استخدام الحاسوب في تصنيف الكلمة العربية غير المشكولة داخل النص اعتماداً على التحليل اللفظي

Applying Computerized Word-Class Tagging to Non vocalized Arabic Texts Based on Morphological Analysis

إعداد

ساري محمد إسماعيل عوّاد

المشرف الرئيسي

الدكتور **سليمان حسين مصطفى**

المشرف المشارك

الدكتور **محمود الديكي**

أعضاء لجنة المناقشة

التوقيع

قدمت هذه الرسالة استكمالاً لمتطلبات الحصول على درجة الماجستير في علم الحاسوب من كلية الأمير الحسين بن عبدالله لتكنولوجيا المعلومات في جامعة آل البيت.

نوقشت وأوصي بإجازتها بتاريخ: ٢٠٠٥/١/١٨

الإهداء

أهدي هذا العمل...

إلـــــى وطني الحبيب الأردن.

إلـــــى جامعة آل البيت، وأخص بالذكر كلية الأمير الحسين بن علي لتكنولوجيا المعلومات.

إلـــــى والدتي التي مازالت مشاعرها متدفقة بالدعاء.

إلـــــى زوجتي وابنتي، وإلى كل من شجعني على اكمال دراستي.

إلـــــى كل غيور على اللغة العربية.

إلـــــى كل طموح بارتقاء الأمة العربية وتطورها وإلى كل من بذل جهـــدا فـــي ســـبيل نهضتها.

الشكر

أوجه شكري الجزيل وتقديري إلى الدكتور سليمان مصطفى المشرف الرئيسي على الرسالة، لما بذله من جهد خالص وعطاء ونصيحة، حتى اكتملت على الوجه المرجو واللائق، وشكري موصولٌ أيضاً إلى الدكتور محمود الديكي المشرف المشارك على الرسالة، لما بذله من إرشادات ونصائح لسانية ولغوية ومتابعته المتواصلة لي.

كما أشكر الدكتور إسماعيل عبابنة على إرشاداته ونصائحه الهامّة خلال دراستي في الجامعة، وأشكر العاملين في مجمع اللغة العربية الأردني على ما أبدوه من مساعدة ونصيحة، وأشكر كل من ساهم من قريب أو بعيد في ظهور هذه الأطروحة على هذا الشكل، كما أشكر إدارة جامعة آل البيت لإتاحتها الفرصة لي بالموافقة على دراسة موضوع الأطروحة، وعلى ما قدمته من تسهيلات في هذا الصدد، وكل هذا كان من شأنه أن يدفعني قدما للمثابرة والجد في طلب العلم، وأشكر الجامعة الهاشمية التي ابتعثتني لإكمال دراستي في جامعة آل البيت.

قائمة المحتويات

الموضوع	الصفحة
• الإهداء.....	ب
• الشكر.....	ج
• قائمة المحتويات.....	د
• فهرس الجداول.....	و
• فهرس الأشكال.....	ز
• فهرس الملاحق.....	ح
• الملخص بالعربية.....	ط
• **الفصل الأول: موضوع الدراسة وأهدافها وأهميتها**.....	١
١:١ مقدمة.....	٢
٢:١ موضوع الدراسة.....	٤
٣:١ أهداف الدراسة.....	٦
٤:١ أهمية الدراسة.....	٧
٥:١ خلاصة.....	٩
• **الفصل الثاني: الدراسات والأعمال السابقة ذات الصلة**.....	١٠
١:٢ دراسات تتصل باللغات غير العربية.....	١١
٢:٢ الدراسات المتصلة باللغة العربية.....	١٢
أ.الدراسات التي اهتمت بالتحليل اللفظي.....	١٣
ب.الدراسات التي اهتمت بتصنيف المفردات.....	١٤
٣:٢ الأنظمة التي اهتمت بتصنيف المفردات.....	١٧
٤:٢ خلاصة.....	١٨
• **الفصل الثالث: الزوائد اللفظية ومضامينها الدلالية**.....	١٩
١:٣ البنية التركيبية للكلمة العربية.....	٢٠
٢:٣ الزوائد النحوية.....	٢٢
١,٢:٣ السوابق النحوية.....	٢٢
٢,٢:٣ اللواحق النحوية.....	٢٥

٣,٢:٣ علاقة السوابق النحوية باللواحق النحوية..... ٢٧

٣:٣ الزوائد الصرفية..... ٢٨

١,٣:٣ السوابق الصرفية..... ٢٨

٢,٣:٣ اللواحق الصرفية..... ٢٩

٣,٣:٣ الوسطيات الصرفية..... ٣٠

٤,٣:٣ الأوزان الصرفية..... ٣١

٤:٣ علاقة الزوائد النحوية بالزوائد الصرفية..... ٣١

٥:٣ خلاصة..... ٣٤

- **الفصل الرابع: منهجية التحليل اللفظي لتصنيف المفردات**..... ٣٥

١:٤ الإطار العام لأسلوب التحليل اللفظي..... ٣٦

٢:٤ القوائم المستخدمة في التحليل..... ٣٨

٣:٤ عملية التحليل اللفظي..... ٤١

٤:٤ مثال على عملية التحليل..... ٤٥

٥:٤ خلاصة..... ٤٧

- **الفصل الخامس: تطبيق المنهج وتحليل النتائج**..... ٤٨

١:٥ عينة الدراسة..... ٤٩

٢:٥ تحليل نتائج الدراسة..... ٥٢

أ.تصنيف المفردات..... ٥٣

ب.دور الزوائد المختلفة في النتائج..... ٥٧

١. الأخطاء المتصلة بالزوائد النحوية..... ٥٨

٢. الأخطاء المتصلة بالزوائد الصرفية..... ٥٩

٣:٥ مناقشة النتائج..... ٦٠

٤:٥ خلاصة..... ٦٢

- قائمة المراجع..... ٦٣
- الملاحق..... ٦٦
- الملخص باللغة الإنجليزية..... ١٠٩

قائمة الجداول

رقم الجدول	العنوان	صفحة
١	حروف وأدوات السوابق النحوية	٢٣
٢	السوابق النحوية الثنائية	٢٣
٣	السوابق النحوية الثلاثية	٢٤
٤	أهمية السوابق النحوية في تحديد نوع الكلمة	٢٤
٥	حروف وأدوات اللواحق النحوية	٢٥
٦	مجموعة اللواحق النحوية المركبة	٢٦
٧	أهمية اللواحق النحوية في تحديد نوع الكلمة	٢٧
٨	العلاقة بين السوابق النحوية واللواحق النحوية	٢٧
٩	أهمية السوابق الصرفية في تحديد نوع الكلمة	٢٩
١٠	أهمية اللواحق الصرفية في تحديد نوع الكلمة	٣٠
١١	أهمية الوسطيات الصرفية في تحديد نوع الكلمة	٣١
١٢	علاقة السوابق النحوية بالسوابق الصرفية	٣٢
١٣	علاقة اللواحق الصرفية باللواحق النحوية	٣٢
١٤	أهمية العلاقة بين السوابق النحوية والسوابق الصرفية في تحديد نوع الكلمة	٣٣
١٥	أهمية العلاقة بين اللواحق الصرفية واللواحق النحوية في تحديد نوع الكلمة	٣٣
١٦	توزيع لأطوال المفردات الواردة في النص وأنواع المفردات من كل فئة	٤٩
١٧	توزيع مفردات عينة الدراسة من حيث النوع	٥١
١٨	توزيع مفردات عينة الدراسة من حيث وجود زوائد نحوية وصرفية	٥٢
١٩	تحليل نتائج البرنامج في تحديد أنواع المفردات الرئيسية	٥٣
٢٠	تحليل نتائج البرنامج في تحديد الفعل	٥٤
٢١	تحليل نتائج البرنامج في تحديد الاسم	٥٥
٢٢	تفصيل نتائج البرنامج الصحيحة التي تعطي أكثر من احتمال	٥٦
٢٣	أنواع الزوائد التي تم الاعتماد عليها آلياً في تحديد نوع الكلمة	٥٧
٢٤	تحليل الأخطاء التي وقع فيها البرنامج عند تحديد السوابق النحوية	٥٨
٢٥	تحليل الأخطاء التي وقع فيها البرنامج عند تحديد اللواحق النحوية	٥٩
٢٦	النتائج المتصلة بتحديد الزوائد والأوزان الصرفية	٦٠
٢٧	عينة من الأخطاء أثناء تطبيق الخوارزمية	٦١

قائمة الأشكال

رقم الشكل	العنوان	صفحة
١	العلاقة بين مختلف مستويات المعالجة الآلية	٥
٢	الإطار العام لمعالجة الصرف العربي آلياً	٦
٣	توضيح مستويات الكلمة في اللغة العربية	٢٠
٤	البنية الهيكلية للكلمة العربية الواردة في النص	٢١
٥	قواعد بناء الكلمات النصية العربية	٢٢
٦	الخطوط العريضة لمنهج المعالجة اللفظية	٣٨
٧	المعلومات التي توفرها قائمة السوابق	٣٩
٨	المعلومات التي توفرها قائمة اللواحق	٣٩
٩	المعلومات التي توفرها قائمة الأوزان الصرفية	٤٠
١٠	تمثيل قائمة الأدوات في منهجية الدراسة	٤٠
١١	قائمة تمثل علاقة السوابق باللواحق (صرفية ونحوية)	٤١
١٢	خطوات التحقق من أن الكلمة واحدة من الأدوات أو الضمائر	٤٢
١٣	خطوات التحليل باستخدام الزوائد النحوية	٤٣
١٤	توضيح خطوات التحليل باستخدام الزوائد الصرفية	٤٤
١٥	توضيح خطوات التحقق من العلاقات	٤٤
١٦	التحليل اليدوي لأطوال الكلمات الواردة في النص	٥٠
١٧	توزيع للزوائد المختلفة الواردة في مفردات النص	٥٢
١٨	توضيح نتيجة البرنامج في تمييز الأفعال	٥٤
١٩	توضيح نتيجة البرنامج في تمييز الأسماء	٥٦
٢٠	الزوائد المختلفة التي تم الاعتماد عليها في تصنيف المفردات أو في الحصول على معلومات أخرى	٥٧

قائمة الملاحق

رقم الملحق	العنوان	صفحة
١	قائمة الزوائد الصرفية المختلفة	٦٦
٢	قائمة الزوائد النحوية المختلفة وعلاقتها بالزوائد الصرفية	٦٨
٣	قائمة الأدوات والضمائر	٨١
٤	مثال على تنفيذ البرنامج وجزء من عينة النص	٨٥
٥	خوارزمية المنهجية	٨٧
٦	نص برنامج المنهجية	٨٩

الملخص بالعربية

يعد حقل اللسانيات الحاسوبية من أوسع حقول اللسانيات التطبيقية، فهو يعنى بمعالجة اللغات الطبيعية. تختلف اللغة العربية عن اللغات الطبيعية الأخرى بأبنيتها وتراكيبها، فهي لغة اشتقاقية وليست إلصاقية. وقد كانت الإنجليزية من أولى اللغات التي تم معالجتها بالحاسوب.

يرتكز موضوع هذه الدراسة على تصنيف الكلمة العربية غير المشكولة اعتمادا على التحليل اللفظي، حيث أن معالجة اللغة العربية آليا تمر بأربع مراحل هي: التحليل المعجمي فالصرفي فالنحوي فالدلالي، حيث تهدف هذه الدراسة إلى معرفة مدى الاعتماد على الزوائد اللفظية والصرفية في تحديد نوع الكلمة النصية.

تنطلق الدراسة من فرضية أساسية مفادها أنّ البنية التركيبية للكلمة العربية داخل النص تنطوي على مضامين دلالية يمكن الاعتماد عليها لدرجة كبيرة في تحديد نوعها وخصائصها الوظيفية ضمن النص العربي. وتكمن أهمية هذه الدراسة في أهمية التحليل الصرفي لكثير من التطبيقات الحاسوبية لمعالجة اللغة العربية، وهي توفر المساحة وترفع من الكفاءة بالتقليل من استخدام المعاجم.

يقوم منهج هذه الدراسة على أربعة محاور أساسية. أولها: التحقق من إمكانية أن تكون الكلمة التي ننوي معالجتها تنتمي إلى مجموعة الأدوات والضمائر في اللغة العربية. وثانيها: التحليل اللفظي القائم على استخدام الزوائد النحوية على اختلاف أنواعها (كحروف العطف، وحروف الجر وأل التعريف وغيرها مما ذكره النحاة). وثالثها: التحليل اللفظي القائم على استخدام الأوزان الصرفية، وفق ما تنطوي عليها من زوائد صرفية على اختلاف أشكالها فهناك أوزان خاصة بالأفعال، وأوزان خاصة بالأسماء، وأوزان خاصة بالصفات، وغيرها، وهناك أوزان مشتركة بين هذه الأنواع. وللقيام بهذا النوع من التحليل، كان لا بد من تحديد الأوزان الصرفية وحصر الزوائد الصرفية التي تدخل فيها وحصر العلاقات القائمة بينها وبين الزوائد النحوية التي تشترك معها في أوائل الكلمات أو أواخرها. ورابعها: الاعتماد على طول الكلمة في التحليل حيثما كان ذلك ضروريا. فهناك كلمات ثلاثية ورباعية، لا تتضمن أية زوائد لفظية، نحوية أو صرفية. وبذلك كان لابد من اللجوء إلى طول الكلمة لبيان الفئة التي يمكن أن تنتمي إليها من الأسماء أو الأفعال أو للتحقق من نتائج المعالجة اعتمادا على الزوائد اللفظية.

وقد أظهرت نتائج الدراسة أن المنهج التحليلي الذي استخدم في هذه الدراسة قد نجح في تحديد انتماء الغالبية العظمى من المفردات إلى فئتي الأسماء والأفعال. فقد كان البرنامج قادرا

على إدراك ٨٨,٤٦% من الأسماء و٨٤,١١% من الأفعال، ولكن هذه النسبة تقل عند تحديد النوع الدقيق للكلمة من الأسماء أو الأفعال حيث نجح البرنامج في تحديد النوع الدقيق للمفردات لما يقارب ٧٩% من المفردات التي يضمها النص. ومن هنا نستنتج أن بنية الكلمة داخل النص العربي غير كافية في بعض الحالات لتحديد نوع الكلمة. وذلك لعدّة أسباب، أولاً الاعتماد على زوائد في الكلمة لتحديد نوعها وهي من أصل الكلمة، وكان لهذا السبب ما يقارب فيما نسبته ٧,٧٥% من المفردات التي أخطأ البرنامج في تحديد نوعها الصحيح، ثانياً خلو الكلمات من الزوائد النحوية والصرفية والتي شكلت ١١,٤% من مفردات النص، فلا يوجد ما يدل في بنية هذه المفردات ما يدلّ على نوعها، ثالثاً احتواء الكلمة على زوائد نحوية لا تحدد نوع الكلمة، فهي تلتصق بالأسماء والأفعال، وجود أوزان تخلط بين نوعين من المفردات.

الفصل الأول

موضوع الدراسة وأهدافها وأهميتها

١. مقدمة

٢. موضوع الدراسة

٣. أهداف الدراسة

٤. أهمية الدراسة

٥. خلاصة

موضوع الدراسة وأهدافها وأهميتها

يهدف هذا الفصل التمهيدي إلى التعريف بموضوع الدراسة الحالية ضمن إطار المجال العلمي الذي تنتمي إليه وتوضيح أهدافها وإبراز أهميتها العلمية. يبدأ الفصل بمقدمة عامة تعرض أهم خصائص ومشكلات معالجة اللغة العربية. ثم يحدد القسم الثاني موضوع الدراسة وعلاقة ذلك بالإطار العلمي الذي تقع ضمن حدوده فيما يتصل بالجوانب المختلفة لمعالجة اللغة العربية. أما القسم الثالث، فيعالج أهداف الدراسة وتحديد الأسئلة العلمية التي تسعى للإجابة عليها. وفي الجزء الأخير من الفصل بيان لأهمية الدراسة في ضوء الأهداف التي تسعى إليها والمنهجية العلمية التي تقوم عليها.

١. مقدمة

منذ ظهور الحاسوب تنبه غير واحد من الباحثين في مجالات اللسانيات وعلوم الحاسوب إلى أهمية العلاقة التي يمكن أن تنشأ بين الدراسات اللغوية والدراسات الحاسوبية، وبخاصة في مجال الذكاء الاصطناعي. وقد كانت الترجمة الآلية من أولى المحاولات في هذا الباب. وقد أدى الاهتمام المتزايد في تطوير هذه العلاقة إلى بروز علم جديد يهتم بالنظريات والأساليب المتصلة بالتطبيقات الحاسوبية في مجالات تحليل اللغة، والترجمة، وبناء المعاجم، ومعالجة النصوص، وتخزين النصوص العلمية واسترجاعها وغير ذلك، فيما يعرف الآن اصطلاحا باسم "معالجة اللغات الطبيعية".

تعد معالجة اللغة العربية آليا بوساطة الحاسوب أحد المقومات الرئيسة لإعداد اللغة العربية لكي تلحق بعصر المعلومات، حيث يتوقف على هذه المعالجة مدى نجاح جهود تطوير برامج عربية في العديد من التطبيقات العلمية والتعليمية والمكتبية والثقافية وغيرها، وفي تطوير أساليب البحث اللغوي وتحديث البنية الأساسية للغة العربية.

تختلف اللغة العربية عن اللغات الطبيعية الأخرى، كالإنجليزية وغيرها، بتراكيبها وأبنيتها الصرفية ودلالاتها. فالغالبية العظمى من كلمات اللغة العربية تُبنى من خلال جذور الكلمات بإتباع الأوزان وإضافة السوابق واللواحق للكلمة النصية وفق قواعد صرفية ونحوية محددة. فهي تتسم بخصائص عديدة تجعل من أمور معالجتها آليا بوساطة الحاسوب موضوعا مثيرا للاهتمام سواء على المستوى اللغوي أو المستوى الحاسوبي.

ولكن من المهم، من جانب آخر، الإشارة إلى أن العلاقة الوطيدة التي تربط بين بنية الكلمة العربية وصرفها ومعناها وإعرابها، تفرض على الباحث، في تناوله لهذه القضية، أن يضع نصب عينيه علاقات التداخل المتشابكة التي تربط بين الفروع اللغوية المختلفة داخل منظومة اللغة العربية. وبينما تكون هذه العلاقات أحيانا عونا للباحث في تحديد أساليب ومناهج المعالجة الحسابية للغة، تفرض في بعض الحالات قيودا يصعب تجاوزها أو التغلب عليها دون اللجوء إلى بعض الأساليب الخاصة.

ومنذ ظهور الحاسوب اهتم الباحثون بمعالجة اللغة الإنجليزية آلياً، وذلك بهدف خدمة الأبحاث اللغوية، وكانت بداياتهم متعثرة، ولكن ما إن جاءت الثمانينات حتى كانت الأبحاث قد تطورت في هذا المجال ووصلت إلى مستويات متقدمة. أمّا بالنسبة للغة العربية فقد كان الأمر مختلفا. إذ تأخرت الأبحاث التي تعنى بمعالجة اللغة العربية آلياً إلى بداية الثمانينات حيث بدأت أولى الخطوات الأساسية على هذا الطريق.

"وقد تركزت الجهود الأولى في مجال اللغة العربية على إمكانية إدخال المعلومات وطباعتها بحروف عربية، وما أن أخذت هذه المشكلة طريقها إلى الحل حتى صارت البحوث والدراسات العلمية تتوجه نحو معالجة النصوص العربية، وتخزينها واسترجاعها بالحاسوب. وبذلك بدأت عملية المعاجم المحوسبة، والترجمة الآلية، وبناء قواعد البيانات، ومعالجة الكلمات والفهرسة الآلية...وغيرها" (مصطفى،١٩٩٦).

ولو اطلعنا على الأبحاث التي درست معالجة اللغة العربية لوجدناها قاصرة في كثير من الجوانب مقارنة مع الأبحاث التي درست اللغات الأخرى مثل الإنجليزية، إذ "نجد الأبحاث التي درست معالجة اللغة العربية إما أن تكون عامة أو أن تكون في جزئية محددة من اللغة".(معتصم،٢٠٠٢). ولعل مرد ذلك هو الصعوبات التي تنطوي عليها معالجة اللغة العربية. فبالإضافة إلى حداثة الموضوع وقلة البحوث العلمية فيه، هناك العديد من المشكلات والتحديات التي تواجه العاملين في مجال حوسبة اللغة العربية تتصل بخصائص العربية وسماتها التي تختلف عن اللغات الأخرى. وقد أوجز علي(١٩٨٨) أهم هذه المشكلات والتحديات بالنقاط الآتية:

أ. تعدّد حالات اللبس وتداخلها (كاللبس الصرفي واللبس المعجمي) التي تعدّ من أكبر المشاكل التي تواجه التحليل الصرفي.

ب. إغفال علامات الضبط في معظم النصوص العربية، وقصور المعجم العربي من حيث المعطيات اللغوية (الصوتية والنحوية والصرفية والدلالية) التي تعترض المعالجة الآلية في مختلف المستويات.

ج. الصعوبات المتعلقة بالأخطاء التطبيقية لاستخدام اللغة العربية كالأخطاء النحوية والإملائية.

وقد وجه الكثير من الباحثين اهتمامهم إلى هذه المشكلات وحاولوا إبراز أهم الأوجه التي يمكن إحراز تقدم ملموس فيها. حيث شهدت السنوات العشرة الأخيرة، بشكل خاص، نشاطا علميا ملحوظا في هذا المجال وكانت اللغة العربية محور اهتمام العديد من الدراسات الأكاديمية والبحوث العلمية في الدوريات والمؤتمرات الإقليمية والعالمية. فعلى الرغم من الصعوبات المذكورة أعلاه، فاللغة العربية مبنية على ميزان صرفي اشتقاقي يعطيها طابعا حوسبيا خاصا يميزها عن الكثير من اللغات. ولعل هذا يفسر توجيه الكثير من هذه الجهود العلمية إلى الجانب الصرفي والقليل منها في باب التحليل النحوي، ويندر وجود دراسة في باب التحليل الدلالي.

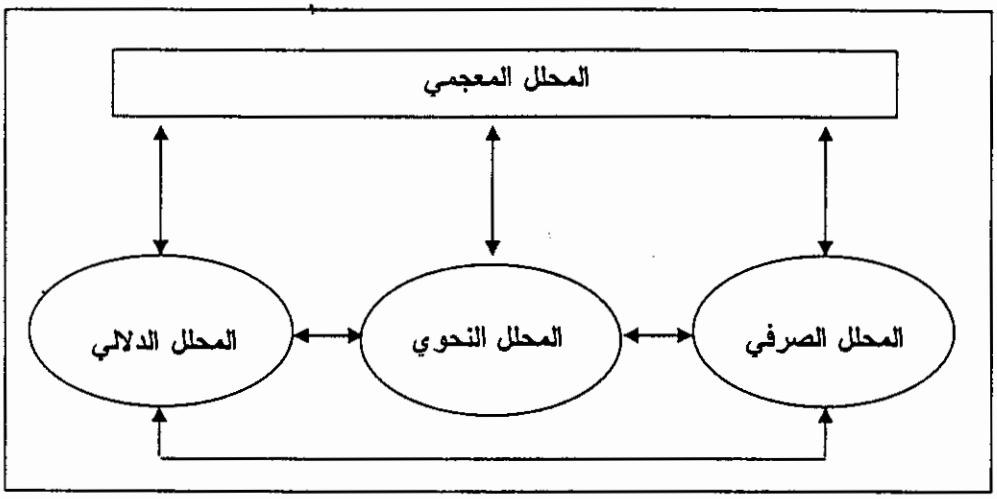
٢. موضوع الدراسة

يتركز اهتمام هذه الدراسة على تحديد نوع الكلمة العربية غير المشكولة داخل النص اعتماداً على التحليل اللفظي للكلمة النصية، وهو موضوع يتصل بأحد المحاور المهمة التي تقوم عليها المعالجة الآلية للنصوص المكتوبة باللغة العربية، ألا وهو محور التحليل الصرفي. فمعالجة اللغة العربية بالحاسوب يمكن أن تمرّ بأربعة مستويات أساسية، على النحو الموضح في الشكل (١)، وهي: التحليل المعجمي، والتحليل الصرفي، والتحليل النحوي، والتحليل الدلالي. وبينما يعالج المستويان الأول والثاني مفردات اللغة فإن الآخرين يركزان على تركيب الجملة ودلالاتها.

فالتحليل المعجمي يقوم على وجود قاعدة بيانات تتركب من جزئيات تحتاجها المعالجة الآلية في مستوياتها الثلاثة الأخرى، كالمعلومات الصرفية والبنائية للكلمات المستخدمة في اللغة. أمّا في التحليل الصرفي فتتم دراسة بنية الكلمة نفسها بغض النظر عن موقعها داخل الجملة. ويهدف التحليل الصرفي إلى ربط الكلمة بعناصرها الأولية التي تدخل في تكوينها. وللتحليل الصرفي وظائف متنوعة مثل إعادة الكلمات إلى جذورها، ومعرفة السوابق واللواحق

النحوية والصرفية في الكلمة، ودراسة وزن الكلمة، وبناء علاقات بين هذه الأجزاء (معتصم،٢٠٠٢).

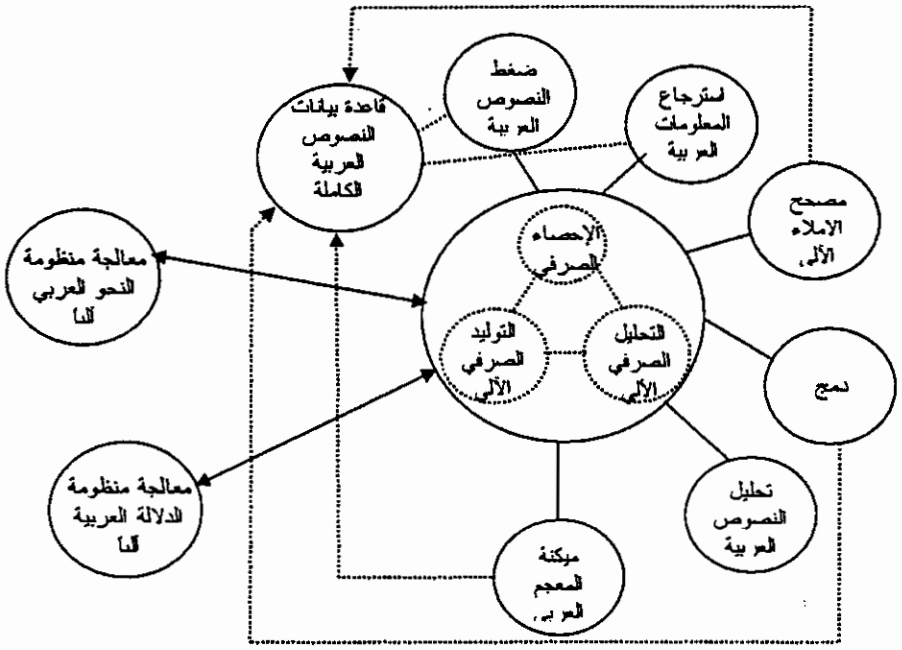
وفي المقابل فإن التحليل النحوي يهتم بتحديد الموقع الإعرابي للكلمة داخل الجملة، ويهتم هذا الجزء بتقسيمات الجملة الفعلية والاسمية، وبالترتيب النحوي لها. أما التحليل الدلالي فيُعنى بوجوه المعاني الممكنة التي يمكن أن تؤول إليها المفردات المكونة للجملة. ويمكن الاعتماد عليه في فك اللبس الحاصل بين بعض المفردات (معتصم، ٢٠٠٢).



الشكل (١): العلاقة بين مختلف مستويات المعالجة الآلية (مقتبس من معتصم، ٢٠٠٢، ص ٢٤)

ينصب اهتمام التحليل الصرفي الآلي، بشكل أساسي، على معالجة المفردات وتحديد مكوناتها الاشتقاقية والصرفية، وتحديد التغيرات التي تطرأ عليها لتعكس الخصائص النحوية والدلالية المختلفة، كالزمان، والجنس، والعدد، والمتكلم أو المخاطب، وغير ذلك مما تفرضه الضرورة في سياق النص.

ويمكن الإفادة من هذا التحليل من زاويتين (Khoja,1999): أولاهما من خلال الجذر وثانيتهما من خلال الزوائد. إذ يمكن استخدام الجذور الناتجة عن التحليل في الكثير من التطبيقات كضغط النصوص، والبحث في النصوص واسترجاعها، والتدقيق الإملائي، وبناء المعاجم، وتحليل النصوص، والشكل (٢) يوضح الإطار العام لمعالجة التحليل الصرفي آليا. أما الزوائد المستخلصة من التحليل فيمكن الإفادة منها في تحديد البناء النحوي للمفردات. وكما يتضح من أهداف الدراسة، التي سيرد ذكرها في القسم التالي، فإن الاهتمام في هذا البحث يتركز حول الجانب الثاني من فوائد التحليل الصرفي.



الشكل (٢): الإطار العام لمعالجة الصرف العربي آليا (مقتبس من: علي، ١٩٨٨، ص ٣٠١)

٣. أهداف الدراسة

أي نظام حاسوبي يقوم بمعالجة احدى اللغات الطبيعية يحتاج إلى الإلمام بكلمات تلك اللغة، من حيث بنية الكلمات وأنواعها (اسم، فعل، حرف...الخ)، وسماتها الأخرى، كالعدد (جمع، مفرد، مثنى)، والجنس (مذكر، مؤنث)، وبعض المعلومات الأخرى اللازمة في التحليل. وكما هو معلوم فإن البناء اللفظي لكلمات اللغة العربية يقوم على قواعد نحوية وصرفية محددة، ويحمل في طياته مضامين دلالية تتحدد من خلال الميزان الصرفي الذي تقوم عليه والزوائد اللفظية الأولية والنهائية التي تضاف إليه.

إن الغرض الأساسي من هذه الدراسة هو معرفة مدى إمكانية الاعتماد على بنية الكلمة النصية في تحديد نوعها دون اللجوء إلى معاجم وقوائم تضم مفردات اللغة (باستثناء الأدوات والضمائر)، ودون اللجوء إلى سياق النص لتحديد نوع الكلمة أو التأكد من نتائج التحليل. ومن هنا، فقد سعت الدراسة إلى توفير الإجابة على الأسئلة البحثية الثلاثة التالية:

١. إلى أي مدى يمكن الاعتماد على كل نوع من أنواع الزوائد اللفظية النحوية والدلالية في تحديد نوع الكلمة وخصائصها اللغوية ؟

٢. متى نحتاج إلى الرجوع إلى الميزان الصرفي والزوائد الصرفية لتحديد نوع الكلمة وبيان خصائصها وإلى أي مدى يمكن الاعتماد على ذلك؟

٣. متى يفشل المحلل اللفظي في تحديد المفردات اعتمادا على البنية اللفظية وما هي طبيعة المشكلات المصاحبة لذلك؟

وبناءً على ذلك تهدف الدراسة إلى تحقيق أمرين أساسيين هما: بناء نموذج تحليلي للكلمة العربية غير المشكولة وفق ورودها داخل النصوص العربية، واستخدام هذا النموذج في تحديد نوع الكلمة العربية غير المشكولة بما يتوافق مع التقسيمات المتبعة في تصنيف الكلمات في النصوص العربية.

فقد درج النحاة العرب القدماء على تقسيم الكلمة العربية إلى ثلاثة أنواع رئيسية هي: الأسماء وأفعال وحروف، وكل من هذه الأقسام تقسم على النحو الآتي:

١. تقسيم الفعل إلى ثلاثة أنواع من حيث الزمان هي الماضي، والمضارع، والأمر.
٢. تقسيم الاسم إلى عشرة أنواع هي: المصدر، والمصدر الصناعي، واسم الفاعل، واسم المفعول، واسم الآلة، والصفة المشبهة، واسم الزمان والمكان، والاسم المنسوب، واسم التفضيل، واسم العلم، والاسم العام(أسماء الجنس ، وبعض الأسماء المشتقة من الأوزان الصرفية).
٣. تم اعتبار أسماء الإشارة، والأسماء الموصولة، والظروف، والضمائر، وحروف المباني والمعاني، والخوالف وغيرها مما لا تنطبق عليها قواعد علم الصرف مجموعة واحدة، أطلق عليها مجموعة الأدوات والضمائر (انظر ملحق رقم ٣).

وبناء على ذلك، اعتمدت الدراسة الحالية على تقسيم المفردات إلى أفعال، وأسماء على النحو الوارد أعلاه. أما بالنسبة للأدوات والضمائر، فلم يتم التمييز بينها، حيث تم التعامل معها كمجموعة واحدة وتم استبعادها من التحليل اللفظي لأنها لا تخضع لقواعد الصرف.

٤. أهمية الدراسة

لا تزال معالجة اللغة العربية حاسوبياً تحتاج لمزيدٍ من الدراسات والأبحاث، وذلك لسعة اللغة العربية وكثرة قواعدها، ولاختلاف منهجية البحث وتشتتها في معالجتها، إضافة إلى عصر المعلوماتية الذي يرتكز على التبادل الكوني للمعلومات المختلفة. وهناك من الدراسات التي ركزت على تصنيف المفردات، وهذه الدراسات كانت في لغات مختلفة كالإنجليزية والألمانية والاسبانية وغيرها من اللغات، أما بالنسبة للدراسات التي ركزت على تصنيف المفردات في

اللغة العربية، فاتسمت باعتمادها على المعاجم بشكل رئيسي إضافة إلى سياق النص. وعليه تكمن أهمية الدراسة في ثلاثة جوانب رئيسية:

أولها: أنها تركز على جانب رئيسي من جوانب معالجة اللغة العربية بالحاسوب وهو التحليل الصرفي الذي تعتمد عليه جميع مستويات التحليل المحوسب. فإدراك نوع الكلمة أساس في تحليل الجمل من الناحيتين النحوية والدلالية. إذ تعدّ نتائج التحليل الصرفي مدخلات للتحليل النحوي والتحليل الدلالي، كما يتضح من الشكل (٢) الذي ورد ذكره سابقا. فالسوابق المتصلة ببداية الكلمة، مثلا، تبين للمحلل النحوي العطف والجر وغيرها من الأمور التي تساعد المحلل النحوي في تحديد الموقع الإعرابي للكلمة.

