاح حمداني

معهد الدراسات والأبحاث للتعريب(المغرب)

ملخص:

نهدف من خلال هذه الدراسة إلى عرض مفصل لنماذج مختلفة من أساليب التعرف الآلي على جموع التكسير في اللغة العربية مع تصنيفها وتقديم أهم الأعمال التي اعتمدته ا، ثم محاولة تقديم عرض لآليات رفع اللبس.

وأهم هاته الأساليب والمقاربات ثلاثة : المقاربة المعتمدة على الأوزان (أوزان جمع التكسير)، المقاربة المعتمدة على المعجم، ثم المقاربة الجامعة بين الأوزان والمعجم. وكذا آليات إزالة اللبس وهي: الآلية اللسانية- الألية- الإحصائية- و الآلية الهجينة. ثم في الأخير عرضنا بعض أنظمة معالجة جمع التكسير: نظام مورف-نظام أراجن- و مورف أراب، ونظام الخليل.

الملخص باللغة الأجنبية:

L'objectif principal de cette étude à la présentation détaillée des différents modèles d'identification automatique sur le pluriel brisé de la langue arabe, avec la classification et de fournir des œuvres les plus importants qui ont adopté, puis tenter de présenter les mécanismes pour lever l'ambigüité.

Le plus important de ces trois méthodes et approches : une approche basée sur les patrons, une approche basé sur le dictionnaire, et l'approche associe entre le dictionnaire et les patrons. Et aussi les outils de lever l'ambigüité : l'approche linguistique- et statistique et hybride .Enfin on a présenté quelque systèmes de traitement de pluriel brisé : système morph –Arajen-morph Arab et Khalil.

الكلمات المفاتيح: التعرف ألآلي على جمع التكسير، التحليل الصرفي

مقدمة :

يطرح موضوع التعرف الآلي على جموع التكسير تحديا كبيرا في ميدان المعالجة الآلية للغة العربية وخصوصا في محور التحليل الصرفي الآلي للنصوص العربية، حيث يعتبر هذا النوع من الجمع من أبواب الصرف المهمة الذي يضم أوزان سماعية وقياسية، ولا زالت هناك مجموعة من الصعوبات قائمة أمام موضوع التعرف على جمع التكسير واشتقاقه انطلاقا من مفرده أو توليد المفرد انطلاقا من جمعه، ومرد ذلك من جهة إلى عدم وجود قواعد تحويلية لبعض جموع التكسير يعتمد عليها للوصول إلى المفرد انطلاقا من الجمع أو العكس، أو أوزان محددة له تسهل عمليتي الاشتقاق والتوليد بخلاف الأنواع الأخرى من الجموع القياسية(جمع المذكر والمؤنث

السالمين). وتعتبر المحللات الصرفية والتركيبية ومحركات البحث من أهم التطبيقات اللغوية المحوسبة في حاجة ماسة إلى وجود آليات للتعرف على جمع التكسير.

نتعرف في هاته الدراسة على أهم المقاربات المستعملة للتعرف الآلي على جموع التكسير في النصوص العربية ، وكذلك على رفع آليات اللبس فيها ونذكر أهم الأعمال التي اعتمدت هاته المقاربات.

مقاربات معالجة جموع التكسير آليا: تنقسم الأعمال التي عالجت جمع التكسير آليا إلى قسمين :

- قسم اهتم بالتعرف على جمع التكسير في النص والحصول على مفرده.
- قسم آخرر كز على اشتقاق جمع التكسير انطلاقا من مفرده.

نهتم في هذه الدراسة بالقسم الأول، ونتطرق فيما يلي إلى أهم المقاربات المتداولة للتعرف على جموع التكسير في النص العربي والتي اعْتُمدت في مجموعة من المشاريع:

-مقاربة الأوزان، مقاربة المعجم، مقاربة الأوزان والمعجم.

وقبل الخوض في هاته المقاربات سنعرف جمع التكسير وأنواعه وخصائصه.

1/ جمع التكسير: يعتبر جمع التكسير أحد أنواع الجموع في اللغة العربية بجانب جمع المذكر والمؤنث السالمين، ويدلّ على أكثر من اثنين وله مفرد يشاركه في معناه وأصله مع تغيير ضروري يحدث لمفرده عند الجمع.

يشترط في الأسماء التي تجمع جمع تكسير أن تكون ثلاثية أو رباعية، أما ما زاد على أربعة أحرف فتجمع بحذف حرف، أو حرفين من حروفه مثل: سفرجل / سفارج، حذفت اللام.عنكبوت/ عناكب، حذفت الواو والتاء.عندليب/ عنادل، حذفت الياء والباء.

أما الصفات بأنواعها (اسم فاعل، اسم مفعول، صفة مشبهة، اسم تفضيل) فالأصل فيها أن تجمع جمع مذكر سالما، لأن القياس مطرد فيها، وجمعها جمع تكسير ضعيف وشاذ حسب النحاة.

