

منهج التحليل الآلي في مجال: لغويات المدونة الحاسوبية

CORPUS LINGUISTICS

عبد القادر ميسوم*

جامعة حسيبة بن بوعلي بالشلف (الجزائر) arabeuhbc@gmail.com

الملخص :

تهدف هذه الدراسة إلى تقديم قراءة معلوماتية منهجية في كتاب: (لغويات المدونة الحاسوبية، تطبيقات تحليلية على العربية الطبيعية) بهدف التعريف بالكتاب شكلاً ومضموناً وتوضيح أهميته العلمية ضمن نطاق "التحليل الآلي في اللسانيات الحاسوبية". باستخدام برامج في التحليل الآلي من مثل: (غواص ، أنتكونك ، وسحب الكلمات). بالنظر إلى ذلك ، فإن هذا يفرض نفسه بقوة اليوم في عدة مجالات حيث نلمس فيها بونا شاسعا عن الجهود البشرية في تحليل المعطيات والنصوص ، وعلى هذا الأساس سنحاول الكشف عن معالجة النصوص العربية تلقائياً وفقاً لعناصر هذا الكتاب.

الكلمات المفتاحية: المدونة الحاسوبية؛ التحليل الآلي؛ العربية الطبيعية؛ برنامج غواص؛ أنتكونك؛ سحب الكلمات.

Abstract:

This study aims to provide a systematic informative reading of the book: (Corpus Linguistics, Analytical Applications on Natural Arabic language) with the aim of introducing the book in form and content and clarifying its scientific importance within the scope of "Automated Analysis in Computational Linguistics". Using programs in automated analysis such as: (Ghawwas, Antconc, and word clouds). Given that, this imposes itself strongly today in several areas where we see a huge gap in human efforts in analyzing data and texts, and on this basis, we will try to detect the processing of Arabic texts automatically according to the elements of this book.

Keywords : Corpus Linguistics, Automated Analysis, Natural Arabic language, Ghawwas, Antconc,

تصدير :

صدر أخيراً عن مركز الملك عبد الله بن عبد العزيز الدولي، الرياض 2016 ، كتاب رقمي جديد لمجموعة من الباحثين- يتصدرهم على قائمة التحرير د / سلطان بن ناصر المحيول- في مجال لغويات المدونة الحاسوبية، تطبيقات تحليلية على العربية الطبيعية، يقع في 266 صفحة. يعتبر هذا المؤلف الرقمي إضافة نوعية في مجال اللسانيات الحاسوبية، ومعالجة العربية آلياً، إذ أصبحت الدراسات اللغوية اليوم مسايرة لركب العالم الرقمي في مجالات مختلفة، مما جعل كل ما هو غير حاسوبي إما خاضعاً لها أو مهمشاً، وهذا هو شأن اللسانيات العربية اليوم التي تحاول أن تقتحم مجالات البحث في حوسبة اللغة العربية كنظيراتها من اللغات الأخرى. فقد أدرك الباحثون في الجامعات الغربية أهمية معالجة اللغات الطبيعية آلياً فطوروا أدوات حاسوبية ومصادر وبرمجيات بداية من اللغة الإنجليزية، ثم توسع مجال البحث ليشمل اللغات الأخرى منها اللغة العربية والكتاب الذي بين أيدينا يعد إسهاماً متميزاً في خدمة لسانيات المدونات العربية.

دواعي اختيار هذا الكتاب:

يتناول هذا الكتاب قضايا مختلفة في مجال اللسانيات الحاسوبية وبالتقريب جداً - كيفية التعامل مع المدونات الحاسوبية بالتحليل والدراسة باستخدام مناهج آلية في التحليل وهو المحور العام الذي يطرحه هذا الكتاب، وسيأتي تفصيل ذلك في عرض المحتوى. وأعتقد أن ما قام به فريق البحث هنا يعد خطوة هامة يقل نظيرها في الوطن العربي لأنها أبحاث ركزت على التطبيق أكثر من التنظير، قد تشكل أرضية معرفية للنهوض باللغة العربية في العالم الرقمي. وليس الغرض من تقديم هذا الكتاب حصر المعالجات الآلية، كونها واسعة المجال ومتشعبة الموضوعات بتشعب أغراض المعالجة، وليس حصر كذلك إشكالية المنهج في اللسانيات الحاسوبية بل هدفنا هو التعريف بطبيعة التطبيقات التي قدمها الباحثون في هذا المجال.