وثانيها: أنها أساس لكثير من التطبيقات الحاسوبية العربية مثل استرجاع المعلومات، وفهرسة المصطلحات والمفاهيم العلمية في الموسوعات والمصادر العلمية، وميكنة المعاجم، وفي البرامج التعليمية للغة العربية، وغيرها من التطبيقات التي بدأت تأخذ طريقها في مجال البرمجيات العربية.

وثالثها: أنها تبرز قضية هامة في أسلوب االتحليل اللفظي للمفردات العربية حيث يقوم في أساسه على العلاقات القائمة بين الزوائد النحوية والزوائد الصرفية، بدلاً من الاستعانة بمعاجم لغوية يصعب تخزينها، ويقلل الرجوع إليها من درجة كفاءة المحلل الصرفي. ولو ألقينا نظرة عامة على الدراسات السابقة في تصنيف مفردات النصوص العربية نجد أنها استندت في التحليل على المعاجم بشكل رئيسي لمقارنة النتائج بها. وعلى الرغم من أن ذلك يساعد في التحقق من صحة النتائج، إلا أنه قد يزيد من كلفة عملية التحليل بشكل ملموس قد لا يتناسب مع المكاسب التي نحققها من ذلك. وانطلاقا من ذلك، كان الافتراض في الدراسة الراهنة بأنه قد يكون من الممكن تحقيق نتائج مرضية دون اللجوء إلى فكرة المعاجم إلا في حدود ضيقة لا تتجاوز الأدوات لأنها لا تخضع لقواعد الاشتقاق. إذ تعتمد منهجية هذه الدراسة على مجموعة الزوائد النحوية والدلالية والصرفية، وما يقوم بينها من علاقات قد تكون في كثير من الأحيان كافية للاستدلال منها على تحديد أنواع المفردات. ولعل ذلك المنهج هو ما يميزها عن غيرها من الدراسات السابقة.

٥.خلاصة

- تختلف اللغة العربية عن اللغات الطبيعية الأخرى بأبنيتها وتراكيبها، فهي لغة اشتقاقية وليست إلصاقية. وكانت الإنجليزية من أولى اللغات التي تم معالجتها بالحاسوب، وتأخرت الأبحاث بالنسبة لمعالجة اللغة العربية آليا. وهناك صعوبات تواجه معالجة اللغة العربية آليا كحالات اللبس وتداخلها وإغفال علامات التشكيل في معظم النصوص العربية وغيرها من الصعوبات المتعلقة بالأخطاء التطبيقية كالنحوية والإملائية.
- يرتكز موضوع هذه الدراسة على تصنيف الكلمة العربية غير المشكولة اعتماداً على التحليل اللفظي، حيث أن معالجة اللغة العربية آليا تمر بأربع مراحل وهي التحليل المعجمي والصرفي والنحوي والدلالي.
- تهدف هذه الدراسة إلى التعرف على مدى الاعتماد على الزوائد اللفظية والصرفية في تحديد نوع الكلمة النصية ومتى يفشل ذلك، وتم اعتماد التصنيف المذكور سابقاً للاسم والفعل.
- تكمن أهمية هذه الدراسة في أهمية التحليل الصرفي لكثير من التطبيقات الحاسوبية لمعالجة اللغة العربية، وتتميز بالتقليل من استخدام المعاجم.

الفصل الثاني

الدراسات والأعمال السابقة ذات الصلة

١. دراسات تتصل باللغات غير العربية

٢. الدراسات المتصلة باللغة العربية

أ.الدراسات التي اهتمت بالتحليل اللفظي

ب.الدراسات التي اهتمت بتصنيف المفردات

٣. الأنظمة التي اهتمت بتصنيف المفردات

٤. خلاصة

الدراسات والأعمال السابقة ذات الصلة

لقد حظيت اللغة العربية باهتمام العديد من الدراسات والتقارير العلمية في السنوات العشر الماضية. وتركز هذا الاهتمام بشكل خاص على التحليل اللفظي والصرفي لمفردات اللغة العربية. ولعل أهم ما يهمنا في هذا المقام تلك الدراسات والأعمال التي وجهت جهدها أو بعض جهدها إلى موضوع تصنيف المفردات. ولوضع هذه الدراسات في إطارها العلمي المناسب يبدأ الباحث بالإشارة إلى الدراسات التي تناولت هذا الموضوع في اللغات الأخرى غير العربية ثم يتناول بعد ذلك الدراسات والأعمال العلمية ذات الصلة المباشرة بالتحليل اللفظي للمفردات العربية وبتصنيف هذه المفردات. واستكمالا لذلك، ستتم الإشارة إلى بعض الجهود البرمجية العربية في هذا المجال.

١. دراسات تتصل باللغات غير العربية

هناك العديد من الدراسات التي عالجت موضوع التحليل اللفظي في الكثير من اللغات وعلى رأسها اللغة الإنجليزية، إلا أن ما يهمنا في هذا المقام تلك الجهود العلمية المتصلة بتصنيف المفردات في هذه اللغات بشكل خاص. ولعل من المفيد أن نذكر هنا بأن القواعد الصرفية لبناء المفردات في هذه اللغات لا تنسجم مع خصائص المفردات العربية مما يقيد من إمكانية تعميم الأساليب المتبعة في هذه الدراسات على التحليل اللفظي في اللغة العربية.

ومن أمثلة هذه الدراسات تلك التي قام بها إيكلوند (Eklund,1993) ، وتناول فيها تصنيف الكلمات الإنجليزية. وقد اعتمد في منهجيته على تخزين الكلمات وسماتها في معجم خاص لهذه الغاية، واستخدم ما سمي بأسلوب اليد اليسرى (left-stripping) عند عدم العثور على الكلمة في هذا المعجم، وهذا الأسلوب يعتمد على حذف أول حرف من جهة اليسار وبعد ذلك يقوم بالبحث عن بقية الكلمة داخل المعجم. وفي حالة التطابق يصنف الكلمة اعتماداً على المعلومات الواردة في المعجم. وفي حالة عدم العثور عليها يتم البحث في المعجم عن كلمات لها نفس تردد الكلمة. وإذا لم يجد كلمات لها نفس التردد يقوم بحذف آخر حرف بالكلمة للبحث عن بقيتها داخل المعجم. وقد أشارت نتائج الدراسة إلى أن هذا الأسلوب قد تمخض عن نجاح في تحديد نوع المفردات الإنجليزية تراوحت نسبته بين ٩٤% و ٩٩%.

أما يونج وزملاؤه (Yung et al.,1996) فقد استخدموا منهجية مختلفة عن تلك التي استخدمتها الدراسة السابقة لتحديد نوع الكلمة الإنجليزية داخل النص، وتقوم على ثلاث مراحل.

الأولى، إزالة الزوائد الملتصقة بنهاية الكلمة مثل "ing" و "ed" وغيرهما من الزوائد النهائية. فإذا لم يكن لهذه الحروف دلالة على نوع الكلمة، يتم الانتقال إلى المرحلة الثانية، وهي النظر في إمكانية تحديد نوع الكلمة بناءً على نوع الكلمة السابقة لها في النص. فإذا كانت الاحتمالية ضعيفة، يجري تطبيق المرحلة الثالثة، وهي البحث عن الكلمة في المعجم بعد إزالة الحروف الملتصقة بالكلمة. وفي حالة التطابق، يقوم بقراءة نوع الكلمة وخواصها من المعجم الذي تم إعداده لهذه الغاية. وقد استنتج الباحث أن هذه الطريقة تقدم نتائج تفوق في نسبة نجاحها تلك التي يمكن الحصول عليها من خلال الاعتماد على المعجم بشكل مباشر.

ومن جانب آخر، حاول جاكسون (Jackson, 2002) في الدراسة التي قام بها لتصنيف الكلمات في اللغة الألمانية، أن يستفيد من الجانب الصرفي بالإضافة إلى الاعتماد على المعجم والزوائد. وقد قام منهجه على ثلاث خطوات رئيسية. أولها، إزالة الزوائد من الكلمة، وثانيها، المقارنة بالكلمات المخزنة في المعجم دون الزوائد، وفي حال التطابق، يقوم بقراءة نوع وخواص الكلمة من المعجم. وفي حال فشل ذلك يقوم بالخطوة الثالثة، وهي اللجوء إلى وزن الكلمة حيث يبحث عن وزن الكلمة في قائمة الأوزان، وفي حال التطابق، يقوم بقراءة النوع والخواص الذي يدل عليه ذلك الوزن. وقد أشار الباحث إلى أنّ منهجه هذا قد حقق نجاحا في تحديد نوع الكلمات وصلت نسبته ٩٢,٤٧%.

وكما رأينا في دراسة يونج وزملاؤه، لجأ ترومر ودالينا(Trommer and Dalina, 1997)، في دراستهما التي ركزت على تصنيف الكلمات في اللغة الألبانية، إلى الاستعانة بسياق النص بالإضافة إلى المعلومات التي يمكن الحصول عليها من خلال بنية الكلمة. فقد اعتمد الباحثان في منهجهما على أسلوبين أساسيين لمعرفة نوع الكلمة. الأول، هو أسلوب التحليل اللفظي للكلمات المكونة للنص وعلى المعجم لمعرفة نوع كل كلمة. حيث قاما بتحليل بنية الكلمة وبيان زوائدها. فإذا لم يكن لهذه الزوائد دلالات على نوع الكلمة أو لم يتم العثور على الكلمة في المعجم، يتم في هذه الحالة الاستعانة بالأسلوب الثاني وهو الاعتماد على سياق الكلمة في النص لتحديد نوعها. وقد أشارت نتائج الدراسة إلى أن هذا المنهج كان ناجحا في تحديد ما نسبته ٨٣,٦% من مجموع كلمات النص الذي تم فحصه.

٢. الدراسات المتصلة باللغة العربية

لقد قامت دراسات التحليل الصرفي في اللغة العربية على عدة محاور منهجية. فمنها ما استند إلى بعض تقنيات الذكاء الاصطناعي مثل تقنية التعلم من الأمثلة (Darwish, 2002)، والشبكات الانتقالية (غيث ومشهور، ١٩٨٥)، ومن الدراسات التي جمعت بين التحليل اللفظي

والمعجمي، ما ارتكز على أسلوب التخزين الشامل لمفردات اللغة العربية أو جذوعها (Al-jlayl and Frieder, 2002). ومنها أيضا ما تبنى طريقة إجرائية تقوم على أساس تخزين القليل من المعلومات اللغوية، وتحليل الكلمات وفق خطوات متسلسلة تجرى على الكلمة تنتج تحليلاً صرفيا (Mustafa and Masoud, 2000).

وفيما يلي نحاول استعراض أهم الدراسات التي اهتمت بالتحليل اللفظي لغايات المعالجة اللغوية وليس لغايات تطبيقية مثل التحليل اللفظي الموجه لاسترجاع المعلومات مثلا. وستتم الإشارة أولا إلى أهم الدراسات العامة التي وجهت جهودها إلى تحديد الجذور وإيجاد الأوزان الصرفية، مع الاكتفاء في ذلك ببعض الدراسات التي تميزت بمنهجية خاصة. إذ ليس من أهداف هذه الدراسة استعراض جميع الجهود التي تناولت موضوع التحليل الصرفي. ثم تتلوها تلك الدراسات التي وجهت اهتمامها، أو جانبا من اهتمامها، إلى موضوع تصنيف المفردات.

أ. الدراسات التي اهتمت بالتحليل اللفظي

من أوائل الدراسات التي اهتمت بالتحليل اللفظي تلك الدراسة التي قام بها كل من الفداغي والعنزي (Al-Fedaghi and Al-Anzi, 1989). وقد بنيت هذه الدراسة على خوارزمية تقوم بتوليد الجذور والوزن الصرفي للكلمة العربية المعطاة استنادا إلى الأوزان الصرفية. حيث يتم تحديد مواقع الزوائد في الوزن الصرفي ومطابقتها مع ما يقابلها في الكلمة المعطاة. وفي حالة التطابق تتم إزالة الزوائد واستخلاص الجذر.

وفيما بعد قام الشلبي وإيفنس (Al-Shalabi and Evens, 1998) ببناء محلل صرفي، يهدف إلى إيجاد الجذور اعتمادا على إزالة الزوائد والتحقق من بقية الكلمة في معجم للجذور يضم معظم جذور المفردات العربية. يقوم المنهج الذي تم اتباعه في هذه الدراسة على ثلاثة خطوات: إزالة زوائد الكلمة، ومن ثم البحث عن الحرف الأول والثاني والثالث من الكلمة المتبقية في معجم الجذور، وإذا لم يتم العثور على ما يقابلها في المعجم يتم البحث عن الحرف الأول والثاني والرابع من الكلمة المتبقية في معجم الجذور، وإذا لم يتم العثور على ما يقابلها في المعجم يتم البحث عن الحرف الأول والثاني والخامس من الكلمة المتبقية في معجم الجذور، وإذا لم يتم العثور على ما يقابلها في المعجم يتم البحث عن الحرف الأول والثالث والرابع من الكلمة المتبقية في معجم الجذور. وقد تم تطبيق الخوارزمية على ١٩١٦٧ كلمة، ونجحت في

إخراج الجذور لها، باستثناء ١٢٦٨ كلمة قيل إمّا أنها أسماء جامدة مرتجلة أو أن أصلها أعجمي.

وخلافا لكل الدراسات السابقة التي اهتمت بالتحليل اللفظي، قدم كل من مصطفى ومسعود (Mustafa and Masoud, 2000) منهجا مغايرا يقوم على معالجة المفردات وتحليلها بطريقة عكسية، بدءا من نهاية الكلمة وليس من أولها. وقد انطلقت هذه الدراسة من الافتراض بأن نهايات كلمات اللغة العربية أقل عرضة للتغيير من أوائلها. وبناء على ذلك، يتم التحليل على مرحلتين: تختص الأولى بمعالجة الزوائد النهائية للكلمة، بينما تستخدم المرحلة الثانية، مجموعة من القواعد المشتقة من الأوزان الصرفية لتحديد جذر الكلمة. وقد أشار الباحثان إلى دراستهما قد نجحت بنسبة تزيد عن ٩٠% في استخلاص الجذور.

وكما هو الحال بالنسبة لدراسة الفداغي والعنزي المذكورة آنفا، قام أديورايز وآخرون (Adurize et al. 2000) ببناء وتصميم محلل لفظي، بهدف اشتقاق جميع الكلمات الممكنة من خلال الكلمة المعطاة. واعتمد في منهجيته على إيجاد الوزن الصرفي للكلمة، وفي حال التطابق، يقوم باشتقاق الكلمات الممكنة من هذه الكلمة من خلال الأوزان الصرفية المشتقة من الوزن الصرفي للكلمة. حيث تكون هذه الأوزان مخزنة ضمن قائمة يوجد بها كل وزن صرفي والأوزان الصرفية المشتقة من هذا الوزن. وقد أشارت نتائج الدراسة إلى نسبة نجاح في التحليل تصل إلى ٨٥,٣%.

ب. الدراسات التي اهتمت بتصنيف المفردات

لعل أقدم المحاولات لدراسة تصنيف المفردات العربية هو ما جاء ضمن الدراسات التي تناولت جانب المعالجة النحوية للغة العربية كالدراسة التي قام بها إبراهيم (Ibrahim, 1986) لنيل درجة الدكتوراه. فقد أفرد لهذا الموضوع فصلا خاصا تناول فيه الأسس التي بنى عليها تحديد نوع الكلمات أثناء التحليل النحوي للجمل. فقد تحدث عن الزوائد النحوية التي تلتصق بالكلمة وبيّن أن لهذه الزوائد دلالات يمكن الاعتماد عليها في تصنيف الكلمات في اللغة العربية، كما بيّن أهمية الأوزان الصرفية في هذا الجانب، ووضّح أنّه يوجد من الأوزان ما هو خاص بالأسماء ومنها ما هو خاص بالأفعال، ومنها ما هو مشترك بينهما، كما تحدّث عن الشبكات الانتقالية في تمثيل حروف الكلمات وتكوينها وأثرها في تقليل الوقت المستغرق أثناء عملية البحث، وبيّن الباحث كيفية الاعتماد عليها في تصنيف الكلمات في اللغة العربية.

ومن الدراسات الأخرى التي حاولت الدمج بين التحليل الصرفي والتحليل النحوي ما قام به مكالستر (McAalester, 1994) بهدف توليد الجمل من خلال أنواع الكلمات التي يتم تحديدها في الخطوة الأولى، ومن ثم يتبع قواعد سياق الجمل في اللغة، لتركيب هذه الكلمات مع بعضها لتكوين الجملة.

واهتم أبو الليل وإيفان (Abuleil and Evens,1998)، في دراسة لهما بتصنيف الكلمات العربية داخل النص اعتمادا على الكلمات السابقة في سياق النص وعلى الأوزان الصرفية. إذ اعتمدا في منهجيتهما على بعض مفردات اللغة العربية كمفاتيح داخل النص للدلالة على نوع الجملة ونوع الكلمة. ومن أمثلة تلك المفردات كلمة "السيد" حيث تشير إلى أن ما يأتي بعدها هو اسم علم. كما اعتمدا أيضا على الأوزان الصرفية لإيجاد نوع الكلمة.

أما خوجا (Khoja, 1999) فقد قامت ببناء محلل لفظي يهتم بمعرفة أنوع الكلمات من أجل إيجاد الجذور لها. إذ استخدمت في ذلك خوارزمية تقوم بإزالة السوابق واللواحق من الكلمة، ثم تقوم بفحص الكلمة لبيان ما إذا تمت إزالة حروف أصلية منها، وذلك بمقارنتها مع الأوزان الصرفية في قاعدة بيانات. وبعد إيجاد الوزن الصرفي للكلمة، تتم إزالة الزوائد الداخلية من الكلمة الموجودة في الوزن الصرفي، وبذلك تكون قد حصلت على الجذر للكلمة. وقد أشارت الباحثة إلى أن الأسلوب المتبع في الدراسة قد نجح بنسبة ٨٠%.

وخلافا لذلك، نجد أن دراسة الفارس ورويك (Alfares and Roeck, 2000) قد سعت إلى تحديد أنواع الكلمات من خلال الزوائد. إذ قام ببناء محلل لفظي، ركز فيه على تحليل بنية الكلمات العربية بهدف استخراج الأسماء المشتركة بالجذر مع الأفعال. وكان معتمداً في منهجيته على تحليل بنية الكلمة ليحصل على السوابق واللواحق ويستدل من خلالها على نوع الكلمة. وعندما يتم بيان نوع الكلمة على أنه من باب الاسم، يبحث عن الكلمة بعد تجريدها من الزوائد في قائمة الجذور. وفي حال التطابق تكون الكلمة اسما مشتقا من فعل. وقد أشار الباحث إلى أن نتائج دراسته قد أظهرت أن هذا الأسلوب كان قادرا على تحقيق نسبة نجاح وصلت إلى ٨٨%.

وكما هو الأمر بالنسبة للدراسة السابقة، قام فريمان (Freeman, 2001) ببناء محلل لفظي لتصنيف الكلمات معتمداً في ذلك على الزوائد التي تحتويها الكلمة. وفي حالة أنّ الزوائد غير موجودة في الكلمة، أو لم يستدل من خلالها على نوع الكلمة، يقوم بالبحث عن الكلمة في المعجم بعد تجريدها من الزوائد لقراءة نوع الكلمة وخواصها. ولم يشر الباحث إلى مدى نجاح دراسته في تصنيف المفردات.

وخلافا للدراسات الواردة آنفا، قصر تالمون وونتنر (Talmon and Wintner, 2001) دراستهما لتصنيف المفردات على القرآن الكريم. حيث قاما ببناء محلل لفظي معتمدين في ذلك على معجم مكون من ثلاث أجزاء، وهي: قاموس بالأدوات والضمائر ، وقاموس لجذور الكلمات المذكورة في القرآن، وقائمة بالكلمات الجامدة. ويقوم التحليل على تجريد الكلمة من الزوائد، ليتم البحث عنها بعد ذلك في المعجم المذكور ويتم تصنيف الكلمات وفق المعلومات الواردة في المعجم. وقد أشار الباحثان إلى أن أسلوبهما هذا قد نجح في تصنيف المفردات بنسبة تزيد عن ٩٥%.

وقام بكوالتر(Buckwalter, 2002) ببناء محلل لفظي للكلمة العربية بهدف تصنيف الكلمات، منطلقا في منهجيته على مجموعة من الفرضيات، أهمّها: أنّ الكلمة تتكون من سوابق وجذر ولواحق، وأن السوابق يبلغ طولها في حدّها الأقصى أربعة أحرف، وأن طول الجذر غير محدّد، وأنّ طول اللواحق يبلغ في حده الأقصى ستة أحرف. وبناءً على ذلك، يقوم بتجزيء الكلمة إلى الأجزاء الثلاثة، إذا كان ذلك ممكنا. ثم يسأل ثلاثة أسئلة: هل سوابق الكلمة موجودة في قاموس السوابق؟ إذا كانت الإجابة بنعم، يسأل هل جذر الكلمة موجود في قاموس الجذور؟ وإذا كانت الإجابة بنعم، يسأل هل لواحق الكلمة موجودة في قاموس اللواحق؟ وفي حالة التحقق من وجود السوابق والجذر واللواحق في القواميس المستخدمة يقوم بفحص نوع الكلمة عن طريق مطابقة النوع في قاموس السوابق مع النوع في قاموس الجذور، وكذلك مطابقة النوع في قاموس السوابق مع النوع في قاموس اللواحق. وفي حالة التطابق، يتم إيجاد نوع الكلمة وخواصها وفقا للمعلومات الواردة في المعجم.

وبينما حاولت جميع الدراسات المذكورة أن تتناول جميع أنواع المفردات، سعى أبو الليل وزملاؤه (Abuleil et al., 2002) إلى بناء محلل لفظي يقتصر على استخراج الأسماء غير المشتقة من أفعال. وقد اعتمدوا في منهجيتهم على تحليل بنية الكلمة ليحصلوا على الزوائد النحوية فيها للاستدلال من خلالها على نوع الكلمة. وعند تحديد الاسم، يتم البحث عن الوزن الصرفي لهذه الكلمة ضمن قائمة الأوزان الصرفية المستخدمة في الدراسة. وفي حالة تعذر إيجاد وزن صرفي للاسم، تكون النتيجة، وفقا للدراسة، هي أنّ الكلمة اسماً غير مشتق من فعل. وقد أشار الباحثون إلى منهجهم هذا قد حقق نسبة نجاح في تحديد الأسماء غير المشتقة وصلت إلى ٩٠,٢%.

وقد حاول فرغلي وسنيلارت (Farghali and Senellart, 2003) تطوير برنامج يقوم على تصنيف كلمات اللغة العربية بقصد حل مشكلة بناء القاموس يدويا. ويهدف البرنامج إلى توليد

جميع الكلمات الممكنة لبناء القاموس اعتماداً على جذور الكلمات والأوزان الصرفية التي يتم إدخالها يدوياً للبرنامج. وقد استند منهج الباحثين على تجريد الكلمة من الزوائد أولا ومن ثم البحث عن بقية الكلمة في معجم للجذور. وفي حالة عدم التطابق تتم الاستعانة بقائمة الأوزان الصرفية لتحديد نوع الكلمة وخصائصها الأخرى.

وقام يوسف طاهر وزملاؤه (Tahir et al.,2003) باقتراح محلل لفظي للنص العربي يهدف إلى تحديد نوع كل كلمة في النص وخواصها اعتمادا على بناء قاعدة بيانات تخزن فيها جميع كلمات النص بجميع أشكالها، إضافة إلى السوابق واللواحق وجميع أدوات الربط بين الكلمات. حيث يقوم المحلل اللفظي المستخدم في هذه الدراسة بمقارنة الكلمة والسوابق واللواحق مع الكلمات والسوابق واللواحق في قاعدة البيانات. وفي حالة التطابق تتم قراءة نوع الكلمة وخواصها من قاعدة البيانات المذكورة.

ولعل أحدث الدراسات التي تمت في موضوع التحليل اللفظي لغايات تصنيف المفردات العربية هو ما قام كل من مارك فان و بولوسين (Van Mol and Paulussen , 2004) لبناء محلل لفظي لتصنيف الكلمات الواردة في النص. وقد اعتمدا في ذلك على منهجية تتكون من خطوتين رئيسيتين، أولهما، تخزين جميع الأوزان الصرفية بما تحتويه من زوائد صرفية. وقد ألصقا بها الزوائد النحوية بجميع أشكالها الممكنة (مثل "كال" حيث ألصقاها بالوزن "فعل" لتصبح "كالفعل"). وفعلا ذلك مع كل زائدة نحوية، فأصبحت قائمة الأوزان تحتوي على الزوائد الصرفية والنحوية، حيث تتم مقارنة الكلمة بهذه الأوزان، وفي حال التطابق تتم قراءة نوع الكلمة وخواصها من تلك القائمة. أما الخطوة الثانية فهي الاعتماد على سياق النص، والتي يتم اللجوء إليها عندما تفشل الخطوة الأولى. إذ ينظر إلى نوع الكلمة السابقة في النص لتحديد نوع الكلمة التي تتم معالجتها. وقد أشار الباحثان إلى أن هذا المنهجية قد نجحت بشكل كبير في تحديد نوع الكلمات.

وكما فعل فرغلي وسنيلارت في محاولة البناء الآلي للمعجم، قام حبش (٢٠٠٤)، بتصنيف الكلمات بقصد توليد اللغة. وقد اعتمد في منهجيته على قاعدة بيانات واسعة النطاق تحتوي على جذور الكلمات وجميع السوابق واللواحق التي يمكن أن تتصل بالجذور. إذ يتم تحليل الكلمة بتجريدها أولا من زوائدها، ثم يجري البحث عنها في معجم الجذور لمعرفة نوعها وخصائصها.

٣. الأنظمة التي اهتمت بتصنيف المفردات

وبالإضافة إلى ما ذكر، تجدر الإشارة إلى أن هناك بعض الجهود التي وجهت من جانب بعض الشركات إلى بناء أنظمة تجارية تختص بالتحليل اللفظي للكلمات العربية. ولعل من أهم

هذه الأنظمة ما يعرف باسم المحلل الصرفي العربي (RDI ArabMorpho©)، الذي أنتجته الشركة الهندسية لتطوير نظم الحاسبات. حيث يقوم النظام بتحليل أجزاء الكلمة النصية المشكولة إلى السوابق واللواحق النحوية، واستنتاج الوزن الصرفي لكل كلمة، حيث اعتمدت الشركة تقسيم الكلمة إلى الأسماء (أسماء مشتقة وأسماء جامدة) والأفعال (ماض، مضارع، أمر) والصفات والأدوات والضمائر.

كما أنتجت الشركة نفسها نظاماً آخر يختصّ بتشكيل النصوص العربية واسمه المشكل الآلي للنص العربي(RDI ArabDiac©) معتمدة في ذلك على تحليل الكلمة إلى أجزائها ومعرفة نوعها ومن ثم استنتاج موقعها الإعرابي داخل الجملة.

ومن الشركات أيضاً التي أنتجت أنظمة تحليل لفظي لكلمات اللغة العربية مركز البحث الأوروبي لكسيروكس (١٩٩٦). حيث قام ببناء محلل لفظي للغة العربية العصرية وتمت من خلاله معالجة النصوص المشكولة وغير المشكولة، ولم توضح الشركة تقسيم الكلمة التي اعتمدته.

٤. خلاصة

لو ألقينا نظرة عامة على الدراسات السابقة في تصنيف مفردات النصوص العربية نجد أنها استندت في التحليل على تخزين الجذور لتتم مقارنة النتائج بها. وعلى الرغم من أن ذلك يساعد في التحقق من صحة النتائج، إلا أنه قد يزيد من كلفة عملية التحليل بشكل ملموس قد لا يتناسب مع المكاسب التي نحققها من ذلك. وانطلاقا من ذلك، كان الافتراض في الدراسة الراهنة بأنه قد يكون من الممكن تحقيق نتائج مرضية دون اللجوء إلى فكرة المعاجم إلا في حدود ضيقة لا تتجاوز الأدوات لأنها لا تخضع لقواعد الاشتقاق. إذ تعتمد منهجية هذه الدراسة على مجموعة الزوائد النحوية والدلالية والصرفية، وما يقوم بينها من علاقات قد تكون في كثير من الأحيان كافية للاستدلال منها على تحديد أنواع المفردات. ولعل ذلك المنهج هو ما يميزها عن غيرها من الدراسات السابقة.

الفصل الثالث

الزوائد اللفظية ومضامينها الدلالية

١. البنية التركيبية للكلمة العربية

٢. الزوائد النحوية

١,٢ السوابق النحوية

٢,٢ اللواحق النحوية

٣,٢ علاقة السوابق النحوية باللواحق النحوية

٣. الزوائد الصرفية

١,٣ السوابق الصرفية

٢,٣ اللواحق الصرفية

٣,٣ الوسطيات الصرفية

٤,٣ الأوزان الصرفية

٤. علاقة الزوائد النحوية بالزوائد الصرفية

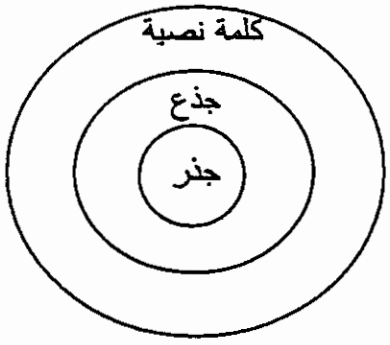
٥. خلاصة

الزوائد اللفظية ومضامينها الدلالية

تنطلق هذه الدراسة من فرضية أساسية مفادها أنّ البنية التركيبية للكلمة العربية داخل النص تنطوي على مضامين دلالية يمكن الاعتماد عليها لدرجة كبيرة في تحديد نوعها وخصائصها الوظيفية ضمن النص العربي. ومن هنا، يتناول هذا الفصل البنية الأساسية للكلمة العربية، كما ترد في النصوص، والأشكال المختلفة للزوائد اللفظية وما تتضمنه من معلومات يمكن الاستدلال منها على نوع الكلمة. يتكون الفصل من ثلاثة أقسام، يعالج الأول منها البنية التركيبية للكلمة، ويعالج القسمان الثاني والثالث الزوائد النحوية والزوائد الصرفية على التوالي ومضامينها الدلالية ومدى إمكانية الإفادة منهما في تحديد نوع الكلمة.

١. البنية التركيبية للكلمة العربية

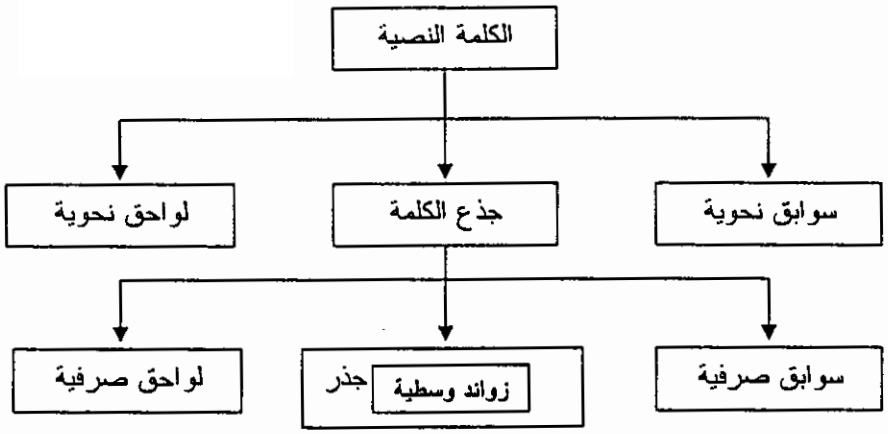
تقوم الكلمة في اللغة العربية، كما يتضح من الشكل (٣) على ثلاثة مستويات هي الجذر، والكلمة المعجمية، والكلمة النصية. فالجذر هو أدنى هذه المستويات. وهو يحوي الأحرف الأصلية ويكون مجردا من جميع أنواع الزوائد. أما الكلمة المعجمية (أو ما يوصف أحيانا بالجذع) فإنها تمثل أحد الأوزان الصرفية المعروفة في اللغة، بما ينطوي عليه ذلك من زوائد أولية ووسطية ونهائية. أما الكلمة النصية فإنها تمثل الوضع الذي تكون عليه الكلمات في سياق النصوص، بما ينطوي عليه ذلك من زيادات تمليها الضرورات النحوية والدلالية للجمل وسياق النص.



الشكل رقم (٣): توضيح مستويات الكلمة في اللغة العربية.

وبناء عليه، نجد أنّ هناك نوعين من الزوائد اللفظية ، وهي زوائد صرفية وزوائد نحوية. ولكل منها وظيفته الخاصة. فالزوائد الصرفية تلتصق بالجذر لتكوّن جذع الكلمة، وترد في ثلاثة مواقع من الكلمة، كما يتضح من الشكل (٤): في بدايتها ويشار إليها بالسابقة الصرفية (مثل "است" في كلمة استقبل) ، وفي أوسطها ويشار إليها بالوسطية الصرفية (مثل الألف الوسطى في كلمة استقبال)، وفي آخرها ويشار إليه باللاحقة الصرفية (مثل "اء" في كلمة بيداء).

أما الزوائد التي تلتصق بجذع الكلمة لتكون لنا الكلمة النصية، فإنها ترد في موقعين من الكلمة: في أولها (أي قبل الزائدة الصرفية الأولية) ويشار إليها بالسابقة النحوية (مثل الواو في كلمة واستقبال)، وفي آخرها (أي بعد اللاحقة الصرفية) ويشار إليها باللاحقة النحوية (مثل "هم" في كلمة استقبالهم).



الشكل (٤): البنية الهيكلية للكلمة العربية الواردة في النص

ويمكن التعبير عن هذه الزوائد اللفظية بالأسلوب العلمي على النحو الوارد في الشكل (٥)، مع التنبيه على أن ذلك يقتصر على الجذور الثلاثية والرباعية لأن الجذور الأخرى نادرة الاستعمال في اللغة. وأقل مكونات الكلمة النصية هو ما جاء على صيغة الجذر المجرد، وأكبرها هو ما تضمن جميع الزوائد وفق القواعد المذكورة في الشكل المذكور. وبناء على ذلك وإذا أخذنا بالاعتبار ما يحدث على بعض المفردات من إعلال، يمكن القول بأن الطول الافتراضي للكلمة يتراوح بين ما يمكن أن يطلق عليه الحد الأدنى والحد الأقصى (مصطفى، ٢٠٠٠)، حيث:

الحد الأدنى = طول الجذر(المجرد)

الحد الأقصى = الطول الأقصى للسوابق النحوية + الطول الأقصى للوزن الصرفي + الطول الأقصى للواحق النحوية

ومن الناحية العملية، لا يوجد هناك كلمات عربية تصل في طولها إلى الحد الافتراضي الأعلى. وقد لا يتجاوز الطول الفعلي للكلمات حد ثلاثة عشر حرفا (كما هو الحال في كلمة مثل "وباستخداماتها").