2/ أنواع جمع التكسير تنقسم جموع التكسير من حيث الدلالة على العدد إلى نوعين:**جموع قلة، وجموع كثرة،** ولكل من هاذين النوعين أوزان وقواعد، حيث حرص الصر فيون جاهدين العمل بها والتقيد بضوابطها، ومما تجدر الإشارة إليه أن نذكر بأنّ هذه الجموع منها السماعي الذي لا بد من الرجوع لمعرفته إلى معاجم اللغة، وكتب النحو للوقوف عليها، ومنها القياسي

الذي وضعت له قواعد وأوزان تطبق على كل ما توفرت فيه الشروط عند الجمع، وسنركز في هذه الدراسة على الأوزان القياسية.

أولا/ جموع القلة :هو ما دلّ على أكثر من اثنين ولا يزيد على العشرة. وقد وضع له الصرفيون صيغا وقواعد أطلقوا عليها جمع القلة، وله أربعة أوزان هي:

النوع	الجمع		المفرد	
	وزن	مثال	وزن	مثال
جمع قلة	أفْعُل	أبْحُر- أنْهُر	فَعْل	بَحر- نَهر
	أفْعِلة	أفئِدة	فُعال	فؤاد
	فِعلة	فِتية	فَعل	فتى
	أفْعَال	أبواب	فاعل	باب

ثانيا/ جموع الكثرة: هو ما دلّ على أكثر من اثنين وزاد عن العشرة إلى ما لا نهاية، وجموع الكثرة متعددة وكثيرة، والغالب فيها السماع، وقد يكون للاسم المفرد الواحد أكثر من جمع، وللجمع الواحد أكثر من مفرد.

وقد استطاع الصرفيون أن يضيقوا حدود السماع في هذه الجموع ووضعوا لبعضها قواعد وأوزانا (ما يقارب 39 وزن) يقاس عليها نظائرها، و فيما يلي بعض هذه الأوزان:

النوع	الجمع		المفرد	
	وزن	مثال	وزن	مثال
جمع كثرة	فُعَل	زُمر	فُعْلة	زُمْرة
	فُعَلة	سَحرة	فاعل	ساحر
	فَعَلة	قِردة	فِعل	قِرد
	فَعْلى	مَرضى	فَعيل	مريض
	فُعَّال	صُوّام	فاعل	صائم
	فُعُول	قُلوب	فَعل	قَلب
	فِعْلان	غِلمان	فُعال	غُلام

أهم المقاربات للتعرف على جموع التكسير في النص:

1 – المقاربة المعتمدة على الأوزان: من بين المقاربات التي اعتمدت على الأوزان نجد كلا من المقاربة التواصلية البسيطة، والمقاربة التواصلية المقيدة، والمقاربة المعتمدة على آلة الترجمة، وهناك مقاربات أخرى اشتقت المفرد انطلاقا من جمع التكسير (ماكرتي و برنس 1990) (بيسلي 1989) باستعمال الأوزان.

1–1/ المقاربة التواصلية البسيطة(L'approche de correspondance simple): تعتمد هاته المقاربة(كويدر 2004) على إعداد لائحة أوزان جموع التكسير والتي تظم 39 وزنا، كما تقوم باستخراج جذع الكلمة المدخلة وذلك بإزالة السوابق واللواحق، بعد ذلك يقارن الجذع المتبقي مع جميع أوزان اللائحة الأوزان للنظر في مدى توافق الجذع مع أحد الأوزان، ولكي يتم هذا التوافق يجب أن يكون الجذع والوزن من نفس الطول، كما يجب أن يكونا لهما نفس الأحرف بعد إزالة الأحرف الصامتة (ف، ع، للجذر). وإذا تطابق وزن الجدع مع أحد أوزان اللائحة اعْتُبِرت الكلمة جمع تكسير.

مثال:جـو ا هـر←فـو ا عـل

كلمة'جواهر' تتوافق مع الوزن 'فواعل' حيث لهما نفس الطول ونفس الأحرف الزائدة (واو– لام) عن الحروف الأصلية (ف، ع، ل). إلا أن هاته المقاربة يمكن أن تقدم لنا كلمات ليست بجموع تكسير،حيث أن هناك بعض أنواع الأسماء تتشارك مع بعض جموع التكسير في الوزن، مثل كلمات: خداع، قتال، نفاق،لها وزن 'فعال' الذي يعتبر أحد أوزان جموع التكسير ولكن هذه الكلمات لا تعتبر جموع تكسير. كما يمكن أن تعتبر بعض الحروف الأصلية في الكلمة كزوائد، مثلافي كلمة "بواخر" عندما تحذف 'ب' على أساس أنها سابقة نحصل على 'واخر'.

في الختام تعد المقاربة التواصلية البسيطة مقاربة ضعيفة لأنها لا تمكن دائما من التعرف على جمع التكسير وبالتالي تقدم لنا كنتائج كلمات ليست بجموع تكسير. مكنت هذه السلبيات من تطوير هذه المقاربة وإضافة قيود أثناء المعالجة للتوصل إلى مقاربة تواصلية مقيدة والتي سنتطرق إليها فيما يلي.