قراءة في العنوان:

مما لا شك فيه، أن انتقاء عنوان على هذه الشاكلة " لغويات المدونة الحاسوبية- تطبيقات تحليلية على العربية الطبيعية- مغامرة علمية مخفوفة بالتحدي، لسبب بسيط هو أنه عنوان عريض وإشكالي في الآن نفسه، إذ يصعب أن يحاط بكافة الجوانب التقنية التي تخص المدونات اللسانية، ثم إن

محاولات إنشاء برمجيات تخص اللغة العربية بالمستوى المطروح اليوم، كأن تكون البرامج في أساسياتها لغة آلية تستمد منطقها من الرياضيات ثم تأتي عملية الاسقاط على اللغات الأجنبية، وعليه ندخل في مرحلة المحاكاة والاسقاط ونجرب على اللغة العربية. والأساس في نظرنا غريب عن خصوصية اللغة العربية، والسؤال المطروح إلى أي حد استطاع هؤلاء الباحثين الإحاطة بموضوع الكتاب؟ وما هي الأدوات الإجرائية والمناهج المستخدمة في تحليل المدونة اللغوية الحاسوبية؟ أسئلة كذه تحتاج منا تصفح الكتاب لتبيان خيطه الأبيض من خيطه الأسود.

كشف المتن:

إن إلقاء نظرة فاحصة على محتويات الكتاب، يسمح لنا بالقول إن جهود هؤلاء الباحثين أمر بالغ الأهمية، بدليل أن كل فصل من فصول الكتاب لا يخرج عن إطار العنوان الكبير للكتاب، بصورة تجعل كل فصل ممثلاً بمنهج النظرية والتطبيق من عدة مناهج ونظريات مختلفة، بمعنى أن كل باحث انتقى المنهج المناسب لطبيعة موضوع البحث، وإن شئنا الاحتكام إلى الكتاب فحري بنا القيام بمسحة شمولية في فهرس الموضوعات الذي جاءت عناوينه كما يلي:

في الفصل الأول: (المعالجة الآلية للصحف العربية: تحليل الخطاب بمناهج BCU) يطرح الدكتور سلطان المجيول طبيعة المعالجة الآلية للخطاب، والمعنية بوعاء الصحف العربية في المدونة العربية التابعة لمدينة الملك عبدالعزيز للعلوم والتقنية، مع شرح المعالجة القائمة على تحليل الأنماط الخطابية. ولم يكن الغرض من الفصل تقديم تحليل شمولي لكل موضوعات تحليل تواليات الخطاب (الفوخطابي metadiscourse) بسبب محدودية الفصل، بل كان لغرض تسليط الضوء على طبيعة الاتجاهات التحليلية الخطابية القائمة على المعالجة الآلية: تحليل الخطاب بالاعتماد على مناهج باير وكورنر وأبتون المصطلح عليها approaches BCU.

وفي الفصل الثاني (المعالجة الآلية للعربية الإدارية: دراسة مدونات الإدارة وأخبار الأسهم والاقتصاد في الوطن العربي، للدكتور محمود عثمان الحاج) جاءت تطبيقات المعالجة الآلية على نصوص الإدارة وأخبار الأسهم والاقتصاد في الوطن العربي. ويفيد هذا الفصل بمنهج تحليلي آلي يعتمد على أدوات حاسوبية مهمة لمعالجة نصوص ضخمة خاصة بالتقارير المالية للشركات في الوطن العربي، مكوناً من ذلك ثلاث مدونات أساسية. وقد عالج مؤلف الفصل هذه المدونات بطرائق عديدة: توسيم

مدونات الدراسة آليا، وأدوات معالجتها، ومنهج مقروئية النص المعتمد على الكشف عن الكلمات الطويلة والمعقدة في النصوص، وأدوات معالجة اللغة (غواص وأنت-كوك)، وسحابات الكلمات clouds word ونتائجها، وتصنيف الوثائق بواسطة تعلّم الآلة.

