



المملكة العربية السعودية
وزارة التعليم
جامعة الإمام محمد بن سعود
الإسلامية
معهد تعليم العربية
قسم اللُّغويات التطبيقية العربيّة

بناء خوارزمية لفك اللبس الصّرفي الحاسوبي في جمع الكثرة (الأوزان الرباعيّة)

رسالة مقدّمة لنيل درجة الدكتوراه في تخصص اللُّغويات التطبيقية العربيّة

إعداد الطالبة

هاجر بنت عبد الرحمن بن عبدالله القبيلي

المشرف العلمي

د. عبدالله بن يحيى الفيحي

رئيس قطاع الحوسبة اللُّغوية بمجمع الملك سلمان العالمي للُّغة العربية

أستاذ مشارك في اللُّغويات الحاسوبية

العام الجامعي

1446هـ / 2025م

ملخص الرسالة

تُناقش هذه الدراسة مشكلة اللَّبَس الصَّرْفِي في الأوزان الرباعيَّة لجمع الكثرة: فُعُول، وفِعَال، وفُعَل، وفَعَلَى، وفَعَلَة، وفُعَلَة، وفِعْلَة، كَنُوم، وجَمال، وصُوم، وقَتْلَى، وسَحَرَة، ورُماة، وقِرْدَة في اثنين من المحللات الصَّرْفِيَّة العربيَّة، وهما: محلل مداميرا (MADAMIRA)، ومحلل كامليرا (Camelira)؛ وذلك بهدف بناء خوارزمية حاسوبية تُسهم في رفع دقة أداء المحللات الصَّرْفِيَّة العربيَّة في معالجة تلك الأوزان.

وقد انتهجت الباحثة في هذه الدراسة المنهج الوصفي المعتمد على الإحصاء، واستخرجت عيَّنتها المكونة من 175 جملة من المدونة العربيَّة المتاحة على منصة فلك للمدونات اللُّغوية، ثم حللتها في محلي الدراسة، واستعرضت النتائج وناقشتها، وحددت بناء عليها المشكلات التي ستعالجها في هذه الدراسة، وهي: مشكلة استخراج نوع الكلمة، ومشكلة معرفة جنس الكلمة، ومشكلة معرفة دلالة العدد في الكلمة.

وبعد دراسة تلك المشكلات ومناقشتها وضعت الباحثة العديد من القواعد اللُّغوية التي بُنيت على أساسها الخوارزمية الحاسوبية، ثم حُكِّمت لدى أساتذة متخصصين في اللغة العربيَّة واللُّغويات الحاسوبية. وقد أجاز المحكِّمون تلك الخوارزمية في جانبها اللُّغوي والحاسوبي، وأكدوا على وضوحها، وقدرتها على الإسهام في معالجة مشكلات الدراسة.

وقد أوصت الباحثة في ختام دراستها بأهمية إجراء المزيد من الدراسات اللُّغوية الحاسوبية في موضوع جموع التكسير بمختلف جوانبه، وضرورة ربط نتائج المعالجة الصَّرْفِيَّة بالمعالجة النَّحْوِيَّة؛ نظراً للارتباط الشديد بين المستويين. كما أوصت بإضافة التحليل الدلالي لمعالجة مشكلات التعرف على جموع التكسير؛ إذ إنَّ ذلك سيسهم بشكل كبير في رفع دقة تعرف المحللات عليه.

الكلمات المفتاحية: جمع التكسير، جمع الكثرة، اللَّبَس الصَّرْفِي، التحليل الصَّرْفِي، المحللات الصَّرْفِيَّة.

Abstract

In this thesis, titled “Building an Algorithm for Arabic Morphological Disambiguation of the Plural of Multitude (Jam‘ al-Kathrah)”, the researcher addresses the issue of morphological ambiguity within Quadruple patterns: Fu‘ūl (فُعُول), Fi‘āl (فِعَال), Fu‘aal (فُعَل), Fa‘lá (فُعَلَى), Fa‘alah (فُعَلَة), Fu‘alah (فُعَلَة) and Fi‘alah (فِعَلَة) using MADAMIRA and Camelira, two Arabic morphological analyzers. The primary aim of this thesis is to develop an algorithm that enhances the accuracy of Arabic morphological analyzers.

A descriptive statistical approach was adopted for this study. The researcher used 175 sentences as a linguistic sample from the Arabic Corpus that available on Falak platform for Arabic language corpora. Subsequently, she analyzed the sample using both MADAMIRA and Camelira, followed by a review and discussion of the results. After that, the researcher specified three key problems for discussion: First, the challenge of specifying part of speech; Second, the challenge of determining the words' gender; and third, the challenge of determining the number that the word represent.

After studying and discussing those problems, the researcher built an algorithm based on a set of linguistic rules. She then collaborated with specialized professors, and co-professors in Arabic Language and Computational Linguistics to arbitrate the linguistic rules and algorithm. Afterwards, the arbitrators authorized the linguistic rules and the algorithm, and they emphasized their clarity and ability to help solving the problems.

In conclusion, the researcher highlighted the necessity for further computational linguistic studies on Broken Plurals (Jumū‘ al-taksīr) in all their aspects. She also stressed the importance of linking the outcomes of morphological processing with syntactic processing, given the strong correlation between these two linguistic levels. Additionally, she recommended incorporating semantic analysis to address the challenges associated with recognizing broken plurals (Jumū‘ al-taksīr), as this approach could significantly improve their accuracy by Arabic analyzers.