1–2/ المقاربة التواصلية المقيدة(L'approche de correspondance restrictive): تعتبر المقاربة التواصلية المقيدة (كويدر 2004) تطويرا وامتدادا للمقاربة السابقة، وتتميز عنها بإضافة مرحلة تتضمن قيدا. ولتبيان القيد الإضافي على المقاربة الأولى نأخذ:

المثال الأول الذي جاء على وزن'فعال'، نجد أن المقاربة السابقة قدمت كلمات ليست جموع تكسير (نفاق) ولكن للتمييز بين الوزن(فعال)، الذي هو جمع تكسير، ونفس الوزن الذي ليس جمع تكسير(اسم)إذا كان وزن المفرد (مفرد لجمع التكسير)جاء على وزن فعيل أو فعيلة فهو جمع تكسير إذا لم يكن كذلك فهو ليس جمع تكسير.

إن المقاربة التواصلية المقيدة لم تحل مشكلة الجذور(إيجاد الجذر الصحيح), ما دمنا نحصل على كلمات خاطئة، وكذا على أوزان مشتركة بين جمع التكسير وأسماء أو صفات، من هنا ظهرت المقاربة المعتمدة على آلة الترجمة.

1-3 /المقاربة المعتمدة على آلة الترجمة(L'approche basée sur la machine de traduction): تعتمد المقاربة المعتمدة على آلة للترجمة (كويدر 2008) على خمسة مراحل هي:

- ✓ التواصل الأصلي.
- ✓ إزالة العلامات المدخلة من اللغة الأصل إلى اللغة الفرع.
- ✓ إزالة علامات الترقيم .
- ✓ تحويل الكلمة الجمع التي في لغة الأصل إلى مفردها.
- ✓ ترجمة الكلمة المفرد التي في لغة الأصل (الانجليزية مثلا) إلى اللغة الهدف (العربية).

المرحلة الأولى تسمح بالبحث عن الكلمة المدخلة في قائمة الأوزان الموجودة لجمع التكسير قبل أن نستخرج جذرها. وإذا تطابق وزن هاته الكلمة المدخلة مع الأوزان الموجودة والخاصة بجمع التكسير فان الكلمة تترجم مباشرة إلى اللغة الأصل، بعد ذلك تزال علامات الترقيم إذا كانت موجودة، وفي المرحلة ما قبل الأخيرة يتم تحويل الكلمة الجمع التي في اللغة الفرع إلى الكلمة المفرد، ثم يتم ترجمة هذه الكلمة المفرد إلى العربية بمساعدة نظام الترجمة.

ولتوضيح هذه المقاربة، نستعمل اللغة الانجليزية كلغة فرعية، ونقدم **مثالا** 'بواخر' وهو جمع 'باخرة' .

الجدول-1-

المرحلة1	المرحلة2	المرحلة3	المرحلة4	المرحلة5
بواخر-فواعل	Steamers	لا شيء	Steamer	باخرة

الكلمة المدخلة 'بواخر' ترجمت مباشرة عن طريق المترجم 'غوغل'، وهكذا حول الكلمة الجمع إلى المفرد 'باخرة'.

عندما نلاحظ النتائج نلاحظ أن بعض الكلمات لم تعالج بطريقة صحيحة، مثال جمع التكسير 'أروقة' مفرده 'رواق'، فعندما يتم إدخال هذا الجمع ترجمت الكلمة المفرد، ونحصل في المرحلة الأخيرة على 'ممر'، وهاته الكلمة ليست مفرد 'أروقة'، كذلك اعتبرت هاته الطريقة أسماء وصيغا للمبالغة جموعا للتكسير 'مثل العباس' .

وقد اعتبرت هاته الأخطاء ضعفا في هاته المقاربة وفي تقديم حلول جدية، ولذا ظهرت مقاربة المعجم، ولكن قبل التطرق إليها سنقدم إطلالة على منهجية 'ماكرتي'.

منهجية 'ماكرتي' لاشتقاق جمع التكسير من المفرد باستعمال أوزان جموع التكسير :

العميق Iambique	الترويشي Trochaïque	أحادي التفعيلة Monosyllabique	أخر Autre
Cicaac	Cicac	Cucc	Cuc1c1ac
Cicaac+at	Cicac+at	Cicc+at	Cuc1c1aac
Cucuuc	Cacac	Cicc+aan	
Cucuuc+at	Cacac+at	Cucc+aan	
Cacaac	Cucac	Cacc	
Caciic2	Cucac+at		
Cawaacic	'Cucac+aa		
Cacaa'ic	Cucuc		
Cacaacic	Cacic+at		
Cacaaciic	Cacic+aa		
	Cucuc		

تعتمد منهجية (ماكرتي و برنس 1990) على اشتقاق جمع التكسير انطلاقا من مفرده، وحسب (ماكارتي) فان أوزان جمع التكسير تفوق 70 وزنا يتضمنها 31 وزنا الأكثر استعمالا، بالنسبة للأوزان المستعملة قسمها

إلى أربعة أقسام (رايت 1971)، العميق و(الترويشي)، و(أحادي التفعيلة)، و(أخر)، ويوضح الجدول أعلاه هاته الأقسام (رايت 1971) .