وفي الفصل الثالث (المعالجة الآلية للكلمات المميزة للمدونات: قضايا تقنية للدكتور عبدالمحسن بن عبيد الشبتي) سلّط الباحث الضوء على قضايا تقنية متعلقة بحساب الكلمات المميزة keywords ، وكيف أن لحساب هذه الكلمات بين المدونة الرئيسة والمدونة المرجعية أثرا غير خاضع فقط لحجم المدونة المرجعية الذي يتطلب أن يكون أكبر بكثير من حجم المدونة الرئيسة (كما هو مقرر في أدبيات لغويات المدونة الحاسوبية)، بل برهن الباحث على وجود تأثيرات أخرى مرتبطة بالمنطقة الجغرافية، وبزمن إنتاج النص الطبيعي، وبالموضوعات.

وفي الفصل الرابع (المعالجة الآلية لمدونات المتعلمين: تحليل الأخطاء في المدونة اللغوية لمتعلمي اللغة العربي للدكتور عبدالله بن يحيى الفيقي) الباحث طرائق الاستفادة من مدونة المتعلمين التي صممها، وذلك عن طريق وضع الأسس الأولية لمفاهيم المعالجة الآلية لمدونات المتعلمين، ومناهج تحليل الأخطاء فيها، وتحديد نوعها، ووصفها، وتفسيرها، وتقويمها بشكل علمي ومنهجي.

وفي الفصل الخامس (تقدم الدكتورة هند بنت مطلق العتيبي المعالجة الآلية للمدونات المتوازية واستخداماتها في تعليم اللغات وتدريب المترجمين) مفهوم المدونات المتوازية وأهميتها في تدريب المترجمين، وتعرض مشروع AEPC الذي تشرف الباحثة عليه في قسم اللغة الإنجليزية بكلية اللغات والترجمة في جامعة الملك سعود، وبالتعاون مع مجموعة إيوان البحثية، مع تبيان مراحل الأولى التي استأنفت، ومراحل الثانية التي مازالت في طور التطوير وإكمال البناء والتصميم.

ويحلل الدكتور عقيل بن حامد الشمري والدكتور محمود بن عبدالله المحمود في الفصل السادس المعالجة الآلية لوعاء الأخبار: تحليل الخطاب النقدي المعتمد على المدونة الحاسوبية (نصوص) الأخبار في كل من قناتي الجزيرة والعربية حول أحداث الحرب على غزة، وقد وظفا منهج تحليل الخطاب النقدي بأدوات معالجة اللغة والتحليل الإحصائي الكمي، مفيدتين حقل التحليل بمناهج معتمدة على أداة من أدوات المدونة الحاسوبية لمجال تحليل الخطاب العربي.

التعريف بالمعضلة:

سيتيح لنا هذا الكتاب فرصة شائعة ونحن نمنع النظر في ضروب المقاربات التي أثارت إشكالية المنهج في اللسانيات الحاسوبية، وهي تتطور في اتجاه العمق والتدقيق. ولا نريد أن نسترسل هنا في تحديد كل الإسهامات التي ذكرها الباحث في كتابه، إنما سيكون تركيزنا على ما طرحه من أفكار تصل من قريب بمجال الدرس اللساني الحاسوبي ونخص تحديداً " تحليل الأنماط الخطائية بمناهج BCU للباحث " سلطان بن ناصر المجبول" معتمداً مناهج حديثة في تحليل الخطاب discours analysis ومطبقة على خطاب الصحف العربية كمدونة حاسوبية .