كلمة نصية = سابقة نحوية(٠-٤) + جذع + لاحقة نحوية(٠-٤)
جذع = سابقة صرفية + بؤرة صرفية + لاحقة صرفية
جذر = حرف(٣-٤)
بؤرة صرفية = جذر١ + واسطة أولى (٠-٢) + جذر٢ + واسطة ثانية(٠-٢) + جذر٣
بؤرة صرفية = جذر١ + جذر٢ + واسطة أولى(٠-١) + جذر٣ + واسطة ثانية(٠-١) + جذر٤
سابقة صرفية = "مست، من، مت، م، ت، است، افت، ان،....."
جذرس = "أ..ي"
واسطة أولى = "ا، و، ي، وا، وي، ت، ط، د"
واسطة ثانية = "ا، و، ي، ائـ"
سابقة نحوية = أداة.٣،٠ | أداة.٢،٠ + ضمير
أداة = "و، ال، ل، ف، أ، ك، ب، وا، س، أ "
ضمير = "أ، إ، ا، ن، ي، ت"
لاحقة نحوية = "وا، هم، ت، نا، ياتهما ...."
لاحقة صرفية = "ة، اء، وت، ان، ى"

الشكل (٥): قواعد بناء الكلمات النصية العربية (مقتبس مع التعديل من مصطفى، ٢٠٠٠)

من هذه البنية الهيكلية للكلمة العربية، بما فيها من جذر ووزن صرفي وزوائد، تتشكل الأسماء والأفعال، ومنها تتشكل بعض الدلالات المختلفة المصاحبة لها، كالماضي والمضارع والأمر، والعدد، والجنس وغيرها. وعلية، فإن التحليل اللفظي المعتمد في هذه الدراسة قد بني على أساس هذه المكونات التي تقوم عليها البنية الهيكلية للكلمة وما تقدمه من مضامين دلالية نفرق من خلالها بين الأصناف المختلفة للكلمات. وفي القسمين التاليين نبين أهمية الزوائد في هذا التحليل.

٢. الزوائد النحوية

٢,١ السوابق النحوية

تنقسم الزوائد النحوية إلى سوابق ولواحق، كما ذكر سابقا، ولكل من الفئتين حروف وأدوات تتركب منها. فقائمة السوابق النحوية تضم الأحرف والأدوات التالية وعددها ثلاثة عشر وفقا لما هو وارد في الجدول رقم (١): "أ، ا، ب، ت، س، ف، ك، ن، ل، و، ي، ال، وا". وهذه الحروف والأدوات الأساسية تأتي مفردة أو مركبة، ولكل منها في تركيب السوابق النحوية معنى تدل عليه (الدحداح،١٩٩٩).

الجدول رقم (١): حروف وأدوات السوابق النحوية

السوابق النحوية	معانيها ودلالاتها
أ	استفهام، تصريف، مصدرية، نداء
ا	تصريف
ب	توكيد، ظرفية، غاية، قسم
ت	تصريف، تفسير، زيادة
س	استقبال، تصريف
ف	ابتداء، جواب، عطف
ك	تشبيه، تفسير
ن	تصريف، تفسير، وقاية
ل	ابتداء، أمر، تصريف، تعجب، تعليل، زيادة، جواب، قسم، نفي، جر
و	ابتداء، تفسير
ي	تصريف، تفسير
ال	أداة تعريف
وا	أداة استغاثة

تنشأ السوابق النحوية المركبة من خلال تركيب هذه الأحرف والأدوات مع بعضها وفقا لعلاقات الوصل التي تحددها اللغة. وقد حاولت هذه الدراسة تتبع هذه العلاقات على نحو يخدم المنهج الذي قامت عليه. والجدول رقم (٢) يمثل السوابق النحوية التي تتشكل من أداتين متمثلتين في التقاء الأداة في الصف مع الأداة في العمود. كما يقدم الجدول رقم (٣) السوابق النحوية الثلاثية والرباعية. ويقرأ الجدول من اليمين إلى اليسار وفق ترتيب الأرقام. فالسابقة "أبال" على سبيل المثال، ممثلة بالجدول وفق الترتيب التالي: أ(١)، ب(٢)، ال(٣). وبالإضافة إلى ما ذكر، فإن هناك عدداً محدوداً جداً من السوابق الرباعية مثل "أفبال" وهي نادرة الاستعمال.

الجدول رقم (٢): السوابق النحوية الثنائية

السوابق		أ	ا	ب	ت	س	ف	ك	ل	ن	و	ي	ال	وا
أ	x	x		x	x	x	x	x	x	x	x	x		
ب	x	x											x	
ت	x													
س	x	x			x					x		x		
ف	x	x		x	x	x		x	x	x	x	x	x	
ك	x	x											x	
ل	x	x			x					x		x	x	
ن	x		x											
و	x	x		x	x	x	x	x	x	x		x	x	
ي	x													
ال	x													
وا	x													

الجدول رقم (٣): السوابق النحوية الثلاثية

١	٣ / ٢	ال	أ	ب	ت	س	ل	ن	ي
أ	ب	x	x						
	س		x		x			x	x
	ف		x	x	x	x	x	x	x
	ل	x	x		x			x	x
	و		x	x	x		x	x	x
ف	ب	x	x						
	ك		x						
	ل	x	x		x				x
و	ب	x	x						
	ك	x	x						
	ل	x	x		x			x	x

كما يلاحظ من هذين الجدولين، أن هناك ما يزيد على مائة زائدة نحوية أولية، بعضها كثير الاستعمال وبعضها الآخر نادر الاستعمال. ومن هذه الزوائد ما يشير إلى الأسماء (مثل أداة التعريف "ال") ومنها ما يشير إلى الأفعال (مثل "سيـ"). ويقدم الجدول رقم (٤) ملخصا لهذه الدلالات. أما تفاصيل المضامين الدلالية لهذه السوابق ومدى إمكانية الاعتماد عليها في تصنيف المفردات فيمكن الرجوع إليها في الملحق (٢) الوارد في نهاية هذه الرسالة العلمية. وكما يتضح من هذا الجدول، هناك العديد من السوابق يمكن الاستدلال منها في المعالجة على نوع الكلمة، اسماً أو فعلاً، بدلالة مؤكدة، و في المقابل هناك الكثير من السوابق تحتمل الدلالة على نوع معين ولكنها غير كافية بحد ذاتها، وذلك بسبب تشابه بعض السوابق مع مقاطع أصلية في المفردات، أو أنها تدلّ على أكثر من نوع في غياب التشكيل .

الجدول رقم (٤): أهمية السوابق النحوية في تحديد نوع الكلمة

نوع الكلمة	دلالة مؤكدة	دلالة غير مؤكدة	العدد
فعل	سأ، ست، سن، سي، سن، أت، ألت، ألي، أوت ، سأ، ست، سي، فسأ، فست، فسي، وسأ، وست، وسي	ا، أت، أف، أو، ف، ل، ول،و، ن، ون، ي، وي، ولي، ولن، ت، أي، أوي، س، لن، وسن، أن، فلن، فن، فسن، ألن، أوأ، وأ، أفا، أأ، ولأ، لأ، ولت، فلأ، فلت، فلي، في، فت، أفت، فأ، لت، لي	٦١
اسم	أل، أبا، أيال، أكال،أل، با، بال، فال، فكال،فل، فأل، كأ، كال، لل، ويال، وكال، وال	أل، أف، أو، ف، ل، و، ول، ك، وك، أك، فك، وال، ، ب، وب، فك، فب، أول ، أقب، أب، أوب،	٣٦
المجموع	٣٥	٦٢	٩٧

٢,٢ اللواحق النحوية

تتضمن اللواحق النحوية واحدة أو أكثر من الأحرف وضمائر الوصل التالية "ا، ات، ان، ت، تم، تما، تن، ة، ك، كم، كما، كن، ن، نا، هـ، هم، هما، هن، و، وا، ون، ي"، وفق الدلالة أو الوظيفة التي ينبغي أن تؤديها. والجدول رقم (٥) يبين الدلالات التي تستخدم لها كل من هذه الزوائد النحوية.

الجدول رقم (٥): حروف وأدوات اللواحق النحوية

اللواحق النحوية	معانيها ودلالاتها
ا	تصريف
ات	جمع مؤنث
ان	مثنى
ت	تصريف، زيادة
ة	للتأنيث
تم	جمع مذكر
تما	مثنى
تن	جمع مؤنث
ك	خطاب
كم	خطاب جمع مذكر
كما	خطاب مثنى
كن	خطاب جمع مؤنث
ن	تصريف، وقاية، توكيد، تأنيث
ه	تصريف، سكت
هم	للغائب(جمع مذكر)
هما	للغائب(مثنى)
هن	للغائب(جمع مؤنث)
و	ابتداء، تصريف
وا	للجمع
ون	للجمع أو ملحق بالجمع
ي	تصريف

وهذه الزوائد تأتي مفردة أو مركبة. إذ تنشأ اللواحق النحوية المركبة من دمج اثنتين أو ثلاثة من الحروف والأدوات المذكورة أعلاه. والجدول رقم (٦) يعرض اللواحق النحوية الثنائية والثلاثية ممثلة بتقاطع اللاحقة في الصف مع اللاحقة في العمود.

الجدول رقم (٦): مجموعة اللواحق النحوية المركبة

لاحقة		ا	ت	ك	ن	ه	ة	و	ي	تم	تما	تن	ها	هما	هن	هم	كم	كما	كن
ا	x		x	x	x	x				x	x	x		x	x	x	x	x	x
ت	x				x	x			x				x	x	x	x	x	x	x
ك	x																		
ن	x	x				x			x				x	x	x	x			
ه	x																		
ة	x																		
و	x	x		x	x	x							x	x	x	x	x	x	x
ي	x				x	x	x						x	x	x	x			
ات	x					x								x	x	x	x	x	x
هما	x																		
هن	x																		
هم	x																		
كما	x																		
كن	x																		
كم	x																		
تم	x	x						x											
تما	x																		
تن	x																		
تا	x				x	x			x					x	x	x	x	x	x
نا	x			x		x								x	x	x	x	x	x
ون	x					x							x	x	x				
ين	x					x							x	x	x				

وكما هو الحال بالنسبة للسوابق النحوية، تتضمن هذه اللواحق مضامين دلالية يمكن الاستعانة بها في تحديد نوع الكلمة وبعض سماتها الأخرى. فهناك لواحق نحوية لا تلتصق إلا بالأسماء (مثل "ـة")، وهناك لواحق لا تلتصق إلا بالأفعال (مثل "وا")، ومنها ما يحدد العدد (مثل اللاحقة "ون" للجمع)، ومنها يوضح الجنس (مثل اللاحقة "ات" للمؤنث). ويقدم الجدول رقم (٧) ملخصا لهذه الدلالات من حيث الأسماء والأفعال. أما تفاصيل المضامين الدلالية لهذه اللواحق ومدى إمكانية الاعتماد عليها في تصنيف المفردات فيمكن الرجوع إليها في الملحق (٢) الوارد في نهاية هذه الرسالة العلمية. وكما يتضح من هذا الجدول، هناك عدد من اللواحق يمكن الاستدلال منها في المعالجة على نوع الكلمة بدلالة مؤكدة، وهناك أيضا بعض اللواحق التي تحتمل الدلالة على نوع معين ولكنها غير كافية بحد ذاتها لنفس الأسباب التي ذكرت عند الحديث عن السوابق النحوية.

الجدول رقم (٧): أهمية اللواحق النحوية في تحديد نوع الكلمة

نوع الكلمة	دلالة مؤكدة	دلالة غير مؤكدة	العدد
فعل	تما، تمو، تن، تك، تكم، تكما، تكن، ته، تا، ت‌تك، تكم، تكما، تكن، ته، نا، ه، ها، هم، هما، هن، ون، ي، ين، ان، ته ،تهم، تهما، تهن، اه، وك، وكم، وكما، وكن، وه، وها، وهم، وهما، وهن، ك، لكم، لكما، لكن، اها، اهم، اهما، اهن، ن	٧٢
اسم	اتك، اتكم، اتكما، اتكن، اته، اتها، اتهم، اتهما، اتهن، تاك، تاكم، تاكما، تاكن، تان، تاهم، تاهما، تاهم، تاهن، تاي	ا، ان، ون، ين، تا، تاه، تاها، ك، كم، كما، كمو، كن، كي، ات، ة، ية	٣٥
المجموع	٥٣	٥٤	١٠٧

٣,٢ علاقة السوابق النحوية باللواحق النحوية

لاحظنا من الجداول السابقة أنه يوجد من السوابق واللواحق النحوية ما ليس لـه دلالـة مؤكدة على نوع الكلمة (أو حتى غير مؤكدة) في الإشارة إلى نوع الكلمة. ومن هنا، قـد يكـون من المفيد أن ننظر إلى علاقة السوابق لتحديد مدى إمكانية الإفادة من هذه العلاقة في تحديد نوع الكلمة. فهناك بعض السوابق التي تأتي مع بعض اللواحق ولا تأتي مع غيرها كما يتضـح مـن الجدول رقم (٨). ولمعرفة مدى أهمية الدلالات التي تتضمنه هذه العلاقة، يمكن الرجـوع إلـى الملحق (٢) في نهاية هذه الدراسة.

الجدول رقم (٨): العلاقة بين السوابق النحوية واللواحق النحوية

[redacted]	لواحق ١	ات | ت												[redacted]	[redacted]	[redacted]
سوابق ١	لواحق٢ / سوابق٢	ي	هما	هم	هن	ها	كم	كما	كن	كي	ك	ين	ان	ون
أ| ف| و	بال	x										x	x	x
	كال	x										x	x	x
	ال											x	x	x
	لت		x	x	x	x	x	x	x		x	x	x	x
	ست		x	x	x	x	x	x	x		x	x	x	x
	لت		x	x	x	x	x	x	x		x	x	x	x
	ت											x	x	x
	ي		x	x	x	x	x	x	x	x	x		x	x
	ت		x	x	x	x	x	x			x	x	x	x
	ك											x	x	x
	ل											x	x	x
	ب	x										x	x	x

٣. الزوائد الصرفية

تنقسم الزوائد الصرفية، كما أسلفنا، إلى ثلاثة أجزاء: السوابق الصرفية، واللواحق الصرفية، والوسطيات الصرفية. وهي تتكون من مجموعة من الحروف أوجزها النحاة في حروف كلمة "سألتمونيها". وقد تتشكل الزائدة الصرفية من حرف واحد أو من أكثر من حرف تبعا للوزن الصرفي الذي تقوم عليه. ولكل من هذه الأوزان لفظا يبنيه ومعنى يؤديه. فهناك أوزان خاصة بالأسماء، وأخرى خاصة بالأفعال، وثالثة خاصة بأنواع أخرى من الألفاظ كالصفات وغيرها. وفي الملحق (١) يجد القارئ قائمة بالأوزان الصرفية ومضامينها الدلالية ومدى الاعتماد عليها في تحديد أنواع المفردات وسماتها المختلفة. وفيما يلي بيان لأهمية كل من السوابق واللواحق الصرفية كلا على حدة.

١,٣ السوابق الصرفية

يتراوح طول السابقة الصرفية بين حرف واحد وثلاثة أحرف (مثل "است" في الوزن الصرفي استفعل). وتتكون مجموعة السوابق الصرفية من الزوائد التالية: "أ، ا، است، ان، ت، تم، م، مت، مست، متم، من".

ولكل من هذه السوابق الصرفية دلالتها الخاصة بحيث يمكن من خلال ذلك تحديد أنواع بعض الكلمات وبيات سماتها. إلا أنها قد لا تكون كافية في حد ذاتها، في الكثير من الأحيان. حيث لا بد أن تقترن بواسطة صرفية حتى يمكن الاعتماد عليها في ذلك. فالسابقة "مست"، على سبيل المثال، قد تشير إلى اسم فاعل أو اسم مفعول والسابقة "ت" قد تشير إلى اسم أو فعل. والجدول رقم (٩) يبين الأوزان الصرفية التي تبدأ بسوابق ويمكن الاستعانة بها في تحديد نوع الكلمة. وكما يتضح من هذا الجدول، فإن جميع هذه السوابق توفر دلالات غير مؤكدة على نوع الكلمة التي قد تشير إليها.

الجدول رقم (٩): أهمية السوابق الصرفية في تحديد نوع الكلمة

نوع الكلمة	دلالة مؤكدة	دلالة غير مؤكدة	العدد
فعل	–	تم، است، أ، ان، ت	٥
مصدر	–	است، أ، ان	٣
مصدر صناعي	–	است،ان	٢
اسم مفعول	–	مت، متم، من، م	٤
اسم فاعل	–	مت، متم، من، م	٤
صفة مشبهة	–	أ، ت	٢
اسم تفضيل	–	أ	١
اسم منسوب	–	–	
زمان ومكان	–	م	١
اسم آلة	–	م	١
الأعلام	–	–	
الاسم العام	–	م	١
المجموع	٠	٢٤	٢٤

٢,٣ اللواحق الصرفية

تتكون اللاحقة من حرف أو حرفين (مثل الزائدة "اء" في الوزن الصرفي فعلاء، والزائدة "ان" في الوزن الصرفي فعلان). وتضم قائمة الزوائد النحوية عددا قليلا من الزوائد، بعضها يشترك مع الزوائد النحوية (مثل التاء المربوطة في الوزن الصرفي "فعالة"). وهذه اللواحق هي: "اء، ة، ى، وت، ان". والجدول رقم (١٠) يوضح أهمية اللواحق الصرفية في تصنيف المفردات. وكما يتضح من هذا الجدول، فإنه لا يمكن أن تقدم لنا هذه الزوائد دلالات مؤكدة في تصنيف المفردات.

الجدول رقم (١٠): أهمية اللواحق الصرفية في تحديد نوع الكلمة

نوع الكلمة	دلالة مؤكدة	دلالة غير مؤكدة	العدد
فعل	-	-	
مصدر		وت، ة، ى	٣
مصدر صناعي	-	ة	١
اسم مفعول	-	-	
اسم فاعل	-	-	
اسم تفضيل	-	-	
اسم منسوب	-	-	
صفة مشبهة	-	ة، اء، ان	٣
زمان ومكان	-	-	
اسم آلة	-	ة	١
الأعلام	-	-	
الاسم العام	-	ان، ة، اء	٣
المجموع	٠	١١	١١

٣,٣ الوسطيات الصرفية

هناك فئتان من الوسطيات الصرفية: أولاهما يرد تاليا للحرف الأول من الجذر الثلاثي أو الحرف الثاني من الجذر الرباعي، وثانيهما تلك التي ترد بعد الحرف الثاني من الجذر الثلاثي أو الثالث من الجذر الرباعي. والواسطة الواحدة تتكون من حرف أو حرفين، وغالبا ما يكون ذلك من أحرف المد المعروفة وهي والألف والواو والياء. وبصفة عامة، تضم الفئة الأولى من هذه الزوائد ما يلي: "ا، ت، و، وا، ي". وقد تتحول التاء أحيانا إلى "ط" في مثل كلمة "اصطدم". أما الفئة الثانية فتضم الزوائد التالية: "ا، و، ي، ان".

أما من حيث الاعتماد عليها في تحديد نوع الكلمات، فالوسطيات الصرفية غير كافية بحد ذاتها لهذه الغاية إلا في حالات قليلة. إذ لا بد من ربطها بقرينة أخرى كالسوابق أو اللواحق للاستعانة بها. والجدول رقم (١١) يوضح لنا مدى إمكانية الاستعانة بالواسطة الصرفية في تصنيف الكلمات.

الجدول رقم (١١): أهمية الوسطيات الصرفية في تحديد نوع الكلمة

نوع الكلمة	دلالة مؤكدة	دلالة غير مؤكدة	العدد
فعل	–	ا	١
مصدر	–	ا، ت، ي، و	٤
مصدر صناعي	–	ا، ت، ي، و	
اسم مفعول	–	ا، و	٢
اسم فاعل	–	ا	١
صفة مشبهة	–	ي، و	٢
اسم منسوب	–	–	
اسم تفضيل	–	–	
زمان ومكان	–	–	
اسم آلة	–	ت، ا	٢
الأعلام	–	–	
الاسم العام	–	ا، ي، و	٣
المجموع	٠	١٥	١٥

٤,٣ الأوزان الصرفية

كما يلاحظ من الجداول السابقة، إن الاعتماد على السوابق الصرفية لوحدها أو اللواحق الصرفية لوحدها أو الوسطيات الصرفية لوحدها غير كافٍ في تصنيف الكلمات. ولكن اعتمادا على علاقة الزوائد الصرفية ببعضها والمتمثلة بالوزن الصرفي، نستطيع تصنيف الكثير من المفردات كما هو مبين في الملحق رقم (١).

٤. علاقة الزوائد النحوية بالزوائد الصرفية

إن العلاقات القائمة بين الزوائد النحوية والزوائد الصرفية على قدر كبير من الأهمية في التحليل. فهناك بعض السوابق النحوية التي قد تأتي متبوعة ببعض السوابق الصرفية ولا تأتي مع غيرها. وكذا الأمر بالنسبة للواحق الصرفية وما يمكن أن يأتي بعدها من لواحق نحوية. والجدولان رقم (١٢) ورقم (١٣) يبينان المركبات التي تنشأ من هذه العلاقة. كما أنّ الملحق رقم (٢) يبين أهمية هذه العلاقات القائمة بين الزوائد النحوية والزوائد الصرفية. كما أنّ الجدولين رقم(١٤) ورقم (١٥) يلخصان هذه الأهمية في تحديد أنواع الكلمات.

السوابق الصرفية

السوابق	أ	است	ان	ت	م	مت	من
كال	x	x	x		x	x	x
ال	x	x	x		x	x	x
ك	x	x	x	x	x	x	x
ب	x	x	x	x	x	x	x
بال	x	x	x		x	x	x
فك				x	x	x	x
فكال	x	x	x		x	x	x
فل	x			x	x	x	x
فال	x	x	x		x	x	x
فب	x	x	x	x	x	x	x
وك				x	x	x	x
وكال	x	x	x		x	x	x
ول	x			x	x	x	x
وال	x	x	x		x	x	x
أك				x	x	x	x
أكال	x	x	x		x	x	x
أل	x			x	x	x	x
أب	x	x	x	x	x	x	x
أوب	x	x	x	x	x	x	x

السوابق النحوية

الجدول رقم (١٢): علاقة السوابق النحوية بالسوابق الصرفية

اللواحق النحوية

اللواحق	نا	ه	هم	هما	هن	ك	كم	كما	كن
اء	x	x	x	x	x	x	x	x	x
ة	x	x	x	x	x	x	x	x	x
ية	x	x	x	x	x	x	x	x	x
وت	x	x	x	x	x	x	x	x	x

اللواحق الصرفية

الجدول رقم (١٣): علاقة اللواحق الصرفية باللواحق النحوية

الجدول رقم (١٤) أهمية العلاقة بين السوابق النحوية والسوابق الصرفية في تحديد نوع الكلمة

نوع الكلمة	دلالة مؤكدة	دلالة غير مؤكدة	العدد
فعل	–	لأ، لت	٢
مصدر		بالم كالاست، كالان، الاست، الان، كأ، كاست، كان، باست، بان، بالان، بالاست	١٢
مصدر صناعي	–	كالاست، كالان، الاست، الان، كأ، كاست، كان، باست، بان، بالان، بالاست	١١
اسم مفعول	–	كالم، كالمن، كالمت، كم، كمن، كمت، المن، المت، لم، بالمت، بالمن، بالم	١٢
اسم فاعل	–	كالم، كالمن، كالمت، كم، كمن، كمت، المن، المت، لم، بالمت، بالمن، بالم	١٢
صفة مشبهة	كالأ، كت	المن، المت	٢
زمان ومكان	–	كالم، لم، كم	٣
اسم آلة	–	كالم، كم، بم	٣
اسم تفضيل	–	–	
اسم منسوب	–	–	
الأعلام	–	–	
الاسم العام	–	–	
المجموع	٢	٥٥	٥٧

الجدول رقم (١٥): أهمية العلاقة بين اللواحق الصرفية واللواحق النحوية في تحديد نوع الكلمة

نوع الكلمة	دلالة مؤكدة	دلالة غير مؤكدة	العدد
فعل	–	تنا، ته، تهم، تهما، تهن، تك، تكم، تكما، تكن	٩
مصدر	وتنا، وته، وتهما، وتهم، وتهن، وتك، وتكم، وتكما، وتكمن	تنا، ته، تهم، تهما، تهن، تك، تكم، تكما، تكن	١٨
اسم مفعول	–	–	
اسم فاعل	–	–	
صفة	–	اعنا، اعه، اعهم، اعهما، اعهن، اعك، اعكم، اعكما، اعكن	
زمان ومكان	–	–	
اسم آلة	–	ونا، ونه، ونهم، ونهما، ونهن، ونك، ونكم، ونكما، ونكن	٩
اسم تفضيل	–	–	
اسم منسوب	–	–	
المجموع	٨	٣٧	٤٥

٥.خلاصة

- تقوم هذه الدراسة على فرضية أساسية مفادها أنّ البنية التركيبية للكلمة العربية داخل النص تنطوي على مضامين دلالية يمكن الاعتماد عليها لدرجة كبيرة في تحديد نوعها وخصائصها الوظيفية ضمن النص العربي.
- تقسم الكلمة في اللغة العربية إلى ثلاثة مستويات هي الكلمة النصية، وجذع الكلمة، وجذر الكلمة، ومن خلال هذه الدراسة تمّ بناء قواعد للكلمة العربية داخل النص.
- تم حصر السوابق النحوية التي تتصل ببداية جذع الكلمة من خلال جداول تمثل السوابق النحوية المركبة، ولهذه السوابق النحوية دلالات على النوع الرئيسي للكلمة تم توضيح دلالاتها من خلال جدول سواء بدلالة مؤكدة أو غير مؤكدة.
- تم حصر اللواحق النحوية التي تتصل بنهاية جذع الكلمة من خلال جداول تمثل اللواحق النحوية المركبة، ولهذه اللواحق النحوية دلالات على النوع الرئيسي للكلمة تم توضيح دلالاتها من خلال جدول سواء بدلالة مؤكدة أو غي مؤكدة.
- تم بناء علاقات بين السوابق النحوية واللواحق النحوية للاستفادة من دلالات هذه العلاقة على نوع الكلمة.
- تم توضيح دلالات السوابق الصرفية واللواحق الصرفية والوسطيات الصرفية على نوع الكلمة من خلال جداول، وتبين أن الاعتماد على السوابق الصرفية لوحدها أو اللواحق الصرفية لوحدها أو الوسطيات الصرفية لوحدها غير كافٍ في تصنيف الكلمات. ولكن اعتماداً على علاقة الزوائد الصرفية ببعضها والمتمثلة بالوزن الصرفي، نستطيع تصنيف الكثير من المفردات.
- تم بناء علاقات بين الزوائد النحوية والزوائد الصرفية للاستفادة من دلالات هذه العلاقة على نوع الكلمة.

الفصل الرابع

منهجية التحليل اللفظي لتصنيف المفردات

١. الإطار العام لأسلوب التحليل اللفظي

٢. القوائم المستخدمة في التحليل

١,٢ قائمة السوابق

٢,٢ قائمة اللواحق

٣,٢ قائمة الأوزان الصرفية

٤,٢ قائمة الأدوات والضمائر

٥,٢ قائمة العلاقات

٣. عملية التحليل اللفظي

٤. مثال على عملية التحليل

٥. خلاصة

منهجية التحليل اللفظي لتصنيف المفردات

تقوم منهجية هذه الدراسة على الافتراض الأساسي الذي تمّ ذكره فيما سبق والقائل بأن من الممكن الاعتماد بدرجة عالية على الزوائد اللفظية (النحوية والصرفية) في تحديد نوع الكلمة النصية وتحديد خصائصها. فمن خلال تحليل بنية الكلمة داخل النص، يتم الحصول على معلومات يتم استخدامها كمؤشرات ودلائل نحدد بها خواص الكلمات المكونة للنص، كنوع الكلمة وسماتها من حيث الجنس والعدد والحالة الزمانية وغيرها.

يهدف هذا الفصل إلى التعريف بالأسلوب الذي تم اتباعه في تحليل المفردات وصولا إلى التحقق من هذا الافتراض. ويتكون الفصل من ثلاثة أقسام هي على التوالي: أولها يعرض الإطار العام لأسلوب التحليل اللفظي الذي بنيت عليه هذه الدراسة، وثانيها يستعرض القوائم التي استخدمت في تخزين البيانات المساندة للتحليل، وثالثهما يبين بالتفصيل الخطوات الأساسية لعملية التحليل وتصنيف المفردات. ويختتم الفصل بمثال يوضح خطوات عملية التحليل.

١. الإطار العام لأسلوب التحليل اللفظي

يقوم منهج هذه الدراسة على أربعة محاور أساسية. أولها، هو التحقق من إمكانية أن تكون الكلمة التي ننوي معالجتها تنتمي إلى مجموعة الأدوات والضمائر في اللغة العربية (كحروف الجر، والجزم، والضمائر المنفصلة، وما إلى ذلك). ونظرا لأن هذه الأدوات والضمائر لا تخضع لقواعد الاشتقاق، فقد تم حصرها وتخزينها بشكل يسهل الرجوع إليها أثناء التحليل.

وثانيها، هو التحليل اللفظي القائم على استخدام الزوائد النحوية على اختلاف أنواعها (كحروف العطف، وحروف الجر وأل التعريف وغيرها مما ذكره النحاة). ولتحديد مدى إمكانية الاعتماد على هذه الزوائد في تحديد نوع الكلمة، كان لا بد من حصر هذه الزوائد النحوية، وبيان مضامينها الدلالية في الإشارة إلى أنوع المفردات على النحو الذي ورد بيانه في الفصل السابق. ذلك أنّ الكثير من هذه الزوائد يدل دلالة مؤكدة أو غير مؤكدة على الأسماء أو الأفعال.

وثالثها، هو التحليل اللفظي القائم على استخدام الأوزان الصرفية، وفق ما تنطوي عليها من زوائد صرفية على اختلاف أشكالها في ضوء ما تم بيانه في الفصل السابق، وعلى العلاقة التي تقوم بين الزيادات الصرفية والزيادات النحوية. فهناك أوزان خاصة بالأفعال، وأوزان خاصة بالأسماء، وأوزان خاصة بالصفات، وما إلى ذلك، وهناك أوزان مشتركة بين هذه

الأنواع. وللقيام بهذا النوع من التحليل، كان لا بد من تحديد الأوزان الصرفية وحصر الزوائد الصرفية التي تدخل فيها وحصر العلاقات القائمة بينها وبين الزوائد النحوية التي تشترك معها في أوائل الكلمات أو أواخرها.

ورابعها، الاعتماد على طول الكلمة في التحليل حيثما كان ذلك ضروريا. فهناك كلمات ثلاثية أو رباعية، لا تتضمن أية زوائد لفظية، نحوية أو صرفية. وبذلك كان لابد من اللجوء إلى طول الكلمة لبيان الفئة التي يمكن أن تنتمي إليها من الأسماء أو الأفعال أو للتحقق من نتائج المعالجة اعتمادا على الزوائد اللفظية.

ويقدم الشكل (٦) توضيحا للخطوات الأساسية التي ينطوي عليها هذا الإطار العام للتحليل. وهذه الخطوات هي على النحو التالي:

١. اقرأ الكلمة التالية من النص.

٢. ابحث عن الكلمة في قائمة الأدوات، فإن وجدت فيها، يتم الانتقال إلى الخطوة الأخيرة.

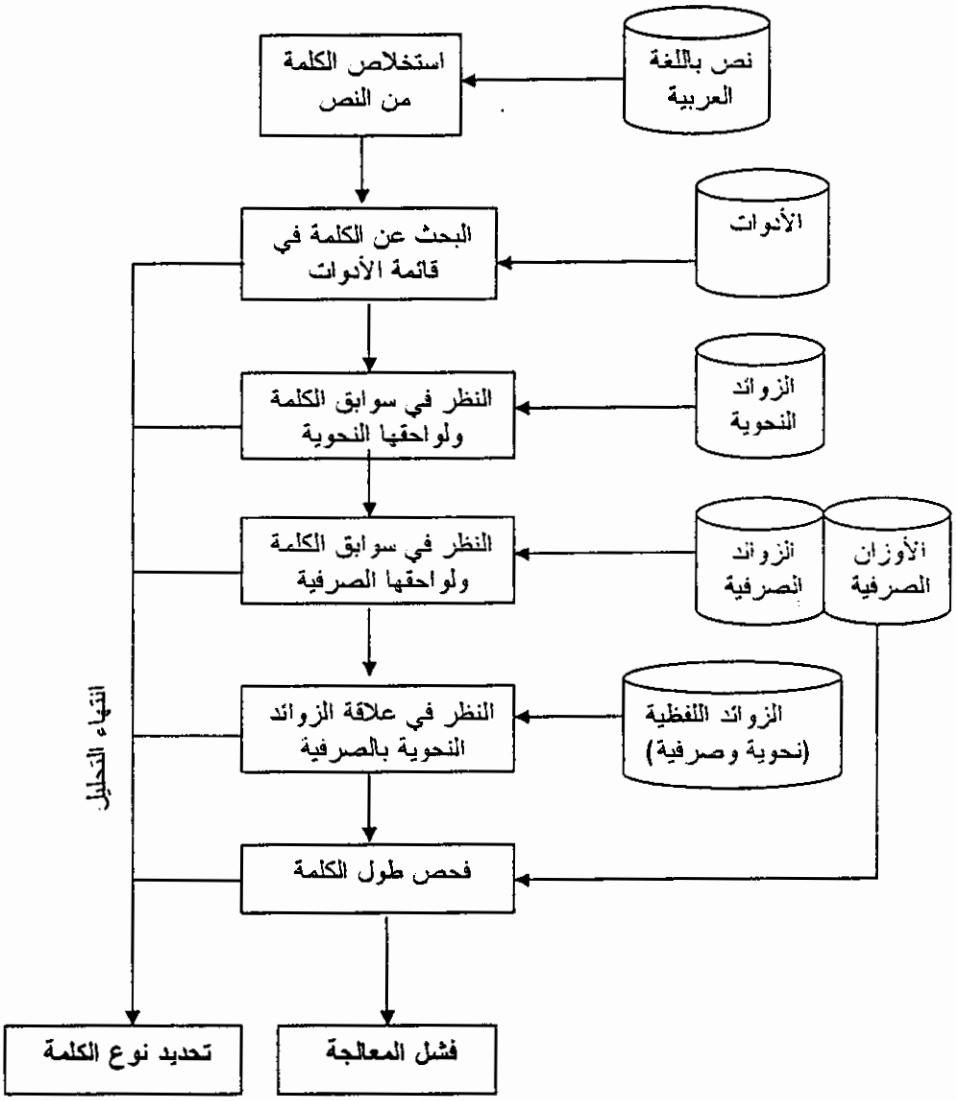
٣. حدد السابقة النحوية واللاحقة النحوية في الكلمة المقروءة، ومن ثم استخدم كلا من هاتين الزائدتين والعلاقة بينهما للتأكد من إمكانية تحديد نوع الكلمة. فإن أمكن ذلك، ننتقل إلى الخطوة الأخيرة.