وحسب (Wright,1971) فإن كل قسم من الأقسام الأربعة له خصائص محددة، وسميت هاته الأقسام بأسمائها بناء على خصائص الاسم العروضية. مثلا قسم (العميق) خصص للجموع التي تبتدئ بالجزء الثابت (حرف حركة)، مثلا جمع (نفس/نفوس)، الذي يحتوي على حرف (النون) كحرف أول، وحرف (الفاء) كحرف ثاني، وحركة (الضمة) كحركة قصيرة أولى وحركة الضمة الطويلة (الواو)، و بهذا تشكلت القدم الأولى (نفو).

القسم الثاني (الترويشي) سمي بهذا الاسم انطلاقا من المقطع الذي ابتدأ به جمع تكسير ينتمي لهذا القسم، كل نموذج يتكون من البنية الثابتة (حرف حركة حرف حركة حرف)، مثل (ركب) جمع (ركبة).

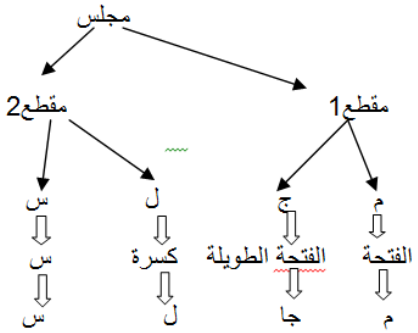
بالنسبة للقسم الثالث، فصنفها في الجموع التي تبتدئ ب (حرف حركة حرف حرف)، مثل (حمر- خضر)، بالنسبة لباقي الأوزان الخاصة بجمع التكسير فصنفها في الأوزان الخاصة بالقسم (الأخر)، وهي أوزان قليلة الاستعمال، لان ليس لها بنية شكلية ثابتة.

بالنسبة للتقسيم الذي قسمه (ماكرتي)، فتختلف تقسيماته في الاستعمال، فالقسم الأول هو الأكثر استعمالا، يليه القسم الثاني، والقسمين الأخيرين هما الأقل استعمالا.

بعدما صنف (ماكرتي) أوزان جموع التكسير في أربعة أقسام، اقترح (ماكرتي) خوارزميات احتمالية لمنهجيته، هاته الخوارزميات قدمت طريقة للحصول على الجمع من المفرد.

لنأخذ على سبيل المثال، الكلمة التي تتكون من أربعة أحرف (مجلس)، والتي تتكون من مقطعين اثنين (مج) و (لس)، فالمقطع الثاني يتكون من (حرف حركة حرف)، وبهذا ينتمي إلى القسم الأول، ومن ثم يجب تعويض كل حركة ب (الفتحة) في المقطع الأول، وتعويض كل حركة في المقطع الثاني (الكسرة). وتصبح جمع (مجلس) (مجالس).

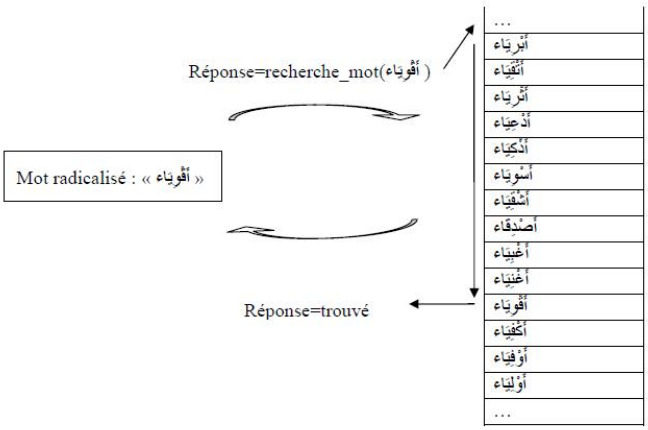
و الشكل التالي يبين طريقة الانتقال من المفرد إلى الجمع في المثال (كلمة مجلس الرباعية).



إن منهجية (ماكرتي)، عالجت فقط جمع التكسير الأكثر استعمالا(بمعنى الأوزان الأكثر استعمالا) ، وليس الأكثر احتمالا(أي الأوزان المهملة)، وهذا هو الجانب السلبي فيها، كذ لك اقتصرت على القسم الأول 'العميق' فقط.

إن ضعف نتائج مختلف المقاربات المعتمدة على الأوزان أدت إلى ظهور مقاربة جديدة وهي مقاربة المعجم.

