Ain Shams University Information Network
جامعة عين شمس
شبكة المعلومات الجامعية
@ ASUNET

بسم الله الرحمن الرحيم

شبكة المعلومات الجامعية

جامعة عين شمس

التوثيق الالكتروني والميكروفيلم

قسم

نقسم بالله العظيم أن المادة التي تم توثيقها وتسجيلها

علي هذه الأفلام قد أعدت دون أية تغيرات

يجب أن

تحفظ هذه الأفلام بعيدا عن الغبار

في درجة حرارة من ١٥-٢٥ مئوية ورطوبة نسبية من ٢٠-٤٠%

To be Kept away from Dust in Dry Cool place of 15-25- c and relative humidity 20-40%

شبكة المعلومات الجامعية

التوثيق الالكتروني والميكروفيلم

Ain Shams University Information Network
شبكة المعلومات الجامعية
@ ASUNET

بعض الوثائق الاصلية تالفة

بالرسالة صفحات لم ترد بالاصل

جامعة القاهرة

كليـة دار العلـوم

قسم علم اللغة والدراسات السامية والشرقية

اللبس الدلالي في المعالجة الآلية للغة العربية المعاصرة المكتوبة

رسالة مقدمة لنيل درجة الدكتوراه

إعداد الطالب

حسين محمد علي البسومي

إشراف

أ.د. محمد حسن عبد العزيز
أستاذ علم اللغة بكلية دار العلوم جامعة القاهرة
وعضو مجمع اللغة العربية بالقاهرة
مشرفا

أ.د. سلوى السيد حمادة
أستاذ الإلكترونيات والاتصالات بمعهد
بحوث الإلكترونيات بالمركز القومي للبحوث
مشرفا مشاركا

1431هـ - 2011م

شكر وعرفان

أحمد الله تعالى أن هيأ لي أسباب طلب العلم، فرزقني التعلم على يدي أستاذين كريمين، أفاضا علي من علمهما وفضلهما ما يجعلني أدين لهما دائما بجليل معروف وعظيم امتنان، فما من حسنة في هذه الرسالة إلا وهي ثمرة لغرسهما الطيب، ورعايتهما الكريمة. ووفاء ببعض حقهما علي، وعرفانا بفضلهما فإني أتقدم بوافر شكري وتقديري إليهما:

الأستاذ الدكتور/ محمد حسن عبد العزيز

الذي علمني كيف أبحث بعلم الأستاذ وحكمة المربي، فمهد لي السبيل بتوجيهاته ونصائحه، وتعهدني بالتقويم والإرشاد، وبذل لي كثيرا من جهده ووقته، فأسأل الله تعالى أن يجعل له من علمه وفضله نورا في الدنيا والآخرة، وأن يرزقه ما يرضيه.

الأستاذة الدكتوره/ سلوى السيد حمادة

التي أجلّت لي بكريم أخلاقها وسعة علمها وخبرتها كثيرا من صعوبات الجانب الحاسوبي لهذه الرسالة، فأسأل الله تعالى لقاء ما قدمت وأنفقت من وقتها وجهدها أن يجري الخير على يديها، وأن يرزقها ما يرضيها.

كما أتقدم بوافر الشكر والامتنان إلى أستاذيَّ عضوي لجنة المناقشة **الأستاذ الدكتور إبراهيم الدسوقي رئيس قسم علم اللغة بالكلية، والأستاذ الدكتور أحمد شرف الدين الأستاذ بكلية الحاسبات والمعلومات جامعة حلوان** على قبولهما مناقشة البحث، وتوجيه الباحث وتقويمه، فأسأل الله تعالى أن يجزيهما خيرا على ما ينفقانه من علم ووقت وجهد.

فهرس الموضوعات

- **المقدمة** أ – ك
- **الفصل الأول: مباحث تمهيدية** 1–37
  - المبحث الأول: المعنى: مفهومه وأنواعه 2
  - المبحث الثاني: المقاصد العامة للغة العربية 5
    - المقصد الأول: الإفادة 6
    - المقصد الثاني: الاقتصاد 7
    - المقصد الثالث: الثراء 10
  - المبحث الثالث: اللبس الدلالي في المعالجة الآلية للغة العربية: مفهومه وخصائصه 11
    - أولا: مستويات المعالجة الآلية للغة العربية 11
    - ثانيا: تحديد مفهوم اللبس الدلالي 12
    - ثالثا: اللبس الدلالي بين العقل البشري والحاسوب 16
  - المبحث الرابع: التمثيل الحاسوبي للمعلومات اللغوية 19
    - المنهج الإحصائي **Statistical Based** 20
    - المنهج القاعدي **Rule Based** 22
  - المبحث الخامس: منهج العلاقات الدلالية في فك اللبس الدلالي آليا 24