٤. حدد السابقة الصرفية وحدد اللاحقة الصرفية في الكلمة المقروءة، ومن ثم استخدمهما والوزن الصرفي الذي ينتميان إليه، مع ما تم الحصول عليه من معلومات في الخطوة السابقة، للتأكد من إمكانية تحديد نوع الكلمة. فإن كان ذلك كافيا، ننتقل إلى الخطوة الأخيرة.

٥. استخدم العلاقة القائمة بين السابقة واللاحقة (الصرفية والنحوية مجتمعتين)، والميزان الصرفي، وما تم الحصول عليه من معلومات في الخطوتين السابقتين للتأكد من إمكانية تحديد نوع الكلمة. فإن كان ذلك كافيا، ننتقل إلى الخطوة الأخيرة.

٦. إذا لم يكن في الكلمة أية سوابق أولواحق نحوية أو صرفية، استخدم طول الكلمة وما فيها من زوائد وسطية (إن وجدت) لتحديد نوع الكلمة وبيان سماتها. وفي كل الأحوال ينبغي أن لا يقل طول الكلمة المستخلصة بعد تحديد الزوائد عن حرفين.

٧. إذا نجحت أي من الخطوات السابقة، اعرض نوع الكلمة وسماتها المختلفة الواردة في القوائم المعتمدة للسوابق واللواحق والأوزان الصرفية والأدوات. وإن كان غير ذلك، أعلن فشل عملية التحليل في تحديد نوع الكلمة وأعلن انتهاء العملية.



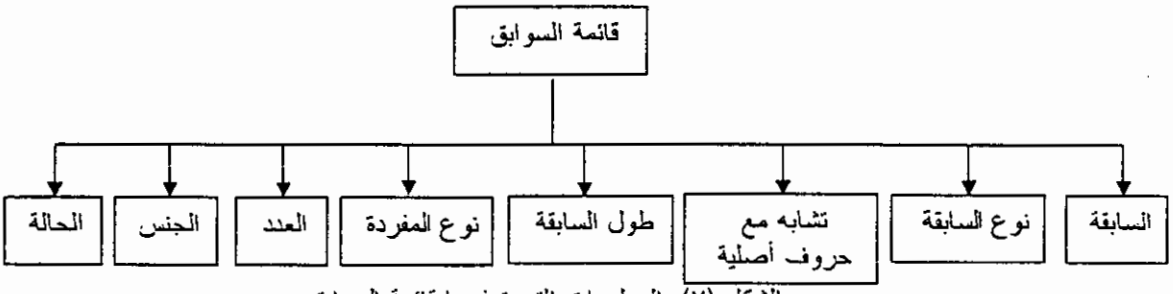
الشكل (٦): الخطوط العريضة لمنهج المعالجة اللفظية

٢. القوائم المستخدمة في التحليل

يرتكز أسلوب التحليل اللفظي، على النحو الموضح أعلاه، على مجموعة من قوائم البيانات تضم المعلومات اللازمة لتحديد نوع الكلمة وسماتها الدلالية في ضوء التحليل الذي ورد في الفصل السابق. وقد تم تمثيلها من الناحية التخزينية بشكل يتيح إمكانية الوصول إلى عناصرها بكفاءة عالية. وفيما يلي عرض لأهم هذه القوائم والمعلومات التي توفرها.

٢,١. قائمة السوابق

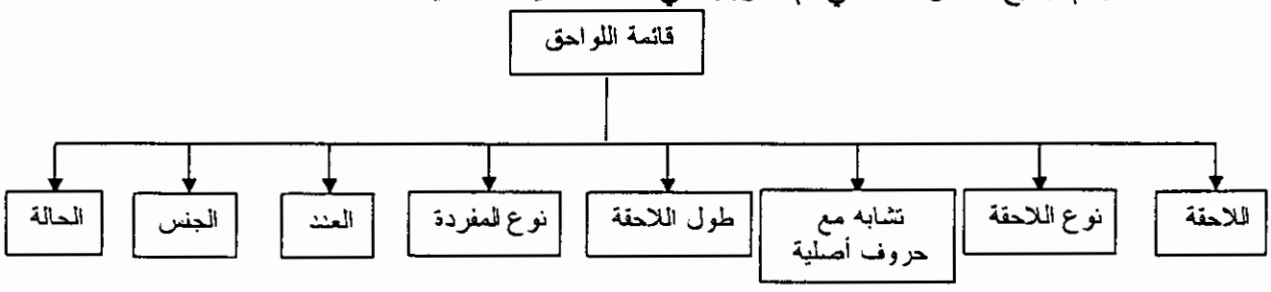
تحتوي هذه القائمة على جميع السوابق النحوية والسوابق الصرفية والسوابق المركبة التي تجمع بين النحوية والصرفية (مثل:"والاستـ"). وقد تم التمييز بين هذه السوابق بتزويد كل منها برمز يدل على نوعها. فقد أشير الى السابقة النحوية بالرقم "10"، وإلى الصرفية بالرقم "20"، وإلى ما يحتمل الاثنين بالرقم "30" وإلى المركبة من كليهما بالرقم "21". ولاعتبارات تتعلق بالكفاءة، تم تخزينها وفق ترتيبها الهجائي بما يسمح باستخدام البحث الثنائي في الوصول إلى معلوماتها. ويتم الرجوع إلى قائمة السوابق في معظم خطوات التحليل الذي ذكرت في القسم السابق. والشكل (٧) يبين البيانات التي توفرها هذه القائمة. كما أن الملحق (٢) يقدم جميع المعلومات التي تم تخزينها في القائمة لغايات التحليل.



الشكل (٧): المعلومات التي توفرها قائمة السوابق

٢,٢. قائمة اللواحق

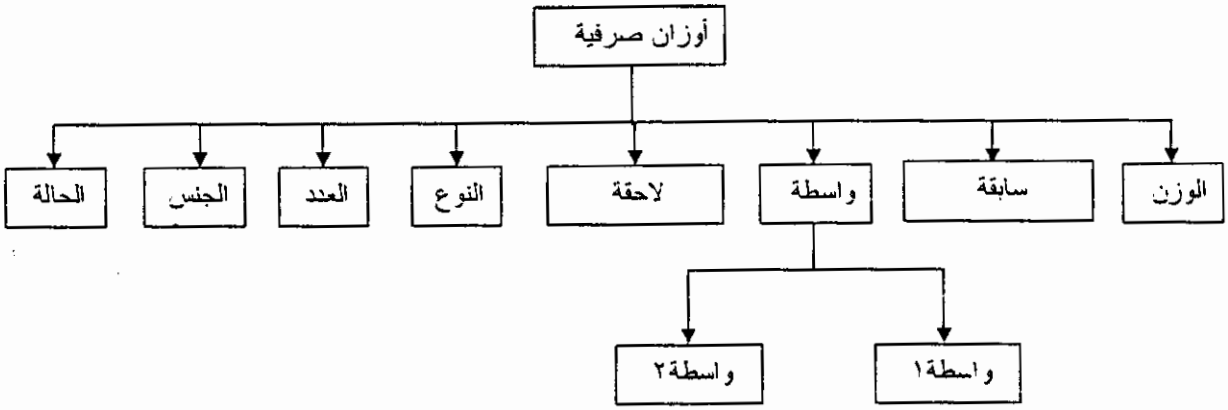
تتكون هذه القائمة من جميع اللواحق النحوية واللواحق الصرفية واللواحق المركبة التي تجمع بين الصرفية والنحوية (مثل: "اءهم") المفردة والمركبة. وقد تم التمييز بين هذه اللواحق بتزويد كل منها برمز يدل على نوعها. فقد أشير الى اللاحقة الصرفية بالرقم "40"، وإلى النحوية بالرقم "50"، وإلى ما يحتمل الاثنين بالرقم "60" وإلى المركبة من كليهما بالرقم "54". وكما هو الحال بالنسبة للسوابق، تم تخزين هذه اللواحق هجائيا لنفس الغاية، كما تستخدم في معظم مراحل التحليل. والشكل (٨) يبين البيانات التي توفرها هذه القائمة. كما أن الملحق (٢) يقدم جميع المعلومات التي تم تخزينها في القائمة لغايات التحليل.



الشكل (٨): المعلومات التي توفرها قائمة اللواحق

٣,٢. قائمة الأوزان الصرفية

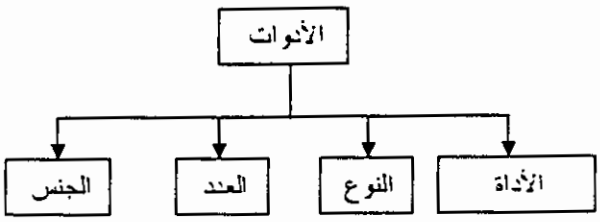
الهدف من هذه القائمة هو تخزين المعلومات المصاحبة لكل وزن من الأوزان الصرفية، كالزوائد ومواقعها من الكلمة المعجمية، ونوع الكلمة الذي يدل عليه ذلك الوزن، ومعلومات أخرى تتصل بالعدد والجنس والزمان. وكبقية القوائم، رتبت هذه القائمة هجائيا وفق الأوزان الصرفية. والشكل (٩) يبين طبيعة المعلومات التي احتوتها هذه القائمة. كما أن الملحق رقم (١) يقدم جميع المعلومات التي تم تخزينها في القائمة لغايات التحليل.



الشكل (٩): المعلومات التي توفرها قائمة الأوزان الصرفية

٤,٢. قائمة الأدوات والضمائر

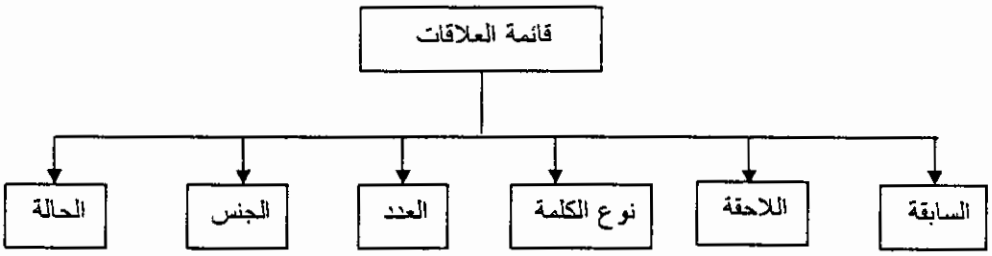
تضم هذه القائمة الأدوات بأنواعها من حروف وضمائر منفصلة وظروف وغيرها من المفردات التي لا تخضع لقواعد الاشتقاق ويمكن أن نعاملها معاملة الأدوات، حيث أن بعضها لا يخضع للتعريف المتداول لدى النحاة بالنسبة للأدوات. وتستخدم هذه القائمة في أول مرحلة من مراحل التحليل، حيث يتم التأكد أولا من إمكانية أن تكون الكلمة المعنية بالمعالجة من هذه الفئة. وقد تم تخزين جميع الأشكال التي ترد فيها هذه الأدوات من حيث دخول الزوائد عليها. وكما هو الحال بالنسبة لبقية القوائم، فقد تم تنظيمها هجائيا. والشكل (١٠) يبين طبيعة المعلومات التي تم تخزينها في هذه القائمة، وهي معلومات نجدها مفصلة في الملحق (٣). ونظرا لغياب الحالة الزمانية بالنسبة للكلمة الجامدة، فقد تم إعطاء كلمة "مجهول" للمعلومة التي لا يتمكن المحلل اللفظي من تحديدها.



الشكل (١٠): تمثيل قائمة الأدوات في منهجية الدراسة

٥,٢. قائمة العلاقات

تضم هذه القائمة المعلومات التي يمكن أن يقدمها اجتماع السابقة واللاحقة في الكلمة الواحدة في ضوء ما تم الحديث عنه في الفصل السابق. ويتم الرجوع إليها عندما لا تكون الزوائد النحوية أو الزوائد الصرفية كافية بمفردها للوصول إلى قرار يحدد نوع الكلمة المعنية. والشكل (١١) يبين نوع المعلومات التي تقدمها هذه القائمة، كما أن الملحق (٢) يقدم البيانات التي خزنت في هذه القائمة أثناء التحليل.



الشكل (١١): قائمة تمثل علاقة السوابق باللواحق (صرفية ونحوية)

٣. عملية التحليل اللفظي

تقوم عملية التحليل اللفظي لتصنيف المفردات على مراحل ستة تم ذكرها في الجزء الأول من هذا الفصل وهي: استخلاص الكلمة من النص، والتحقق من الأدوات والضمائر، والتحليل باستخدام الزوائد النحوية، والتحليل باستخدام الزوائد الصرفية، والتحليل باستخدام العلاقة القائمة بين السوابق واللواحق، وأخيرا التحليل باستخدام طول الكلمة والأوزان الصرفية. وفي كل مرحلة من هذه المراحل يتم الإفادة من المعلومات التي يتم الحصول عليها من خلال المرحلة السابقة عليها. ويتم الانتقال إلى المرحلة التالية في حالة عدم كفاية المرحلة الحالية في تحديد نوع الكلمة، وذلك لعدم كفاية المعلومات المتوفرة لها. وفيما يلي نبين تفاصيل كل مرحلة من هذه المراحل.

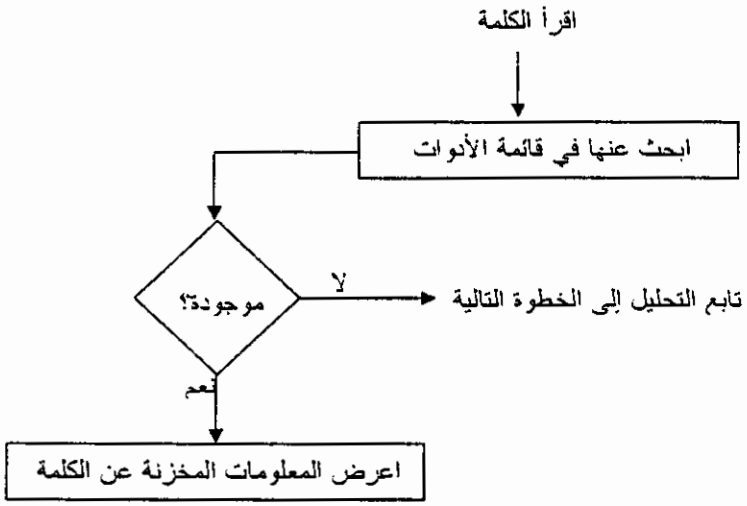
١,٣. استخلاص الكلمة من النص

تقوم الفكرة الأساسية لهذا البحث على التعامل مع مفردات اللغة كما ترد في النصوص العربية غير المشكولة. ونظرا لأن الكلمات تأتي مسبوقة بفراغ، فإن العملية تتم بطريقة بسيطة. إلا أن هناك بعض الرموز التي قد تأتي ملتصقة بأواخر الكلمات كعلامات التشكيل والأقواس

وغيرها، الأمر الذي يستوجب التعامل معها. أما بالنسبة لعلامات التشكيل، فقد تمت إزالتها من النص قبل المعالجة.

٢,٣. التحقق من الأدوات والضمائر

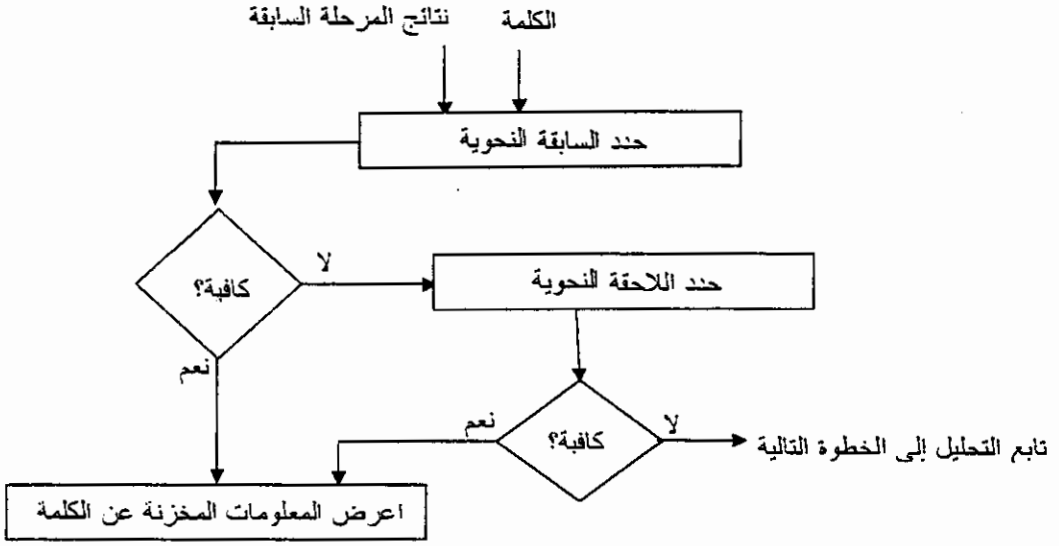
بعد استخلاص الكلمة من النص، يتم البحث عنها أولا في قائمة الأدوات التي تضم المفردات التي لا تنطبق عليها قواعد علم الصرف. وفي حالة العثور على الكلمة تتم استرجاع المعلومات اللازمة من القائمة وعرضها كما هو موضح في الشكل (١٢).



الشكل (١٢) خطوات التحقق من أن الكلمة واحدة من الأدوات أو الضمائر

٣,٣. التحليل باستخدام الزوائد النحوية

يتم استخلاص السابقة النحوية أولا من خلال تتبع الحروف الأولى في الكلمة، حرفا بعد حرف، ومقارنتها بما هو موجود في قائمة السوابق حتى يتم الوصول إلى آخر حرف فيها. وكذا نفعل بالنسبة للاحقة، حيث يتم استخلاصها بطريقة عكسية اعتمادا على قائمة اللواحق. وبناء على المعلومات المخزنة في قائمة السوابق أو قائمة اللواحق أو قائمة العلاقات بين السوابق واللواحق، يتم النظر في إمكانية معرفة نوع الكلمة. فإذا كانت هناك سابقة أو لاحقة أو كلاهما ونجحت هذه الخطوة، يتم عرض النتائج، وإلا سيتم الانتقال إلى المرحلة التالية لمتابعة التحليل، مع الاحتفاظ بما تم التوصل إليه من معلومات قد تفيدنا في المرحلة التالية، وذلك على النحو الموضح في الشكل (١٣).



الشكل (١٣) خطوات التحليل باستخدام الزوائد النحوية

٤,٣. التحليل باستخدام الزوائد الصرفية

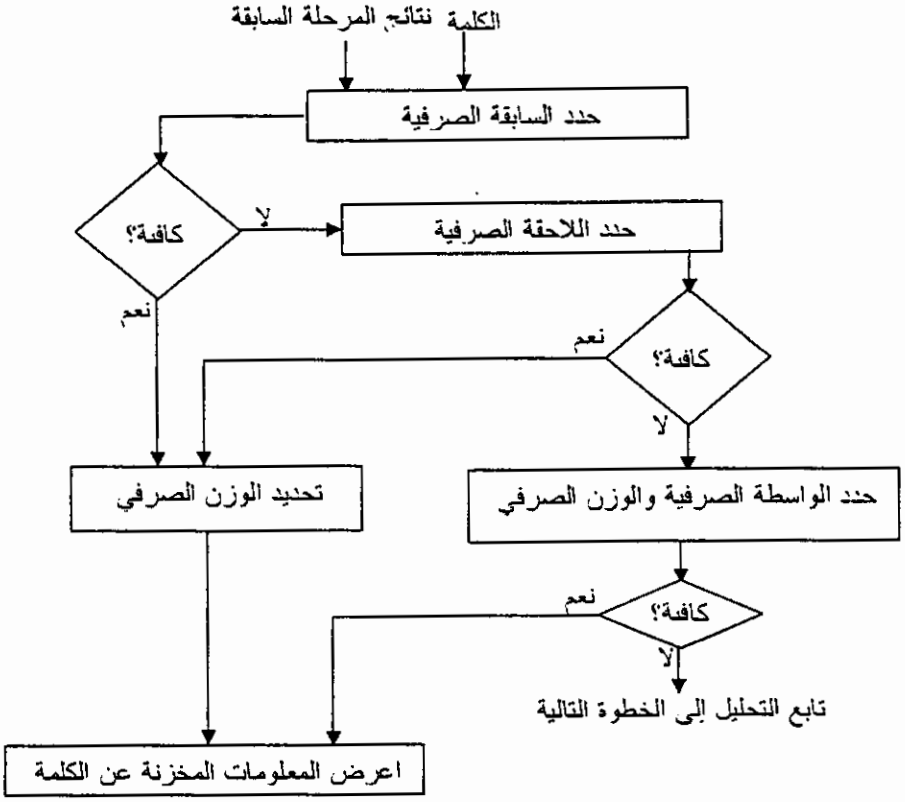
يتم استخلاص السابقة الصرفية من خلال تتبع الحروف الواردة بعد السابقة النحوية ومقارنة ذلك تدريجيا بقائمة السوابق. وفي حالة عدم وجود سابقة نحوية، فإن الأمر يتم بشكل بسيط، أما في حالة وجود سابقة نحوية فإن تحديد السابقة الصرفية يتم على خطوتين كما يلي:

أولا: تحديد السابقة المركبة التي تجمع بين السابقة النحوية والصرفية.

ثانيا: التتبع العكسي لحروف السابقة المركبة حتى يتم تحديد بداية السابقة الصرفية.

وكما هو الحال بالنسبة للسابقة الصرفية، فإن استخلاص اللاحقة الصرفية يتم من خلال تحديد اللاحقة المركبة أولا، فمن خلال متابعة التحليل من النقطة التي انتهت إليها خطوة التحليل باستخدام الزوائد النحوية، ومن ثم يتم تحديد نهاية الزائدة الصرفية واستخلاصها اعتمادا على اللاحقة المركبة إن وجدت. أما في حالة عدم وجود لاحقة نحوية، فإن استخلاص الزائدة الصرفية يتم من خلال التتبع العكسي لحروف الكلمة (أي ابتداءً من نهايتها) ومطابقة ذلك مع قائمة اللواحق.

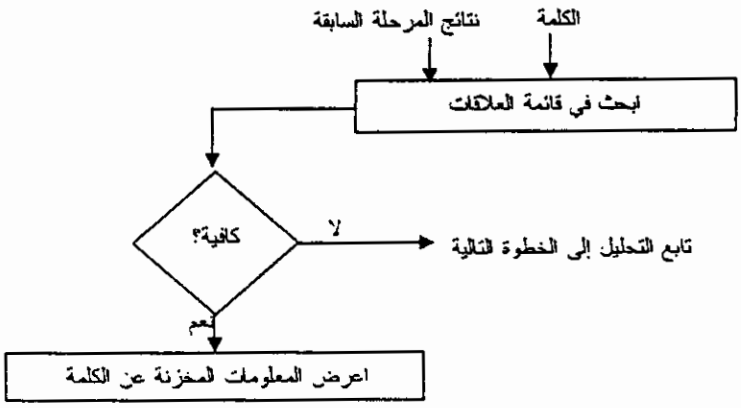
ثم يتم استخدام هاتين الزائدتين لتحديد الميزان الصرفي. ويتم ذلك من خلال مقارنة طول الكلمة مع طول الوزن الصرفي ومن ثم مقارنة مواقع الزوائد الصرفية التي في الكلمة مع مواقع الزوائد الصرفية التي يتضمنها الوزن الصرفي. وكما هو الحال بالنسبة للسوابق النحوية، إذا نجحت هذه الخطوة، يتم عرض النتائج، وإلا سيتم الانتقال إلى المرحلة التالية لمتابعة التحليل وذلك عل النحو الموضح في الشكل (١٤).



الشكل (١٤) توضيح خطوات التحليل باستخدام الزوائد الصرفية

٥,٣. التحليل باستخدام العلاقة القائمة بين السوابق واللواحق

إذا لم يكن بالإمكان تحديد نوع الكلمة من خلال المراحل السابقة، يتم الرجوع إلى علاقة السوابق باللواحق كما عبرت عنها قائمة العلاقات التي ورد ذكرها في القسم السابق. ويتم البحث عن البيانات في هذه القائمة باستخدام السابقة (نحوية كانت أو صرفية أو مركبة) التي تم تحديدها سبق. فإذا نجحت العملية في المرحلة، يتم عرض النتائج، وإلا سيتم الانتقال إلى المرحلة التاليـــة لمتابعة التحليل وذلك عل النحو الموضح في الشكل (١٥).



الشكل (١٥) توضيح خطوات التحقق من العلاقات

٦,٣. التحليل باستخدام طول الكلمة والأوزان الصرفية

هناك كلمات لا تتضمن أية زوائد وهناك كلمات تتشابه مع الأجزاء الأصلية للكلمة. ومن هنا، إذا لم تنجح عملية التحليل في المراحل السابقة في الوصول إلى نتيجة محددة فيما يتصل بتحديد نوع الكلمة، يتم اللجوء إلى طول الكلمة والأوزان الصرفية الأساسية. فإذا كان طول الكلمة أقل أو يساوي أربعة حروف يتم ترجيح الفعل على الاسم، أما إذا كان طول الكلمة أكثر من ذلك فيتم ترجيح الاسم على الفعل.

٤. مثال على عملية التحليل

لو أخذنا كلمة "وباستخداماتها"، على سبيل المثال، وطبقنا عليها خطوات المنهج، فإن عملية التحليل ستكون كما يلي:

١. يتمّ البحث عن كلمة "وباستخداماتها" في قائمة الأدوات، فتكون النتيجة: أنّ الكلمة غير موجودة، فنذهب إلى الخطوة الثانية.

٢. يتمّ تحديد السابقة النحوية في الكلمة وذلك من خلال ما يلي:

البحث عن أول حرف في كلمة "وباستخداماتها" وهو حرف "و" في قائمة السوابق، فتكون النتيجة:أنّها موجودة، ونوعها سابقة نحوية، فتصبح السابقة النحوية = "و" .

ثمّ يبحث عن "وبـ" في قائمة السوابق، فتكون النتيجة: أنها موجودة، ونوعها سابقة نحوية، فتصبح السابقة النحوية = "وبـ".

ثم يبحث عن "وبا" فتكون النتيجة: غير موجودة، ثم يبحث عن "وباسـ" فتكون النتيجة غير موجودة، ثم يبحث عن "وباستـ" فتكون النتيجة أنها موجودة، ونوعها سابقة مركبة من نحوية وصرفية، فتصبح السابقة المركبة = "وباست". وهكذا يستمر البحث حتى نهاية الكلمة وفي حالة وجودها في قائمة السوابق يتم تعديل السابقة، وهنا بعد البحث عن السابقة من بداية الكلمة حتى نهايتها، تكون السابقة النحوية= "وبـ"، والسابقة المركبة = "وباستـ".

فتكون المعلومات التي تدل عليها السابقة النحوية "وبـ":

نوع الكلمة = "اسم"، العدد = "مجهول"، الجنس = "مجهول"، الحالة = "مجهول"

والسابقة المركبة "وباستـ" تدل على أن نوع الكلمة = "مصدر" ولكن بدلالة غير مؤكدة.

٣. يتمّ تحديد اللاحقة النحوية في الكلمة وذلك ابتداءً من نهاية الكلمة، فيتم البحث ابتداءً من أخر حرف في كلمة " وباستخداماتها" وهو حرف "ا" في قائمة اللواحق، فتكون النتيجة: أنها موجودة. ونوعها لاحقة نحوية، فتصبح اللاحقة النحوية = "ا".

فيبحث عن "ها" في قائمة اللواحق، فتكون النتيجة أنها موجودة، ونوعها لاحقة نحوية فتصبح اللاحقة النحوية = "ها"، ثمّ يبحث عن "تها" في القائمة، فتكون النتيجة أنها موجودة، ونوعها لاحقة نحوية فتصبح اللاحقة النحوية = "تها".

ويستمر البحث عن اللاحقة في الكلمة من نهايتها إلى بدايتها، فتكون اللاحقة النحوية بعد انتهاء عملية البحث = "اتها"، ونوع هذه اللاحقة في قائمة اللواحق هي لاحقة نحوية، وكما تشير قائمة اللواحق، هذه اللاحقة النحوية تدل على المعلومات التالية:

نوع الكلمة = "اسم" العدد= "مفرد"، الجنس=" مؤنث"، الزمان= "مجهول"، اللاحقة النحوية = "اتها"

إضافة إلى ما دلّت عليه الخطوة السابقة من معلومات أن النوع = "مصدر"، ولكن كانت دلالتها غير مؤكدة، فنتابع إلى المرحلة التالية مع المعلومات التي دلّت عليها هذه الخطوة من حيث العدد والجنس.

٤. يتم تحديد السابقة الصرفية في الكلمة، لكن في حالة وجود سابقة مركبة من سابقة نحوية وسابقة صرفية، يتم فصل السابقة الصرفية عن السابقة النحوية باتباع ما يلي:

يتم البحث عن السابقة الصرفية في السابقة المركبة "وباستـ" ابتداء من الخلف، فيتم البحث في قائمة السوابق عن "تـ" كسابقة صرفية، فتكون النتيجة أنها غير موجودة، ثم يبحث عن "ستـ" كسابقة صرفية، فتكون النتيجة غير موجودة، ثم يبحث عن "استـ" كسابقة صرفية فتكون النتيجة أنها موجودة، فتصبح السابقة الصرفية = "استـ".

ثم يبحث عن "باستـ" فتكون النتيجة غير موجودة، ويستمر البحث من نهاية السابقة حتى بدايتها، فتكون النتيجة أن السابقة الصرفية = "استـ".

يتم في هذه الخطوة قراءة المعلومات التي تدلّ عليها السابقة الصرفية التي تمّ الحصول عليها في الكلمة. والسابقة الصرفية "استـ" تدلّ أن الكلمة ممكن أن تكون مصدر أو فعل فهي غير كافية للتمييز بين المصدر والفعل من خلالها، ولا يوجد لاحقة صرفية بالكلمة فنلجأ لتحديد الوزن الصرفي الذي تنتمي إليه هذه الكلمة بعد إزالة السوابق واللواحق النحوية من الكلمة.

٥. يتم مطابقة طول كلمة "استخدام" مع طول الأوزان الصرفية ويتم معرفة الوزن الصحيح بمطابقة السابقة الصرفية بالوزن من خلال موقع بدايتها ونهايتها بالكلمة مع موقع بداية السابقة الصرفية ونهايتها بالوزن الصرفي وكذلك الأمر بالنسبة للوسطية الصرفية واللاحقة الصرفية إن وجدوا، فيتمّ الحصول على الوزن "استفعال"، فيتم قراءة ما يدل عليه هذه الوزن من حيث النوع إضافة إلى المعلومات التي تم الحصول عليها من الخطوات السابقة من حيث العدد والجنس والزمان، فيدلّ هدا الوزن أنّ نوع الكلمة هو "مصدر" وهو كافٍ لتحديد نوع الكلمة. فيعرض المعلومات المخزنة عن الكلمة:

الكلمة: "وباستخداماتها", النوع: مصدر، العدد: مفرد، الجنس: مؤنث، الزمان: مجهول
السابقة النحوية: "وب"، السابقة الصرفية: "است"، السابقة المركبة" "وباست"
اللاحقة النحوية: "اتها"، اللاحقة الصرفية: لا يوجد، اللاحقة المركبة: لا يوجد
الوزن:استفعال

٥. خلاصة

- تعتمد منهجية الدراسة على الزوائد النحوية والصرفية وعلاقة هذه الزوائد ببعضها للكلمة داخل النص لتحديد نوع الكلمة.
- يتم اتباع خطوات متسلسلة لتحديد نوع الكلمة ابتداءً من التحقق أن الكلمة تنتمي لمجموعة الأدوات والضمائر، فإذا كانت الكلمة تنتمي للأدوات والضمائر يتم انهاء العملية.
- يتم تحديد السابقة النحوية واللاحقة والنحوية في الكلمة، فإذا كان هناك ما يدل على نوع الكلمة يتم انهاء العملية ويعرض النتائج.
- يتم تحديد السابقة الصرفية واللاحقة الصرفية في الكلمة، فإذا كان هناك ما يدل على نوع الكلمة يتم انهاء العملية ويعرض النتائج.
- يتم تحديد الوزن الصرفي الذي تنتمي إليه الكلمة وما يدل عليه ذلك الوزن، فإذا تم الوصول إلى نوع الكلمة من خلال الوزن يتم انهاء العملية ويعرض النتائج.
- يتم النظر في العلاقة بين الزوائد النحوية والزوائد الصرفية وما تدل عليه هذه العلاقة لتحديد نوع الكلمة، فإذا تم الوصول إلى نوع الكلمة من خلال العلاقة، يتم انهاء العملية ويعرض النتائج.

الفصل الخامس

تطبيق المنهج وتحليل النتائج

١. عينة الدراسة

٢. تحليل نتائج الدراسة

أ. تصنيف المفردات

ب. دور الزوائد اللفظية والأوزان الصرفية

١. الأخطاء المتصلة بالزوائد النحوية

٢. الأخطاء المتصلة بالزوائد الصرفية

٣. مناقشة النتائج

٤. خلاصة

تطبيق المنهج وتحليل النتائج

يعالج هذا الفصل الجوانب المتصلة بترجمة منهج التحليل اللفظي، على النحو الموضح في الفصل السابق، إلى برنامج بإحدى لغات البرمجة وتطبيقه على أحد النصوص العربية وتحليل النتائج ومناقشتها. يبدأ الفصل بالتعريف بالعينة التي تم اختيارها، ثم يتلو ذلك استعراض للنتائج التي توصلت إليها هذه الدراسة وتحليل لها في ضوء الخطوات المتبعة في منهج التحليل. ويختتم الفصل بمناقشة النتائج وإبراز بعض الملاحظات المتصلة بها وبالمنهج الذي استخدم بالدراسة والدراسات المقترحة للمزيد من البحث في هذا المجال.