2- مقاربة المعجم: اعتمدت مقاربة المعجم (درور 2004)، و(كودر 2004) استعمال جدول مقابلات يضم كل جذوع جموع التكسير، وتبدأ المقاربة بإزالة السوابق واللواحق من الكلمة المدخلة للحصول على الجذع، بعدها يتم البحث عن هذا الجذع في جدول الجذوع، ويتم الاحتفاظ بالكلمة في حال العثور على جدعها في الجدول. مثلا كلمة (أوفياء) جمع (وفي)، يتم البحث فيها على الشكل التالي:



يظهر من منهجية البحث في هاته المقاربة أنها منهجية بسيطة في البحث، ما دام البحث يتم فيها البحث عن الجدع الذي حصلنا عليه في جدول جموع التكسير كما أنها تقدم كلمات ليست بجموع تكسير وخاصة تلك الكلمات التي تبدأ بحروف أصلية حيث يعتبرها النظام سوابق.

إن مقاربة المعجم سريعة في البحث ولكن تستغرق وقتا كبيرا لجرد جميع جموع التكسير في اللغة العربية، (مثل الكلمات التي تبدأ بحرف الفاء أو الكاف، مثال فوانيس جمع فانوس، كوابيس جمع كابوس).

إن ثغرات مقاربة المعجم والتي تتمثل في الوقت والمجهود الكبير لتحقيق جدول شامل وكامل يضم كل جذوع جمع التكسير المستعملة والمهملة، أدت إلى اقتراح مقاربة تعتمد على جرد المعجميات (الليمات) التي تضم جمع تكسير مع تعدد قواعد الاشتقاق لهذه الجموع من الليمات. سنتطرق بتفصيل في المحور التالي لمبادئ هاته المقاربة التي تعتمد المعجم (الذي يظم الليمات) والأوزان (التي تسمح بالاشتقاق من الليمات).

3/ مقاربة الأوزان والمعجم: جمعت مقاربة الأوزان والمعجم (المصفر 2008) و(عباس2004) بين أوزان جمع التكسير وإعداد معجم خاص بليمات(المعجميات) الأسماء التالية (المصادر- الأسماء المهملة-المشتقات(اسم الفاعل- اسم المفعول)-جموع التكسير التي لا مفرد لها)، وتتمثل منهجية هاته المقاربة كالتالي:

- المرحلة الأولى لهاته المقاربة تهم دائما البحث عن جذع الكلمة المدخلة بالاعتماد على قواعد إزالة السوابق واللواحق للحصول على الجذع الصحيح، بعدها يتم البحث عن هذا الجدع في قائمة المعجم الخاص بالليمات المذكورة سابقا. فإذا وجد الجدع فسيعرف انتسابه إلى أحد ليمات الجدول، أما إذا لم يكن موجودا فيتم مقارنة وزنه مع أوزان جموع التكسير.

هذه المقاربة لا تعاني كباقي المقاربات السابقة من الخطأ في تحديد جذر الكلمة، ولكنها تتطلب الوقت والمجهود الأكبر في جمع أكبر عدد من الكلمات في صيغة المفرد، وتحديد الانتساب الاشتقاقي الذي له علاقة بهذا الاسم المفرد. وتعتبر اهم نقاط ضعف هاته المقاربة هي المجهود المبذول فيجمع أكبر عدد من الكلمات المفردة التي تنتمي إلى أحد أوزان جموع التكسير، والتمييز بينها وبين الكلمات التي ليست جمع التكسير، ولها أوزان هذا الجمع.

يرجع السبب الرئيسي في صعوبة معالجة جمع التكسير بالإضافة إلى بعض الظواهر اللغوية، إلى طبيعة اللغة العربية باعتبارها لغة اشتقاقية، وتحتاج إلى تشكيل الكلمات.

بعدما تعرفنا على المقاربات التي عالجت جمع التكسير، سنعرض مقاربات إزالة اللبس في جموع التكسير وهي ثلاثة:

- الآلية اللسانية l'approche linguistique
- الآلية الإحصائية l'approche statistique
- الآلية الهجينة l'approche hybride

قبل التفصيل في هاته المقاربات لا بد من تعريف مفهوم اللبس، حيث يقصد باللبس الغموض الذي يقع في الكلام نتيجة مجموعة من الظواهر اللغوية الصرفية والتركيبية.

و قد قسم اللغويون اللبس تبعا للمستوى اللغوي الذي ينتمي إليه إلى الأنواع التالية:

– **اللبس الصوتي**: ويقصد به الغموض الناشئ عن احتمال نطق أكثر من حركة صوتية في النص المكتوب على الحرف الواحد في الكلمة، كاحتمال الكسر والفتح مثل: أديت(ت)ما عليك(ك)والعامل في هذا الغموض هو غياب التشكيل بالحركات القصيرة بالإضافة إلى ذلك اتساع السياق لقبول الاحتماليين.

– **اللبس الصرفي**: ويقصد به دلالة الكلمة على أكثر من معنى ووظيفية صرفية، كدلالة بعض المشتقات على معنى الوصفية والاسمية معا مثل: عادل –صادق–حازم التي يمكن أن تكون من الصفات أو من الأسماء أو جموع التكسير كاحتمالية دلالة الضمير غير المشكول على معنى المخاطب أو المتكلم.