قائمة المحتويات

أ	 حقوق النشر
ب	 إهداء
ج	 شكر وتقدير
د	 مستخلص الرسالة
هـ	 المستخلص باللغة الإنجليزية (Abstract)
و	 قائمة المحتويات
ح	 قائمة الجداول
ط	 قائمة الأشكال
1	 الفصل الأول: المقدمة
2	 1. مشكلة الدراسة
2	 1.1. أسئلة الدراسة
2	 2.1. حدود الدراسة
3	 2. أهمية الدراسة وأهدافها
3	 1.2. أهمية الدراسة
3	 2.2. أهداف الدراسة
4	 3. منهجية الدراسة
4	 1.3. منهج الدراسة
4	 2.3. مجتمع الدراسة
4	 3.3. إجراءات الدراسة
4	 4. مصطلحات الدراسة
5	 5. الدراسات السابقة
6	 1.5. الدراسات العربية
9	 2.5. الدراسات الأجنبية
13	 3.5. الإضافة العلمية للدراسة
15	 6. فصول الدراسة ومباحثها
17	 الفصل الثاني: الإطار النظري
18	 المبحث الأول: المعالجة الآلية للغة العربية

29	المبحث الثاني: اللَّبْس الصَّرْفِي والمحللات الصرفيَّة العربيَّة
42	المبحث الثالث: معالجة جموع التكسير حاسوبياً
50	الفصل الثالث: منهجيَّة الدراسة وإجراءاتها
51	1. منهج الدراسة
51	2. مجتمع الدراسة
51	3. عينة الدراسة
59	4. أدوات الدراسة
62	5. إجراءات الدراسة
63	الفصل الرابع: نتائج الدراسة ومناقشتها وتفسيرها
64	المبحث الأول: تحليل البيانات
64	المطلب الأول: اختبار العينة في محلل الدراسة
72	المطلب الثاني: تحديد المشكلات الظاهرة في محلل الدراسة
76	المبحث الثاني: عرض نتائج الدراسة، ويشمل ثلاثة مطالب:
76	المطلب الأول: معالجة المشكلات الظاهرة في محلل الدراسة
96	المطلب الثاني: بناء خوارزمية حاسوبية
107	المطلب الثالث: تقييم الخوارزمية
125	المبحث الثالث: مناقشة النتائج وتفسيرها
134	الفصل الخامس: الخاتمة
135	1. ملخّص نتائج الدراسة
138	2. توصيات الدراسة ومقترحاتها
140	قائمة المصادر والمراجع
140	المصادر والمراجع العربيَّة
143	المصادر والمراجع الأجنبية
145	الملاحق

قائمة الجداول

الصفحة	العنوان	رقم
13	مقارنة الدراسات السابقة بالدراسة المقترحة	1-1
52	نسبة كل وزن من الأوزان الرباعية من جمع الكثرة بالنسبة لكلمات اللُّغة كلها، وبالنسبة للأوزان الرباعية من جموع الكثرة في أكثر 10.000 كلمة وروداً في المدونة العربيّة	1-3
53	عيّنة الدراسة	2-3
68	نتائج تحليل عينة الدراسة في محلل مداميرا (MADAMIRA)	1-4
68	نتائج تحليل عينة الدراسة في محلل كامليرا (Camlira)	2-4
81	قاعدة بيانات الكلمات السّماعية لوزن (فَعْلَى)	3-4
83	الضمائر والحروف والأدوات المستخدمة في توصيف القواعد اللُّغوية	4-4
92	إعادة تقييم عينة الدراسة من ناحية الجنس اعتماداً على منهجيّة المحلل في التحليل	5-4
97	التّعابير النّمطية Regular Expression المستخدمة في توصيف القواعد اللُّغوية	6-4
99	ترميز القواعد اللُّغوية لتمييز الأسماء الثلاثيّة عن الأفعال الثلاثيّة	7-4
100	ترميز القواعد اللُّغوية الخاصّة بوزن (فَعْلَى)	8-4
101	ترميز القاعدة اللُّغوية الخاصّة بوزن (فَعْلَة)	9-4
101	ترميز القاعدة اللُّغوية الخاصّة بوزن (فُعْلَة)	10-4
101	ترميز القواعد اللُّغوية العامّة	11-4
104	القواعد اللُّغوية الخاصّة بوزن (فِعْلَة)	12-4
115	قاعدة بيانات الكلمات السّماعية لوزن (فَعْلَى) بعد التحكيم	13-4
118	ترميز القواعد اللُّغوية لتمييز الأسماء الثلاثيّة عن الأفعال الثلاثيّة بعد التعديل	14-4
119	القواعد اللُّغوية العامّة بعد التعديل	15-4
122	القواعد اللُّغوية الخاصّة بوزن (فِعْلَة) بعد التعديل	16-4

قائمة الأشكال

الصفحة	العنوان	رقم
62	واجهة محلل كامليرا (Camelira)	1-3
69	المقارنة بين محلي مداميرا (MADAMIRA) وكامليرا (Camlira)	1-4
106	خوارزمية الدراسة	2-4
124	خوارزمية الدراسة النهائية بعد التعديل	3-4

قائمة الملاحق

الصفحة	العنوان	رقم
145	الكلمات الواردة في كتب الصَّرف العربي على وزن فُعْل، وورودها في مدونة الدراسة	1
147	الكلمات الواردة في كتب الصَّرف العربي على وزن فِعْلَة، وورودها في مدونة الدراسة	2
148	نتائج تحليل عينة الدراسة في محلل مداميرا (MADAMIRA)	3
170	نتائج تحليل عينة الدراسة في محلل كاميليرا (Camelira)	4