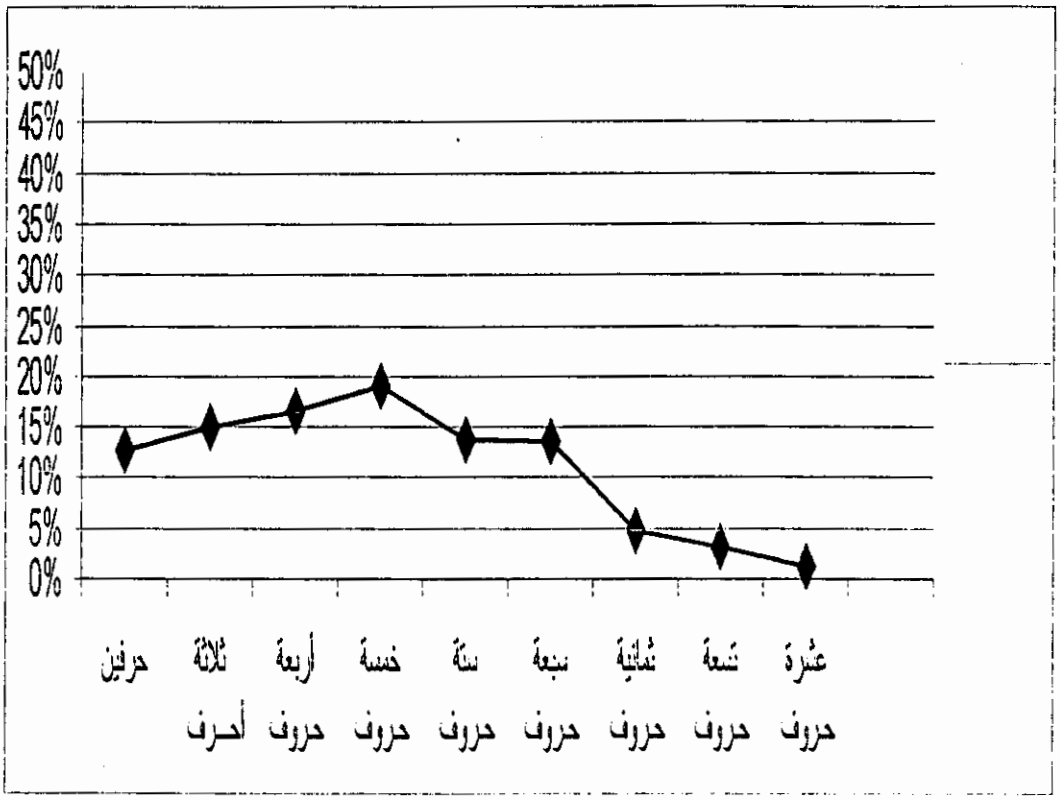
١. عينة الدراسة

لقد تمت ترجمة منهج التحليل اللفظي الوارد شرحه في الفصل السابق وتضمينه في برنامج متكامل كُتب بلغة c++ وذكرت تفاصيله في الملحق رقم (٦) من هذه الدراسة. ولاختبار الافتراضات التي بني عليها المنهج، تم تطبيق هذا البرنامج على عينتين من النصوص، ورد جزء منهما في الملحق رقم (٤). تحتوي أولاهما على ألفي كلمة، بينما تحتوي الثانية على ألف وخمسمائة كلمة.

وللتأكد من صحة النتائج، تم تحليل كلمات النصوص يدويا قبل تطبيق البرنامج عليها. والجداول رقم (١٦) ورقم (١٧) ورقم (١٨) تبين المعالم الأساسية لعينة النصوص التي استخدمت في تجريب المنهج، فالجدول رقم (١٦) يوضّح توزيع مفردات العينة حسب أطوالها بتقسيماتها الأساسية، إلى إسم وفعل وأداة. وكما يتضح من الشكل (١٦) فإن ما يزيد عن ٧٥% من مفردات النص محصورة بين ثلاثة حروف وسبعة حروف.

الجدول رقم (١٦): توزيع لأطوال المفردات الواردة في النص وأنواع المفردات من كل فئة

طول الكلمة	الأسماء		الأفعال		الأدوات	
	العدد	%	العدد	%	العدد	%
حرفين	–	–	–	–	٤٤٥	١٢,٧١
ثلاثة أحرف	١٩	٠,٥٤	٨٨	٢,٥١	٤١٧	١١,٩١
أربعة حروف	٢٣٦	٦,٧٤	٣١١	٨,٨٨	٣٥	١
خمسة حروف	٣٩٩	١١,٤	٢٤٢	٦,٩١	٢٦	٠,٧٤
ستة حروف	٤٢٤	١٢,١١	٥٨	١,٦٥	–	–
سبعة حروف	٤٧٢	١٣,٥	٦	٠,١٧	–	–
ثمانية حروف	١٦٥	٤,٧١	–	–	–	–
تسعة حروف	١١٢	٣,٢	–	–	–	–
عشرة حروف	٤٥	١,٢٩	–	–	–	–
المجموع	١٨٧٢	٥٣,٤٨	٧٠٥	٢٠,١٤	٩٢٣	٢٦,٣٧



الشكل (١٦) التحليل البنوي لأطوال الكلمات الواردة في النص

أما الجدول رقم (١٧) فيبين توزيع مفردات العينة على الأنواع المختلفة من الأسماء والأفعال التي تم اعتمادها لغايات تصنيف المفردات في هذه الدراسة. ويلاحظ من هذا الجدول أنّ نسبة المصدر والفعل الماضي والإسم العام هي الأكثر تكراراً داخل النص، حيث تراوحت نسبة هذه الأنواع بين ١٢% إلى ١٤% من مفردات النص، بينما نجد أن اسم التفضيل واسم الزمان والمكان والمصدر الصناعي وفعل الأمر هي الأقل تكراراً في النص، حيث تراوحت نسبة هذه الأنواع بين ٠% إلى ١,٥%.

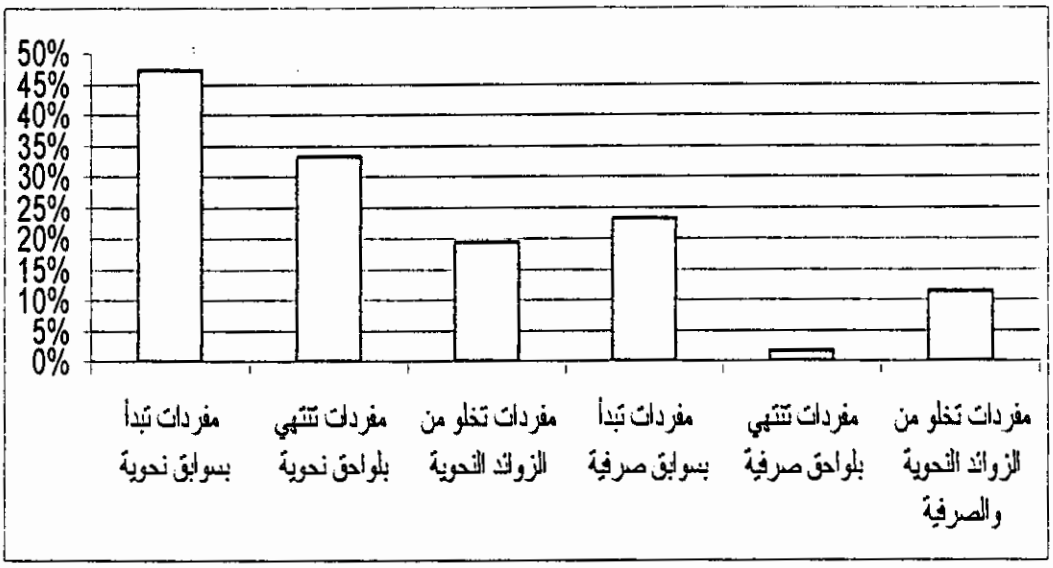
الجدول رقم (١٧): توزيع مفردات عينة الدراسة من حيث النوع

نوع المفردات	التحليل اليدوي (عدد المفردات)	%
فعل ماض	٤٥٥	١٣
فعل مضارع	٢٥٠	٧,١٤
فعل أمر	٠	٠
مصدر	٤٦٤	١٣,٢٦
اسم_عام	٤٣٢	١٢,٣٤
اسم_علم	٢١٧	٦,٢
اسم_منسوب	٢٩٧	٨,٤٨
صفة مشبهة	١٣٦	٣,٨٨
اسم_فاعل	١٧٨	٥,٠٨
اسم_مفعول	٨٨	٢,٥١
اسم_تفضيل	١٨	٠,٥١
اسم_زمان_مكان	٣٨	١,٠٨
مصدر صناعي	٤	٠,١١
الأدوات والضمائر	٩٢٣	٢٦,٣٧
المجموع الكلي	٣٥٠٠	١٠٠

امّا بالنسبة للجدول رقم (١٨) فيبين توزيع مفردات العينة على الأنواع المختلفة من الزوائد التي وردت فيها. وكما يتضح من هذا الجدول، فإن السوابق النحوية قد وردت في ما نسبته حوالي ٤٧,٤% من مفردات النص وكانت الأكثر تكرارا من بين أنواع الزوائد المختلفة، أما بالنسبة للواحق الصرفية فقد وردت في ما نسبته حوالي ١,٥٧% من مفردات النص حيث كانت الأقل تكرارا من بين الزوائد. كما يوضّح الجدول نسبة المفردات التي لم تضمّ زوائد نحوية(سوابق نحوية ولواحق نحوية) والتي بلغت ١٩,٣% من مفردات النص ومن بين هذه المفردات ١١,٤% خلت من الزوائد النحوية والصرفية معا، وتمّ توضيح ذلك من خلال الشكل (١٧).

الجدول رقم (١٨): توزيع مفردات عينة الدراسة من حيث وجود زوائد نحوية وصرفية

نوع المفردات	عدد المفردات	%
مفردات تبدأ بسوابق نحوية	١٦٥٩	٤٧,٤
مفردات تنتهي بلواحق نحوية	١١٦٥	٣٣,٢٨
مفردات تخلو من الزوائد النحوية	٦٧٦	١٩,٣
مفردات تبدأ بسوابق صرفية	٧٠٩	٢٠,٦
مفردات تنتهي بلواحق صرفية	٥٥	١,٥٧
مفردات تخلو من الزوائد النحوية والصرفية	٣٩٧	١١,٤٠



الشكل (١٧) توزيع الزوائد الواردة في مفردات النص

٢. تحليل نتائج الدراسة

فيما يلي نستعرض الملامح الأساسية للنتائج التي تمخضت عنها الدراسة ومقارنة ذلك مع نتائج التحليل اليدوي. إذ نعرض أولاً نتائج البرنامج في تصنيف المفردات، ويتلو ذلك بيان لدور الزوائد اللفظية والأوزان الصرفية التي تمّ الاعتماد عليها في تصنيف المفردات. وقد تمّ تحليل هذه الزوائد وتوضيح نسبة الأخطاء التي حصلت أثناء تنفيذ البرنامج على مفردات النص.

أ.تصنيف المفردات

لقد أظهرت النتائج، كما يتضح من الجدول رقم (١٩)، أن المنهج التحليلي الذي استخدم في هذه الدراسة قد نجح في تحديد انتماء الغالبية العظمى من المفردات إلى فئتي الأسماء والأفعال. فقد كان البرنامج قادراً على إدراك ٨٨,٤٦% من الأسماء بشكل صحيح بينما أخفق في إدراك ١١,٥٤% منها. أما بالنسبة للأفعال، فقد كانت نسبة الدقة أقل من ذلك حيث بلغت ٨٤,١١%، بينما أخفق البرنامج في ١٥,٨٩% منها. وكما يتضح من الجدول فإن هناك ٣,٨% من الأسماء والأفعال لم يستطع البرنامج التمييز بينها. أمّا فيما يتعلّق بالأدوات والضمائر فكانت نسبة إدراكها من خلال البرنامج ١٠٠% ويعود سبب ذلك إلى كون هذه الأدوات والضمائر مخزّنة كما هي في قائمة خاصة بها.

الجدول رقم (١٩): تحليل نتائج البرنامج في تحديد أنواع المفردات الرئيسية

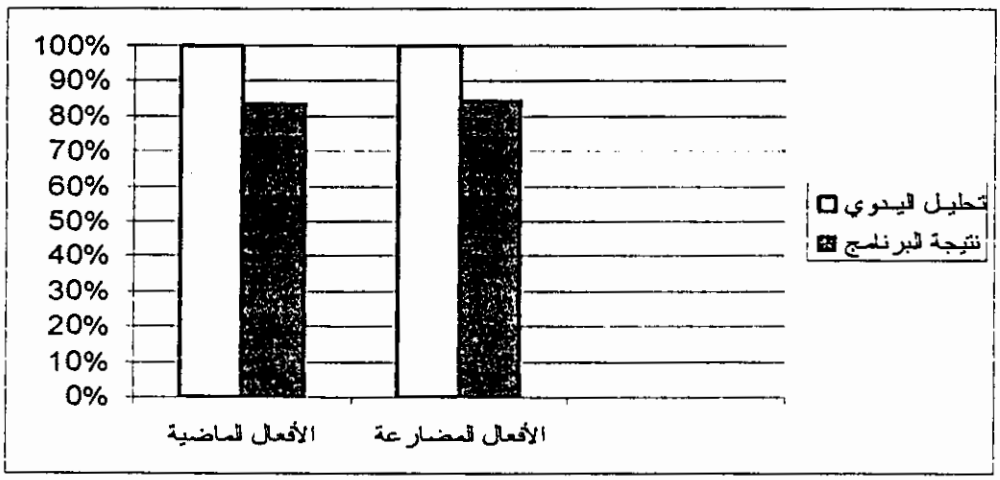
نوع المفردات	التحليل اليدوي	نتيجة البرنامج	
	عدد المفردات	عدد المفردات	نسبة الدقة %
الأسماء	١٨٧٢	١٦٥٦	٨٨,٤٦
الأفعال	٧٠٥	٥٩٣	٨٤,١١
اسم أو فعل	–	٩٨	٣,٨
الأدوات والضمائر	٩٢٣	٩٢٣	١٠٠

وفي خطوة تالية تمّ تفصيل النتائج الواردة في الجدول رقم (١٩) بحيث تمّ فصل نتائج الأفعال عن الأسماء على النحو الموضّح في الجدولين رقم (٢٠) و رقم (٢١). وكما يتضح من الجدول رقم (٢٠) فإنّ نسبة الأفعال الماضية التي أدركها البرنامج بشكل صحيح ومؤكّد قد بلغت ٧٩,١٢% من العدد الفعلي للأفعال الماضية. كما يبين الجدول أنّ نسبة الأفعال المضارعة التي أدركها البرنامج بشكل صحيح ومؤكد قد بلغت ٨٤,٤% من العدد الفعلي للأفعال المضارعة. اذ ينبغي الإشارة هنا إلى أن هناك نسبة من الأفعال قد اختلطت مع أنواع أخرى من المفردات حيث لم يتمكن البرنامج من تحديد هويتها على نحو مؤكد.

وبالإضافة إلى ذلك يبين الجدول أن هناك نسبة قليلة من الأفعال لم يكن البرنامج قادراً على تأكيدها بدقة. وكما يظهر من هذا الجدول، فإنّ هناك ما لايزيد عن خمسة بالمئة من المجموع الكلي للأفعال الماضية قد اختلط الأمر فيه على المحلل مع اسم الفاعل والمصدر. وكذا الأمر بالنسبة للفعل المضارع حيث اختلط الأمر فيه مع اسم التفضيل، حيث كان هناك مالا يزيد عن نصف بالمئة من المفردات كانت تحتمل الفعل المضارع أو اسم التفضيل. هذا مع العلم بأن جميع الكلمات التي وردت ضمن هذه الفئة تنتمي فعليا إلى أسماء التفضيل. والشكل (١٨) يوضح علاقة النتائج التي توصل إليها البرنامج مقارنة مع نتائج التحليل اليدوي، مع ملاحظة أن العينة قد خلت من أفعال الأمر.

الجدول رقم (٢٠): تحليل نتائج البرنامج في تحديد الفعل

الأفعال	نتيجة التحليل اليدوي	نتيجة البرنامج بدلالة مؤكدة		نتيجة البرنامج بدلالة غير مؤكدة		المجموع	
	عدد المفردات	عدد المفردات	نسبة الدقة %	عدد المفردات	نسبة الدقة %	عدد المفردات	النسبة %
فعل ماض	٤٥٥	٣٦٠	٧٩,١٢	٢١	٤,٦١	٣٨١	٨٣,٧٣
فعل مضارع	٢٥٠	٢١٢	٨٤,٨	٠	٠	٢١٢	٨٤,٨
فعل أمر	٠	٠	٠	٠	٠	٠	٠
المجموع	٧٠٥	٥٧٢	٨١,١٣	٢١	٤,٦١	٥٩٣	٨٤,١١

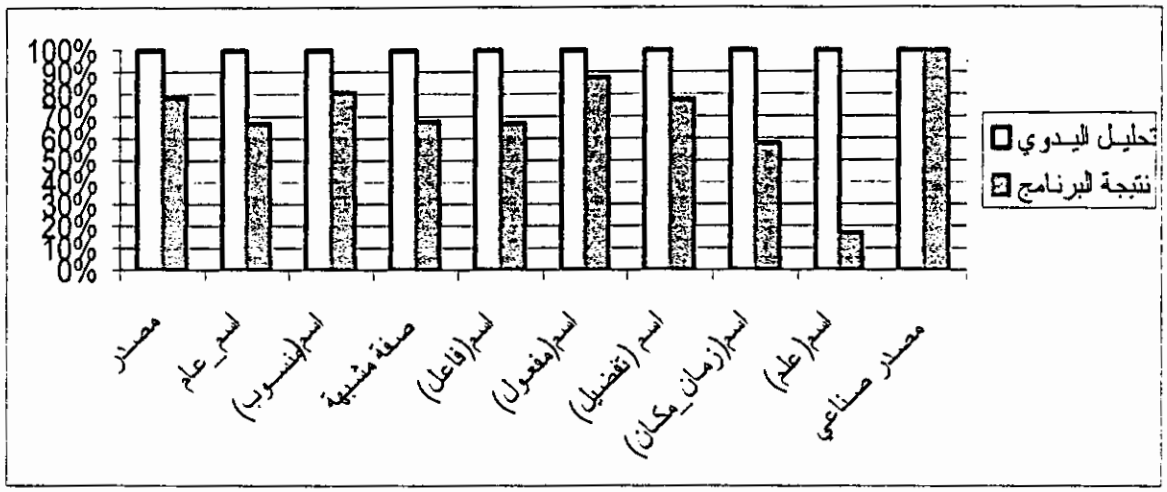


الشكل (١٨) توضيح نتيجة البرنامج في تمييز الأفعال

أما فيما يتعلق بالأسماء، على النحو المبين في الجدول رقم (٢١)، فإن النتائج تشير إلى أن المحلل لم يكن موفقا في الكثير من الحالات في تحديد الهوية الدقيقة لكل اسم من الأسماء بدلالة مؤكدة. اذ لم تتجاوز هذه النسبة المؤكدة ما مقداره ٣١% من المفردات التي صنفت أنها من باب الأسماء عند تحليلها يدوياً. وبالإضافة إلى ذلك يبين الجدول نسبة المفردات من كل فئة التي تم تحديد هويتها من خلال البرنامج بدلالة غير مؤكدة، حيث اختلطت مع غيرها من أنواع المفردات. فيتضح من الجدول أن المصدر واسم الزمان والمكان قد حظيا أكثر من غيرها في دقة التصنيف بدلالة مؤكدة، وذلك خلافا للاسم المنسوب واسم التفضيل الذي كان البرنامج أقل دقة بتحديد هويتها بدلالة مؤكدة. والشكل (١٩) يوضح علاقة النتائج التي توصل إليها البرنامج مقارنة مع نتائج التحليل اليدوي.

الجدول رقم (٢١): تحليل نتائج البرنامج في تحديد الاسم

الأسماء	نتيجة التحليل اليدوي	نتيجة البرنامج بدلالة مؤكدة		نتيجة البرنامج بدلالة غيرمؤكدة		المجموع	
	عدد المفردات	عدد المفردات	نسبة الدقة %	عدد المفردات	نسبة الدقة %	عدد المفردات	النسبة %
مصدر	٤٦٤	٢٦٦	٥٧,٣٢	٩٧	٢٠,٩٠	٣٦٣	٧٨,٢٣
اسم(عام(	٤٣٢	٩٤	٢١,٧٥	١٩٥	٤٥,١٣	٢٨٩	٦٦,٨٩
اسم(منسوب)	٢٩٧	٥٩	١٩,٦٨	١٨٢	٦١,٢٧	٢٤١	٨١,١٤
صفة مشبهة	١٣٦	٥١	٣٧,٥	٤١	٣٠١٤	٩٢	٦٧,٦٤
اسم(فاعل)	١٧٨	٢	١,١٢	١١٧	٦٥,٧٣	١١٩	٦٦,٨٥
اسم(مفعول)	٨٨	٤٤	٥٠	٣٣	٣٧.٥	٧٧	٨٧,٥
اسم (تفضيل)	١٨	٠	٠	١٤	٧٧,٧٧	١٤	٧٧,٧٧
اسم(زمان_مكان)	٣٨	٢٢	٥٧,٨٩	٠	٠	٢٢	٥٧,٨٩
اسم(علم)	٢١٧	٣٦	١٦,٥٨	٠	٠	٣٦	١٦,٥٨
مصدر صناعي	٤	٠	٠	٤	١٠٠	٤	١٠٠
المجموع	١٨٧٢	٥٧٤	٣٠,٦٦	٦٨٣	٣٦,٤٨	١٢٥٧	٦٧,١٥



الشكل (١٩) توضيح نتيجة البرنامج في تمييز الأسماء

وفيما يلي نوضح نتائج البرنامج التي أعطت أكثر من احتمال لنوع الكلمة نظرا لوجود وزن صرفي أو زائدة تشترك في نوعين من المفردات داخل النص، ويكون أحد الاحتمالين صحيح، وهذا ما نعتبره نتيجة غير دقيقة أو غير مؤكدة لنوع الكلمة وهذه النتائج موضحة من خلال الجدول رقم (٢٢)، ويوضح هذا الجدول أنّ ما يقارب ٢٠% من المفردات تم اعطاؤها أكثر من احتمال لنوعها، كما يوضح الجدول أن المصدر والاسم العام أكثر نوعين يختلطان ببعضهما.

الجدول رقم (٢٢): تفصيل نتائج البرنامج الصحيحة التي تعطي أكثر من احتمال

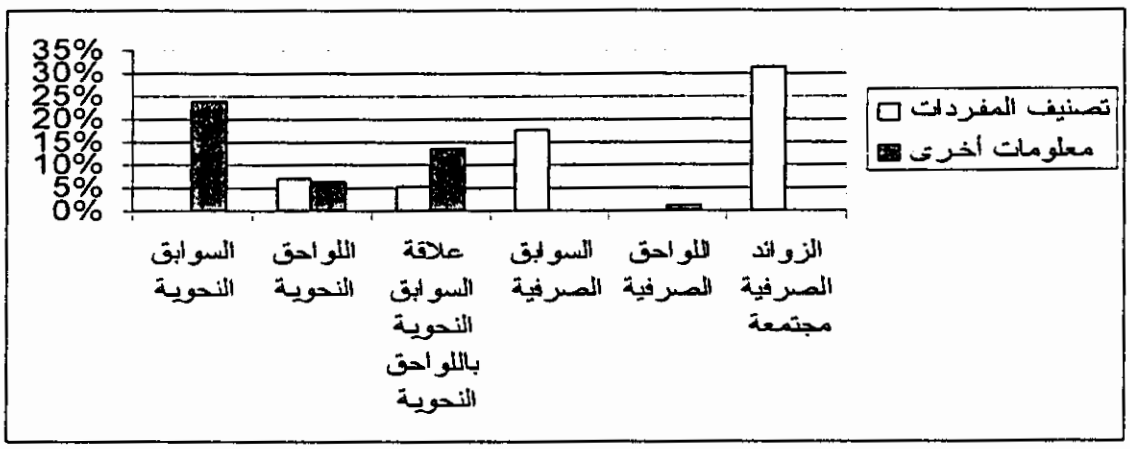
النوع (نتيجة البرنامج)	التحليل اليدوي لنتيجة البرنامج	%	عدد المفردات (نتيجة البرنامج الصحيحة)
فعل_ماض أو مصدر	٧ ١٩	١,٥٣ ٤,٠٩	٢٦
فعل ماض أو اسم_فاعل	١٤ ٤٤	٣,٠٧ ٢٤,٧	٥٨
فعل مضارع أو اسم_تفضيل	٠ ١٤	٠ ٧٧,٧٧	١٤
اسم_فاعل أو اسم_مفعول	٧٣ ٣٣	٤١,٠١ ٣٧,٥	١٠٦
صفة_مشبهة أو اسم_علم	٤١ ١١٢	٣٠,١٤ ٢٥,٩٢	١٥٣
مصدر أو اسم_علم	٧٨ ٨٣	١٦,٨١ ١٩,٢١	١٦١
اسم_منسوب مصدر_صناعي	١٨٢ ٤	٦١,٢٧ ١٠٠	١٨٦

ب. دور الزوائد اللفظية والأوزان الصرفية

لقد قامت منهجية هذه الدراسة على خطوات يتم فيها التأكد من مدى إمكانية الاعتماد على الزوائد اللفظية المختلفة في تحديد نوع الكلمة وذلك على النحو الموضح في الفصل الرابع. والجدول رقم (٢٣) يوضح دور كل نوع من هذه الزوائد في عملية التحليل وتصنيف المفردات، كما يوضح الجدول أن هناك من الزوائد لم يكن له دور في تصنيف المفردات وإنما كان له دور في الحصول على معلومات تفيد في المرحلة التالية في التحليل، كالنوع العام للمفردة مثل الاسم أو الفعل، وكالعدد والجنس، وهذه المعلومات تم تسميتها بمعلومات أخرى داخل الجدول. وكما يتضح من الشكل (٢٠) فإن الاعتماد على الوزن الصرفي، بما فيه من زوائد مختلفة، فقد فاق الاعتماد على بقية الفئات، ويليها في ذلك الاعتماد على السوابق الصرفية ومن ثم اللواحق النحوية. ولكن ينبغي القول أن هذه النتائج لم تكن صحيحة في كل الأحوال. ولعل مرد بعض الأخطاء هو عدم قدرة البرنامج أحياناً على التمييز بين ما هو زائدة لفظية وما هو من أصل الكلمة. وفيما يلي توضيح لطبيعة الأخطاء المتصلة بالزوائد:-

الجدول رقم (٢٣): أنواع الزوئد التي تم الاعتماد عليها آليا في تحديد نوع الكلمة

نوع الزوائد التي تم الاعتماد عليها آليا	تصنيف المفردات	%	معلومات أخرى	%	المجموع	%
السوابق النحوية	–	–	٨٣٨	٢٣,٩٤	٨٣٨	٢٣,٩٤
اللواحق النحوية	٢٤٥	٧	٢١٦	٦,١٧	٤٦١	١٣,١٧
علاقة السوابق النحوية باللواحق النحوية	١٩٢	٥,٤٨	٤٧٣	١٣,٥١	٦٦٥	١٩
السوابق الصرفية	٦٢٢	١٧,٧٧	–	–	٦٢٢	١٧,٧٧
اللواحق الصرفية	٣	٠,٠٩	٥٢	١,٤٨	٥٥	١,٥٧
الزوائد الصرفية مجتمعة	١٠٩٨	٣١,٣٧	–	–	١٠٩٨	٣١,٣٧



شكل (٢٠) الزوائد المختلفة التي تم الاعتماد عليها في تصنيف المفردات أو في الحصول عل معلومات أخرى

١. الأخطاء المتصلة بالزوائد النحوية

عند تحليل مفردات النص يدويا نجد أنّ السوابق النحوية تتكرّر بنسب متفاوتة. فبعضها كثير التكرار وبعضها نادر الحدوث. فمثلا السابقة النحوية "الـ" كانت أكثرها تكرارا داخل النص. أما السابقة النحوية "وكال" فكان وجودها داخل النص يكاد يكون معدوما. والجدول رقم (٢٤) يبين النتائج التفصيلية للأخطاء التي وقع فيها البرنامج عند استخلاص الزوائد النحوية.

ففي هذا الجدول نجد عدد المفردات التي تحتوي على السابقة النحوية، ونسبتها إلى العدد الكلّي لمفردات النص، إضافة إلى عدد الكلمات بالنسبة للعدد الكلي للكلمات التي أخطأ البرنامج بتحليلها باعتبارها تحتوي على هذه السابقة كزائدة نحوية، ولكنها في الحقيقة هي من أصل الكلمة. فكلمة "بصرهم"، على سبيل المثال، تحتوي على "بـ" في بدايتها ولكن حرف الباء هنا هو من أصل الكلمة وليس سابقة نحوية. وتظهر في الجدول نسبة الخطأ من العدد الكلي لمفردات النص، وكانت نسبة الخطأ في تحديد السابقة النحوية ٤,٠٥%.

الجدول رقم (٢٤): تحليل الأخطاء التي وقع فيها البرنامج عند تحديد السوابق النحوية

السابقة النحوية	عدد التكرار	%	عدد المفردات الخاطئة	%
الـ	٨٨١	٢٥,١٧	٤٨	١,٣٧
و	٢٧٩	٧,٩٧	٥	٠,١٤
والـ	١٥٣	٤,٣٧	٨	٠,٢٢
بـ	٧١	٢,٠٢	٣١	٠,٩
يـ	٦٦	١,٨٨	٣	٠,٠٨
تـ	٤٠	١,١٤	١٦	٠,٤٥
لـ	٣٤	٠,٩٧	٥	٠,١٤
للـ	٢٦	٠,٧٤	–	–
فـ	٢٦	٠,٧٤	٧	٠,٢
وبـ	٢٥	٠,٧١	٣	٠,٠٨٥
بالـ	٢٣	٠,٦٥	–	–
كـ	١٥	٠,٤٢	١٦	٠,٤٥
بالـ	١٠	٠,٢٩	–	–
فالـ	٨	٠,٢٢	–	–
كالـ	٢	٠,٠٥	–	–
المجموع	١٦٥٩	٤٧,٤	١٤٢	٤,٠٥

هذا بالنسبة للسوابق النحوية، أما بالنسبة للواحق النحوية فكان أكثرها تكرارا في النص هو اللاحقة "ة". والجدول رقم (٢٥) يعرض نتائج تحليل الأخطاء بالنسبة للواحق التي تمت

مصادفتها داخل النص. وكما يشير الجدول، فإن نسبة الخطأ في تحديد اللواحق النحوية كانت ٣,٧٧% وهي نسبة متقاربة مع نسبة الأخطاء في تحديد السوابق.

الجدول رقم (٢٥): تحليل الأخطاء التي وقع فيها البرنامج عند تحديد اللواحق النحوية

اللاحقة النحوية	عدد التكرار	نسبة التكرار %	عدد المفردات الخاطئة	%
ة	٤٧٢	١٣,٤٨	–	–
يـة	١٨٦	٥,٣١	–	–
ي	١١٣	٣,٢٢	٢٥	٠,٧١
ها	١٠١	٢,٨٨	١٠	٠,٢٨
ه	٨٥	٢,٤٢	٣٩	١,١
ات	٦١	١,٧٤	٣	٠,٠٨
ون	٤٩	١,٤	٢٣	٠,٦٥
اء	٢٩	٠,٨٢	١٣	٠,٣٧
ين	٢٦	٠,٧٤	١٧	٠,٤٨
هم	٢٢	٠,٦٢	٢	٠,٠٥
وا	٥	٠,١٤	–	–
نا	٥	٠,١٤	–	–
هما	٢	٠,٠٥	–	–
المجموع	١١٦٥	٣٣	١٣٢	٣,٧٧

٢. الأخطاء المتصلة بالزوائد الصرفية

لقد تم حصر الأوزان الصرفية التي تمثلها مفردات النص الذي تم تحليله، وتمّ حساب نسبة تكرار هذه الأوزان إلى عدد الكلمات الكلّي في النّص. وفي الجدول رقم (٢٦) نجد عرضا لنسب الأخطاء التي وقع فيها البرنامج خلال التحليل الصرفي لتحديد الوزن المناسب للكلمة المعنية. وكما يشير هذا الجدول فإن نسبة الخطأ في تحديد الوزن الصرفي إلى مجموع عدد تكرارها في النص هي ٢,٨٣%.

الجدول رقم (٢٦): النتائج المتصلة بتحديد الزوائد والأوزان الصرفية

الوزن الصرفي	عدد التكرار في النص	نسبة التكرار إلى عدد مفردات النص %	عدد المفردات الخاطئة	نسبة الخطأ(%) إلى عدد المفردات
استفعال	١٠٧	٣,٠٥	٩	٠,٢٥
افتعال	٩٣	٢,٦٦	٦	٠,١٧
فاعل	٨٨	٢,٥١	١٣	٠,٣٧
مفتعل	٧٥	٢,١٤	٨	٠,٢٢
فعيل	٧٤	٢,١١	–	–
مفعول	٦٥	١,٨٦	–	–
مفاعل	٦٣	١,٨	–	–
تفعيل	٦٢	١,٧٧	–	–
فعول	٥٧	١,٦٢	٣	٠,٠٨
مفعل	٥٦	١,٦	٥	٠,١٤
فعال	٥٦	١,٦	٤	٠,١١
افتعل	٥٣	١,٥١	٥	٠,١٤
مفعلة	٥٢	١,٤٨	٧	٠,٢
افعال	٤٥	١,٢٩	٣	٠,٠٨
استفعل	٣٤	٠,٩٧	–	–
تفاعل	٢٦	٠,٧٤	–	–
أفعل	٢٤	٠,٦٨	–	–
انفعال	٢١	٠,٦	١	٠,٠٢
تفعل	٢٠	٠,٥٧	٤	٠,١١
مستفعل	٢٠	٠,٥٧	–	–
فعائل	١٩	٠,٥٤	–	–
فواعيل	١٧	٠,٤٨	–	–
فعلاء	٣	٠,٠٨	–	–
المجموع	١١٣٠	٣٢,٢	٦٨	١,٩٤

٣. مناقشة النتائج

تشير نتائج التحليل، على النحو الذي تم بيانه في القسم السابق أن المنهج التي بنيت عليه الدراسة قد نجح في تحديد نوع المفردات لما يقارب ٧٩% من المفردات التي يضمها النص. فقد استطاع البرنامج أن يقدم نتائج دقيقة لما نسبته ٥٧,٤٦%، ونتائج غير مؤكدة لما نسبته ٢١,٤٦% من المفردات.