- أولا: العلاقات الدلالية والتعبير عن االمحتوى الدلالي للكلمة 24
- ثانيا: العلاقات الدلالية ومنهج تمثيل المعرفة آليا 25
- ثالثا: واقع العلاقات الدلالية 26
  - العلاقات الدلالية في التراث العربي 27
  - العلاقات الدلالية في البحوث التطبيقية المعاصرة 31

- **الفصل الثاني: العلاقات الدلالية: أنواعها وخصائصها** 38-88
  - المبحث الأول: كيفية رصد العلاقات الدلالية 39
    - الإجراء الأول: الاستبانة اللغوية 39
    - الإجراء الثاني: تحليل عينة لغوية نصية 44
  - المبحث الثاني: الأنواع الدلالية : صورها وخصائصها 47
  - المبحث الثالث: أنواع العلاقات الدلالية 52
- **الفصل الثالث: التحول الدلالي التركيبي** 89-135
  - المبحث الأول: تحول الثنائيات الدلالية إلى ثنائيات تركيبية 90
  - المبحث الثاني: تحليل الثنائيات التركيبية إلى ثنائيات دلالية 116
- **الفصل الرابع: نظام مقترح لفك اللبس الدلالي آليا** 136-207
  - المبحث الأول: تفاعل النظام الدلالي والنظامين الصرفي والتركيبي 137
  - المبحث الثاني: من خصائص التحول الدلالي التركيبي 141
  - المبحث الثالث: تصميم قواعد بيانات النظام 146

- المبحث الرابع: نظام الفهم الآلي للجملة العربية 153

- **الخاتمة** 209

- **الملاحق**

  - **الملحق الأول:** أنواع العلاقات الدلالية 211

  - **الملحق الثاني:** الثنائيات الدلالية للاستبانة اللغوية 232

  - **الملحق الثالث:** نموذج الاستبانة اللغوية 333

  - **الملحق الرابع:** قاعدة البيانات الصرفية 337

- **قائمة المصادر والمراجع** 343

- **الملخص باللغة العربية** 349

- **الملخص باللغة الإنجليزية** 350

المقدمة

علم اللغة الحاسوبي Computational Linguistics هو أحد العلوم اللغوية البينية، التي تهتم بدراسة الظاهرة اللغوية في ضوء مناهج علم اللغة ومناهج العلوم الأخرى ذات الصلة، كعلم الاجتماع، وعلم النفس، وعلم الفيزياء، وعلم الرياضة؛ بهدف إفادة كل علم من نتائج العلم الآخر في تفسير ظواهره وتحقيق أهدافه،[1] ويهدف علم اللغة الحاسوبي إلى حوسبة الملكة اللغوية التي يمتلكها ابن اللغة، وصياغة نماذج حاسوبية لها.[2] ويعتمد في تحقيق ذلك على الإفادة من تفسير مناهج علم اللغة للظاهرة اللغوية، وتقديمه على نحو يتوافق ومنهج الحاسوب الرياضي.

وعلم اللغة الحاسوبي يتألف من جانبين، أولهما اللغة، وهي ظاهرة إنسانية، وليدة العقل البشري، ينضبط نظامها بانضباطه، ويخضع لمنطقه ومبادئه، والعقل محكوم بمجموعة من التراكمات المنطقية والمعرفية التي اكتسبها الإنسان منذ طفولته المبكرة، كما أنه يعمل في ضوء مبدأ توفير الجهد؛ حيث يميل إلى اختزال المعلومات الواردة إليه، واستنباط معلومات جديدة لم ترد إليه من قبل، ويتميز باطراد قواعده وانتظامها، وقد تأثر النظام اللغوي بهذه الخصائص، فجاءت اللغة مفيدة، مختزلة، مطردة بين أبناء الجماعة اللغوية التي تتكلمها. فإذا ما أراد ابن اللغة أن ينشئ تواصلا لغويا ناجحا مع غيره، فعليه أن يلتزم بما استقر في العقل الجمعي لأبناء لغته؛ أي بما تعارفت عليه الجماعة اللغوية التي يعد واحدا منها.[3]

