المملكة العربيَّة السُّعوديَّة
وزارة التعليم
جامعة الإمام محمَّد بن سعود الإسلاميَّة
معهد تعليم اللُّغَة العربية
قسم اللُّغويات التَّطبيقية العربيَّة

بناء خوارزميّة لِفَكِّ اللبْس الصرفي الحاسوبي في اسم الآلة

رسالة مقدمة لنيل درجة الدكتوراة في اللُّغويَّات التَّطبيقية العربيَّة

إعداد
منى بنت قاسم بن سلمان الودعاني

إشراف
د. عبدالله بن يحيى بن جابر الفيفي
الأستاذ المشارك في قسم اللغويات التطبيقية العربية

العام الجامعي
١٤٤٥هـ/ ٢٠٢٤م

مستخلص الدراسة

عنوان الدراسة: بناءُ خوارزميّةٍ لِفَكِّ اللبْسِ الصرفيِّ الحاسوبيِّ في اسمِ الآلةِ

الباحثة: منى بنت قاسم سلمان الودعاني **المشرف:** د. عبدالله يحيى جابر الفيفي

الدرجة العلمية: الدكتوراة **الجامعة والكلية:** جامعة الإمام محمد بن سعود الإسلامية– معهد تعليم اللغة العربية

القسم والتخصص: اللغويات التطبيقية العربية **العام الجامعي:** ١٤٤٥هـ/ ٢٠٢٤م

سعتِ الدراسةُ إلى تحديدِ مواضعِ اللبْسِ الصرفيِّ الحاسوبيِّ في أسماءِ الآلةِ من خلالِ اختبارِ عينةِ الدراسةِ التي تحويْ أسماءَ الآلةِ موضعَ اللبسِ في محللَين صرفيين، هما: محلّلُ فراسة FARASA، ومحلّلُ مداميرا MADAMIRA. وبناءِ خوارزميّةٍ حاسوبيّةٍ تُسهِمُ في حلِّ مشكلةِ اللبسِ الصرفيِّ، ثم تقييمِها.

اعتمدتُ الدراسةُ على المنهجِ الوصفيِّ الإحصائيِّ، وتكونتْ عينةُ الدراسةِ من ٢٤ اسمًا من أسماءِ الآلةِ من الأوزانِ القياسيةِ الآتيةِ: "مِفْعَال، مِفْعَل، مِفْعَلة، فَعّالَة/ فعّال، فَاعوُل، فِعَال"، والوزنِ السماعيِّ: " مُفْعُل"، واستخرجتْ جميعَ جملِ العينةِ التي تحويْ أسماءَ الآلةِ موضعَ اللَّبْسِ من المدونةِ اللغويةِ العربيةِ العالميةِ، وبلغَ عددُ جملِ العينةِ المستخرَجةِ ١٧٢ جملةً، حُلِّلَتْ في محلّلِ مداميرا MADAMIRA ومحلِّلِ فراسة FARASA، ثم حددتِ الباحثةُ مواضعَ اللبسِ الصرفيِّ في المحلّلين، وتمثلتْ في المشكلاتِ الآتيةِ: أولًا: مشكلةُ وسمِ جنسِ الكلمةِ في محلّلي الدراسةِ. ثانيًا: مشكلةُ وسمِ أصلِ الكلمةِ في محلّلي الدراسةِ. ثالثًا: مشكلةُ الوسمِ بالعددِ في محلّلِ مداميرا MADAMIRA. رابعًا: مشكلةُ وسمِ نوعِ الكلمةِ في محلّلِ مداميرا MADAMIRA. ولمعالجةِ المشكلاتِ السابقةِ؛ عمدتِ الباحثةُ إلى الآتي: أولًا: بناءُ قاعدةِ بياناتٍ تحويْ جميعَ أسماءِ الآلةِ (القياسيةِ والسماعيةِ)، وتشتملُ على: الكلمةِ، ووسمِها من حيثُ التذكيرُ والتأنيثُ، ووسمِها من حيثُ العددُ. وصيغةِ الجمعِ من الكلمةِ؛ يليها: وسمُها من حيثُ التذكيرُ والتأنيثُ، ووسمُها من حيثُ العددُ، وأصلُ الكلمةِ، بحيثُ تكونُ هذهِ القاعدةُ قابلةً للتعديلِ والإضافةِ. ثانيًا: اعتمادُ سمةِ الجمعِ للكلماتِ من أوزانِ أسماءِ الآلةِ (القياسيةِ والسماعيةِ) الآتيةِ: "مَفَاعِل، ومَفَاعِيل، وفَواعِيل". ثالثًا: اعتمادُ سمةِ التأنيثِ في جنسِ الكلمةِ الدالّ على الجمعِ من أسماءِ الآلةِ (القياسيةِ والسماعيةِ) من الأوزانِ الآتيةِ: "مَفَاعِل، ومَفَاعِيل، وفَواعِيل". ولزيادةِ دقةِ النتائجِ؛ قامتِ الباحثةُ في خطوةٍ إضافيةٍ بصياغةِ بعضِ القواعدِ اللغويةِ التي قدْ تسهمِ في تقليلِ اللبسِ الصرفيِّ، وتحويلِ هذهِ القواعدِ إلى رموزٍ؛ تسهيلًا لحوسبتِها. تلا ذلك عرضُ القواعدِ اللغويةِ وترميزِها، وقاعدةِ البياناتِ، والخوارزميةِ على المحكّمينَ المتخصّصينَ، ثم تعديلُها حسبَ ملحوظاتِهم؛ لتصلَ الباحثةُ بذلكَ إلى خوارزميةٍ حاسوبيةٍ تسهمُ في التقليلِ من اللبسِ الصرفيِّ في اسمِ الآلةِ.