تعريف المدونة الحاسوبية:

يمكن تعريف المدونة اللغوية (corpus) بأنها مجموعة من النصوص اللغوية المكتوبة -أو الأحاديث الشفهية- الطبيعية، التي تجمع لأغراض محددة، وتحفظ بطريقة قابلة للقراءة والبحث حاسوبياً وتتعدد التسميات العربية للفظة corpus، لكن تسميتها "مدونة لغوية" قد تكون الأكثر شيوعاً وكذلك وردت في مجموعة من معاجم ترجمة المصطلحات اللغوية والحاسوبية بهذه التسمية. والسؤال المطروح: كيف طبق الباحث منهج تحليل الخطاب النقدي على مدونة حاسوبية؟ وأظن أن المسألة هنا تحتاج إلى إمعان نظر. وحال تحليل الخطاب العربي لا يعدو أن يخرج عن أمرين؛ الأول: مدونة ورقية قليلة الحجم، وممكن تحليلها بالحدس والعين واليد، أما الثاني: انصباب ركائزه على موضوعات اللغة، والحجاج، والاتساق، والانسجام، واستراتيجيات الخطاب، والتداولية. فكيف جاء الطرح؟

يطرح " سلطان بن ناصر المجبول" أرضية مصطلحاتية أساسية ليوضح كيفية تناول وتطبيق آليات تحليل الخطاب فيكون حول كثرة التشابهات في هذا المجال، وحتى يتمكن من إبراز الفروقات وتبيان وجهة النظر يتناول المفهوم الجامع بقوله " وتحليل الخطاب ليس حقلاً ذا علم صارم وبمتصورات تأسيسية، ومرحلية، ومن ثم كيانية؛ كونه قد بدأ قديماً في سياق دراسات النصوص البلاغية وفقه اللغة إلى أن تضافرت مدارس وحقول أخرى في تكوين شعبه، وسبقه تحليل المحتوى content analysis الذي ظهر في مطلع القرن العشرين في الولايات الأمريكية المتحدة، ثم تشعبت بتشعب المدارس اللسانية والوظيفية والإثنوجرافية، بدءاً من عام 1952 عند هاريس Harris ثم نال أبعاداً

عديدة متضافرة التداخل التداولي والنفسي والاجتماعي والسلوكي، غير أن حجم الخطاب المحلل في المجمل لا يتجاوز صفحات معدودة سواء كانت مكتوبة أو منقولة نَسْخًا من المنطوق. وبعد تطور أدوات معالجة اللغات الطبيعية ظهرت أبحاث في تحليل الخطاب تقوم على جمع نصوص رقمية تبلغ آلاف الكلمات، أكبر من احتمال العين واليد والورق واعتمدت هذه الأبحاث على تحليل التكرارات والكشافات السياقية بوصفهما المنطلق لأي منهج من مناهج تحليل الخطاب.

المنهج وكيفية الاشتغال:

جدير بالذكر أننا رجعنا إلى الكتاب المصدر بعنوان "**Discours on the move**" في محاولة منا للتوسع في النقاط الأساسية المعتمدة في طريقة التحليل. والنتيجة أن الدراسة وظفت هنا مناهج حديثة في تحليل الخطاب analysis discourse وتحليل الخطاب النقدي critical analysis discourse بالاعتماد على ما قدمه كل من باير Biber وكونر Conner وأبتون Upton والمصطلح عليه بمنهج BCU. ووظفت طرائق معالجات هذه المناهج الآلية على خطاب الصحف العربية، والعمل على تحليل أنماط النسوية واللانسوية أنموذجًا. وللعلم أن منهج BCU هو المرحلة الثالثة بعد مرحلتين من مراحل تحليل الخطاب؛ الأولى: مرحلة العين واليد والخطاب الصغير غير الحاسوبي، والثانية: مرحلة الاعتماد على أدوات معالجة اللغة (التكرارات والكشافات السياقية..). وينطلق هذا الاتجاه من أساس مهم؛ وهو: تنوع النصوص أو المجالات (الأوعية genres أو المجالات domains للنصوص الضخمة)، وإعمال المنهج التنازلي top-down ؛ أي: توزيع الأوعية، ثم إعمال المنهج التصاعدي bottom-up ، ومن بعده يعمد إلى تحليل الخطاب تفصيلًا لكل نص من النصوص، ومن ثمة تجزئة مجالات النصوص، ثم الموضوعات، والعودة من الأخير إلى عموم التشكل النصي للأوعية، وذلك بأدوات تحليلية تساعد على ذلك.