وفي الجانب الآخر من علم اللغة الحاسوبي نجد الحاسوب وقد اعتمد عددا من المناهج الرياضية التي تكفل له الاختزال والاطراد في تعامله مع اللغات الطبيعية، غير أنه يفتقد جميع التراكمات المعرفية والمنطقية التي اختزنها العقل البشري، وتعبر عنها اللغة وتنضبط بها، وليس من سبيل أمام الحاسوب لكي يحاكي فهم العقل البشري للغة

---

[1] -يتعاون علم اللغة مع غيره من العلوم في دراسة الظاهرة اللغوية أو جانب منها، فعلم النفس يدرس اللغة لأن لها علاقة بالتفكير واكتساب اللغة، واضطرابات استخدامها، وعلم الاجتماع يدرس اللغة لأنها شكل من أشكال السلوك، وعلم الطبيعة يدرس الجانب الفزيائي من الكلام الإنساني. للتفصيل انظر: مدخل إلى علم اللغة، د. محمد حسن عبد العزيز، ص 78، 79.

[2] -العربية نحو توصيف جديد، د. نهاد الموسى، ص 34.

[3] - يتمثل العقل الجمعي اللغوي في موقف أبناء اللغة من لغتهم، وسلوكهم اللغوي في المواقف الاتصالية المختلفة مستخدمين الرموز اللغوية للتعبير عما يريدون. لمزيد من التفصيل: انظر اللغة العربية بين المعيارية والوصفية، د. تمام حسان، ص 108.

إلا بتزوده بتلك المعارف والقواعد، وهنا يأتي دور علم اللغة الحاسوبي الذي يهتم بتفسير الظواهر اللغوية، وعرضها على نحو يوافق المنهج الرياضي الذي يعتمده الحاسوب.

وقد اعتمد علم اللغة الحاسوبي في توصيف اللغة حاسوبيا منهجين أساسيين أحدهما هو المنهج الإحصائي **Statistical Based** الذي يمكّن الحاسوب من استنتاج عدد من النماذج اللغوية Language Models من خلال تدريبه على عينة لغوية محللة بمعرفة متخصصين في اللغة، وإسقاطها على نماذج لغوية أخرى لم يتدرب عليها، وقد لقي هذا المنهج قبولا في الدراسات اللغوية الحاسوبية ولاسيما عند الحاسوبيين لاعتماده على الجانب الرياضي أكثر من اعتماده على الجانب اللغوي، أما الآخر فهو المنهج القاعدي **Rule Based** الذي يعتمد على توصيف قواعد النظام اللغوي وتمثيلها على نحو يعبر عن خصائص اللغة ومنطقها، ويوافق منهج الحاسوب، وتنتمي هذه الدراسة إلى هذا المنهج القاعدي.

وقد تنامى في الآونة الأخيرة الاهتمام بالمعالجة الآلية للغة العربية على مستوياتها المختلفة الصوتية والصرفية والدلالية والتركيبية، وقد حققت التطبيقات الحاسوبية الصوتية والصرفية معدلات دقة متقدمة، أما التطبيقات الدلالية والتركيبية فلا تزال في حاجة إلى كثير من الجهود، نظرا لتعقد النظام اللغوي على المستوى الدلالي والتركيبي، وصعوبة تمثيلهما حاسوبيا، ومن أولى الصعاب التي تبطئ تطوير التطبيقات الحاسوبية الدلالية والتركيبية اللبسُ وتعدد المقصود من الوحدة اللغوية في الجملة العربية، وقد اختصت هذه الدراسة بتناول ظاهرة اللبس الذي يجابه البحث الحاسوبي الدلالي للغة العربية المعاصرة المكتوبة.