Abstract

Title of the study: Developing an Algorithm to Resolve the Computing orphological ambiguity in the instrument noun
Researcher's name: Mona Qasem Salman Al-Wadani**.**
supervisor professor: Dr. Abdullah Yahya Jaber Al-Fifi.
University: imam Mohammed bin Saud Islamic University- Arabic language teaching Institute. **Degree:** doctorate **Department**: Applied Linguistics
Academic Year: 1445 – 2024

The study sought to identify the Computing Morphological Ambiguities in the instrument nouns by testing the study sample (which includes the instrument nouns subject matter of the ambiguity) in two morphological analyzers, namely: FARASA Analyzer and MADAMIRA Analyzer, and sought also to develop a computing algorithm contributing to resolving the morphological ambiguity issue, and then evaluating it.

The study relied on the descriptive statistical approach, while the study sample consisted of 24 instrument nouns with the following standard weights: "Mif'al, Mif'l, Mif'la, Fa'ala/Fa'al, Fa'ul and Fi'al" and the auditory weight: "Muf'ul". The entire sample sentences (which contain the ambiguous instrument nouns) are obtained from the International Arabic Linguistic Corpus. Number of the extracted sample sentences is 172 sentences, where such sentences were analyzed by MADAMIRA Analyzer and FARASA Analyzer; then, the researcher identified the morphological ambiguities in both analyzers, where the following issues resulted: First Issue: marking type of the word in terms of the masculine and feminine in the two study analyzers; Second Issue: marking the origin of the word in the two study analyzers; Third Issue: marking the word in numbers in MADAMIRA Analyzer; Fourth Issue: marking the word type in MADAMIRA Analyzer. To address the previous issues, the researcher made the following:

First: Developing a database containing all the standard and auditory, including: the word, marking the word in terms of masculine and feminine, marking the word in terms of the number, and plural form of the word followed by marking the word in terms of masculine and feminine, marking it in terms of number, and marking it in terms of its origin, so that such database is modifiable and addable.

Second: Adopting the plural form of the words in terms of the following weights of standard and auditory instrument nouns: "Mafa'il, Mafa'eel, and Fawa'eel".

Third: Adopting the feminine form of the plural word derivated from the standard and auditory instrument nouns of the following weights: "Mafa'il, Mafa'eel, and Fawa'eel". To increase the results accuracy, the researcher additionally formulated some linguistic rules which may contribute to reducing the morphological ambiguity, and transferred these rules into codes for easy computing, followed by presenting and coding the linguistic rules, presenting the database and the algorithm to the specialized arbitrators, and then modifying the same according to the arbitrators' feedback for the researcher to conclude a computing algorithm contributing to reducing the morphological ambiguity in the instrument noun.