ولو حاولنا بيان الأسباب التي أدت إلى عدم قدرة المنهج على تحديد نوع بعض المفردات بالاعتماد على بنية الكلمة النصية مثل (تلك الواردة في الجدول رقم (٢٧))، فإن مـــن الممكـــن القول بأن هناك أربعة أسباب رئيسية هي كما يلي:

١- الاعتماد على زوائد في الكلمة لتحديد نوعها وهي من أصل الكلمة، وكان لهذا السبب ما نسبته ٧,٧٥% من المفردات التي أخطأ البرنامج في تحديد نوعها الصحيح ومن الأمثلة على ذلك في النص مثل كلمة "يهود" فهنا اعتبر البرنامج حرف الياء سابقة نحوية يستدل من خلالهـــا على أن نوع الكلمة فعل مضارع.

٢- خلو الكلمات من الزوائد النحوية والصرفية والتي شكلت ١١,٤% من مفردات النص، فلا يوجد ما يدل في بنية هذه المفردات ما يدلّ على نوعها ومن الأمثلة على ذلك في النص كلمة "ظهر". فبغياب التشكيل في هذه الكلمة لا نستطيع تحديد نوعها هل هي اسم أم فعل.

٣- احتواء الكلمة على زوائد نحوية لا تحدد نوع الكلمة، فهي تلتصق بالأسماء والأفعـــال مثل اللاحقة النحوية "تهم" في الفعل "وضربتهم" والاسم "ولعبتهم" فيقوم النظـــام بتحليـــل الكلمـــة حسب المنهجية المتبعة كالتالي:

(و + ضرب + تهم) فهنا السابقة النحوية الـــ "و" واللاحقة النحوية "تهم" لا نستطيع من خلالهمـــا تحديد نوع الكلمة. وهذا السبب شكل حوالي ٧% من الأخطاء التي حصلت في نتائج البرنامج.

٤- وجود أوزان تخلط بين نوعين من المفردات مثل الوزن الصـــرفي "مســـتفعل" فهـــذا الوزن الصرفي يمكن أن يكون لاسم فاعل، او لاسم مفعول.

الجدول رقم (٢٧): عينة من الأخطاء أثناء تطبيق الخوارزمية

الكلمة	نوعها الفعلي	النتيجة الخاطئة من حيث (النوع، العدد، الجنس، الزمن)
وذهب	فعل ماض	(مجهول، مجهول، مجهول، مجهول)
أهم	اسم تفضيل	(اسم، مجهول، مجهول، مجهول)
الدين	اسم عام	(اسم، جمع+مثنى، مجهول، مجهول)
جل	اسم عام	(مجهول، مجهول، مجهول، مجهول)
يهود	اسم علم	(فعل مضارع، مجهول، مذكر، مجهول)

وقد أشارت نتائج الدراسة إلى أن الفئات المختلفة من الزوائد اللفظية قد أسهمت بأشكال مختلفة في تحديد نوع المفردات وبيان خصائصها. وفيما يلي بعض الملاحظات التي برزت في تطبيق المنهج الذي بنته هذه الدراسة:

١- إن من الممكن الاعتماد على معظم السوابق النحوية في تحديد النوع الرئيسي للكلمة (اسم، فعل ، أداة) ولا نستطيع من خلال السابقة النحوية تحديد النوع الفرعي للكلمة.

٢- أن معظم اللواحق النحوية ليس لها دلالة على نوع الكلمة بل لها دلالة على خصائص للكلمة مثل العدد والجنس.

٣- نلجأ في كثير من الحالات إلى الزوائد الصرفية والعلاقة القائمة بينها ضمن ما يعرف بالأوزان الصرفية لتحديد نوع الكلمة، فالوزن الصرفي هو غالباً المميز الوحيد في بنية الكلمة ليدل على نوعها من الأسماء والأفعال.

٤- في غياب التشكيل, ليس من السهل التفريق بين بعض الأنواع ومن هنا جاء الخلط أحياناً في الدلالات غير المؤكدة النتائج. إلا أنّ هذا الخلط كان ضمن حدود احتمالين فقط.

٤ .خلاصة

تتميز المفردات العربية بنسبة عالية من الزوائد اللفظية التي ترد في أماكن متفرقة منها، فبعضها يلحق ببداية المفردات وبعضها في نهايتها، وبعضها يأتي في مواقع متوسطة منها كما أوضحت ذلك في بناء الكلمة في اللغة العربية.

وقد برهنت نتائج هذه الدراسة أنه يمكن الاعتماد على بنية الكلمة داخل النص العربي في تحديد نوعها وبيان كثير من خصائصها بما لا يتجاوز حدود النسبة التي أظهرتها الدراسة وفق الظروف التي أعدت فيها وطبيعة العينة التي استخدمتها. فلا نستطيع الاعتماد بشكل كامل على بنية الكلمة في تحديد نوع الكلمة النصية غير المشكولة وذلك للأسباب التي ذكرناها آنفاً في تحليل الأخطاء. وكما يتضح فإنه لابدّ من استخدام أساليب أخرى تقوم على سياق النص مثل الإفادة من الأدوات وعلاقة الكلمات ببعضها.

لقد اعتمدت الدراسات الأخرى التي ورد ذكرها خلال هذه الدراسة على المعاجم وعلى سياق النص بشكل أساسي وهو ما لم تفعله هذه الدراسة. وفي ضوء النتائج التي خلصت إليها الدراسة الحالية، فإن من الممكن القول بأن الأساليب الأخرى، وبشكل خاص سياق النص، يمكن أن تستخدم لدعم المنهج الذي تم تطبيقه في هذا البحث.

قائمة المراجع

1- Abuleil, S., Alsamara, K., and Evens M., **Acquisition System for Arabic Noun Morphology,** 2002 .
http://www.cs.um.edu.mt/~mros/WSL/papers/abuleil:alsamara:evens.pdf.

2- Abuleil, S., and Evens, M., **Discovering Lexical Information by Tagging Arabic Newspaper Text,** 1998.
http://citeseer.nj.nec.com/correct/583651

3- Adurize, I., Agirre, E., Aldezadal, I., Algria, I., Arregi, X., Arriola J., Artola, X., Gojenola, K., Maritxalar, A., and Sarasola, K., **A Word-Grammar Based Morphological Analyzer for Agglutinative Languages,** 2000.
http://acl.ldc.upenn.edu/C/C00/C00-1001.pdf

4- Al-Fares, W., and Roeck, A., **A Morphologically Sensitive Clustering Algorithm for Identifying Arabic Roots,** Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics. Hong Kong, Oct 1-8, 2000. pp.199-206
http://acl.ldc.upenn.edu//P/P00/P00-1026.pdf

5- Al-Fedaghi, S., and Al-Anzi, F., **A New Algorithm to Generate Arabic Root-Pattern Forms.** King Fahd University of Petroleum and Minerals, Saudi Arabia., pp 4-7, 1989.

6- Al-Jlayl, M., and Frieder, O., **Improving The Retrieval Effectiveness Via Light Stemming Approach,** 2002, pp. 340-347.
http://www.iro.umontreal.ca/~douzidif/Reference/CIKM02-Arabic.pdf

7- Al-Shalabi, R., and Evens, M., **A Computational Morphology System for Arabic.** Ph.D. Thesis, Computer Science, Illinois Institute of Technology, USA, 1998.

8- Buckwalter, T., **Buckwalter Arabic Morphological Analyzer Version 1.0,** Linguistic Data Consortium (LDC) catalog number LDC2002L49
http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2002L49

9- Darwish, K., **Building A Shallow Arabic Morphological Analyzer in One Day,** In Proceedings of the Association for Computational Linguistics , University of Pennsylvania, Philadelphia, PP. 47-54 , 2002.
http://www.cs.um.edu.mt/~mros/WSL/papers/darwish.pdf

10- Eklund, R., **A Probabilistic Word Class Tagging Module Based on Surface Pattern Matching,** 1993, pp. 83–95.
http://www.ida.liu.se/~g-robek/nodalida93tag.pdf.

11- Farghali, A., and Senellart, J., **Intuitive Coding of the Arabic Lexicon,** MT summit IX Workshop on Machine Trnaslation for semitic languages, U.S.A., 2003.
http://www.systransoft.com/Technology/2003_MTIX_AR.pdf

12- Freeman, A., **Brill's POS Tagger and A Morphology Parser for Arabic,** Ph.D. thesis, Ann Arbor, Michigan, USA, 2001.
http://www.elsnet.org/arabic2001/freeman.pdf.

13- Habash, N., **Large Scale Lexeme Based Arabic Morphological Generation,** University of Maryland Institute for Advanced Computer Studies, U.S.A, 2004
http://www.nizarhabash.com/publications/taln-04-1.pdf

14- Ibrahim, F., **A Syntactically-Based Preprocessor for A Limited Experimental Arabic Document Retrieval System,** Ph.D. Thesis. Loughborough University of Technology, 1986.

15- Jackson, M., **Part-of-Speech Guessing With Bogus Statistics,** 2002
http://www.reitter-it-media.de/kurse/latex/latex-challenge/latex-ex.pdf.

16- Khoja, S., **Arabic Part-of-Speech Tagger,** 1999.
http://archimedes.fas.harvard.edu/mdh/arabic/NAACL.pdf.

17- McAalester, D., **Natural Language Syntax and Semantics,** 1994
http://citeseer.ist.psu.edu/5387.html

18- Mustafa, S., And Masoud, F., **A Backward Algorithm for Lexical Analysis of Textual Arabic Words,** ABHATH AL-YARMOUK, Basic Sciences and Engineering series, vol 9(1), 2000, pp 91 -125.

19- Sung, Y., **Markov Random Eld Based English Part of Speech Tagging System,** 1996.
http://brainew.com/research/publish/coling96/coling96_proc.pdf

20- Tahir,Y., Chenfour, N., and Harti, M., **Realization of A Morphological Analyzer for Arabic Language Text,** Workshop on Information Technology, Rabat, Morroco, March 2003.

21- Talmon, R., and Wintner, S., **Morphological Tagging of the Qur'an,** 2001.
http://cs.haifa.ac.il/~shuly/publications/talmon-wintner-eacl03.pdf.

22- Trommer, J., and Kallulli, D., **A Morphological Tagger for Standard Albanian,** 1997
http://iit.demokritos.gr/skel/bci03_workshop/papers/SESSION4_2-9_Trommer.pdf.

23- Van Mol, M., and Paulussen, H., **Natural Language Prcessing and Arabic: the Leuven Tandem Apparoach,** 2004.
http://www.ilt.kuleuven.ac.be/ilt/arabic/_pdf/talnfez2004.pdf

24- Xerox Research Center Europe .
http://www.xrce.xerox.com, Access Date: march/15/2004

٢٥- الدحداح، أنطوان. **معجم قواعد العربية العالمية**. مكتبة لبنان، بيروت، ١٩٩٩.

٢٦- مصطفى، سليمان. **الادراك الآلي للزوائد اللفظية في معالجة السلاسل الرمزية العربية بالحاسوب استبعاد الأداة "ال"**. أبحاث اليرموك، سلسلة العلوم الأساسية والهندسية، مجلد ٥، عدد ٣، ١٩٩٦.

٢٧- مشهور، غيث. **نظام لفهم اللغة العربية بمساعدة الحاسب**. ندوة استخدام اللغة العربية في الحاسب الآلي، معهد الكويت للأبحاث العلمية، دار الرازي، ١٩٨٥، ص ص ٢٤٧-٢٦٤.

٢٨- سليمان، معتصم. **نموذج محوسب لمحلل نحوي للجمل الاسمية غير المشكولة في اللغة العربية**. رسالة ماجستير، جامعة آل البيت، المفرق، الأردن، ٢٠٠٢.

٢٩- علي، نبيل. **اللغة العربية والحاسوب**. مجمع اللغة العربية، عمّان، ١٩٨٨.

الملحق رقم(١)
الزوائد الصرفية المختلفة

السوابق	فـ	واسطة١	عـ	واسطة ٢	لـ	اللواحق	النوع	العدد	الجنس	الزمان
أ		ا		0		0	اسم/فعل	جمع_مفرد	مجهول	مجهول
أ		ا		ي		0	اسم_عام	جمع	مؤنث	مجهول
أ		0		ا		0	اسم_عام	جمع	مؤنث	مجهول
أ		0		0		0	فعل/اسم_تفضيل	مفرد	مجهول	مجهول
أ		0		0		اء	صفة_مشبهة	جمع	مذكر	مجهول
إ		0		ا		0	مصدر	مفرد	مجهول	مجهول
است		0		ا		0	مصدر	مفرد	مذكر	مجهول
است		0		0		0	فعل	مفرد	مذكر	ماض_أمر
ا		ت		ا		0	مصدر	مفرد	مذكر	مجهول
ا		ت		0		0	فعل	مفرد	مذكر	ماض_أمر
ا		0		0		ل	فعل	مفرد	مذكر	ماض
ا		0		و		0	فعل	مفرد	مذكر	ماض
ان		0		ا		0	مصدر	مفرد	مذكر	مجهول
ان		0		0		0	فعل	مفرد	مذكر	ماض_أمر
تت		0		0		0	فعل	مفرد	مؤنث	ضارع
ت		ا		0		0	اسم_عام/مصدر/فعل	مفرد	مذكر	مجهول
ت		ا		ي		0	اسم_عام	جمع	مؤنث	مجهول
ت		0		ا		0	اسم_عام	مفرد	مجهول	مجهول
ت		0		0		0	اسم_عام/مصدر/فعل	مفرد	مجهول	مجهول
ت		0		0		0	فعل	مفرد	مذكر_مؤنث	مضارع
ت		0		0		ة	مصدر	مفرد	مؤنث	مجهول
ت		0		ي		0	مصدر	مفرد	مذكر	مجهول
تم		0		0		0	فعل	مفرد	مذكر	ماض_أمر
ست		0		0		0	فعل	مفرد	مجهول	مضارع
0		ا		0		0	اسم_فعل	مفرد	مذكر	مجهول
0		ا		و		0	اسم	مفرد	مذكر	مجهول
0		0		اء		0	اسم	جمع	مؤنث	مجهول
0		0		ا		0	اسم_عام/مصدر	مجهول	مجهول	مجهول
0		0		ا		ى	اسم	جمع	مؤنث	مجهول
0		0		ا		ي	اسم_منسوب	جمع	مؤنث	مجهول
0		0		ا		ية	اسم_منسوب/ مصدر_صناعي	مفرد	مؤنث	مجهول
0		0		ع		0	فعل	مفرد	مذكر	ماض_أمر
0		0		0		اء	اسم_عام/صفة_مشبهة	مجهول	مؤنث	مجهول

0		0		0		ان	اسم_عام	مفرد	مذكر	مجهول
0		0		0		ة	اسم_عام	مفرد	مؤنث	مجهول
0		0		0		ى	اسم_عام	مفرد	مؤنث	مجهول
0		0		و		0	اسم_عام/مصدر	مجهول	مجهول	مجهول
0		0		و		ة	اسم_عام	مفرد	مؤنث	مجهول
0		0		ي		0	صفة	مفرد	مذكر	مجهول
0		وا		0		0	اسم_فاعل	جمع	مؤنث	مجهول
0		و		0		0	فعل	مفرد	مذكر	ماض
0		ي		و		0	صفة	مفرد	مذكر	مجهول
مت		ا		0		0	اسم_فاعل/اسم_مفعول	مفرد	مذكر	مجهول
مت		0		0		0	اسم_فاعل/اسم_مفعول	مفرد	مذكر	مجهول
متم		0		0		0	اسم_فاعل/اسم_مفعول	مفرد	مذكر	مجهول
مست		0		0		0	اسم_فاعل/اسم_مفعول	مفرد	مذكر	مجهول
م		ا		0		0	صفة_مشبهة/اسم_فاعل/اسم_مفعول	مفرد_جمع	مؤنث	مجهول
م		ا		ي		0	اسم_زمان/اسم_مكان	جمع	مؤنث	مجهول
م		ت		0		0	صفة_مشبهة/اسم_فاعل/اسم_مفعول	مفرد	مذكر	مجهول
م		0		ا		0	اسم_آلة	مفرد	مذكر	مجهول
م		0		0		0	اسم_زمان/اسم_مكان	مفرد	مذكر	مجهول
م		0		0		ة	اسم_عام/اسم_آلة	مفرد	مؤنث	مجهول
م		ا		0		ة	مصدر	مفرد	مؤنث	مجهول
م		0		و		0	اسم_مفعول	مفرد	مذكر	مجهول
من		0		0		0	صفة_مشبهة/اسم_فاعل/اسم_مفعول	مفرد	مذكر	مجهول
ن		0		0		0	فعل	جمع	مجهول	مضارع
ي		0		0		0	فعل	مفرد	مذكر	مضارع

جدول رقم (٣٠):الأوزان الصرفية

الملحق رقم (٢)

الزوائد النحوية المختلفة وعلاقتها بالزوائد الصرفية

أ.السوابق النحوية

السوابق	نوع السابقة	تشابه مع حروف أصلية	النوع	الجنس	العدد	الزمن
أأ	١٠	١	فعل	مجهول	مفرد	مضارع
أأل	١٠	١	اسم	مجهول	مجهول	مجهول
أب	١٠	١	اسم	مجهول	مجهول	مجرور
أبأ	١٠	١	اسم	مجهول	مجهول	مجرور
أبأل	١٠	١	اسم	مجهول	مجهول	مجرور
أت	١٠	١	فعل	مجهول	مجهول	مضارع
أس	١٠	١	فعل	مجهول	مجهول	مضارع
أسأ	١٠	١	فعل	مجهول	مفرد	مضارع
أست	١٠	١	فعل	مجهول	مجهول	مضارع
أسن	١٠	١	فعل	مجهول	جمع	مضارع
أسي	١٠	١	فعل	مجهول	مجهول	مضارع
أف	١٠	١	اسم/فعل	مجهول	مجهول	مجهول
أفأ	١٠	١	فعل	مجهول	مفرد	مضارع
أفب	١٠	١	اسم	مجهول	مجهول	مجرور
أفت	١٠	١	فعل	مجهول	مجهول	مضارع
أفن	١٠	١	فعل	مجهول	جمع	مضارع
أفي	١٠	١	فعل	مذكر	مجهول	مضارع
أك	١٠	١	اسم	مجهول	مجهول	مجرور
أكال	١٠	١	اسم	مجهول	مجهول	مجرور
أل	١٠	١	اسم	مجهول	مجهول	مجرور
ألت	١٠	١	فعل	مجهول	مفرد	مضارع
ألل	١٠	١	اسم	مجهول	مجهول	مجرور
ألن	١٠	١	فعل	مجهول	جمع	مضارع
ألي	١٠	١	فعل	مذكر	مجهول	مضارع
أن	١٠	١	فعل	مجهول	جمع	مضارع
أو	١٠	١	اسم/فعل	مجهول	مجهول	مجهول
أوأ	١٠	١	فعل	مجهول	مفرد	مضارع
أوب	١٠	١	اسم	مجهول	مجهول	مجرور

أوت	١٠	١	فعل	مجهول	مجهول	مضارع
أول	١٠	١	اسم	مجهول	مجهول	مجرور
أون	١٠	١	فعل	مجهول	جمع	مضارع
أوي	١٠	١	فعل	مجهول	مجهول	مضارع
أي	١٠	١	فعل	مجهول	مجهول	مضارع
ا	١٠	١	فعل	مجهول	مفرد	أمر
اللا	١٠	١	اسم	مجهول	مجهول	مجهول
ب	١٠	١	اسم	مجهول	مجهول	مجهول
بأ	١٠	١	اسم	مجهول	مجهول	مجرور
بال	١٠	١	اسم	مجهول	مجهول	مجرور
ت	١٠	١	فعل	مجهول	مجهول	مضارع
س	١٠	١	فعل	مجهول	مجهول	مضارع
سأ	١٠	١	فعل	مجهول	مفرد	مضارع
ست	١٠	١	فعل	مجهول	مجهول	مضارع
سن	١٠	١	فعل	مجهول	جمع	مضارع
سي	١٠	١	فعل	مجهول	مجهول	مضارع
ف	١٠	١	اسم/فعل	مجهول	مجهول	مجهول
فأ	١٠	١	فعل	مجهول	مفرد	مضارع
فال	١٠	١	اسم	مجهول	مجهول	مجهول
فب	١٠	١	اسم	مجهول	مجهول	مجرور
فبال	١٠	١	اسم	مجهول	مجهول	مجرور
فت	١٠	١	فعل	مجهول	مجهول	مضارع
فسأ	١٠	١	فعل	مجهول	مفرد	مضارع
فست	١٠	١	فعل	مجهول	مجهول	مضارع
فسن	١٠	١	فعل	مجهول	جمع	مضارع
فسي	١٠	١	فعل	مجهول	مجهول	مضارع
فك	١٠	١	اسم	مجهول	مجهول	مجرور
فكال	١٠	٠	اسم	مجهول	مجهول	مجرور
فل	١٠	١	اسم	مجهول	مجهول	مجرور
فلأ	١٠	٠	فعل	مجهول	مفرد	مضارع
فلت	١٠	١	فعل	مجهول	مجهول	مضارع
فلل	١٠	١	اسم	مجهول	مجهول	مجرور
فلن	١٠	٠	فعل	مجهول	جمع	مضارع

فلي	١٠	١	فعل	مجهول	مجهول	مضارع
فن	١٠	١	فعل	مجهول	جمع	مضارع
في	١٠	١	فعل	مجهول	مجهول	مضارع
ك	١٠	١	اسم	مجهول	مجهول	مجهول
كأ	١٠	٠	اسم	مجهول	مجهول	مجرور
كال	١٠	٠	اسم	مجهول	مجهول	مجرور
ل	١٠	١	اسم/فعل	مجهول	مجهول	مجهول
لأ	١٠	١	فعل	مجهول	مجهول	مضارع
لت	١٠	١	فعل	مجهول	مجهول	مضارع
لل	١٠	١	اسم	مجهول	مجهول	مجرور
لن	١٠	١	فعل	مجهول	جمع	مضارع
لي	١٠	١	فعل	مجهول	مجهول	مضارع
ن	١٠	١	فعل	مجهول	جمع	مضارع
و	١٠	١	اسم/فعل	مجهول	مجهول	مجهول
وأ	١٠	١	فعل	مجهول	مجهول	مضارع
وال	١٠	١	اسم	مجهول	مجهول	مجهول
وب	١٠	١	اسم	مجهول	مجهول	مجرور
وبال	١٠	١	اسم	مجهول	مجهول	مجرور
وت	١٠	١	فعل	مجهول	مجهول	مضارع
وسأ	١٠	١	فعل	مجهول	مفرد	مضارع
وست	١٠	١	فعل	مجهول	مجهول	مضارع
وسن	١٠	١	فعل	مجهول	جمع	مضارع
وسي	١٠	١	فعل	مجهول	مجهول	مضارع
وك	١٠	١	اسم	مجهول	مجهول	مجرور
وكال	١٠	١	اسم	مجهول	مجهول	مجرور
ول	١٠	١	اسم/فعل	مجهول	مجهول	مجرور
ولأ	١٠	١	فعل	مجهول	مفرد	مضارع
ولت	١٠	١	فعل	مجهول	مجهول	مضارع
ولل	١٠	١	اسم	مجهول	مجهول	مجرور
ولن	١٠	١	فعل	مجهول	جمع	مضارع
ولي	١٠	١	فعل	مجهول	مجهول	مضارع
ون	١٠	١	فعل	مجهول	جمع	مضارع
وي	١٠	١	فعل	مجهول	مجهول	مضارع

ي	١٠	١	فعل	مجهول	مجهول	مضارع
أ	٣٠	١	اسم_تفضيل/فعل	مجهول	مجهول	مجهول
است	٢٠	١	مصدر/فعل	مجهول	مجهول	مجهول
الأ	٢١	١	صفة_مشبهة	مجهول	مجهول	مجهول
الاست	٢١	٠	مصدر	مجهول	مجهول	مجهول
الان	٢١	٠	مصدر	مجهول	مجهول	مجهول
الم	٢١	١	مصدر/اسم_فاعل/اسم_مفعول/اسم_آلة	مجهول	مجهول	مجهول
المت	٢١	٠	صفة_مشبهة	مجهول	مجهول	مجهول
المن	٢١	١	صفة_مشبهة	مجهول	مجهول	مجهول
ان	٢٠	١	مصدر/فعل	مجهول	مجهول	مجهول
بأ	٢١	١	اسم_عام	مجهول	مجهول	مجهول
باست	٢١	٠	مصدر	مجهول	مجهول	مجهول
بالأ	٢١	١	صفة_مشبهة	مجهول	مجهول	مجهول
بالاست	٢١	٠	مصدر	مجهول	مجهول	مجهول
بالان	٢١	٠	اسم_مصدر	مجهول	مجهول	مجهول
بالم	٢١	١	مصدر/اسم_فاعل/اسم_مفعول/اسم_آلة	مجهول	مجهول	مجهول
بالمت	٢١	١	صفة_مشبهة	مجهول	مجهول	مجهول
بالمن	٢١	١	صفة_مشبهة	مجهول	مجهول	مجهول
بان	٢١	١	مصدر	مجهول	مجهول	مجهول
بم	٢١	١	مصدر/اسم_فاعل/اسم_مفعول/اسم_آلة	مجهول	مجهول	مجهول
بمت	٢١	٠	اسم_فاعل/اسم_مفعول	مجهول	مجهول	مجهول
بمن	٢١	١	اسم_فاعل/اسم_مفعول	مجهول	مجهول	مجهول
كأ	٢١	١	اسم_عام	مجهول	مجهول	مجهول
كاست	٢١	٠	مصدر	مجهول	مجهول	مجهول
كالأ	٢١	١	صفة_مشبهة	مجهول	مجهول	مجهول
كالاست	٢١	٠	مصدر	مجهول	مجهول	مجهول
كالان	٢١	٠	مصدر	مجهول	مجهول	مجهول
كالم	٢١	٠	مصدر/اسم_فاعل/اسم_مفعول/اسم_آلة	مجهول	مجهول	مجهول
كالمت	٢١	٠	صفة_مشبهة	مجهول	مجهول	مجهول
كالمن	٢١	٠	صفة_مشبهة	مجهول	مجهول	مجهول
كان	٢١	١	مصدر	مجهول	مجهول	مجهول
كت	٢١	١	صفة_مشبهة	مجهول	مجهول	مجهول
كم	٢١	١	مصدر/اسم_فاعل/اسم_مفعول/اسم_آلة	مجهول	مجهول	مجهول

كمت	٢١	١	صفة_مشبهة	مجهول	مجهول	مجهول
كمن	٢١	١	صفة_مشبهة	مجهول	مجهول	مجهول
لأ	٢١	١	فعل	مجهول	مجهول	مجهول
لت	٢١	٠	فعل	مجهول	مجهول	مجهول
م	٢٠	١	مصدر/اسم_فاعل/اسم_مفعول/اسم_آلة	مجهول	مجهول	مجهول
مت	٢٠	١	صفة_مشبهة/اسم_فاعل/اسم_مفعول	مجهول	مجهول	مجهول
مست	٢٠	٠	صفة_مشبهة/اسم_فاعل/اسم_مفعول	مجهول	مجهول	مجهول
من	٢٠	١	صفة_مشبهة/اسم_فاعل/اسم_مفعول	مجهول	مجهول	مجهول

جدول رقم (٣١)

ب- اللواحق النحوية:

اللواحق	نوعها	تشابه مع حروف أصلية	نوع الكلمة	العدد	الجنس	الزمان
ا	٥٠	١	اسم/فعل	مثنى	مذكر/مؤنث	مجهول
اء	٤٠	١	اسم_عام	مفرد	مؤنث	مجهول
اءه	٥٤	١	اسم_عام	مفرد	مذكر	مجهول
انك	٥٤	١	اسم_عام	مفرد	مذكر/مؤنث	مجهول
انكم	٥٤	١	اسم_عام	جمع	مذكر	مجهول
انكما	٥٤	١	اسم_عام	مثنى	مذكر/مؤنث	مجهول
انكن	٥٤	١	اسم_عام	جمع	مؤنث	مجهول
انا	٥٤	١	اسم_عام	جمع	مذكر/مؤنث	مجهول
انهم	٥٤	١	اسم_عام	جمع	مذكر	مجهول
انهما	٥٤	١	اسم_عام	مثنى	مذكر/مؤنث	مجهول
انهن	٥٤	١	اسم_عام	جمع	مؤنث	مجهول
ات	٥٠	١	اسم	جمع	مؤنث	مجهول
اتك	٥٠	١	اسم	مفرد	مذكر/مؤنث	مجهول
اتكم	٥٠	١	اسم	جمع	مذكر	مجهول
اتكما	٥٠	١	اسم	مثنى	مذكر/مؤنث	مجهول
اتكن	٥٠	١	اسم	جمع	مؤنث	مجهول
اته	٥٠	١	اسم	مفرد	مذكر	مجهول
اتها	٥٠	١	اسم	مفرد	مؤنث	مجهول
اتهم	٥٠	١	اسم	جمع	مذكر	مجهول
اتهما	٥٠	١	اسم	مثنى	مذكر/مؤنث	مجهول
اتهن	٥٠	١	اسم	جمع	مؤنث	مجهول
اك	٥٠	١	فعل	مفرد	مفرد/مؤنث	مجهول
اكم	٥٠	١	فعل	جمع	مذكر	مجهول
اكما	٥٠	١	فعل	مثنى	مذكر/مؤنث	مجهول
اكن	٥٠	١	فعل	جمع	مؤنث	مجهول
ان	٥٠	١	اسم/فعل	مثنى	مذكر/مؤنث	مجهول
اه	٥٠	١	فعل	مفرد	مذكر	مجهول
اها	٥٠	١	فعل	مفرد	مؤنث	مجهول

اهم	٥٠	١	فعل	جمع	مذكر	مجهول
اهما	٥٠	١	فعل	مثنى	مذكر/مؤنث	مجهول
اهن	٥٠	١	فعل	جمع	مؤنث	مجهول
ة	٦٠	٠	اسم	مفرد	مؤنث	مجهول
ت	٥٠	١	اسم/فعل	مفرد	مذكر/مؤنث	مجهول
تا	٥٠	١	اسم/فعل	مثنى	مؤنث	مجهول
تاك	٥٠	٠	اسم	مفرد	مذكر	مجهول
تاكم	٥٠	٠	اسم	جمع	مذكر	مجهول
تاكما	٥٠	٠	اسم	مثنى	مذكر/مؤنث	مجهول
تاكن	٥٠	٠	اسم	جمع	مؤنث	مجهول
تان	٥٠	٠	اسم	مثنى	مؤنث	مجهول
تاه	٥٠	٠	اسم/فعل	مفرد	مذكر	مجهول
تاها	٥٠	٠	اسم/فعل	مفرد	مؤنث	مجهول
تاهم	٥٠	٠	اسم	جمع	مذكر	مجهول
تاهم	٥٠	٠	اسم	جمع	مذكر	مجهول
تاهما	٥٠	٠	اسم	مثنى	مذكر/مؤنث	مجهول
تاهن	٥٠	٠	اسم	جمع	مؤنث	مجهول
تاي	٥٠	٠	اسم	مفرد	مذكر/مؤنث	مجهول
تك	٥٠	٠	مجهول	مفرد	مذكر/مؤنث	مجهول
تك	٥٤	٠	اسم/فعل	مفرد	مذكر/مؤنث	مجهول
تكم	٥٠	٠	اسم/فعل	جمع	مذكر	مجهول
تكم	٥٤	٠	اسم/فعل	جمع	مذكر	مجهول
تكما	٥٠	٠	اسم/فعل	مثنى	مذكر/مؤنث	مجهول
تكما	٥٤	٠	اسم/فعل	مثنى	مذكر/مؤنث	مجهول
تكن	٥٠	٠	اسم/فعل	جمع	مؤنث	مجهول
تكن	٥٤	٠	اسم_فعل	جمع	مؤنث	مجهول
تما	٥٠	٠	فعل	مثنى	مذكر/ممؤنث	ماضٍ
تمو	٥٠	٠	فعل	جمع	مذكر	ماضٍ
تمونا	٥٠	٠	فعل	جمع	مذكر/مؤنث	ماضٍ
تن	٥٠	٠	فعل	جمع	مؤنث	ماضٍ
تنا	٥٤	٠	اسم_فعل	جمع	مذكر/مؤنث	مجهول
ته	٦٠	٠	اسم/فعل	مفرد	مذكر	مجهول
تهم	٦٠	٠	فعل	جمع	مذكر	ماضٍ
تهما	٦٠	٠	فعل	مثنى	مذكر/مؤنث	ماضٍ
تهن	٦٠	٠	فعل	جمع	مؤنث	ماضٍ
ك	٥٠	٠	اسم/فعل	مجهول	مذكر/مؤنث	مجهول
كم	٥٠	٠	اسم/فعل	جمع	مذكر	مجهول
كما	٥٠	٠	اسم/فعل	مثنى	مذكر/مؤنث	مجهول
كمو	٥٠	٠	اسم/فعل	جمع	مذكر/مؤنث	مجهول
كن	٥٠	٠	اسم/فعل	جمع	مؤنث	مجهول
كي	٥٠	٠	اسم/فعل	مجهول	مذكر/مؤنث	مجهول