يعبر النظام اللغوي للغة العربية عن المعاني غير المتناهية بعدد متناه من الوحدات اللغوية، وذلك عن طريق تحميل الوحدة اللغوية الواحدة أكثر من دلالة سواء كانت وظيفية أو معجمية، مع توفير القرائن الكافية التي تتيح للنظام التركيبي تخصيص الكلمة بإحدى هذه الدلالات،[4] فإذا مثّل النظام التركيبي هذه القرائن جاءت الجملة خالية من اللبس، وإذا قصّر في ذلك بقصد من منشئ الجملة أو بدون قصد وقع اللبس فيها بمقدار ذلك التقصير. أما الحاسوب فجميع كلمات الجملة بالنسبة له لبسية؛ لأنه يفتقد إلى

---

[4] – مقالات في اللغة والأدب، د. تمام حسان، ص 292

خصائصها البنيوية ودلالاتها الوظيفية والمعجمية، والقرائن التي تخصص المناسب منها في الجملة حال التعدد، ومن ثم يقتضي الفهم الآلي للغة العربية توصيف كل ذلك على نحو يفهمه الحاسوب، وبقدر اكتمال ذلك التوصيف وانتظامه يقترب الحاسوب من العقل البشري في فهم دلالة الجملة وفك لبسها الدلالي، وقد اقترح البحث منهجا لذلك التوصيف يعتمد على العلاقات الدلالية، وذلك لقدرتها على التعبير عن دلالات الكلمات، وتمثيلها في صورة شبكية يفهمها الحاسوب، واطراد تحولها إلى العلاقات التركيبية التي تنبني منها الجملة.

وقد جاءت الدراسة بالعنوان التالي: "**اللبس الدلالي في المعالجة الآلية للغة العربية المعاصرة المكتوبة**" وهو بذلك قد حدد موضوع الدراسة بالأطر التالية:

* اللغة العربية المعاصرة، وهي "لغة فصحى، مكتوبة، تستخدم في التعليم وفي العلم والأدب وفي الصحافة، وهي اللغة الرسمية المشتركة في العالم العربي اليوم"[5]، "وهي اللغة التي تعاصرنا زمنيا"[6]

* المستوى المكتوب من اللغة، دون المنطوق وذلك لتباين أدوات المعالجة الحاسوبية اللازمة لهما.

* المستوى الدلالي للغة، وعدم التعرض للمستويات اللغوية الأخرى إلا بالقدر الذي يكشف عن تداخلها فيما بينها، ويقتضيه إثبات صحة الفرضيات المقترحة، وتحليل الأمثلة التطبيقية.

أهمية الدراسة وأهدافها:

تكتسب الدراسة أهميتها من أنها:

- دراسة بينية تقرب بين مبادئ علم اللغة بوصفه أحد العلوم الإنسانية، ومبادئ علم الرياضة التي يعتمدها الحاسوب، وتوضح دور كل من اللغويين والحاسوبيين في النهوض بمهمة حوسبة اللغة العربية.

---

5 –الوضع اللغوي في الفصحى المعاصرة، د. محمد حسن عبد العزيز، ص 11.

6 –الغموض في الدلالة، أنماطه وعوامله ووسائل التخلص منه في العربية المعاصرة، د. محمد أحمد حمـــاد، ص 3.

- تهتم بالمستوى الدلالي للغة المسئول عن إكساب الجملة القيم المعرفية والمنطقية اللازمة لتمام عملية الفهم والإفهام.

- تحاول تفسير ظاهرة اللبس الدلالي، والإجابة عن إشكالياته، التي تعد من أهم معوقات تقدم الفهم الآلي للجملة العربية، وما يبنى عليه من تقنيات تخدم اللغة والثقافة العربية؛ كالترجمة الآلية، واسترجاع المعلومات، ومحركات البحث الدلالي، وغيرها من التقنيات التي تخدم النص العربي.

وفي ضوء ما سبق تحاول الدراسة تحقيق عدد من الأهداف، منها:

- التفريق بين الفهم البشري للغة العربية المكتوبة والفهم الحاسوبي لها.
- تحديد متطلبات الفهم الحاسوبي للجملة العربية المكتوبة.
- اقتراح منهج يقرب النظام الدلالي إلى المنهج الحاسوبي.
- اقتراح خطوات بناء نظام حاسوبي لفك الالتباس الدلالي للجملة العربية المكتوبة.
- التأكيد على أهمية المنهج القاعدي **Rule Based** في المعالجة الآلية للغة العربية.