المحتويات

شكر وتقدير..... أ

مستخلص الدراسة..... ب

فهرس الجداول..... ح

فهرس الأشكال..... ط

مقدمة..... ١

أولًا: التمهيد..... ٢

ثانيًا: أهمية الموضوع..... ٢

ثالثًا: أهداف الدراسة..... ٣

رابعًا: أسئلة الدراسة..... ٣

خامسًا: حدود الدراسة..... ٣

سادسًا: مصطلحات الدراسة..... ٣

سابعًا: الدراسات السابقة..... ٥

ثامنًا: ما ستضيفه هذه الدِّراسة..... ١١

تاسعًا: الصعوبات التي واجهت إعداد هذه الدراسة..... ١١

الفصل الأول..... ١٢

الإطار النظري..... ١٢

تمهيد..... ١٣

المبحث الأول: التحليل الصرفي العربي..... ١٤

أولًا: طرق التحليل الصرفي..... ١٥

١-متطلبات لغوية..... ١٧

٢-متطلبات تقنية (برمجية)..... ١٧

ثانيًا: أدوات التحليل الصرفي ١٨
الأداة الأولى: تيم باك والتر Tim Buckwalter ١٨
الأداة الثانية: عرب مورف ArabMorph ١٨
الأداة الثالثة: صخر SAKHAR ١٨
الأداة الرابعة: خوجا Khoja ١٨
الأداة الخامسة: MADA + TOKAN ١٩
الأداة السادسة: يمامة YAMAMA ١٩
الأداة السابعة: باما BAMA ١٩
الأداة الثامنة: MADA ١٩
الأداة التاسعة: الخليل AlKhalil ٢٠
الأداة العاشرة: مداميرا MADAMIRA ٢٠
الأداة الحادية عشرة: فراسة FARASA ٢٤
المبحث الثاني: معالجة الصرف العربي حاسوبيًّا ٢٦
المبحث الثالث: اسم الآلة اشتقاقه ومعالجته حاسوبيًّا ٣٠
أولًا: أسماء الآلة القياسية ٣٠
ثانيًا: أسماء الآلة السماعية ٣٠
ثالثًا: أسماء الآلة الجامدة ٣٠
رابعًا: اشتقاق اسم الآلة ٣١
خامسًا: المقيس من أوزان اسم الآلة المشتق ٣١
سادسًا: معالجة اسم الآلة حاسوبيًّا ٣٣
المبحث الرابع: الخوارزميات ٣٤
أولًا: خصائص الخوارزميات ٣٥
ثانيًا: أقسام الخوارزميات ٣٥
ثالثًا: خوارزميات التحليل والتوليد الصرفي الآلي للغة العربية ٣٦
الفصل الثاني ٤٢

تمهيد..... ٤٣
المبحث الأول: إجراءات الدراسة..... ٤٤
أولًا: اختيار المحللات الصرفية..... ٤٤
ثانيًا: منهج الدراسة..... ٤٧
ثالثًا: اختيار المدونات اللغوية..... ٤٧
رابعًا: اختيار برنامج التنقيب في المدونات..... ٤٩
الأداة الأولى: غواص Ghawwas..... ٤٩
الأداة الثانية: اسكتش إنجن Sketch Engine..... ٤٩
الأداة الثالثة: آنت كونك AntConc..... ٤٩
خامسًا: بناء العينة..... ٥١
المبحث الثاني: مجتمع وعينة الدراسة..... ٥٥
أولًا: مجتمع الدراسة..... ٥٥
ثانيًا: عينة الدراسة..... ٥٥
المدونة الأولى: المدونة اللغوية العربية العالمية..... ٥٥
المدونة الثانية: المدونة اللغوية العربية لمدينة الملك عبدالعزيز للعلوم والتقنية (KACST)..... ٥٥
ثالثًا: منهجية استخراج جمل العينة..... ٥٦
المبحث الثالث: أدوات الدراسة..... ٥٧
أولًا: أدوات التحليل الصرفي المستخدمة في الدراسة..... ٥٧
الأداة الأولى: مداميرا MADAMIRA..... ٥٧
الأداة الثانية: فراسة FARASA..... ٥٧
ثانيًا: أداة التنقيب في المدونات..... ٥٩
المبحث الرابع: تحديد مواضع اللبس الصرفي في العينة..... ٦٠
المبحث الخامس: منهجية بناء الخوارزمية وآلية تقييمها..... ٦١
أولًا: مشكلة التذكير والتأنيث..... ٦١
١-مشكلة اللبس في جنس الكلمة في محلل مداميرا MADAMIRA..... ٦٢