– **اللبس المعجمي**: ويقصد به احتمالية دلالة الكلمة على أكثر من معنى مثلا **'أعين'** تدل على مجموعة من المعاني: أعين الإنسان، أعين البئار، أعين الجواسيس...

– وهناك أيضا اللبس النحوي ويقصد به ارتباط البنية السطحية بتمثيليين دلاليين أو أكثر بسبب تركيبي.

من بين الأعمال التي اشتغلت على إزالة اللبس للتعرف على جمع التكسير :

تليليporte 1999 ; Freeman 2001, Al shamsi 2006, Amardeilh 200T, Tlili

.بالإضافة إلى أعمال أخرى، تستعمل هاته الأعمال الآلية اللسانية أو الآلية الإحصائية أو الهجينة

ونبدأ بالحديث عن المقاربة اللسانية مع سرد لأهم الأعمال التي تبنت هاته المقاربة.

1-**الآلية اللسانية**: l'approche linguistique: تعتبر المقاربة اللسانية من أقدم المقاربات التي عالجت مشكلة اللبس الصرفي وهي تعتمد على الجانب اللساني، الذي يتمحور حول إنشاء لائحة من القواعد ترفع اللبس، هاته القواعد يمكن تقسيمها إلى أربعة أقسام: نحوي، بنيوي، دلالي، ومنطقي. رغم أن القيود النحوية(القواعد النحوية)هي الأكثر استعمالا لإزالة اللبس مع أن استعمالها وحدها أحيانا يعاني من الضعف لذلك لابد من إضافة القواعد الدلالية والمنطقيةو كل أنواع القيود السابقة تنقسم إلى ثلاثة أنواع(ألوتي2005),(شانود1995)وهي:

-**القواعد السياقية**: وهي قواعد تقييدية، مثلا 'بعد حرف جر يأتي اسم مجرور.

-**القواعد الاستدلالية**: وهي قواعد ليست عامة، وليست صحيحة دائما، مثلا بعد كل فعل نجد أسم فاعل.

-**القواعد غير السياقية**: تستعمل هاته القواعد لإزالة اللبس التي لم تعالج بالقواعد السياقية والاستدلالية.

ومن بين الأعمال التي استعملت المقاربة اللسانية نجد: Brill; 2001 Freeman (La porte)Elag; 1995 Karlsson et al, 1992

إن إعداد مختلف القواعد لإزالة اللبس لجمع التكسير في هاته المقاربة اللسانية يحتاج للوقت والجهد الأكبر مما جعل الباحثين يفكرون في مقاربات أخرى وهي: المقاربة الإحصائية والمقاربة الهجينة.

2 الآلية الإحصائية: l'approche statistique: إن المقاربة الإحصائية من بين المقاربات التي غالبا ما استعملت لإزالة اللبس في جمع التكسير، ويعتبر العامل الاحتمالي والإحصائي في هذه المقاربة هو الذي يقود الاختيار، ولكي نحدد القيود الإحصائية والاحتمالية يجب البحث في متن لساني لتقييم نسبة ورود كل قيد(معجمي، صرفي، تركيبي، دلالي...)ومقارنة مع القيود الأخرى. وتخزن هاته القيود على شكل سمات رئيسية. والألية الإحصائية آلية تعتمد على نسبة ورود الكلمات والمورفيمات وأصناف الكلمات واحتمالات التآلف بين السوابق واللواحق.

ما لوحظ أن المقاربة الإحصائية مقاربة مكلفة لأنها تحتاج إلى متون نصية ضخمة ,تغطي جميع الظواهر اللغوية ,والنقص فيها أنها استندت على إحصائيات أخذت من متون نصية فقط أي اقتصرت على بعض الظواهر اللغوية فقط وليس قياسا على كل السياقات بصفة عامة وهذا ما بين قصورها مما جعل مقاربة أخرى تظهر هي: المقاربة الهجينة وهي مقاربة جامعة للمقاربتين السابقتين: المقاربة اللسانية–المقاربة الإحصائية وسنفصل في هاته المقاربة في المحور القادم.

3/ الآلية الهجينة(L'approche hybride): تعتبر المقاربة الهجينة محور أغلب الأعمال والبحوث في السنوات الأخيرة حيث استغلت فيها إيجابيات كل من المقاربتين اللسانية والإحصائية، حيث إن أغلب لإزالة اللبس في جمع التكسير وتطبيقاتها في الأنظمة التجارية، تتضمن تقنيات إحصائية ولسانية، للرفع من أداء هاته المقاربة.

ومن بين الأعمال التي اعتمدت المقاربة الهجينة على اللغة العربية، نجد(محمد الحاج 2009)، وقد اعتمد هذه المقاربة وطبقها على القران الكريم، كذلك نجد(خوجا2003 ,2001)، في نظام الواسم الآلي APT (Arabic Part of Speech Tagger)

بعد ما تعرفنا على مختلف المقاربات التي عالجت جمع التكسير، سنتطرق الى بعض انظمة معالجة جمع التكسير.