صعوبات الدراسة:

فرض موضوع الدراسة عددا من الصعاب على الباحث، اقتضت اقتراح أكثر من فرضية، واتخاذ أكثر من إجراء للتأكد من صحتها، وبعض هذه الصعاب يعود إلى ما يختص به النظام الدلالي للغة العربية، وبعضها يعود إلى اختلاف منهج علم اللغة عن المنهج الحاسوبي في تحليل الظواهر اللغوية، والبعض الآخر يعود إلى قلة المصادر التي وازنت بين خصائص المنهج اللغوي والحاسوبي في تناول الظواهر اللغوية عموما والدلالية خصوصا، ومن هذه الصعاب:

أولا: تداخل النظام الدلالي مع غيره من الأنظمة اللغوية مؤثرا فيها ومتأثرا بها، ومتعاونا معها في بناء جملة مفيدة، وهو ما فرض على البحث ضرورة الموازنة بين تناول القضايا اللغوية العامة ذات الصلة بالنظام الدلالي، والاقتصار على القضايا الدلالية الأساسية بما يتسع له جهد الباحث.

ثانيا: نسبية إدراك أبناء اللغة لدلالة الكلمات والجمل لتأثرها بالعوامل النفسية والاجتماعية والثقافية التي يتغايرون فيها، وتسلل هذه النسبية إلى تحليل اللغويين

للظواهر الدلالية المقدم للحاسوب سواء كان هذا التحليل قواعد ومبادئ لغوية أو مدونة محللة يدويا.

ثالثا: اعتماد الحاسوب في التعامل مع المعلومات اللغوية عددا من المبادئ التي تنسجم وطابعه الرياضي كالاطراد المقابل للنسبية والاحتمالية، والشمولية المقابلة للشذوذ، والمنطقية المقابلة للعشوائية، في حين يصعب على علم اللغة بوصفه أحد العلوم الإنسانية التقيد بذلك؛ لأنه يخاطب بتفسيراته للظواهر اللغوية القدرات الإدراكية للعقل البشري التي تَجبُر ما قد يقع من نسبية أو عشوائية، وما كان لعلم اللغة أن يغفل تلك القدرات، وإلا وقع في ذكر مسلمات لغوية لا فائدة منها عند أبناء اللغة.

رابعا: قلة المصادر اللغوية الحاسوبية التي تهتم بدراسة المستوى الدلالي للغة العربية، واتصاف بعضها بواحد أو أكثر مما يلي:

1-التعبير عن تجارب عملية لأحد المشروعات التي استهدفت التحليل اللغوي الحاسوبي للمستوى الدلالي[7]، أو نقل لخبرات إحدى المؤسسات المهتمة بالمعالجة الآلية للغة العربية،[8] ويعد هذا الصنف من الدراسات اللغوية الحاسوبية الأكثر عمقا في تناول القضايا اللغوية عموما والدلالية خصوصا، والأقرب إلى الواقع بما فيه من تحديات وصعاب، وإن قلل من دوره في دعم الجانب النظري من البحث اللغوي الحاسوبي أنها لا تفصح عن تفاصيل كثيرة بداعي الحفاظ على سرية المشروعات أو المؤسسات.

2-بحوث ودراسات جزئية لا تهتم بربط الظاهرة الدلالية بامتداداتها على المستويات اللغوية الأخرى ولاسيما الصرفية والتركيبية، فابتعد بها ذلك عن حل الإشكاليات الواقعية التي تواجه حوسبة اللغة العربية.

3-عدم التفريق الواضح بين متطلبات العقل البشري والحاسوب في فهم دلالة الجملة العربية وفك لبسها، ومن ذلك رسالة الدكتوراه "العلاقات التركيبية في الجملة الفعلية القرآنية. دراسة نحوية حاسوبية" للدكتور مدحت يوسف السبع، فعلى الرغم مما لهذه الرسالة من فضل الأسبقية، فهي من أوائل أطروحات الدكتوراه التي تناولت موضوع حوسبة اللغة العربية، وما لها أيضًا من جودة تناول في جانبها اللغوي فهي لم تراع خصائص الحاسوب في تعامله مع اللغة، ويتبين ذلك من توضيحه لكيفية فهم الحاسوب

---

[7] -من أمثلتها البحوث التي نشرت حول مشروع التنقيب المعلوماتي في النصوص Text Mining، وكان الباحث أحد أعضاء الفريق العامل فيه، انظر الرابط التالي: -http://www.rdi-eg.com/technologies/arabic_nlp.htm

[8] -من أبرز مؤلفات هذا الصنف كتاب "اللغة العربية والحاسوب"، للدكتور نبيل علي، دار تعريب، 1988م.