٢-مشكلة اللبس في جنس الكلمة في محلل فراسة FARASA ٦٤
ثانيًا: مشكلة العدد ٦٧
ثالثًا: مشكلة أصل الكلمة ٦٩
رابعًا: مشكلة نوع الكلمة ٧٤
خامسًا: تحديد آلية تقويم الخوارزمية ٧٥
الفصل الثالث ٧٨
تمهيد ٧٩
المبحث الأول: نتائج اختبار عينة الدراسة في المحللين الصرفيين مداميرا وفراسة ٨٠
المبحث الثاني: نتائج بناء الخوارزمية ٨٥
أولًا: نتائج معالجة مشكلة التذكير والتأنيث ٨٥
ثانيًا: نتائج معالجة مشكلة العدد ٨٨
ثالثًا: نتائج معالجة مشكلة وسم أصل الكلمة ٨٨
رابعًا: نتائج معالجة مشكلة وسم نوع الكلمة ٨٨
المبحث الثالث: نتيجة تقييم الخوارزمية ٩١
أبرز الملاحظات حول الخوارزمية ٩٣
الخاتمة: ملخص الدراسة ٩٦
النتائج ٩٨
التوصيات ٩٩
المراجع ١٠٠
أولاً: المراجع العربية ١٠١
ثانيًا: المراجع الأجنبية ١٠٤
قائمة الملاحق ١٠٥

فهرس الجداول

رقم	العنوان	الصفحة
١-١	مقارنة بين محللي مداميرا MADAMIRA ومدى MADA	٢٣
٢-١	نتائج التقييم لمداميرا MADAMIRA وفراسة FARASA	٢٥
١-٢	الاختبار التجريبي للعينة في محلل مداميرا MADAMIRA	٤٥
٢-٢	الاختبار التجريبي للعينة في محلل فراسة FARASA	٤٦
٣-٢	المدونات اللغوية	٤٨
٤-٢	حصر أسماء الآلة (القياسية والسماعية)	٥٢
٥-٢	أسماء الباحثات المحكمات لمواضع اللبس الصرفي في محللي الدراسة	٦٠
٦-٢	الخلايا المظللة بالأصفر تشير إلى جنس الكلمة الموسوم وسمًا خاطئًا في محلل مداميرا	٦٣
٧-٢	الخلايا المظللة بالأصفر تشير إلى جنس الكلمة الموسوم وسمًا خاطئًا في محلل فراسة	٦٥
٨-٢	الجمل المستخرجة من عينة الدراسة التي تحوي سمات مميزة على وزني مَفَاعِل ومَفَاعِيل	٦٦
٩-٢	الخلايا المظللة بالأصفر تشير إلى الكلمات الموسومة وسمًا خاطئًا من حيث العدد في محلل مداميرا	٦٨
١٠-٢	تكرار أسماء الآلة ذات الصيغ المختلفة في مدونتي الإسكندرية والمدونة العربية	٧٣
١١-٢	مواضع اللبس الصرفي في وسم نوع الكلمة في محلل فراسة FARASA	٧٤
١٢-٢	وسم الكلمات الهدف في محلل مداميرا MADAMIRA	٧٥
١-٣	نسب دقة محلل مداميرا MADAMIRA في وسم العينة	٨٠
٢-٣	نسب دقة محلل فراسة FARASA في وسم العينة	٨٠
٣-٣	توضيح للوسم الخاطئ لجنس الكلمة في محلل فراسة FARASA	٨٣
٤-٣	التعابير النمطية المستخدمة في توصيف القواعد اللغوية	٨٦
٥-٣	توصيف القواعد اللغوية وترميزها	٨٧
٦-٣	أسماء محكمي الجانب اللغوي وقاعدة البيانات	٩١

فهرس الأشكال

رقم	العنوان	الصفحة
١-١	نظرة على هندسة مداميرا MADAMIRA	٢١
٢-١	واجهة موقع مداميرا MADAMIRA (الموقع الرسمي)	٢٢
٣-١	واجهة موقع فراسة FARASA (الموقع الرسمي)	٢٤
١-٢	بعض المشكلات التي ظهرت أثناء استخدام برنامج غواص	٥٠
٢-٢	تحليل جملة من عينة الدراسة في الموقع الرسمي لمحلل فراسة FARASA	٥٨
٣-٢	تحليل الجملة السابقة في محلل فراسة FARASA	٥٩
٤-٢	نسبة أسماء الآلة المذكرة والمؤنثة في عينة الدراسة	٦٢
٥-٢	نسبة الوسوم الصرفية لجنس الكلمة في محلل مداميرا MADAMIRA	٦٣
٦-٢	نسبة وسوم جنس الكلمة في محلل فراسة FARASA	٦٤
٧-٢	نسبة الوسوم الصرفية من حيث العدد في محلل مداميرا MADAMIR	٦٩
١-٣	نسب دقة التحليل في محلل مداميرا ومحلل فراسة في وسم العينة	٨٠
٢-٣	الخوارزمية في صورتها المبدئية	٩٠
٣-٣	خوارزمية فك اللبس الصرفي الحاسوبي في اسم الآلة في صورتها النهائية	٩٤