ن	٥٠	٠	فعل	جمع	مونث	مجهول
نا	٥٠	٠	اسم/فعل	جمع	مذكر/مؤنث	مجهول
ناك	٥٠	٠	فعل	مفرد	مذكر/مؤنث	ماض
ناكم	٥٠	٠	فعل	جمع	مذكر	ماض
ناكما	٥٠	٠	فعل	مثنى	مذكر/مؤنث	ماض
ناكن	٥٠	٠	فعل	جمع	مؤنث	ماض
ناه	٥٠	٠	فعل	مفرد	مذكر	ماض
ناها	٥٠	٠	فعل	مفرد	مؤنث	ماض
ناهم	٥٠	٠	فعل	جمع	مذكر	ماض
ناهما	٥٠	٠	فعل	مثنى	مذكر/مؤنث	ماض
ناهن	٥٠	٠	فعل	جمع	مؤنث	ماض
نه	٥٠	٠	فعل	مفرد	مذكر	مجهول
نها	٥٠	٠	فعل	مفرد	مؤنث	مجهول
نهم	٥٠	٠	فعل	جمع	مذكر	مجهول
نهما	٥٠	٠	فعل	مثنى	مذكر/مؤنث	مجهول
نهن	٥٠	٠	فعل	جمع	مؤنث	مجهول
ني	٥٠	٠	فعل	مفرد	مذكر/مؤنث	مجهول
ه	٥٠	٠	اسم/فعل	مفرد	مذكر	مجهول
ها	٥٠	٠	اسم/فعل	مفرد	مؤنث	مجهول
هم	٥٠	٠	اسم/فعل	جمع	مذكر	مجهول
هما	٥٠	٠	اسم/فعل	مثنى	مذكر/مؤنث	مجهول
هن	٥٠	٠	اسم/فعل	جمع	مؤنث	مجهول
وا	٥٠	٠	فعل	جمع	مذكر	مجهول
وت	٤٠	٠	اسم	مفرد	مذكر	مجهول
وتك	٥٤	٠	اسم	مفرد	مذكر/مؤنث	مجهول
وتكم	٥٤	٠	اسم	جمع	مذكر	مجهول
وتكما	٥٤	٠	اسم	مثنى	مذكر/مؤنث	مجهول
وتكن	٥٤	٠	اسم	جمع	مؤنث	مجهول
وتنا	٥٤	٠	اسم	جمع	مذكر/مؤنث	مجهول
وته	٥٤	٠	اسم	مفرد	مذكر	مجهول
وتهم	٥٤	٠	اسم	جمع	مذكر	مجهول
وتهما	٥٤	٠	اسم	مثنى	مذكر/مؤنث	مجهول
وتهن	٥٤	٠	اسم	جمع	مؤنث	مجهول
وك	٥٠	٠	فعل	مفرد	مذكر/مؤنث	مجهول
وكم	٥٠	٠	فعل	جمع	مذكر	مجهول
وكما	٥٠	٠	فعل	مثنى	مذكر/مؤنث	مجهول
وكن	٥٠	٠	فعل	جمع	مؤنث	مجهول
ون	٥٠	٠	اسم/فعل	جمع	مذكر	مجهول
ونه	٥٠	٠	فعل	مفرد	مذكر	مضارع
ونها	٥٠	٠	فعل	مفرد	مؤنث	مضارع
ونهم	٥٠	٠	فعل	جمع	مذكر	مضارع

ونهما	٥٠	٠	فعل	مثنى	مذكر/مؤنث	مضارع
ونهن	٥٠	٠	فعل	جمع	مؤنث	مضارع
وه	٥٠	٠	فعل	مفرد	مذكر	مجهول
وها	٥٠	٠	فعل	مفرد	مؤنث	مجهول
وهم	٥٠	٠	فعل	جمع	مذكر	مجهول
وهما	٥٠	٠	فعل	مثنى	مذكر/مؤنث	مجهول
وهن	٥٠	٠	فعل	جمع	مؤنث	مجهول
ي	٥٠	١	مجهول	مفرد	مذكر	مجهول
ية	٥٠	١	اسم_منسوب/مصدر_صناعي	مفرد	مؤنث	مجهول
يتك	٥٠	١	اسم_منسوب/مصدر_صناعي	مفرد	مذكر/مؤنث	مجهول
يتكم	٥٠	٠	اسم_منسوب/مصدر_صناعي	جمع	مذكر	مجهول
يتكما	٥٠	٠	اسم_منسوب/مصدر_صناعي	مثنى	مذكر/مؤنث	مجهول
يتكن	٥٠	٠	اسم_منسوب/مصدر_صناعي	جمع	مؤنث	مجهول
يتنا	٥٠	٠	اسم_منسوب/مصدر_صناعي	جمع	مذكر/مؤنث	مجهول
يته	٥٠	٠	اسم_منسوب/مصدر_صناعي	مفرد	مذكر	مجهول
يتهم	٥٠	٠	اسم_منسوب/مصدر_صناعي	جمع	مذكر	مجهول
يتهما	٥٠	٠	اسم_منسوب/مصدر_صناعي	مثنى	مذكر/مؤنث	مجهول
يتهن	٥٠	٠	اسم_منسوب/مصدر_صناعي	جمع	مؤنث	مجهول
ين	٥٠	٠	اسم/فعل	جمع/مثنى	مجهول	مجهول
ينه	٥٠	٠	فعل	مفرد	مذكر	مضارع
ينها	٥٠	٠	فعل	مفرد	مؤنث	مضارع
ينهم	٥٠	٠	فعل	جمع	مذكر	مضارع
ينهما	٥٠	٠	فعل	مثنى	مذكر/مؤنث	مضارع
ينهن	٥٠	٠	فعل	جمع	مؤنث	مضارع
يه	٥٠	٠	فعل	مفرد	مذكر	مجهول
يها	٥٠	٠	فعل	مفرد	مؤنث	مجهول
يهم	٥٠	٠	فعل	جمع	مذكر	مجهول
يهما	٥٠	٠	فعل	مثنى	مذكر/مؤنث	مجهول
يهن	٥٠	٠	فعل	جمع	مؤنث	مجهول

جدول رقم (٣٢)

ج- علاقة السوابق النحوية باللواحق النحوية:

السوابق	اللواحق	النوع	العدد	الجنس	الومن
أكال	ين	اسم	جمع	مذكر	مجهول
أكال	ان	اسم	مثنى	مذكر/مؤنث	مجهول
أكال	ون	اسم	جمع	مذكر	مجهول
فبال	ين	اسم	جمع	مذكر	مجهول
فبال	ان	اسم	مثنى	مذكر/مؤنث	مجهول

فبال	ون	اسم	جمع	منكر	مجهول
فكال	ين	اسم	جمع	منكر	مجهول
فكال	ان	اسم	مثني	منكر/مؤنث	مجهول
فكال	ون	اسم	جمع	منكر	مجهول
وبال	ين	اسم	جمع	منكر	مجهول
وبال	ان	اسم	مثني	منكر/مؤنث	مجهول
وبال	ون	اسم	جمع	منكر	مجهول
وكال	ين	اسم	جمع	منكر	مجهول
وكال	ان	اسم	مثني	منكر/مؤنث	مجهول
وكال	ون	اسم	جمع	منكر	مجهول
أكال	ي	اسم	مفرد	منكر	مجهول
فبال	ي	اسم	مفرد	منكر	مجهول
فكال	ي	اسم	مفرد	منكر	مجهول
وبال	ي	اسم	مفرد	منكر	مجهول
وكال	ي	اسم	مفرد	منكر	مجهول
فلت	هما	فعل	مثني	منكر/مؤنث	مضارع
فلت	كما	فعل	مثني	منكر/مؤنث	مضارع
وست	هما	فعل	مثني	منكر/مؤنث	مضارع
وست	كما	فعل	مثني	منكر/مؤنث	مضارع
ولت	هما	فعل	مثني	منكر/مؤنث	مضارع
ولت	كما	فعل	مثني	منكر/مؤنث	مضارع
ولت	هما	فعل	مثني	منكر/مؤنث	مضارع
فلت	هم	فعل	جمع	منكر	مضارع
فلت	هن	فعل	جمع	مؤنث	مضارع
فلت	ها	فعل	مفرد	مؤنث	مضارع
فلت	كم	فعل	جمع	منكر	مضارع
فلت	كن	فعل	جمع	مؤنث	مضارع
فلت	ين	فعل	مفرد	مؤنث	مضارع
فلت	ان	فعل	مثني	منكر/مؤنث	مضارع
فلت	ون	فعل	جمع	منكر	مضارع
وست	هم	فعل	جمع	منكر	مضارع
وست	هن	فعل	جمع	مؤنث	مضارع
وست	ها	فعل	مفرد	مؤنث	مضارع
وست	كم	فعل	جمع	منكر	مضارع
وست	كن	فعل	جمع	مؤنث	مضارع
وست	ين	فعل	مفرد	مؤنث	مضارع
وست	ان	فعل	مثني	منكر/مؤنث	مضارع
وست	ون	فعل	جمع	منكر	مضارع
ولت	هم	فعل	جمع	منكر	مضارع

ولت	هن	فعل	جمع	مؤنث	مضارع
ولت	ها	فعل	مفرد	مؤنث	مضارع
ولت	كم	فعل	جمع	مذكر	مضارع
ولت	كن	فعل	جمع	مؤنث	مضارع
ولت	ين	فعل	مفرد	مؤنث	مضارع
ولت	ان	فعل	مثنى	مذكر/مؤنث	مضارع
ولت	ون	فعل	جمع	مذكر	مضارع
ولت	هم	فعل	جمع	مذكر	مضارع
ال	تين	اسم	مثني	مؤنث	مجهول
ال	تان	اسم	مثني	مؤنث	مجهول
فت	هما	فعل	مثنى	مذكر/مؤنث	مضارع
فت	كما	فعل	مثنى	مذكر/مؤنث	مضارع
لت	كما	فعل	مثنى	مذكر/مؤنث	مضارع
لت	هما	فعل	مثنى	مذكر/مؤنث	مضارع
لت	كما	فعل	مثنى	مذكر/مؤنث	مضارع
أت	ين	فعل	مفرد	مؤنث	مضارع
أت	ان	فعل	مثنى	مذكر/مؤنث	مضارع
أت	ون	فعل	جمع	مذكر	مضارع
أي	ان	فعل	مثنى	مذكر/مؤنث	مضارع
أي	ون	فعل	جمع	مذكر	مضارع
ست	هم	فعل	جمع	مذكر	مضارع
ست	هن	فعل	جمع	مؤنث	مضارع
ست	ها	فعل	مفرد	مؤنث	مضارع
ست	كم	فعل	جمع	مذكر	مضارع
ست	كن	فعل	جمع	مؤنث	مضارع
ست	ين	فعل	مفرد	مؤنث	مضارع
ست	ان	فعل	مثنى	مذكر/مؤنث	مضارع
ست	ون	فعل	جمع	مذكر	مضارع
فت	هم	فعل	جمع	مذكر	مضارع
فت	هن	فعل	جمع	مؤنث	مضارع
فت	ها	فعل	مفرد	مؤنث	مضارع
فت	كم	فعل	جمع	مذكر	مضارع
فت	كن	فعل	جمع	مؤنث	مضارع
فت	ين	فعل	مفرد	مؤنث	مضارع
فت	ان	فعل	مثنى	مذكر/مؤنث	مضارع
فت	ون	فعل	جمع	مذكر	مضارع
لت	هن	فعل	جمع	مؤنث	مضارع
لت	ها	فعل	مفرد	مؤنث	مضارع
لت	كم	فعل	جمع	مذكر	مضارع

لت	كن	فعل	جمع	مؤنث	مضارع
لت	ين	فعل	مفرد	مؤنث	مضارع
لت	ان	فعل	مثنى	مذكر/مؤنث	مضارع
لت	ون	فعل	جمع	مذكر	مضارع
لت	هم	فعل	جمع	مذكر	مضارع
لت	هن	فعل	جمع	مؤنث	مضارع
لت	ها	فعل	مفرد	مؤنث	مضارع
لت	كم	فعل	جمع	مذكر	مضارع
لت	كن	فعل	جمع	مؤنث	مضارع
لت	ين	فعل	مفرد	مؤنث	مضارع
لت	ان	فعل	مثنى	مذكر/مؤنث	مضارع
لت	ون	فعل	جمع	مذكر	مضارع
لل	ين	اسم	جمع	مذكر	مجهول
لل	ان	اسم	مثنى	مذكر/مؤنث	مجهول
لل	ون	اسم	جمع	مذكر	مجهول
وك	ين	اسم	جمع	مذكر	مجهول
وك	ان	اسم	مثنى	مذكر/مؤنث	مجهول
وك	ون	اسم	جمع	مذكر	مجهول
ال	اء	اسم	جمع	مذكر	مجهول
ال	ات	اسم	جمع	مؤنث	مجهول
ال	ون	اسم	جمع	مذكر	مجهول
ال	ين	اسم	جمع/مثنى	مذكر	مجهول
ال	ان	اسم	مثنى	مذكر	مجهول
ف	ناهم	فعل	جمع	مذكر/مؤنث	مجهول
و	ناهم	فعل	جمع	مذكر/مؤنث	مجهول
ت	هما	فعل	مثنى	مذكر/مؤنث	مضارع
ت	كما	فعل	مثنى	مذكر/مؤنث	مضارع
ا	وا	فعل	جمع	مذكر	أمر
ب	ين	اسم	جمع	مذكر	مجهول
ب	ان	اسم	مؤنث	مذكر/مؤنث	مجهول
ب	ون	اسم	جمع	مذكر	مجهول
ت	هم	اسم/فعل	جمع	مذكر	مضارع
ت	هن	اسم/فعل	جمع	مؤنث	مضارع
ت	ها	اسم/فعل	مفرد	مؤنث	مضارع
ت	كم	اسم/فعل	جمع	مذكر	مضارع
ت	كن	اسم/فعل	جمع	مؤنث	مضارع
ت	ين	فعل	مفرد	مؤنث	مضارع
ت	ان	فعل	مثنى	مذكر/مؤنث	مضارع
ت	ون	فعل	جمع	مذكر	مضارع

ي	ون	فعل	جمع	مذكر	مضارع
ي	ان	فعل	مثنى	مذكر	مضارع
ا	ا	فعل	مثنى	مذكر/مؤنث	أمر
ا	ن	فعل	جمع	مؤنث	أمر
ا	ت	فعل	مفرد	مذكر/مؤنث	ماض
ب	ي	اسم	مجهول	مذكر	مجهول
ت	ت	اسم	مفرد	مؤنث	مجهول

جدول رقم (٣٣)

د- علاقة السوابق النحوية بالسوابق الصرفية:

السوابق النحوية		السوابق الصرفية	النوع	العدد	الجنس	الزمن
و|ف|أ	كال	أ	صفة_مشبهة	مجهول	مجهول	مجهول
		است	مصدر	مجهول	مجهول	مجهول
		ان	مصدر	مجهول	مجهول	مجهول
		م	مصدر/اسم_فاعل/اسم_مفعول/اسم_آلة	مجهول	مجهول	مجهول
		مت	صفة_مشبهة	مجهول	مجهول	مجهول
		من	صفة_مشبهة	مجهول	مجهول	مجهول
	أل	أ	صفة_مشبهة	مجهول	مجهول	مجهول
		است	مصدر	مجهول	مجهول	مجهول
		ان	مصدر	مجهول	مجهول	مجهول
		م	مصدر/اسم_فاعل/اسم_مفعول/اسم_آلة	مجهول	مجهول	مجهول
		مت	صفة_مشبهة	مجهول	مجهول	مجهول
		من	صفة_مشبهة	مجهول	مجهول	مجهول
	ك	أ	اسم	مجهول	مجهول	مجهول
		ت	صفة	مجهول	مجهول	مجهول
		است	مصدر	مجهول	مجهول	مجهول
		ان	مصدر	مجهول	مجهول	مجهول
		م	مصدر/اسم_فاعل/اسم_مفعول/اسم_آلة	مجهول	مجهول	مجهول
		مت	صفة_مشبهة	مجهول	مجهول	مجهول
		من	صفة_مشبهة	مجهول	مجهول	مجهول
	ب	أ	اسم	مجهول	مجهول	مجهول
		ت	اسم_عام	مجهول	مجهول	مجهول
		است	مصدر	مجهول	مجهول	مجهول
		ان	مصدر	مجهول	مجهول	مجهول
		م	مصدر/اسم_فاعل/اسم_مفعول_آلة	مجهول	مجهول	مجهول
		مت	اسم_فاعل/اسم_مفعول	مجهول	مجهول	مجهول
		من	اسم_فاعل/اسم_مفعول	مجهول	مجهول	مجهول
	بال	أ	صفة_مشبهة	مجهول	مجهول	مجهول
		است	مصدر	مجهول	مجهول	مجهول
		ان	مصدر	مجهول	مجهول	مجهول
		م	مصدر/اسم_فاعل/اسم_مفعول/اسم_آلة	مجهول	مجهول	مجهول
		مت	صفة_مشبهة	مجهول	مجهول	مجهول
		من	صفة_مشبهة	مجهول	مجهول	مجهول
	ل	أ	فعل	مجهول	مجهول	مجهول
		ت	فعل	مجهول	مجهول	مجهول

جدول رقم (٣٤)

هـ- علاقة اللواحق الصرفية باللواحق النحوية:

اللواحق الصرفية	اللواحق النحوية	النوع	العدد	الجنس	الزمان
اء	نا	اسم	جمع	مذكر_مؤنث	مجهول
	ه	اسم	مفرد	مذكر	مجهول
	هم	اسم	جمع	مذكر	مجهول
	هما	اسم	مثنى	مذكر_مؤنث	مجهول
	هن	اسم	جمع	مؤنث	مجهول
	ك	اسم	مفرد	مذكر_مؤنث	مجهول
	كم	اسم	جمع	مذكر	مجهول
	كما	اسم	مثنى	مذكر_مؤنث	مجهول
	كن	اسم	جمع	مؤنث	مجهول
ة	نا	اسم/فعل	جمع	مذكر_مؤنث	مجهول
	ه	اسم/فعل	مفرد	مذكر	مجهول
	هم	اسم/فعل	جمع	مذكر	مجهول
	هما	اسم/فعل	مثنى	مذكر_مؤنث	مجهول
	هن	اسم/فعل	جمع	مؤنث	مجهول
	ك	اسم/فعل	مفرد	مذكر_مؤنث	مجهول
	كم	اسم/فعل	جمع	مذكر	مجهول
	كما	اسم/فعل	مثنى	مذكر_مؤنث	مجهول
	كن	اسم/فعل	جمع	مؤنث	مجهول
ية	نا	مصدرصناعي/اسم_منسوب	جمع	مذكر_مؤنث	مجهول
	ه	مصدر_صناعي/اسم_منسوب	مفرد	مذكر	مجهول
	هم	مصدر_صناعي/اسم_منسوب	جمع	مذكر	مجهول
	هما	مصدر_صناعي /اسم_منسوب	مثنى	مذكر_مؤنث	مجهول
	هن	مصدر_صناعي /اسم_منسوب	جمع	مؤنث	مجهول
	ك	مصدر_صناعي /اسم_منسوب	مفرد	مذكر_مؤنث	مجهول
	كم	مصدر_صناعي /اسم_منسوب	جمع	مذكر	مجهول
	كما	مصدر_صناعي /اسم_منسوب	مثنى	مذكر_مؤنث	مجهول
	كن	مصدر_صناعي /اسم_منسوب	جمع	مؤنث	مجهول
وت	نا	اسم	جمع	مذكر_مؤنث	مجهول
	ه	اسم	مفرد	مذكر	مجهول
	هم	اسم	جمع	مذكر	مجهول
	هما	اسم	مثنى	مذكر_مؤنث	مجهول
	هن	اسم	جمع	مؤنث	مجهول
	ك	اسم	مفرد	مذكر_مؤنث	مجهول
	كم	اسم	جمع	مذكر	مجهول
	كما	اسم	مثنى	مذكر_مؤنث	مجهول
	كن	اسم	جمع	مؤنث	مجهول

جدول رقم (٣٥)

الملحق رقم (٣)

الأدوات والضمائر

الأداة	نوعها	العدد	الجنس
أجل	حرف	مجهول	مجهول
ألا	أداة	مجهول	مجهول
أم	حرف	مجهول	مجهول
أما	أداة	مجهول	مجهول
أمام	ظرف	مجهول	مجهول
أن	أداة	مجهول	مجهول
أنت	ضمير	مفرد	مذكر/مؤنث
أنتم	ضمير	جمع	مذكر
أنتما	ضمير	مثني	مذكر/مؤنث
أنتن	ضمير	جمع	مؤنث
أني	أداة	مجهول	مجهول
أو	حرف	مجهول	مجهول
أي	أداة	مجهول	مجهول
أيا	أداة	مجهول	مجهول
أيان	أداة	مجهول	مجهول
أين	أداة	مجهول	مجهول
إنهم	أداة	جمع	مذكر
إذ	أداة	مجهول	مجهول
إذا	أداة	مجهول	مجهول
إذما	أداة	مجهول	مجهول
إذما	أداة	مجهول	مجهول
إذن	أداة	مجهول	مجهول
إلا	أداة	مجهول	مجهول
إلى	حرف	مجهول	مجهول
إن	أداة	مجهول	مجهول
إنك	أداة	مفرد	مذكر_مؤنث
إنكم	أداة	جمع	مذكر
إنكما	أداة	مثني	مذكر_مؤنث
إنكن	أداة	جمع	مؤنث
إنه	أداة	مفرد	مذكر
إنها	أداة	مفرد	مؤنث
إنهما	أداة	مثني	مذكر_مؤنث
إنهن	أداة	جمع	مؤنث
إي	أداة	مجهول	مجهول
الألاء	اسم_موصول	جمع	مذكر
التي	اسم_موصول	مفرد	مؤنث
الذي	اسم_موصول	مفرد	مذكر
الذين	اسم_موصول	جمع	مذكر

اللائي	اسم_موصول	جمع	مؤنث
اللاتي	اسم_موصول	جمع	مؤنث
اللتان	اسم_موصول	مثنى	مؤنث
اللذان	اسم_موصول	مثنى	مذكر
اللواتي	اسم_موصول	جمع	مؤنث
بئس	أداة	مجهول	مجهول
بعد	ظرف	مجهول	مجهول
بعدك	ظرف	مفرد	مذكر_مؤنث
بعدكم	ظرف	مثنى	مذكر
بعدكما	ظرف	مثنى	مذكر_مؤنث
بعدكن	ظرف	جمع	مؤنث
بعده	ظرف	مفرد	مذكر
بعدها	ظرف	مفرد	مؤنث
بعدهم	ظرف	جمع	مذكر
بعدهما	ظرف	مثنى	مذكر_مؤنث
بعدهن	ظرف	جمع	مؤنث
بل	حرف	مجهول	مجهول
بلى	حرف	مجهول	مجهول
بين	ظرف	مجهول	مجهول
بينك	ظرف	مفرد	مذكر_مؤنث
بينكم	ظرف	جمع	مذكر
بينكما	ظرف	مثنى	مذكر_مؤنث
بينكن	ظرف	جمع	مؤنث
بينه	ظرف	مفرد	مذكر
بينها	ظرف	مفرد	مؤنث
بينهم	ظرف	جمع	مذكر
بينهما	ظرف	مثنى	مذكر_مؤنث
بينهن	ظرف	جمع	مؤنث
تلك	اسم_اشارة	مفرد	مؤنث
ثم	حرف	مجهول	مجهول
ثمت	ظرف	مجهول	مجهول
جنوب	ظرف	مجهول	مجهول
حاشا	حرف	مجهول	مجهول
حاشا	أداة	مجهول	مجهول
حتى	حرف	مجهول	مجهول
حيثما	أداة	مجهول	مجهول
حين	ظرف	مجهول	مجهول
خلا	حرف	مجهول	مجهول
خلف	ظرف	مجهول	مجهول
ذلك	اسم_اشارة	مفرد	مذكر
ربّ	حرف	مجهول	مجهول
ريثما	ظرف	مجهول	مجهول

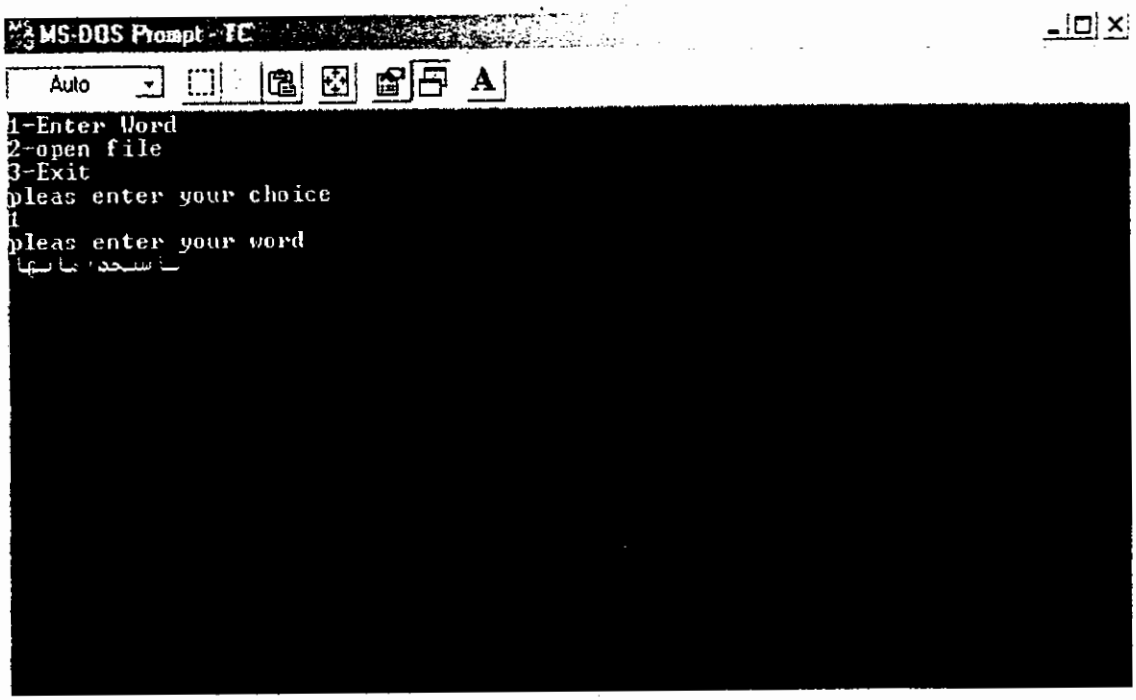
سوف	أداة	مجهول	مجهول
شمال	ظرف	مجهول	مجهول
عدا	أداة	مجهول	مجهول
على	حرف	مجهول	مجهول
عليك	حرف	مفرد	مذكر_مؤنث
عليكم	حرف	جمع	مذكر
عليكما	حرف	مثنى	مذكر_مؤنث
عليكن	حرف	جمع	مؤنث
عليه	حرف	مفرد	مذكر
عليها	حرف	مفرد	مؤنث
عليهم	حرف	جمع	مذكر
عليهما	حرف	مثنى	مذكر_مؤنث
عليهن	حرف	جمع	مؤنث
عن	حرف	مجهول	مجهول
عند	ظرف	مجهول	مجهول
فوق	ظرف	مجهول	مجهول
في	حرف	مجهول	مجهول
قبل	ظرف	مجهول	مجهول
قد	أداة	مجهول	مجهول
قط	ظرف	مجهول	مجهول
كأن	أداة	مجهول	مجهول
كلا	أداة	مجهول	مجهول
كلما	أداة	مجهول	مجهول
كي	حرف	مجهول	مجهول
كيف	أداة	مجهول	مجهول
لا	أداة	مجهول	مجهول
لات	أداة	مجهول	مجهول
لعل	أداة	مجهول	مجهول
لكن	أداة	مجهول	مجهول
لم	أداة	مجهول	مجهول
لما	أداة	مجهول	مجهول
لن	أداة	مجهول	مجهول
لو	أداة	مجهول	مجهول
لولا	حرف	مجهول	مجهول
لولا	أداة	مجهول	مجهول
لوما	حرف	مجهول	مجهول
ليت	أداة	مجهول	مجهول
ليس	أداة	مجهول	مجهول
ما	أداة	مجهول	مجهول
متى	أداة	مجهول	مجهول
مذ	حرف	مجهول	مجهول
من	حرف	مجهول	مجهول

منذ	حرف	مجهول	مجهول
مهما	أداة	مجهول	مجهول
هؤلاء	اسم_اشارة	جمع	مذكر
هتان	اسم_اشارة	مثنى	مؤنث
هذا	اسم_اشارة	مفرد	مذكر
هذان	اسم_اشارة	مثنى	مذكر
هذان	اسم_اشارة	مثنى	مذكر
هذه	اسم_اشارة	مفرد	مؤنث
هذي	اسم_اشارة	مفرد	مؤنث
هل	أداة	مجهول	مجهول
هلا	حرف	مجهول	مجهول
هم	ضمير	جمع	مذكر
هما	ضمير	مثنى	مذكر_مؤنث
هن	ضمير	جمع	مؤنث
هنا	اسم_اشارة	مجهول	مجهول
هناك	اسم_اشارة	مجهول	مجهول
هنالك	اسم_اشارة	مجهول	مجهول
هو	ضمير	مفرد	مذكر
هي	ضمير	مفرد	مؤنث
وراء	ظرف	مجهول	مجهول
يا	أداة	مجهول	مجهول
يسار	ظرف	مجهول	مجهول
يمين	ظرف	مجهول	مجهول

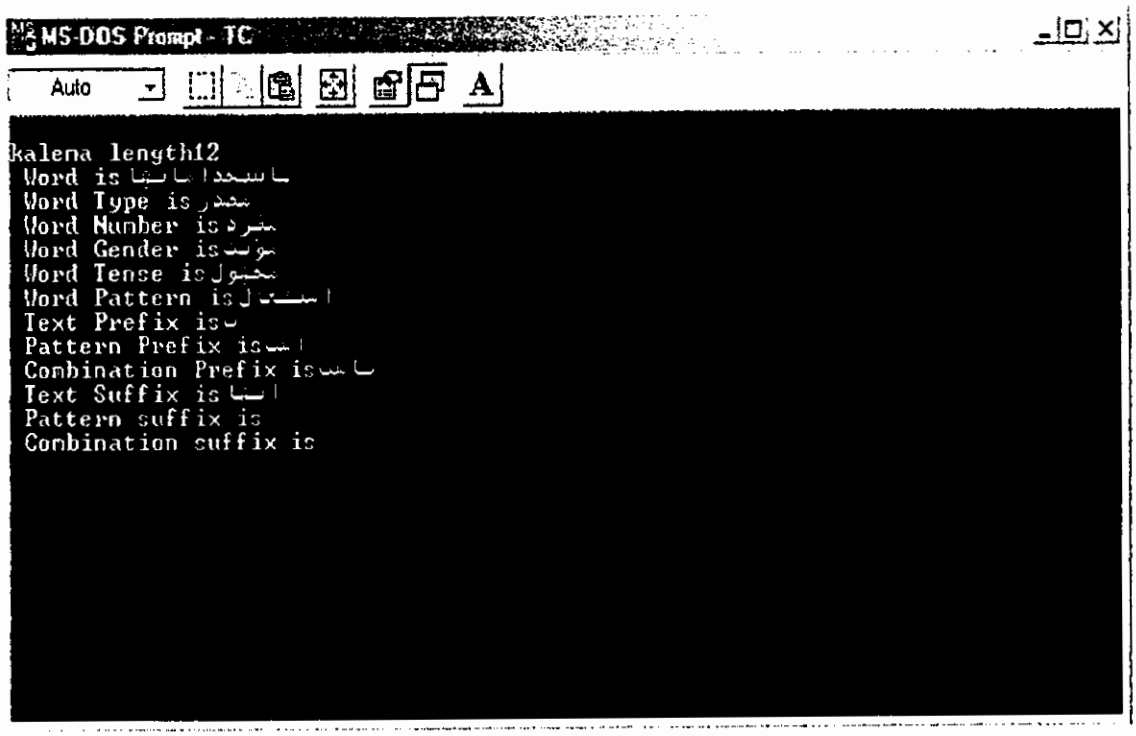
جدول رقم (٣٦)

الملحق رقم (٤)

شكل البرنامج والنتائج وجزء من عينة النص

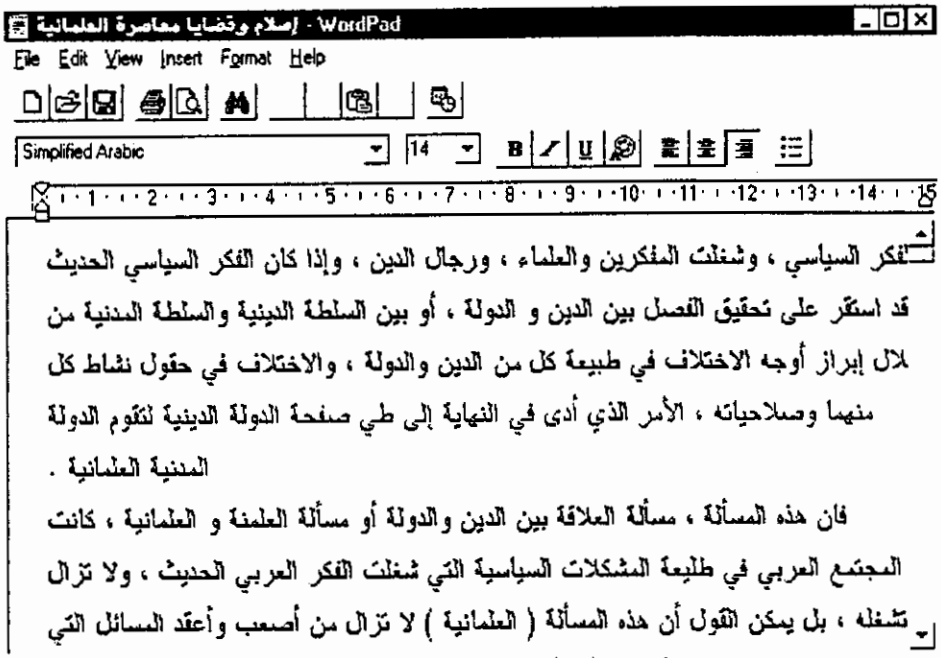


شكل رقم (٢١): شكل بداية البرنامج



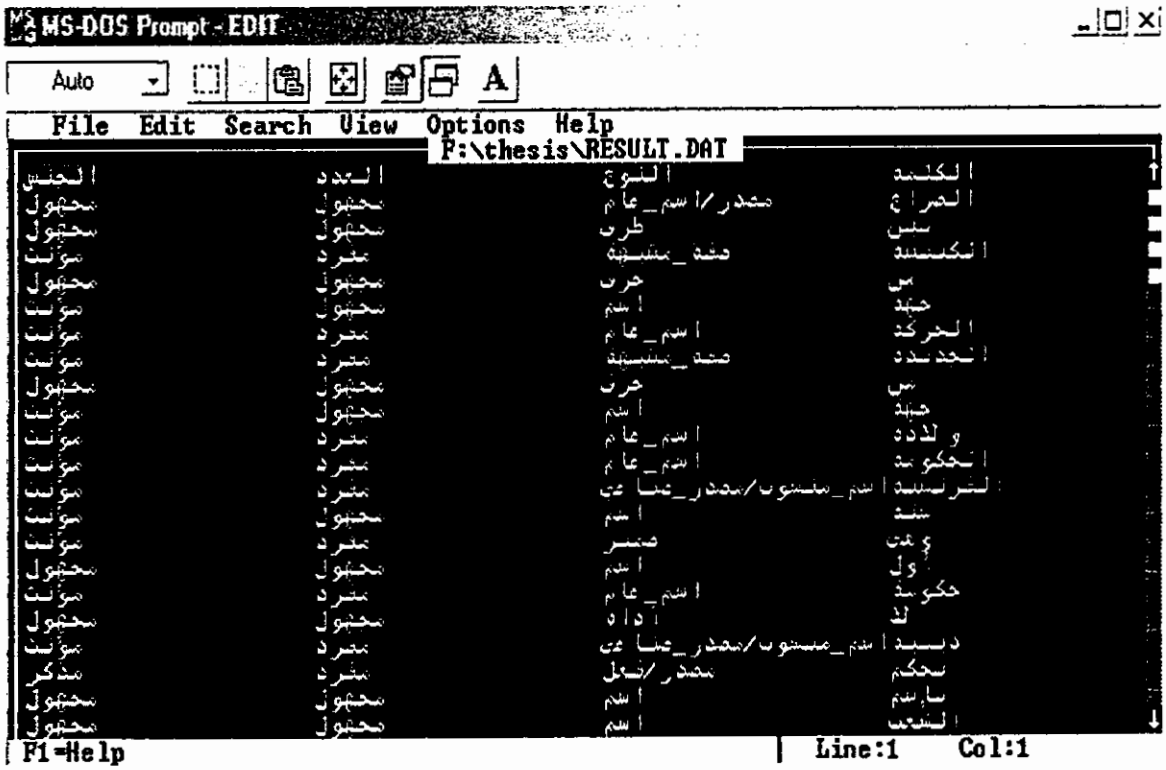
شكل رقم (٢٢): شكل النتائج عند ادخال كلمة

والشكل التالي جزء من النص الذي تم عليه تطبيق الخوارزمية:



شكل رقم (٢٣): عينة من النص الذي تم فحصه

والشكل التالي يوضح جزء من النتائج عند قراءة الكلمات من الملف النصي:



الشكل رقم (٢٤): شكل النتائج

الملحق رقم (٥)

خوارزمية منهجية الدراسة:

1-function find_particles()

open particles file.