– بعض الأنظمة التي عالجت جمع التكسير آليا: تعتبر الأنظمة التي عالجت جمع التكسير بصورة صحيحة قليلة جدا، وأغلب هاته الأنظمة اعتمدت إعداد معجم يظم أغلب جموع التكسير(مقاربة المعجم)، ومن بين الأنظمة العربية التي عالجت جمع التكسير نعرض:

1/ نظام مورف: (Le système Morphe): اعتمد نظام 'مورف' (ليفيت1994)، (سودي وآخرون)، على استعمال قواعد تحويلية صرفية لمعالجة جمع التكسير. ويستغل حاليا النظام كمولد صرفي هذا ما جعل النظام يتطور(كفاليسفورزا2001)ويعتمد مقاربة المعجم.

يقدم نظام 'مورف' نموذجا لتحويل كلمة 'رجل' المفرد، إلى الجمع، مع الأخذ بعين الاعتبار بكل الاحتمالات الممكنة لجمع الكلمة المدخلة من طرف النظام والحصول على الجمع الصحيح .

و يضيف 'نظام مورف' السوابق الضرورية بحسب خصوصيات كل اسم (معرف- غيرمعرف، مبتدأ، خبر، نعت....)مدخل من طرف النظام، وبهذا تحول 'رجل' إلى 'رجال'، وبهذا أضاف النظام التنوين' للكلمة في الجمع.

1/ نظام أراجن(Le systèmeAragen)**:** يعد نظام أراجن (حباش 2004)، نظام تعميم صرفي بالنسبة للغة العربية، وقد أعطى تغطية شاملة تركزت على مفهوم اللكسيم، ويعتمد النظام على قاعدة المعطيات للمحلل الصرفي 'باك والترBuckwallter' مع إضافة محرك مختلف يعتمد على مهمة التوليد والتحليل. ويتكون 'أراجن' من معطيات صرفية ل'باك والتر'و ذلك للحصول على ترسيمة مدعمة بمفهوم 'اللكسيم' مع سماته.

وتضم الكلمة المدخلة في نظام 'أراجن ' مجموعة من السمات'feature set'، بالإضافة إلى اللكسيمات والسمات انطلاقا من قسم محدد من مجموعة من الضمائر الصرفية، هذه السمات تضم العدد، النوع وباقي المحددات الصرفية الأخرى ، كذلك تضم السوابق(مثل حروف الجر والعطف...)، والتي كتبت كجزء من كلمة، معجم 'باك والتر' استعمل في نظام 'أراجن' لفهرسة الكلمات المدخلة و الحصول على سماتها. وبالتالي الخروج 'بالسمات المفاتيح'، التي تستعمل في البحث عن السمات الأساسية في'feature set' للوحدة المعجمية المدخلة. وتوجد كثير من الوحدات المعجمية المدخلة التي تفتقد إلى معلومات خاصة بها، من بينها جمع التكسير الذي يفتقد إلى معلومات مهمة، خاصة أنه لا توجد علامات تبين أنه جمع، معالجة جذر لنوع من أنواع جمع التكسير يتطلب تضمين هذا الجذر مجموعة من سمات(الجمع)، التي يجب أن تكون في 'السمات المفاتيح' للجذر.

استعمال تعليقات باللغة الإنجليزية من بين الحلول التي وضعها النظام لمعالجة جمع التكسير، ولكن هذا الحل غير فعال، وذلك لأنها توجد صفات في اللغة العربية تعد جمع التكسير في الانجليزية لا توجد صفات في صيغة الجمع، كمثال على هذا نأخذ الصفة المفرد 'كبير' والصفة الجمع 'كبار'، كلاهما يعودان على الصفة'big' بالانجليزية، كذلك هناك أسماء في الجمع القياسي (جمع المذكر والمؤنث السالمين) دائما في صيغة الجمع ومن هنا طرق نظام 'باك والتر' غير فعال في معالجة جمع التكسير، لأن نظامه وقع في اللبس لأن هناك كلمات مدخلة ممكن أن تكون أسماء مفردة و جموع التكسير.

ولحل هذا المشكل ، الكلمة المدخلة في النظام إذا كانت مفردا فإنها ستوافق اللكسيم نفسه، أما إذا كانت جمعا فستعرض سلسلة من اللكسيمات الموافقة لجمع التكسير، واعتمد هو أيضا مقاربة المعجم.

2/ نظام مورف أراب: يعتمد نظام 'مورف أراب'، (عباس 2004) على قاموس 'DINAR'، هذا الأخير عبارة عن معجم عربي متخصص، وقاعدة معطيات. ويشارك النظام باقي المحللات الصرفية والمولدات في الاعتماد على 'DINAR'، ووظائفه في التقسيم وتحليل أشكال الكلمات. وتحتوي الموارد المعجمية للغة'DINAR'على:

19.457 مدخل فعلي، 70.702 أسماء أفعال، 39.099 جذور(التي فيها 9.565 جموع للتكسير)، 445 كلمة فارغة، وهي نموذج ل 1384 أسماء أعلام.