وفي ما يلي رصد لأهم الخطوات المتبعة في مجال تحليل الخطاب المتحرك كما هو موضح حيث اعتمد باير في تحليل المدونة (الخطاب) على منهج متعدد الأبعاد، - multidimensional analysis أو ما يعرف ب MD analysis ، الذي يقوم على دراسة أنواع النصوص بالاعتماد على تحليل العامل factor analysis وكشف نسيج الخطاب المتعدد بالأوعية والمجالات. ويعتمد

التحليل-ببساطة:- كشف طبيعة التنوع اللغوي. عبر مجموعة ضخمة من النصوص ويُعتمد على كشف أبعاد طبيعة النصوص بين الأوعية والمجالات والأزمنة والمناطق الجغرافية على النصوص الضخمة. وينطلق هذا المنهج-ببساطة-من أساس حاسوبي يُعالج الخطاب وتحليله بكونه متحرك غير ساكن بمعنى: أن تحليل الخطاب القائم على موقف واحد لا يعني إلا موقفه الخاص، بينما تحليل الخطاب المتحرك، القائم على المنهج المساعد على دراسات الأوعية (أوالأجناس النصوية) genres بوصفه خطابات متحركة ذات وحدات دلالية ووظيفية يمكن تحديدها من أجل تحديد أغراضها الاتصالية وعناصرها اللغوية المتداخلة.

وفي هذه الدراسة، نلاحظ أن تحليل الخطاب يقوم على استعمال أدوات معالجة اللغة، وعلى جمع مدونة حاسوبية ممثلة لخصوصية التحليل ونوعية الخطاب، ثم توظيف الإحصاءات التي تُظهر عدد تكرار الكلمات والمتصاحبات، وعلى الكشافات السياقية الخاصة بالكلمات المركزية الخاضعة للتحليل في المدونة الحاسوبية الضخمة دون المدونة الورقية الصغيرة في حجمها. وعلى مزية التحليل النقدي بالاعتماد على المدونة الحاسوبية وأدوات معالجة اللغة، إلا أن ثمة منهجًا أكثر اتساعا، وهو المنهج الموضوعاتي الذي يقوم على تحليل الخطابات المتنوعة بتنوع الأوعية والمجالات، ويقوم على تجزئة هذه النصوص، وتصنيفها وفقا لموضوعات الممارسات الاجتماعية والاقتصادية والسياسية، وموضوعات التواصل اللغوي بين الثقافات من حيث طبيعة التفاعل الناتج منها، وكل ذلك من أجل إعادة لُمَكة متجزئات الخطابات المتعددة من جديد، وبناء الوحدات الخطابية فيها بشكل متقابل وصولا إلى جسم المدونة الضخمة بصورة مختلفة عن جسمها الأول، لكن لخدمة أهداف وفرضيات مسطرة مسبقا.

تطبيق المنهج على مدونة الدراسة - الصحف العربية:

في هذا العنصر تتضح آليات الاشتغال في خطوات منهجية متبعة ومقسمة إلى حقول دلالية، وبالتالي سيقوم الحاسوب بتحليل وجمع وترصيف الكلمات وفق المعطيات أو المدخلات المبرمج عليها سلفا مستخدما البرنامج الاحصائي R-Studio كما هو مبين فيما يلي:

لقد اختار الباحث - سلطان المحيول - المفردة المركزية المرأة/ النساء من وعاء الصحف العربية في عدة مجالات وموضوعات، من عام 1900 حتى 2013 فيما يخص قضايا المرأة المطروحة في الصحف العربية منها:

الأول: علاقة التشارك وطبيعته وكيفية تشكّل العلاقة النصية في الخطاب بين المرأة والرجل

الثاني: ما للمرأة من حقوق، وطبيعة الأصوات تجاه هذه الحقوق سواء كانت نسوية (إيجابية) كما (في) أو لانسوية (سلبية).