Search the given word for a frozen word

if (a particle word is found) then

go to step 6

2-Function find_prefix()

create temporary prefix

open prefix file

search only about word prefix

if(prefix found and specify all word properties)

go to step 6

3-Function find_suffix()

open suffix file

search only about word suffix

if(suffix found and specify all word properties)

go to step 6

5-if we can not from the previous steps determine all properties for the word we do the following steps:

6-search again about the word after remove prefixes and suffixes in particles file because may remover original letters replacement extra letters.

If word found in particles words then

Go to step 6

4-function find_prefix_and_suffix()

create temporary prefix

create temporary suffix

open file contains prefixes and suffixes together

search for a prefix and suffix together

if(prefix and suffix found and specify all word properties)

go to step 6

5-Function find_pattern()

open patterns file

search about pattern for a given word without remove prefixes

if a pattern not found then

algorithm tries to find a pattern for a given word after remove prefixes from it by pervious steps.

If a pattern not found then

algorithm tries to find a pattern for a given word after remove suffixes

if matches then

go to 6

6-finally print a given word, word type , word number, word gender, word tense, prefixes, and suffixes.

الملحق رقم (٦)

(نص برنامج منهجية الدراسة)

```
#include <iostream.h>
#include <fstream.h>
#include <iomanip.h>
#include <stdlib.h>
#include<conio.h>
#include<iostream.h>
#include<string.h>
#include<stdio.h>
#include<dos.h>

  class prefixes
    {
       public:
       char type[3][90];
       char prefix[3][17];
       char pt[3][2];
       char jender[3][16];
       char number[3][15];
       char tense[3][15];
    };
   class suffixes
    {
       public:
       char type[2][26];
       char suffix[2][17];
       char pt[2][2];
       char jender[2][11];
       char number[2][10];
       char tense[2][10];
    };
  class frozen
    {
       public:
       char frozens[2][15];
       char type[2][15];
       char jender[2][15];
       char number[2][15];
    };
/*****************************************************************/
char
textprefix[6],textsuffix[6],pattemprefix[7],pattemsuffix[7],combination_prefix[11],com
bination_suffix[11];
char
kalema[14],kalematype[35],kalemanumber[15],kalemagender[15],kalematense[15],kale
mapattem[11],temp[4][28];
char infix1[6],infix2[6],words_removed[14],words_removed2[14],word_root[14];
```

```
int prelong,suflong,choice,foundtool=0,foundpattem=0,foundprefix=0,foundsuffix=0;
int pcount=0,scount=0,pscount=0,patcount=0;
/*************************************************************/
void start()
{
 strset(textprefix,'\0');
 strset(textsuffix,'\0');
 strset(words_removed,'\0');
 strset(words_removed2,'\0');
 strset(kalematype,'\0');
 strset(kalemanumber,'\0');
 strset(kalemagender,'\0');
 strset(kalematense,'\0');
 strset(kalemapattem,'\0');
 strset(word_root,'\0');
 strset(temp[0],'\0');
 strset(temp[1],'\0');
 strset(temp[2],'\0');
 strset(temp[3],'\0');
 strset(temp[4],'\0');
 strset(pattemprefix,'\0');
 strset(pattemsuffix,'\0');
 strset(combination_prefix,'\0');
 strset(combination_suffix,'\0');
 strncpy(kalematype,"مجهول",5);
 strncpy(kalemanumber,"مجهول",5);
 strncpy(kalemagender,"مجهول",5);
 strncpy(kalematense,"مجهول",5);
 strncpy(kalemapattem,"مجهول",5);
 foundtool=0;
}/*start function*/
/*************************************************************/
void end()
{

}/*end function*/
/*************************************************************/
void find_frozen()
{
 int count=0,found=0,pos=0,max=148;
 int x=1;
 char sharing[2][10],type2[2][10],gender2[2][10],number2[2][10];
 int count2=0;
 frozen data;

 for(int i=0;i<4;i++)
 {
  if(i==0)
   {
    strcpy(temp[i],kalema);
```

```
 }
if(i==1)
{
 strcpy(temp[i],words_removed);
}
else if (i==2)
{
 strcpy(temp[i],words_removed2);
}
else if(i==3)
{
 strcpy(temp[i],word_root);
}
int tempmax=max,min=0;
 pos=max/2;

while((pos!=0)&&(found!=1)&&(pos!=min)&&(pos!=tempmax))
 {
 ifstream outCredit( "f:/thesis/frozen.dat", ios::in);
 if ( !outCredit )
 {
  cerr << "Tools File could not be opened." << endl;
  getch();
  exit( 1 );
 }
 outCredit.seekg((pos-1)*sizeof(data));
 outCredit.read((unsigned char *) &data ,sizeof( frozen ));
 x=strcmp(temp[i],data.frozens[count]);
 if(x>0)
 {
  min=pos;
  pos=(tempmax+min)/2;
 }
 if(x<0)
 {
  tempmax=pos;
  pos=(tempmax+min)/2;
 }
 if (x==0)
 {

  strset(kalematype,'\0');
  strset(kalemanumber,'\0');
  strset(kalemagender,'\0');
  strset(kalematense,'\0');
  found=1;
  foundtool=1;
  ifstream inClientFile2("f:/thesis/جامدة.txt", ios::in );
  if ( !inClientFile2 )
  {
```

```
        cerr << "Sharing File could not be opened\n";
        getch();
        exit( 1 );
     }//if ( !inClientFile )
   while ( inClientFile2>>sharing[count2]>>type2[count2]>> number2[count2]>>
gender2[count2])
    {
     x=strcmp(temp[i],sharing[count2]);
     if (x==0)
         {
          /*x=strncmp(kalematype,"مجهول",5);
          if(x==0)*/
          strcpy(kalematype,type2[count2]);
          /*x=stmcmp(kalemanumber,"مجهول",5);
          if(x==0)*/
           strcpy(kalemanumber,number2[count2]);
          /*x=strncmp(kalemagender,"مجهول",5);
          if(x==0)*/
           strcpy(kalemagender,gender2[count2]);
           i=4;

         }//if(x==0)
   }//while ( inClientFile2>>sharing[count2]>>type2[count2]>> number2[count2]>>
jender2[count2])
    inClientFile2.close();
  /*x=strncmp(kalematype,"مجهول",5);
  if(x==0)*/
   strcpy(kalematype,data.type[count]);
  /*x=strncmp(kalemanumber,"مجهول",5);
  if(x==0)*/
   strcpy(kalemanumber,data.number[count]);
  /*x=strncmp(kalemagender,"مجهول",5);
 if(x==0)*/
   strcpy(kalemagender,data.jender[count]);
   //strcpy(kalem atense,"مجهول");
  i=4;
  }//if (x==0)
  outCredit.close();
 }//while

  }//for (int i=0...)
 strset(data.frozens[0],'\0');
 strset(data.frozens[1],'\0');
 strset(data.type[0],'\0');
 strset(data.type[1],'\0');
 strset(data.number[0],'\0');
 strset(data.number[1],'\0');
 strset(data.jender[0],'\0');
 strset(data.jender[1],'\0');
 strset(sharing[0],'\0');
```

```
strset(sharing[1],'\0');
strset(type2[0],'\0');
strset(type2[1],'\0');
strset(number2[0],'\0');
strset(number2[1],'\0');
strset(gender2[0],'\0');
strset(gender2[1],'\0');
strset(temp[0],'\0');
strset(temp[1],'\0');
strset(temp[2],'\0');
strset(temp[3],'\0');

}
/*************************************************************/
void find_pattern()
{
 int x,y;
 char infix1_temp[3];
 char infix2_temp[3];
 char prelong[70],infpos1[70],infpos2[70];
 int matchpre=0,matchinf1=0,matchinf2=0,matchsuf=0;
 char pattern[70][10];
 char temp[70][10];
 char type[70][26];
 char prefix[70][10];
 char infix1[70][10];
 char infix2[70][10];
 char suffix[70][10];
 char jender[70][10];
 char number[70][10];
 char tense[70][10];
 clrscr();
strset(temp[0],'\0');
strset(temp[1],'\0');
strset(temp[2],'\0');
strset(temp[3],'\0');
 for(int i=0;i<4;i++)
 {
 ifstream inClientFile("f:/thesis/٣الأوزان.txt", ios::in );
 if ( !inClientFile )
  {
  cerr << "Pattern File  could not be opened\n";
  getch();
  exit( 1 );
 }//if ( !inClientFile )
 if(i==0)
  {
  strcpy(temp[i],kalema);
  }
 if(i==1)
```

```
    {
     strncpy(temp[i],words_removed,strlen(kalema)-strlen(textprefix));
     clrscr();
    }
   else if (i==2)
    {
         strcpy(temp[i],words_removed2);
         //getch();
    }

   else if(i==3)
    {
     strcpy(temp[i],word_root);
    }
   int count=0;
   while (( inClientFile >> pattern[count] >> prefix[count]
>>prelong[count]>>infix1[count]
>>infpos1[count]>>infix2[count]>>infpos2[count]>>suffix[count]>> type[count]>>
number[count]>> jender[count]>> tense[count])&&(count<54))
        {
        if(strlen(temp[i])==strlen(pattern[count]))
        {
          if(prelong[count]=='0')
           matchpre=1;
          else
          {
           matchpre=1;
           for(int kk=0;kk<prelong[count]-48;kk++)
           {
            if(temp[i][kk]!=prefix[count][kk])
            matchpre=0;
           }
          }
          if(infpos1[count]=='0')
           matchinf1=1;
          else
          {
           if(temp[i][infpos1[count]-48]==pattern[count][infpos1[count]-48])
            matchinf1=1;
          }
          if(infpos2[count]=='0')
           matchinf2=1;
          else
          {

           if(temp[i][infpos2[count]-48]==pattern[count][infpos2[count]-48])
            matchinf2=1;
          }

          if (suffix[count][0]=='0')
```

```
            matchsuf=1;
           else
           {
            matchsuf=1;

           for(int kk=strlen(pattem[count])-
strlen(suffix[count]);kk<strlen(pattem[count]);kk++)
           {
            if(temp[i][kk]!=pattem[count][kk])
             matchsuf=0;

           }
           }
           if((matchpre==1)&&(matchinf1==1)&&(matchinf2==1)&&(matchsuf==1))
           {

            foundpattem=1;
            x=stmcmp(pattem[count],"مجهول",5);

            if (x!=0)
             strcpy(kalemapattem,pattem[count]);
            x=stmcmp(type[count],"مجهول",5);
            y=stmcmp(kalematype,"مجهول",5);

            if(i==t)
            {
             strset(textprefix,'\0');
             strset(textsuffix,'\0');
             strcpy(pattemprefix,prefix[count]);
             strcpy(pattemsuffix,suffix[count]);
             x=1;
             y=0;
             strcpy(kalemanumber,"مجهول");
            }

            if(i==1)
            {
             x=1;
             y=0;
             strset(textsuffix,'\0');
             strcpy(pattemsuffix,suffix[count]);
             strcpy(kalemanumber,"مجهول");
            }
            if(i==2)
            {
             x=1;
             y=0;
             strset(textprefix,'\0');
             strcpy(pattemprefix,prefix[count]);
            }
```

```
          if(i==3)
          {
           x=1;
           y=0;
          }
          if((x!=0)&&(y==0))
          {
           x=strcmp(kalematype,"مجهول");
           y=strcmp(kalematype,"اسم");
           if(((i==0)||(i==2))||((x==0)||(y==0)))
           {
              strset(kalematype,'\0');
              strcpy(kalematype,type[count]);
           }
              patcount++;
          }
          x=stmcmp(number[count],"مجهول",5);
          y=stmcmp(kalemanumber,"مجهول",5);
          if((x!=0)&&(y==0))
           strcpy(kalemanumber,number[count]);
          x=stmcmp(jender[count],"مجهول",5);
          y=stmcmp(kalemagender,"مجهول",5);
          if((x!=0)&&(y==0))
           strcpy(kalemagender,jender[count]);
          x=stmcmp(tense[count],"مجهول",5);
          if (x!=0)
           strcpy(kalematense,tense[count]);
          i=4;
          }
        }//if (strlen(kalema.words)==strlen(pattem[count])
       matchpre=0;
       matchsuf=0;
       matchinf1=0;
       matchinf2=0;
    }//while
  inClientFile.close();
 }//for (int i=0...)
}/*function find_pattem*/

/*************************************************************/
void find_pattemprefix()
{
 prefixes data2;
 char temppref[14];
 int max=140,pos=0,found=0;
 int tempmax=max,min=0,x,count=0,i=0,ptype;
   pos=max/2;
   strset(temppref,'\0');
   temppref[0]=combination_prefix[strlen(combination_prefix)-1];
   while((pos!=0)&&(found!=10)&&(pos!=min)&&(pos<=tempmax))
```

```
{
 ifstream outCredit( "f:/thesis/prefix.dat", ios::in );
if ( !outCredit )
{
 cerr << "Prefix File could not be opened." << endl;
 getch();
 exit( 1 );
}
 outCredit.seekg((pos-1)*sizeof(data2));
 outCredit.read((unsigned char *) &data2 ,sizeof(prefixes));
 x=strcmp(temppref,data2.prefix[count]);
 if(x==0)
 {
     ptype=strcmp(data2.pt[count],"1");
     if(ptype==0)
      strcpy(patternprefix,temppref);
     found=0;
     max=140;
     pos=max/2;
     min=0;
     tempmax=max;
     i++;
     int g=0;
     strset(temppref,'\0');
     for(int pre=strlen(combination_prefix) -i;pre<strlen(combination_prefix);pre++)
 {
     temppref[g]=combination_prefix[pre];
     g++;
 }
 for(int sl=i+1;sl<14;sl++)
     temppref[sl]='\0';

 }
 else
 {
 if(x>0)
 {
     min=pos;
     pos=(tempmax+min)/2;
 }
 if(x<0)
 {
     tempmax=pos;
     pos=(tempmax+min)/2;
 }
 found++;
 }
if ((pos==0)||(found==9)||(pos==min)||(pos>tempmax))
{
     found=0;
```

```
        max=140;
        pos=max/2;
        min=0;
        tempmax=max;
        i++;
        int g=0;
        for(int pre=strlen(combination_prefix)-i;pre<strlen(combination_prefix);pre++)
     {
        temppref[g]=combination_prefix[pre];
        g++;
     }
     for(int sl=i+1;sl<14;sl++)
         temppref[sl]='\0';
   }
    if (i>strlen(combination_prefix))
     found=10;
    outCredit.close();
   }/*while*/
  for(int ii=0;i<2;i++)
  {
  strset(data2.prefix[ii],'\0');
  strset(data2.pt[ii],'\0');
  strset(data2.type[ii],'\0');
  strset(data2.number[ii],'\0');
  strset(data2.jender[ii],'\0');
  strset(data2.tense[ii],'\0');
  }

 }
/*************************************************************/
void find_prefix()
{
 prefixes data;
 char temppre[14];
 int max=140,pos=0,found=0,ptype;
 int tempmax=max,min=0,x,count=0,i=0;
  pos=max/2;
  strset(temppre,'\0');
  temppre[0]=kalema[0];
 while((pos!=0)&&(found!=10)&&(pos!=min)&&(pos!=tempmax))
  {

   ifstream outCredit( "f:/thesis/prefix.dat", ios::in );
  if ( !outCredit )
  {
      cerr << "Prefix File could not be opened." << endl;
      getch();
      exit( 1 );
  }
  outCredit.seekg((pos-1)*sizeof(data));
```

```
    outCredit.read((unsigned char *) &data ,sizeof( prefixes ));
    x=strcmp(temppre,data.prefix[count]);
     /*cout<<"\n temppre is "<<temppre<<"and table is "<<data.prefix[count];
     cout<<"\n kalema is "<<kalema<<" and pos is "<<pos<<" tool of
temp"<<strlen(temppre);
     getch();*/
    /*cout<<"\n data "<<data.type[count]<<" pos is"<<pos;
    getch();*/
    if(x==0)
    {
        /*cout<<"\n temppre is2 "<<temppre<<"and table is2 "<<data.prefix[count];
        getch();*/
        x=strncmp(data.type[count],"مجهول",5);
        if(x!=0)
        {
         pcount++;
        }
         if(x!=0)
         {
         /*cout<<"\n kalema type before "<<kalematype;

         cout<<"\n data "<<data.type[count];
         cout<<"\n kalema type after "<<kalematype;*/
         strcpy(kalematype,data.type[count]);
         //getch();
        }
         ptype=strcmp(data.pt[count],"0");
         if(ptype==0)
          strcpy(textprefix,temppre);
        ptype=strcmp(data.pt[count],"1");
        if(ptype==0)
         strcpy(patterntemprefix,temppre);
        ptype=strcmp(data.pt[count],"10");
        if(ptype==0)
        {
         strcpy(textprefix,temppre);
         strcpy(patterntemprefix,temppre);
        }
        ptype=strcmp(data.pt[count],"2");
        if(ptype==0)
         strcpy(combination_prefix,temppre);


         x=strncmp(data.number[count],"مجهول",5);
       //if(x!=0)
        strcpy(kalemanumber,data.number[count]);
       x=strncmp(data.jender[count],"مجهول",5);
       //if(x!=0)
        strcpy(kalemagender,data.jender[count]);
       x=strncmp(data.tense[count],"مجهول",5);
```

```
        //if (x!=0)
         strcpy(kalematense,data.tense[count]);
        found=0;
        max=140;
        pos=max/2;
        min=0;
        tempmax=max;
        //strcpy(textprefix,temppre);
        i++;
        temppre[i]=kalema[i];
        for(int sl=i+1;sl<14;sl++)
         temppre[sl]='\0';
    }
    else
    {

     if(x<0)
     {
         tempmax=pos;
         pos=(tempmax+min)/2;
     }
     if(x>0)
     {
         min=pos;
         pos=(tempmax+min)/2;
     }
     found++;
    }
    if ((pos==0)||(found==15)||(pos==min)||(pos==tempmax))
    {
         found=0;
         max=140;
         pos=max/2;
         min=0;
         tempmax=max;
         i++;
         temppre[i]=kalema[i];
         for(int sl=i+1;sl<14;sl++)
          temppre[sl]='\0';
    }
    if (i>strlen(kalema))
     found=10;
    outCredit.close();

   }/*while*/

   int lo=0;
   stmset(words_removed,'\0',strlen(kalema));
   for(int h=strlen(textprefix);h<strlen(kalema);h++)
   {
```

```
   words_removed[lo]=kalema[h];
   lo++;
   }
    for(int hh=0;hh<strlen(kalema)-strlen(textprefix);hh++)
    {
    cout<<words_removed[hh];
    }
  if(strlen(combination_prefix)>0)
   {
   find_patternprefix();
   }
  for(int ii=0;ii<2;ii++)
 {
  strset(data.prefix[ii],'\0');
  strset(data.pt[ii],'\0');
  strset(data.type[ii],'\0');
  strset(data.number[ii],'\0');
  strset(data.jender[ii],'\0');
  strset(data.tense[ii],'\0');
 }
}
/*************************************************************/
void find_suffix()
{
suffixes data;
char tempsuf[14];
int max=144,pos=0,found=0;
int tempmax=max,min=0,x,count=0,i=0,ptype;
  pos=max/2;
  strset(tempsuf,'\0');
  tempsuf[0]=kalema[strlen(kalema)-1];
  while((pos!=0)&&(found!=10)&&(pos!=min)&&(pos<=tempmax))
  {
   ifstream outCredit( "f:/thesis/suffix.dat", ios::in );
  if ( !outCredit )
  {
   cerr << "Suffix File could not be opened." << endl;
   getch();
   exit( 1 );
  }
   outCredit.seekg((pos-1)*sizeof(data));
   outCredit.read((unsigned char *) &data ,sizeof( suffixes ));
   x=strcmp(tempsuf,data.suffix[count]);
   /*cout<<"\n pos is "<<pos;
   cout<<"\n temp suf is "<<tempsuf<<"from table is "<<data.suffix[count];
   getch();*/
   if(x==0)
   {
       x=strncmp(data.type[count],"مجهول",5);
       if(x!=0)
```

```
        {
         scount++;
        }
         ptype=strcmp(data.pt[count],"0");
          if(ptype==0)
           strcpy(textsuffix,tempsuf);
          ptype=strcmp(data.pt[count],"1");
          if(ptype==0)
           strcpy(patternsuffix,tempsuf);
          ptype=strcmp(data.pt[count],"10");
          if(ptype==0)
          {
           strcpy(textsuffix,tempsuf);
           strcpy(patternsuffix,tempsuf);
          }
          ptype=strcmp(data.pt[count],"2");
          if(ptype==0)
           strcpy(combination_suffix,tempsuf);

        //if(strlen(textprefix)==0)
        x=strncmp(kalematype,"مجهول",5);
        if(x==0)
         strcpy(kalematype,data.type[count]);

        x=strncmp(data.number[count],"مجهول",5);
        //if(x!=0)
         strcpy(kalemanumber,data.number[count]);
        x=strncmp(data.jender[count],"مجهول",5);
        //if(x!=0)
         strcpy(kalemagender,data.jender[count]);
        x=strncmp(data.tense[count],"مجهول",5);
         strcpy(kalematense,data.tense[count]);
         cout<<"\n 4suffix2 tense is"<<kalematense;
        //getch();
        found=0;
        max=144;
        pos=max/2;
        min=0;
        tempmax=max;
        strcpy(textsuffix,tempsuf);
        i++;
        int g=0;
        strset(tempsuf,'\0');
        for(int suf=strlen(kalema) -i;suf<strlen(kalema);suf++)
    {
        tempsuf[g]=kalema[suf];
        g++;
    }
    for(int sl=i+1;sl<14;sl++)
        tempsuf[sl]='\0';
```

```
 }
 else
 {
  if(x>0)
  {
     min=pos;
     pos=(tempmax+min)/2;
  }
  if(x<0)
  {
     tempmax=pos;
     pos=(tempmax+min)/2;
  }
  found++;
 }
 if ((pos==0)||(found==9)||(pos==min)||(pos>tempmax))
 {
     found=0;
     max=144;
     pos=max/2;
     min=0;
     tempmax=max;
     i++;
     int g=0;
     for(int suf=strlen(kalema)-i;suf<strlen(kalema);suf++)
  {
     tempsuf[g]=kalema[suf];
     g++;
  }
  for(int sl=i+1;sl<14;sl++)
      tempsuf[sl]='\0';
 }
  if (i>strlen(kalema))
   found=10;
 outCredit.close();
 }/*while*/


int lo=0;
strnset(words_removed2,'\0',strlen(kalema));
cout<<"\n kalema is"<<kalema;
cout<<"\n text suffix long is" <<strlen(textsuffix);
cout<<"\n kalema long is" <<strlen(kalema);

for(int h=0;h<strlen(kalema)-strlen(textsuffix);h++)
{
 words_removed2[lo]=kalema[h];
 lo++;
}
```

```
  for(int ii=0;ii<2;ii++)
 {
  strset(data.suffix[ii],'\0');
  strset(data.pt[ii],'\0');
  strset(data.type[ii],'\0');
  strset(data.number[ii],'\0');
  strset(data.jender[ii],'\0');
  strset(data.tense[ii],'\0');
 }
}
/*************************************************************/
 void find_pre_and_suf()
{
   char type[10][30];
   char prefix[10][30];
   char suffix[10][30];
   char jender[10][30];
   char number[10][30];
   char tense[10][30];
   int index,max=0;
   int max2=0;
   char temp[15];
   char temp2[15];
   int y;
   // ifstream constructor opens the file

   ifstream inClientFile( "f:/thesis/لو لحقوسو ابق.txt", ios::in );

   if ( !inClientFile )
   {
        cerr << "Prefix_Suffix File could not be opened\n";
        getch();
        exit( 1 );
   }//if(!inClientFile)

   int x,count=0;
   x=1;
   while (inClientFile >> prefix[count] >>suffix[count]>>type[count]>>
number[count]>> jender[count]>> tense[count])
        {
         if ((strlen(prefix[count])>max)&&(strlen(suffix[count])>max2))
         {
          x=(strncmp(textprefix,prefix[count],strlen(prefix[count])));

          y=(strncmp(textsuffix,suffix[count],strlen(suffix[count])));
          if((x==0)&&(y==0))
          {
           x=strncmp(type[count],"مجهول",5);
           if(x!=0)
           {
```

```
            strcpy(kalematype,type[count]);
            pscount++;
           }
           x=strncmp(number[count],"مجهول",5);
           if(x!=0)
            strcpy(kalemanumber,number[count]);
           x=strncmp(jender[count],"مجهول",5);
           if(x!=0)
            strcpy(kalemagender,jender[count]);
           x=strncmp(tense[count],"مجهول",5);
           if (x!=0)
            strcpy(kalematense,tense[count]);
           max=(strlen(prefix[count]));
           max2=(strlen(suffix[count]));
         }//if(x==0)&&(Y=0)
        } //if(strlen(prefix[count])>max)
       }//while
    int rootind=0;
    for(int j=strlen(textprefix);j<strlen(kalema)-strlen(textsuffix);j++)
    {
       word_root[rootind]=kalema[j];
       rootind++;
    }
    inClientFile.close();
}//function find_pre_suf

/*************************************************************/
void print()
{
  cout<<"\n Word is"<<kalema;
  cout<<"\n Word Type is"<<kalematype;
  cout<<"\n Word Number is"<<kalemanumber;
  cout<<"\n Word Gender is"<<kalemagender;
  cout<<"\n Word Tense is"<<kalematense;
  cout<<"\n Word Pattern is"<<kalemapattern;
  cout<<"\n Text Prefix is"<<textprefix;
  cout<<"\n Pattern Prefix is"<<patternprefix;
  cout<<"\n Combination Prefix is"<<combination_prefix;
  cout<<"\n Text Suffix is"<<textsuffix;
  cout<<"\n Pattern suffix is"<<patternsuffix;
  cout<<"\n Combination suffix is"<<combination_suffix;



}
/*************************************************************/
void process(int choice)
{

 switch(choice)
```

```
{

 case 1:
        start();
        cout<<"pleas enter your word \n";
        cin>>kalema;
        find_frozen();
        if(foundtool!=1)
        {
         find_prefix();
        if(foundprefix!=1)
          find_suffix();
        if(foundsuffix!=1)
         find_pattern();
        if(foundpattern!=1)
         find_pre_and_suf();

        }
      cout<<"\nkalema length"<<strlen(kalema);
        print();
        getch();
        end();
        break;

 case 2:
        start();
        int wordcount=0;
        char
words[2000][20],wordtype[2000][20],wordtense[2000][20],wordnumber[2000][20];
        char wordgender[2000][20],wordprefix[2000][10],wordsuffix[2000][10];
        ifstream inClientFile("f:/thesis/test7.txt", ios::in );
        if ( !inClientFile )
        {
         cerr << "Test File could not be opened\n";
         getch();
         exit( 1 );
        }
        while(inClientFile>>kalema)
        {
         find_frozen();
        if(foundtool!=1)
        {
         find_prefix();
        if(foundprefix!=1)
          find_suffix();
        if(foundsuffix!=1)
         find_pattern();
        if(foundpattern!=1)
         find_pre_and_suf();
        }
```

```
            strcpy(words[wordcount],kalema);
            strcpy(wordtype[wordcount],kalematype);
            strcpy(wordnumber[wordcount],kalemanumber);
            strcpy(wordgender[wordcount],kalemagender);
            strcpy(wordtense[wordcount],kalematense);
            strcpy(wordprefix[wordcount],textprefix);
            strcpy(wordsuffix[wordcount],textsuffix);
            wordcount++;
            start();
            }
            inClientFile.close();
            ofstream outClientFile( "f:/thesis/result.dat", ios::out );
            if ( !outClientFile )
            { // overloaded ! operator
             cerr << "File could not be opened" << endl;
             exit( 1 );   // prototype in stdlib.h
            }
             outClientFile <<"الكلمة"<< "النوع"
                 <<"    " << "الجنس" <<"   "<<"العدد"<<"    "<<"اللواحق"
                 <<"   "<<"الزمان"<<"   "<<"الوزن"<<"              "<<"السوابق"<<"    "'\n';
            for(int counter=0;counter<=wordcount;counter++)
            {
             outClientFile<<words[counter]<<"    " <<wordtype[counter]<<"    "
<<wordnumber[counter]
                          <<"    " <<wordgender[counter]<<"    "
<<wordtense[counter]<<"    " <<wordprefix[counter]
                          <<"    " <<wordsuffix[counter]<<'\n';

            }
            cout<<"\n prefixes is:"<<pcount;
            cout<<"\n suffixes is:"<<scount;
            cout<<"\n prefixes_suffixes is:"<<pscount;
            cout<<"\n patterns is:"<<patcount;
            getch();
            outClientFile<<"end-to-file";
            outClientFile.close();
            wordcount=0;
           break;
   case 3:
            choice=-1;
            break;
  }//switch


}

/*************************************************************/
int main()
{
```

```
while(choice!=-1)
  {
    clrscr();
    cout<<"1-Enter Word\n";
    cout<<"2-open file\n";
    cout<<"3-Exit\n";
    cout<<"pleas enter your choice\n";
    cin>>choice;
    process(choice);

  }
return 0;
}/*end main*/
```

Abstract

The computer linguistics field is considered as one of the widest fields that are related with the artificial intelligence, and it is called the applied side for this field, which is related with the natural languages. However, Arabic language is different from other natural languages in terms of its structure and composition; it is a language of inflection but not a suffix one. Moreover, English language has been the first among other languages which is processed by computer.

The topic of this thesis relies on the classification of the Arabic word that is not vocalizes depending on the pronunciation analysis, where the processing of Arabic automatically passes through four stages of analysis: dictionary, syntax, grammar and indicative. Therefore, this study aims to explore the extent of syntax and vocal suffixes reliance on defining the type of text word.

However, this study starts from a basic assumption stipulating that the structural text of the Arabic word within a text includes indicative meanings that enable reliance on them for a large degree in defining there type and functional features within an Arabic text.

Consequently, the significance of this study comes from the importance of the syntactic analysis for many computer application to process Arabic language, in terms of providing more space and raising their efficiency by minimizing the use of dictionaries.

However, this study used four basic methods: First: verifying the possibility that a word which we intend to process, may be a member of the tools and pronouns in Arabic language, Second: the vocal analysis that relies on using the syntax prefixes with all kind (such as conjunctions, prepositions, articles and so on). Third: the vocal analysis depending on using syntax rhymes in accordance with the contents of these prefixes, since there are special rhymes for the verbs, names, adjectives and so on, and there are common rhymes among these types. In order to perform such analysis, it has been necessary to define the linguistic rhymes and decide the used syntactic prefixes as well as highlighting the relationships among them with the linguistic prefixes, which share with them in beginnings and ends of words. Forth: relying on the length of the word in the analysis process where necessary there are three and four- root words, that do not include any linguistic or syntax suffixes. Accordingly, it is

necessary to rely on the length of the word to show its class in terms of nouns or verbs, or to verify the results of processing relying on the vocal prefixes.

Consequently, the results of the study indicated that the analytical methodology used in this study has succeeded in proving that most of the words are referred to the nouns and verbs categories; the program has been able to recognize 88.46% of the nouns and 84.11% of the verbs, but this percentage decreases when defining the accurate type of the word in terms of nouns or verbs as the program could accurately define the words for about 79% of the words included within the text. Hence, it is clear that the word structure within an Arabic text can not be enough in some cases to define the word type for various reasons; first: relying on word prefixes in the word to decide its type from the root of the word, and therefore, there were 7.75% of the words which the program was mistaken in deciding their correct type; second: the words are void of linguistic and syntax prefixes which formed 11.4% of the text words, and thus there was nothing to indicate the type of these words; third: the word contains syntax prefixes that do not define the type of the word, thus they are prefixed with the nouns and verbs, and there are rhymes that mix between two types of words (vocabularies).