'DINAR' يربط كل وحدة معجمية بمجموعة من المحددات الصرفية والتركيبية ، تناسب علاقات نحوية معجمية للكلمة(حسون 1987) (ديشي1990)، هذه المحددات تشمل أيضا روابط اشتقاقية بين أشكال صرفية :الرابط الاشتقاقي بين الاسم المفرد وجمعه التكسير.

يشتغل المحلل 'مورف أراب' على ترجمة الاسم باللغة العربية، ويتعرف على السمات الصرفية التركيبية الموجودة في البنية للاسم المعالج، مثل تعدد الفعل ...و هو أيضا تبنى مقاربة المعجم.

3/ نظام المحلل الصرفي الخليل: يعد نظام 'برنامج الخليل الصرفي': نظام حاسوبي مفتوح المصدر، يقوم بتحليل الكلمات العربية إلى مكوناتها الأساسية كالجذر والجدع واللواحق، وتعريف كل منها بطريقة خوارزمية.

النظام الأول من نوعه من حيث كونه مفتوح المصدر بعد النظام الشهير 'باكوالتر' الواسع الاستخدام في مراكز العلوم ولكونه مفتوح المصدر، إلا أن النظام الحالي يزيد عليه في كونه قادر على إضافة مستوى أعمق من التحليل ، وذلك باستخراج الجذر والوزن من جدع الكلمة. وهو من إصدار مدينة الملك عبد العزيز للعلوم و التقنية بالتعاون مع المنظمة العربية للتربية والثقافة والعلوم وجامعة محمد الأول بالمغرب، واعتمد النظام مقاربة الأوزان.

خاتمة:

يتضح مما سبق من خلال هذه الدراسة، أن المقاربة الأنجع في التعرف الآلي لجمع التكسير هي مقاربة الأوزان، والآلية الهجينة الأكثر فعالية لإزالة اللبس في هذا النوع من الجمع.

المراجع

Abdelhadi Saudi -'Arabiccomputationalmorphology' – Antal Van den BoshGunterNeumann.thése de doctorat UniversityTunisia 2002.

- *Alexis AmidNemea, Éric Laporte a,b ''Pattern-and-rootinflectionalmorphology: the Arabic broken plural'b Université Paris-Est, Laboratoire d'informatique Gaspard-Monge CNRS UMR 8049, F77454, France.*
- *Ellouze Samira -'Etude et analyse du pluriel brisé arabe avec la plateforme Nooj' Mémoire de Master-2010.*
- *-Goweder et al…(2008)-'Arabic Broken Plural Recognition using a machine translation technique'dans les actes de la conference'The International Arab Conference On Information Technology'Tunusie ,2008.*
- *Goweder A .A Almerhag I Ennakoa A ''Arabic broken plural in Unvowelised Arabic Text'' in Procedeedings of EMNLP; Barcelona Spainm 2004.*
- *George Kiraz'(1996)' Analysis of the Arabic Broken Plurals and Diminutive In proceding of the 5th.International Conference and Exhibition on multi-lingual comuting Cambridge' university of Cambridge(s-t John's college) Computer Labratory Prembroke street cb2Itp*
- *Habash(2004)'Large Scale Lexeme Based Arabic Morphological Generation'In proceedings of traitement Automatique du Langage Naturel(TALN-04)Fez-Morocco-2004*
- *Hadad Youssef' Pseudometathesis in three Standart Arabic Broken Plural Templates'*
- *McCarthy and Prince(1990) 'Foot and Word in Prosodic Morphology:The Arabic Broken Plural' Natural Language and Linguistic''Theory 8 pp 209-282,1990.*
- *Mesafer 2008'-Analyse Morphosyntaxique Automatique et Reconnaissance des Entités Nommées En Arabe Standard' Thése.Ecole Doctorale « Langages.Espaces.Temps.Sociétés » Paris 2008*
- *Vicente Melgoza , Amanda Pogue, David Barner ' ' A Broken Fork in the Hand is Worth Two in the Grammar: A Spatio-Temporal*

- *Bias in Children's Interpretation of Quantifiers and Plural Nouns'), University of Toronto, Department of Psychology, 100 St. George Street, Toronto, ON, Canada, M5S 3G3*
- *Alexis Amid Nemea, Éric Laporte a,b ''Pattern-and-root inflectional morphology: the Arabic broken plural'b Université Paris-Est, Laboratoire d'informatique Gaspard-Monge CNRS UMR 8049, F77454, France*
- *Universidade Federal do Espírito Santo, Departamento de Línguas e Letras, Vitória, ES, Brazi*

wright (1971) 'A Grammar of the Arabic Language' Cambridge, University Press, Cambridge, 1971.