الثالث: المرأة والمشاركة المجتمعية أو السياسية، ودلالاتهما.

الرابع: الأنماط الخطابية لطبيعة المرأة العربية في نواة المجتمعات الصغيرة (الأسرة).

الخامس: النصوص المتعلقة بالمرأة والعمل.

السادس: النمط المعجمي للسجل اللغوي الخاص بمنطقة جغرافية محددة و من تقارير دولية تخص هذا الشأن.

السابع: تحليل التصاحبات اللفظية من حيث طبيعة معانيها الناتجة (إيجابية، سلبية)، أو من حيث وضع الفاعلين في كشافاتها السياقية (هيمنة، سلطة، عادات) وفقا للموضوع theme والحمول rhyme.

ولا يقتصر الأمر في قوائم التصاحب اللفظي على تحديد طبيعة البحث التحليلي (لغوي، أو اجتماعي، أو سياسي، أو ثقافي، أو نفسي، أو طبي)، بل يتجاوز تحليل التصاحب اللفظي تحليلات أكثر أهمية؛ ومنها: تحليل التفضيلات دلاليا. وكل هذا قائم على التحليلات الاحصائية المعمول بها في مجال لغويات المدونة الحاسوبية.

وحرري القول إن هذه الدراسة - كما أشرنا إلى ذلك في الملخص والمقدمة - ستركز على أصناف الممارسات الاجتماعية النسوية والانسوية في سياق الموضوعات الاجتماعية الخاصة بالمرأة العربية، والمحددة مسبقا (الزواج، والطلاق، والخلع، والحضانة، والعنف، والتعليم، والتمييز، والحجاب، والرياضة، والسفر بدون محرم، وقيادة السيارة، والتحرش، ونزع ولاية الرجل؛ كونها الأكثر ضخامة من حيث الحجم في الأوساط الاعلامية المكتوبة والمسموعة والمرئية. كما أن تحليل هذا الكم من

الخطابات بالاعتماد على العين واليد مستحيل، أما بالاعتماد على برامج حاسوبية سيكون الأمر أكثر دقة وسهولة وعلمية.

استنتاج:

واقع تحليل الخطاب بمناهج باير وكونر وأبتون Upton Biber, Connor and (2007) ، وبالاعتماد على أدوات وأساليب المعالجة الآلية مهما كانت مصادرها الموثوق بها تحليلًا ومنهجيًا (وهي عديدة وكثيرة)، يشير إلى أن الاتجاه إلى النصوص الضخمة، ومعالجة أنماط لغوية وحجاجية وخطابية وتواصلية ضمن عدد ضخم من الوحدات الخطابية المتضمنة في بنية نصية واحدة، وضمن عدد ضخم من البنى النصية المتضمنة في مدونة التحليل الخطابي أو التحليل الخطابي النقدي، قد بات ملحقًا؛ لأنها أي: هذه الاتجاهات (تفيد البحث والتحليل للخطاب المتعدد الأوعية والمجالات بعدة نتائج دقيقة وشاملة .

بيبلوغرافيا البحث:

- ¹ يمكن الرجوع إلى أبحاث جامعة ليدز في مجال لسانيات المدونة الموقع :- <https://elhadinlp.wordpress.com>
- ² - علي القاسمي، علم المصطلح أسسه النظرية وتطبيقاته العملية، 2008، مكتبة لبنان . ناشرون، ص 683
- ³ - سلطان بن ناصر المحيول وآخرون ، لغويات المدونة الحاسوبية - تطبيقات تحليلية على العربية الطبيعية، دار وجوه للنشر والتوزيع، الرياض، 2016 ، ص: 13¹
- ⁴ - المرجع نفسه ، ص: 14
- ⁵ - Douglas. Biber, Ulla. Connor, Thomas. Aupton. Discours on the move. Amesterdam. USA. 2007

6. خطابي، محمد. لسانيات النص، مدخل إلى انسجام الخطاب. المركز الثقافي العربي،¹ بيروت، 2006 ، الطبعة الثانية ، ص: