BIBLIOTHECA ALEXANDRINA
مكتبة الإسكندرية

19 ESOLEC

CERTIFICATE OF APPRECIATION

This certificate is presented to

Prof. Wafaa Kamel

For her presentation as keynote speaker in

The 19th Language Engineering Conference "ESOLEC'19"

September 26-29, 2020

Bibliotheca Alexandrina, Alexandria, Egypt

Conference co-chair

Prof. Sameh Alansary

الجمعية المصرية لهندسة اللغة

المعجم ونظم المعالجة الآلية

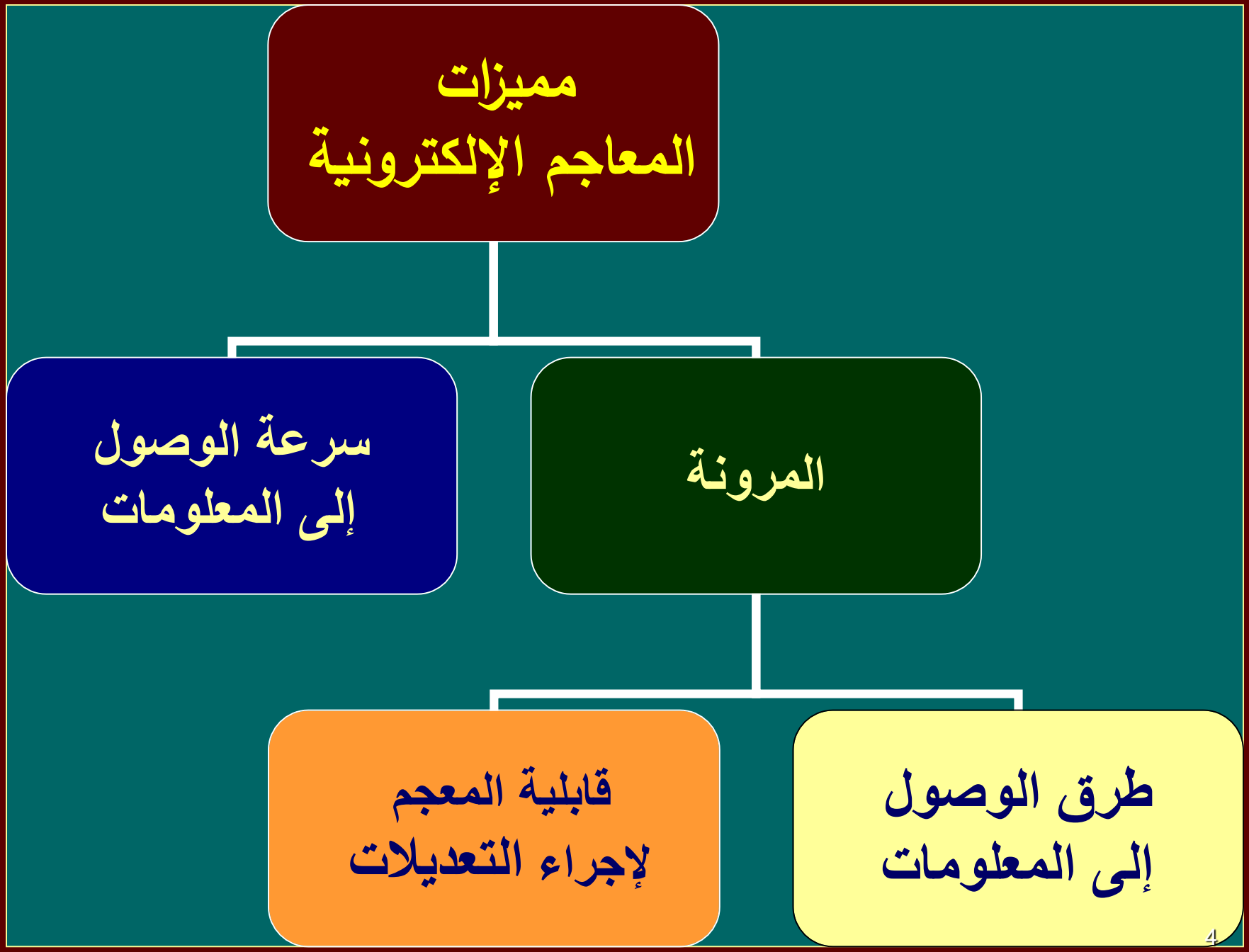
د. وفاء كامل فايد
أستاذة اللسانيات بآداب القاهرة
وعضو مجمعي اللغة بالقاهرة ودمشق

تمهيد: علاقة المعجم بنظم المعالجة الآلية

- المعجم ركيزة أساسية يعتمد عليها بناء نظم المعالجة الآلية للغات الطبيعية N L P
- الأداء النهائي لنظم المعالجة الآلية يعتمد على كفاءة المعجم المرتبط بها.
- تؤدي المعلومات الصرفية، والنحوية، والدلالية - التي يُزَوِّد المعجمُ النظامَ بها- دورًا مهمًّا في أداء نظام المعالجة.
- قد تتنوع مستويات التفصيل، ونوعية المعلومات في المعجم باختلاف نظام المعالجة.

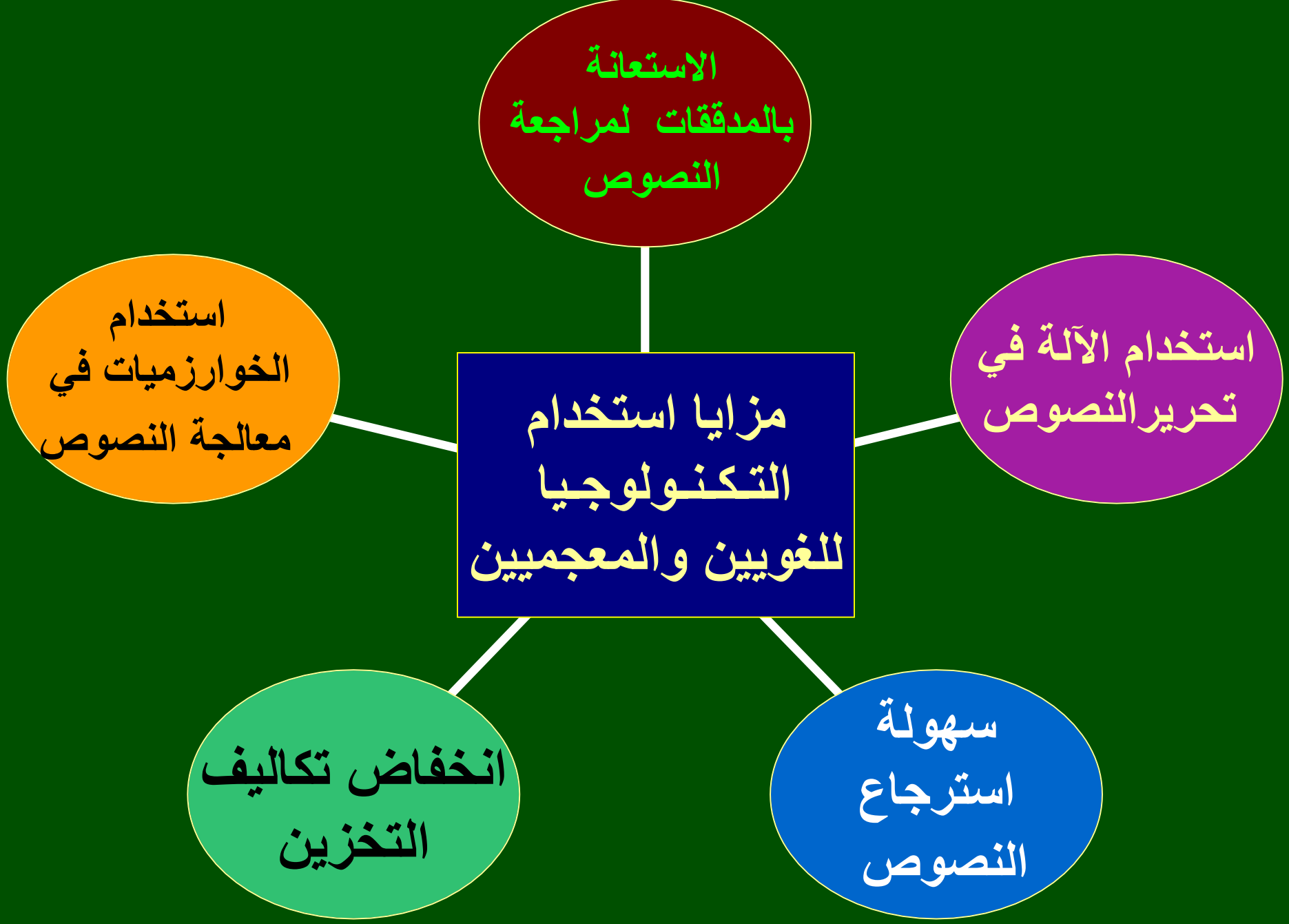
المعجم العربي والحاسب الآلي

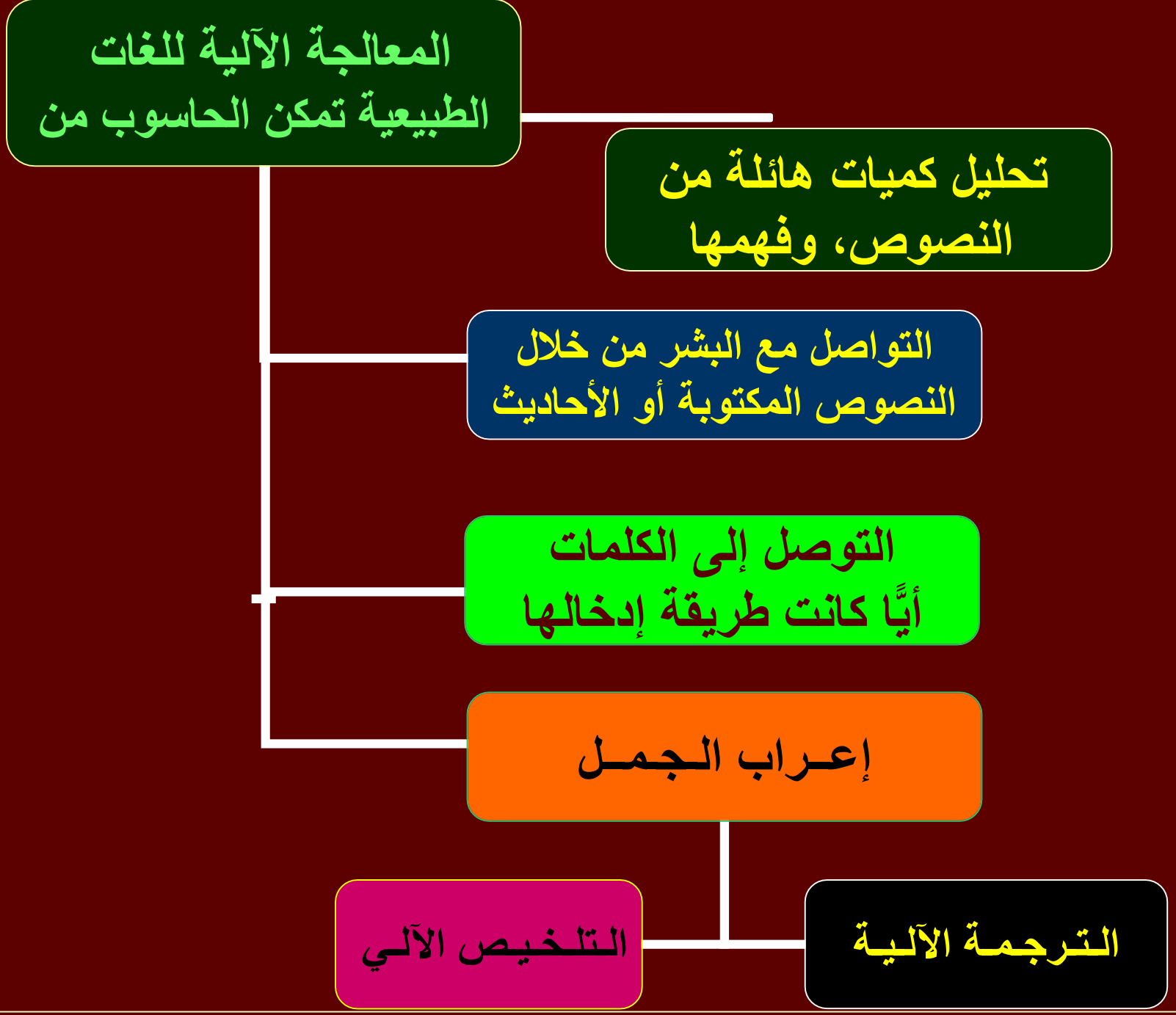
- عدت الإفادة من الحاسب الآلي في صناعة المعاجم من أهم مجالات علم اللغة الحاسوبي.
- أحدث اختراع الحاسب الآلي نقلة نوعية في الدراسات العلمية في مختلف المجالات ، ومنها المعجمية.
- بدأ دخول المعجم العربي عصر الحاسوب عام 1971م بدراسة إحصائية لجذور اللغة العربية في معجم الصحاح، ثم في لسان العرب لابن منظور وتاج العروس للزبيدي 1973. وقد أجراها أستاذنا الراحل علي حلمي موسى.
- تعاني المعجمية الحاسوبية العربية مشكلات جوهرية؛ بسبب اعتماد معظم المصادر المعجمية المحوسبة على المعاجم التقليدية.
- ارتبط الاتجاه نحو حوسبة المعجم العربي بالمميزات التي تقدمها التكنولوجيا للمتعاملين مع المعاجم الحاسوبية.


مميزات
المعاجم الإلكترونية
سرعة الوصول
إلى المعلومات
المرونة
قابلية المعجم
لإجراء التعديلات
طرق الوصول
إلى المعلومات

الحاسوب والتعامل مع الظواهر الدلالية في العربية

- لم يحقق الحاسوب نجاحا كبيرا في الفهم الآلي للغة العربية وتحليلها وفي ترجمتها الآلية.
- من أهم أسباب انخفاض نسبة نجاح الحاسوب في هذه المجالات اشتمال العربية على عدد من الظواهر اللغوية التي تتعلق بالدلالة.
- من هذه الظواهر التجمعات اللفظية، والتعابير الاصطلاحية والمجاز.
- من الضروري أن تتضافر جهود اللغويين والحاسوبيين ومهندسي اللغة في العمل على الفصل بين أنواع التجمعات اللفظية، وتقديم المعلومات التي تتيح للحاسوب التمييز بينها .
- بذلك يمكننا أن نحقق نسبة كبيرة من النجاح في فهم اللغة العربية وترجمتها آليا.


مزايا استخدام التكنولوجيا للغويين والمعجميين
الاستعانة بالمدققات لمراجعة النصوص
استخدام الآلة في تحرير النصوص
استخدام الخوارزميات في معالجة النصوص
سهولة استرجاع النصوص
انخفاض تكاليف التخزين


المعالجة الآلية للغات الطبيعية تمكن الحاسوب من
تحليل كميات هائلة من النصوص، وفهمها
التواصل مع البشر من خلال النصوص المكتوبة أو الأحاديث
التوصل إلى الكلمات أيًّا كانت طريقة إدخالها
إعـراب الـجمـل
الـتـرجـمـة الآلـيـة
الـتـلـخـيـص الآلـي

العربية المعاصرة وتحديات المعالجة الآلية

- يكشف الواقع اللغوي للعربية المعاصرة عن تنوع في مستويات الأداء عند متحدثيها ، وهو ما يعرف بالازدواج اللغوي Diglossia .
- يقصد به وجود مستويين لغويين عند أبناء الجماعة اللغوية الواحدة يمثِّل أحدهما اللغة العليا ، ويمثل الآخر اللغة الدنيا ، ولكل منهما وظيفة اجتماعية ، واستعمال في مواقف خاصة.
- تواجه العربية المعاصرة صعوبات وتحديات تتعلق بالمعالجة الآلية على المستويين النظري والحاسوبي.
- ترتبط هذه التحديات بمستويات اللغة العربية المختلفة، بدءًا بالكتابة ومرورًا بالصرف والنحو، وانتهاءً بالدلالة.

اللغة العربية الفصيحة وصعوبات المعالجة الآلية

- تبدأ صعوبات المعالجة الآلية للغة العربية بمستوى الكتابة ، التي تتميز بسمات تختلف عن سمات الكتابة الأجنبية ، أبرزها ما يلي:
- تختلف أشكال الحروف اختلافًا تامًّا عن أشكال الحروف في اللغات الأخرى.
- تحتوي العربية على حوالي 60 رمزًا مستخدَمًا في الكتابة.
- تتوزع هذه الرموز بين حروف ، وعلامات ضبط بالشكل ، وعلامات ترقيم ، وأرقام .

تابع: اللغة العربية الفصيحة وصعوبات المعالجة الآلية

- تأخذ الكتابة العربية هيئة متشابكة ، لذلك قد يختلف شكل الحرف بناءً على كيفية اتصاله بالحروف المجاورة . وهو ما يجعل لشكل الحرف الواحد أربعة احتمالات : الحرف مستقلًا، والحرف في أول الكلمة ، وفي وسطها ، وفي آخرها.
- تُكْتَب اللغة العربية من اليمين إلى اليسار، خلافا لطريقة الكتابة اللاتينية .
- لا تَستخدم النصوص العربية علامات التشكيل بصورة منتظمة، بل إنها تغيب في أغلب النصوص العربية الحديثة، مما قد يؤدي إلى إحداث قدر من الغموض فيما يتعلق بمعنى الكلمة، أو وظيفتها النحوية.
- تخلو اللغة العربية من الحروف الكبيرة capital letters التي تؤدي وظائف مهمة في بعض تطبيقات المعالجة الآلية ، كتَعرُّف الأسماء آليًّا.

أنواع المعاجم الآلية وأهميتها في الدراسات اللغوية الحاسوبية

- **المعاجم الإلكترونية** electronic dictionaries:

معاجم آلية مصممة للاستخدام البشري.

الهدف منها تزويد المستخدم البشري بالمعلومات المعجمية، بحسب نوع المعجم.

- **المعاجم الحاسوبية** computational lexicons

معاجم مصممة لاستخدامها في نظم معالجة اللغات الطبيعية.

الهدف من بنائها توصيل المعلومات إلى النظم التي تقوم بتحليل اللغة الطبيعية ـ نصوصا أو كلاما ـ وإنتاجها.

موازنة بين نوعي المعاجم الحاسوبية

- على الرغم من اشتمال المعاجم الإلكترونية على قدر من المعلومات القيمة فإن هذه المعلومات لا تصلح للمعاجم الحاسوبية.
- تحتاج المعاجم الحاسوبية إلى معلومات تفصيلية على المستويات الصرفية، والنحوية، والدلالية؛ والسياقية.

لذا يلزم أن يكون المعجم الحاسوبي أكثر شموليةً وتنظيمًا وتصريحًا وتماسكًا ، وشكليةً ، من معاجم الاستخدام البشري.

- اعتبرت النظرةُ التقليدية المعاجمَ الحاسوبية مستودعاتٍ للمعلومات المعجمية لمهام محددة، كالإعراب والتوليد، والترجمة.

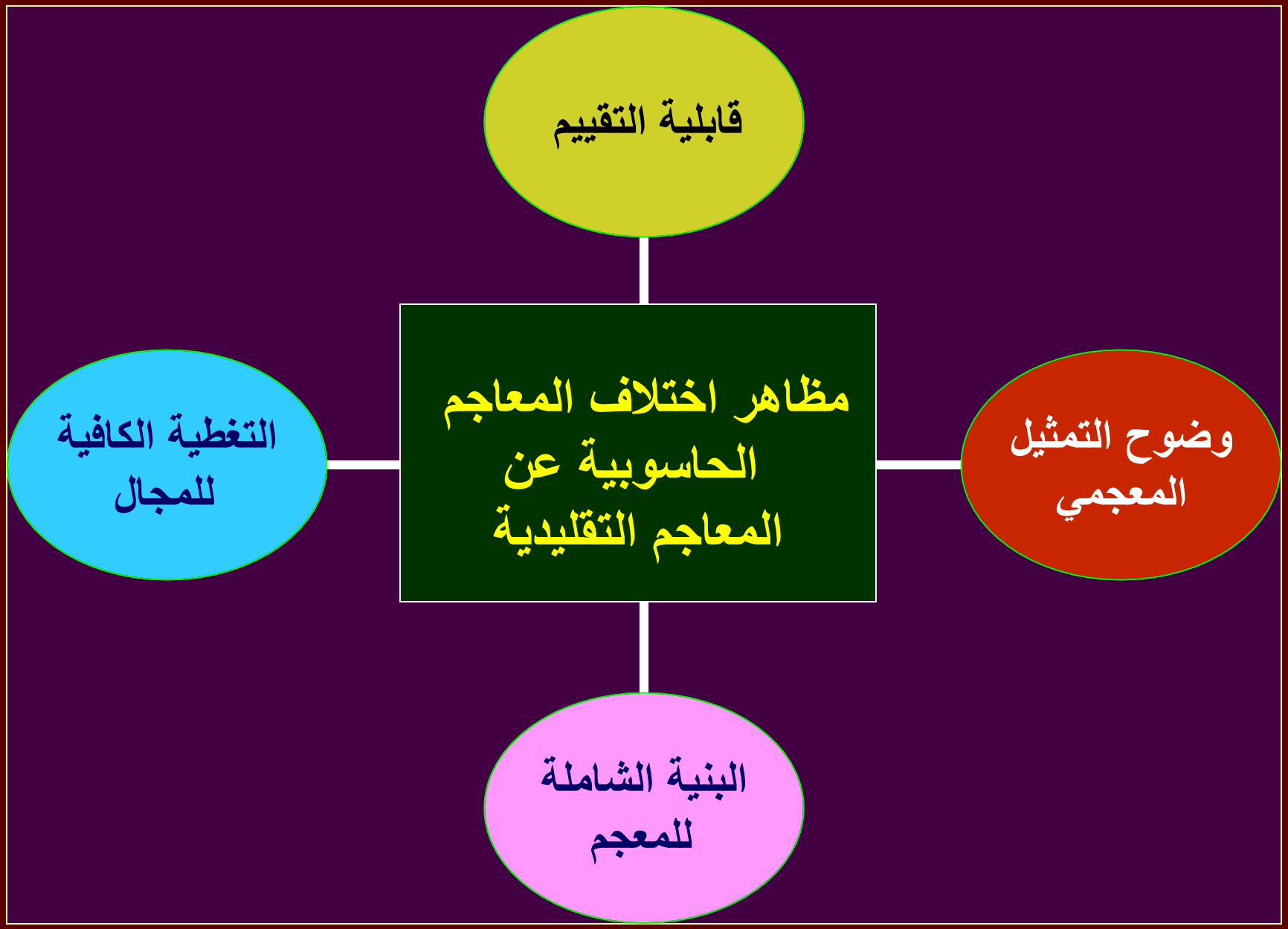
دور المعجم في نظم المعالجة الآلية

- تحتاج نظم المعالجة الآلية للغات الطبيعية إلى أن تتزود بمعجم يغطي الاستخدام الفعلي لمفردات اللغة ، ويمدها بالمعلومات اللغوية المتعلقة بها. لذا فتصميم المعجم وتطويره من القضايا الرئيسية في المعالجة الآلية للغات الطبيعية .

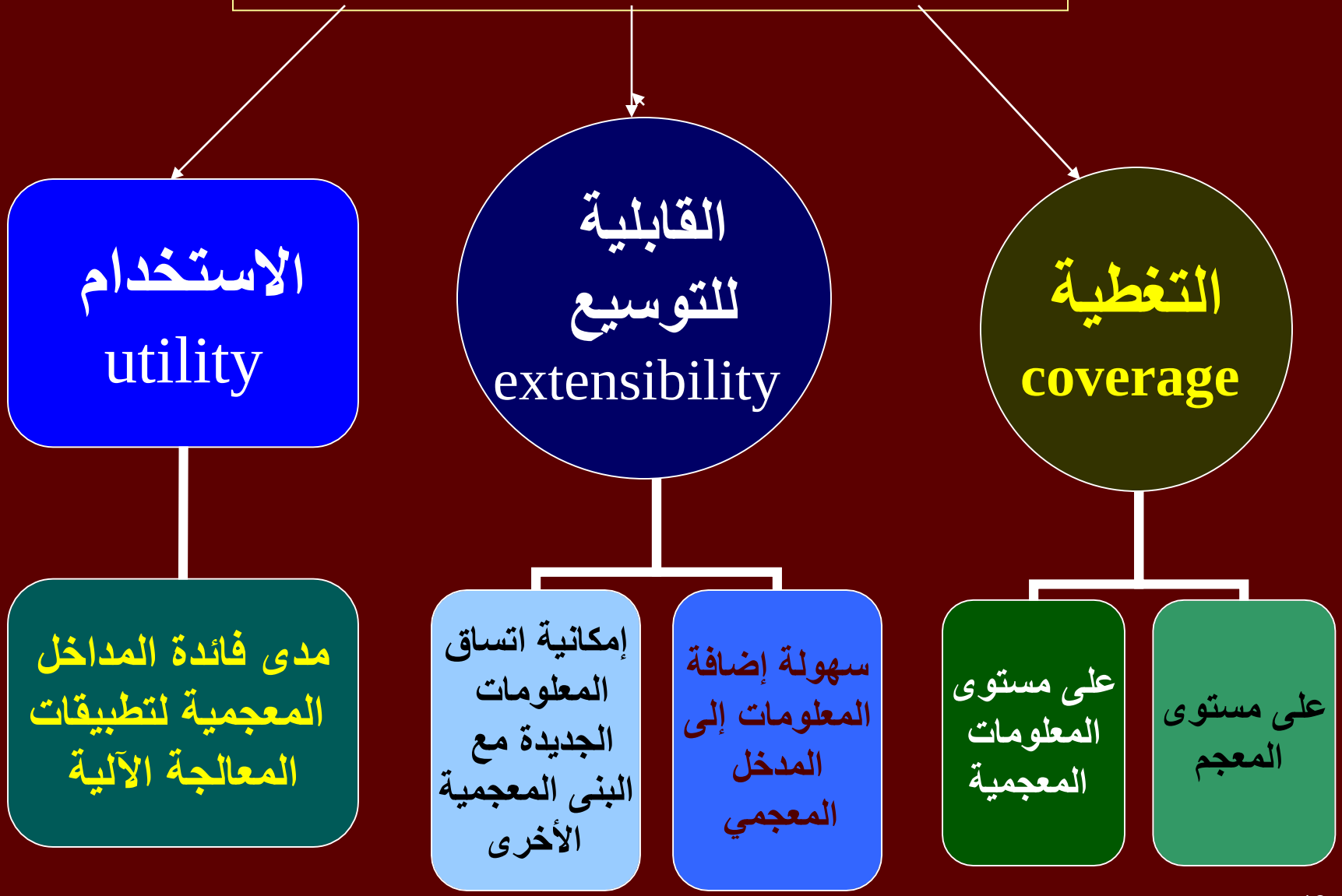
- ويعتمد الأداء النهائي لنظم المعالجة الآلية على كفاءة المعجم المرتبط بها ؛ لأن المعلومات الصرفية، والنحوية ، والدلالية ، والسياقية - التي يزوِّد المعجمُ النظامَ بها - تؤدي دورًا مهمًّا في هذا الأداء .

تابع : دور المعجم في نظم المعالجة الآلية

- إذا قارنا بين بناء معجم يقتصر استخدامه على التصنيف النحوي ، وآخر يستخدم في الإعراب الآلي أو الترجمة الآلية نجد أن معجم المصنف النحوي يجب أن يكون نظامه قادرًا على تعيين كل الصيغ الممكنة للكلمة في اللغة ،
- في حين تحتاج بعض نظم الفهم الآلي للغات الطبيعية كالترجمة واسترجاع المعلومات إلى معلومات تتعلق بمعاني الكلمات؛ وعلى هذا يحتاج المعجم المرتبط بهذه النظم إلى أن يعكس العلاقات الدلالية كالترادف والتضاد والعلاقات الهرمية بين الكلمات ؛ كي تقوم هذه النظم بتحليل (معنى) الجملة.


قابلية التقييم
التغطية الكافية للمجال
مظاهر اختلاف المعاجم الحاسوبية عن المعاجم التقليدية
وضوح التمثيل المعجمي
البنية الشاملة للمعجم

أسس تقَـيـيـم المعاجم الحاسوبية



خاتمة

- من أهم أسباب انخفاض نسبة نجاح الفهم الآلي للغة العربية وتحليلها، وترجمتها الآلية، اشتمالها على عدد من الظواهر اللغوية التي تتعلق بالدلالة، ومن هذه الظواهر المجاز، والتجمعات اللفظية، والتعابير الاصطلاحية.
- من الضروري أن تتضافر جهود اللغويين ومهندسي اللغة والحاسوبيين، في العمل على الفصل بين أنواع التجمعات اللفظية، وتقديم المعلومات التي تتيح للحاسوب التمييز بينها ؛ حتى يمكن أن نحقق نسبة كبيرة من النجاح في فهم اللغة العربية وترجمتها آليا.

شكرا لحسن متابعتكم

المعجم ونظم المعالجة الآلية

أ.د. وفاء كامل فايد

يمثل المعجم ركيزة أساسية يعتمد عليها بناء النظم المختلفة للمعالجة الآلية للغات الطبيعية؛ فالأداء النهائي لنظم المعالجة الآلية يعتمد على كفاءة المعجم المرتبط بها؛ إذ تؤدي المعلومات الصرفية، والنحوية، والدلالية– التي يُزَوِّد المعجمُ النظامَ بها– دورًا مهمًّا في هذا الأداء، ومع ذلك قد تتنوع مستويات التفصيل، ونوعية المعلومات التي يحتوي عليها المعجم باختلاف نظام المعالجة.

في العصر الحديث، أحدث اختراع الحاسب الآلي نقلة نوعية في الأبحاث والدراسات العلمية في مختلف المجالات والتخصصات. ولم تكن الدراسات اللغوية– ومنها المعجمية– بمنأى عن الإفادة من قدراته الهائلة؛ إذ عُدَّت الإفادة من الحاسب الآلي في صناعة المعجم من أهم مجالات علم اللغة الحاسوبي، وأكثرها تلبية للمتطلبات العلمية والثقافية في العالم المعاصر.[1]

أما المعجم العربي، فقد بدأ دخول عصر الحاسوب في سبعينيات القرن الماضي ليبدأ مرحلة جديدة بتحديات جديدة، وآمال عريضة في مواكبة التطور الكبير في الدراسات المعجمية في أوروبا وأمريكا. وقد بدأ دخوله عصر الحاسوب عام 1971 بدراسة إحصائية لجذور اللغة العربية المدونة بمعجم الصحاح، توضح تردد حروف كل جذر في المواقع المختلفة للكلمة، مع إحصائيات تحدد تكوين الثنائيات في تلك الجذور. وهي دراسة أجريت بجامعة الكويت، وقام بها علي حلمي موسى– أستاذ الفيزياء بجامعة الكويت آنذاك– إذ اسْتخدم الحاسب الآلي لإحصاء الجذور الثلاثية[2] وغير الثلاثية لمعجم الصحاح.[3] وانتهت هذه الدراسة إلى "عرض جذور مفردات اللغة العربية المستخدَمة والواردة في المعجم في صورة جداول توضح تردد الحروف في المواقع المختلفة من الكلمة، وتتابع تلك الحروف، واختلافها من حرف لآخر" [4]. وقد تبع ذلك دراسات أخرى لم تخرج عن هذا الإطار الإحصائي، منها دراسة إحصائية لجذور معجم لسان العرب لابن منظور عام 1973م، ودراسة أخرى لإحصاء جذور معجم تاج العروس للزبيدي عام 1973م.

وارتبط الاتجاه نحو حوسبة المعجم العربي– ارتباطًا واضحًا– بمجموعة المميزات التي تقدمها التكنولوجيا لكل الأطراف ذات الصلة بالمعاجم مقارنة بنظائرها الورقية.

---

(1) حجازي، محمود فهمي. (د.ت). البحث اللغوي. القاهرة: دار غريب للطباعة والنشر. ص 71.

(2) صدرت الدراسة الخاصة بالجذور الثلاثية في مايو 1971: موسى، علي حلمي. (1973): دراسة إحصائية لجذور معجم الصحاح باستخدام الكمبيوتر. مطبوعات جامعة الكويت رقم 33- ص5.

(3) صدرت الدراسة الخاصة بالجذور غير الثلاثية في يناير 1972 : المرجع السابق ، والصفحة نفسها.

(4) موسى، علي حلمي. (1973): دراسة إحصائية لجذور معجم الصحاح باستخدام الكمبيوتر. المرجع السابق- ص 10.

مميزات المعاجم الإلكترونية:

إذا أردنا المقارنة بين المعاجم الورقية والإلكترونية وجدنا أن **المرونة** تُعَدّ أولى مميزات المعاجم الإلكترونية التي تتعلق بمستخدم المعجم؛ إذ تتضح أهمية هذه الميزة من جهتين: أولاهما تتعلق بطرق الوصول إلى المعلومات داخل المعجم. ففي الوقت الذي تسمح فيه المعاجم الورقية بطريقة بحث أحادية – وفقًا للترتيب الأبجدي في أغلب الأحيان– نجد المستخدم يمتلك طرقًا بحثية متعددة للوصول إلى المعلومات في حالة استخدام المعجم الإلكتروني.

أما الجهة الثانية فترتبط بقابلية المعجم لإجراء التعديلات التي تتماشى مع التطور اللغوي؛ فالمعاجم الورقية تتسم– في هذا الصدد– بنوع من الجمود يجعلها تتوقف عند لحظة النشر، في حين تتيح المعاجم الإلكترونية إجراء التعديلات التي تتماشى مع التطورات اللغوية، على نحو يسمح بإضافة المفردات الجديدة، وتصحيح الأخطاء بشكل دائم ومتجدد.[5]

أما الميزة الثانية للمعاجم الإلكترونية فهي **سرعة الوصول إلى المعلومات**؛ إذ تتيح تقنيات البحث الآلي الوصول إلى معلومات المدخل المعجمي بمجرد النقر على كلمة على الشاشة[6].

الإمكانات التي تتيحها التكنولوجيا للمعجمي، منها[7]:

- تحرير النصوص باستخدام الآلة.
- استرجاع النصوص بسهولة وسرعة.
- انخفاض تكاليف التخزين.
- استخدام الخوارزميات في معالجة النصوص.
- الاستفادة من المدققات الإملائية والنحوية في مراجعة النصوص.

ويبدو أن المعجم العربي اتكأ على هذه المميزات، منذ ثمانينيات القرن العشرين ، حين بدأ يشيّد قواعد البيانات المعجمية Lexical databases، والمعاجم الآلية؛ كي تُستخدم في المعالجة الآلية للغة العربية، أو يستخدمها المستعمل البشري. كما تزايد الاهتمام بمعالجة المعجم العربي آليًّا داخل الوطن العربي وخارجه، سواء أكان هذا الاهتمام جامعيًّا أكاديميًّا، أو في شكل مؤسسات تجارية.[8]

---

(5) Sterkenburg, P. (2003). P. 5 -6 .

(6) Ibid. P. 7.

(7) مهديوي، عمر. (2007). المعجم العربي في ضوء اللسانيات الحاسوبية. الندوة الدولية الأولى عن الحاسب واللغة العربية. مدينة الملك عبدالعزيز للعلوم والتقنية. الرياض، 2007.

(8) أشارت وفاء كامل إلى بعض المؤسسات المعنية بالعمل المعجمي الحاسوبي العربي داخل الوطن العربي وخارجه، ومن هذه المؤسسات جامعة عمان، وجامعة محمد الخامس، وجامعة بالماند لبنان، والشركة = الهندسية لتطوير أنظمة الحاسب (RDI)، وشركة صخر. ومن خارج الوطن العربي جامعة كوبنهاجن، وجامعة ليون 2، وجامعة أترخت، وغيرها. راجع:
فايد، وفاء كامل. (2007). واقع المعجمية الحاسوبية العربية. ندوة (العربية في عصر العولمة)، المجلس الأعلى للثقافة، ومجمع اللغة العربية بالقاهرة (في إطار الاحتفال باليوبيل الماسي لمجمع اللغة العربية)، القاهرة ، مارس 2007.

وعلى الرغم من الاهتمام الذي اكتسبته المعالجة الآلية للمعجم العربي، فإن ثمة مشكلات جوهرية تعانيها المعجمية الحاسوبية العربية؛ إذ اعتمدت "معظم المصادر المعجمية المحوسبة على المعاجم التقليدية، التي تفتقر بدورها إلى الأسس النظرية البحثية التي تجعلها مادة كافية لمطالب معالجة اللغات الطبيعية.[9]

مفهوم المعالجة الآلية للغات الطبيعية: Natural Language Processing

يُستخدم مصطلح المعالجة الآلية للغات الطبيعية للإشارة إلى "معالجة المواد النصية في اللغات الطبيعية معالجة حاسوبية؛ بهدف ابتكار تقنيات خاصة بالتحليل الآلي لكميات كبيرة من النصوص المكتوبة أو المنطوقة، بطرق مشابهة لما يقوم به البشر في هذا الصدد"[10].

وقد يكون فهم المصطلح أسهل بالنظر إلى مجموعة التطبيقات والنتائج التي يُتَوَصَّل إليها من خلال هذه المعالجة؛ فالمعالجة الآلية للغات الطبيعية تُمَكِّن الحاسوب من تحليل كميات هائلة من النصوص، وفهمها. كما تُمَكِّنه من التواصل مع البشر من خلال النصوص المكتوبة أو الأحاديث. وتمكنه أيضا من التوصل إلى الكلمات أيًّا كانت طريقة إدخالها: عن طريق لوحة المفاتيح، أو من خلال جهاز تعرُّف الكلام آليًّا speech recognition، كما تُمَكِّنه من الإعراب النحوي، أي إعراب الجمل syntactic parsing ، الذي يؤدي دورا كبيرا في تطبيقات أهمّ كالترجمة الآلية machine translation، والتلخيص الآلي automatic summarization ، وتمكنه أيضا من فهم الكلام، والإجابة عن الأسئلة، وإجراء المناقشات مع البشر[11].

العربية المعاصرة وتحديات المعالجة الآلية:

يكشف الواقع اللغوي للعربية عن تنوع واضح في مستويات الأداء عند متحدثيها. وقد تناول فرجسون ظاهرة التنوع اللغوي في اللغة العربية في العصر الحديث، وذلك في ضوء ما أطلق عليه "الازدواج اللغوي Diglossia"، وقصد به وجود مستويين لغويين عند أبناء الجماعة اللغوية الواحدة يمثِّل أحدهما اللغة العليا High language، ويُرمز له بالرمز (H)، ويمثل الآخر اللغة الدنيا Low language، ويُرمز له بالرمز (L)، ولكل منهما وظيفة اجتماعية.[12]

وتُعدّ الوظيفة– وفقًا لفرجسون– من أبرز سمات الازدواج اللغوي؛ "فهناك مواقف لا تُستخدم

---

(9) السابق نفسه.

(10) Crystal, D. (2008). A Dictionary of Linguistics and Phonetics. Sixth edition. Blackwell Publishing Ltd. P. 348.

(11) Nugues, P. (2006). An Introduction to Language Processing with Perl and Prolog. Springer-Verlag Berlin Heidelberg. P. 2.

(12) Ferguson, C. (1959). Diglossia. Word. Vol. 15. pp. 325 – 340.

فيها سوى اللغة العليا، ومواقف أخرى لا يصلح فيها سوى استخدام اللغة الدنيا"[13].

وقد قسم السعيد بدوي مستويات اللغة العربية المعاصرة في مصر إلى خمسة أقسام هي[14]:

1- فصحى التراث : وهي فصحى تقليدية غير متأثرة بشيء نسبيا.

2- فصحى العصر: وهي فصحى متأثرة بالحضارة المعاصرة بوجه خاص.

3- عامية المثقفين: وهي عامية متأثرة بالفصحى والحضارة المعاصرة معا.

4- عامية المتنورين: وهي عامية متأثرة بالحضارة المعاصرة.

5- عامية الأميين: وهي عامية غير متأثرة بشيء نسبيا.

اللغة العربية الفصيحة وتحديات المعالجة الآلية:

تكشف طبيعة اللغة العربية عن جملةٍ من الصعوبات التي تفرض تحديات كبيرة – فيما يتعلق بمعالجتها آليًّا– على المستويين النظري والحاسوبي. وهذه التحديات ترتبط بمستويات اللغة العربية المختلفة، بدءًا بالكتابة، ومرورًا بالصرف والنحو، وانتهاءً بالدلالة.

تبدأ صعوبات المعالجة الآلية للغة العربية بمستوى الكتابة التي تتميز بعدد من السمات التي تختلف اختلافًا كبيرًا عن سمات الكتابة اللاتينية. وتتمحور هذه الاختلافات في عدد من النقاط، منها:

- أشكال الحروف التي تختلف اختلافًا تامًّا عن أشكال الحروف التي تُكتب بها لغات أخرى كالإنجليزية، والفرنسية، والألمانية، وغيرها. وتحتوي العربية على حوالي ستين رمزًا مستخدَمًا في الكتابة؛ وتتوزع هذه الرموز بين حروف، وعلامات ضبط بالشكل، وعلامات ترقيم، وأرقام. كما تأخذ الكتابة العربية هيئة متشابكة –حتى في صورتها المطبوعة– وعلى هذا قد يختلف شكل الحرف بِناءً على كيفية اتصاله بالحروف المجاورة، وهو ما يجعل لشكل الحرف الواحد أربعة احتمالات: الحرف مستقلًا، والحرف في أول الكلمة، وفي وسطها، وفي آخرها، على النحو الموضح بالشكل التالي.

هـ هـ هـ هـ	هـ هـ هـ هـ	هـ هـ هـ هـ	هـ هـ هـ هـ

- تُكْتَب اللغة العربية من اليمين إلى اليسار، بخلاف طريقة الكتابة اللاتينية التي تكتب من اليسار إلى اليمين.
- تتسم النصوص العربية بالاستخدام غير المنتظم لعلامات التشكيل، بل إنها تغيب– رغم

(13) Ibid.

(14) السعيد محمد بدوي: مستويات العربية المعاصرة في مصر- دار المعارف- القاهرة – 1973- ص 89.

تأثيرها في حركة الحرف، وأحيانًا في طريقة نطقه– في أغلب النصوص العربية الحديثة، مما قد يؤدي إلى إحداث قدر من الغموض فيما يتعلق بمعنى الكلمة، أو وظيفتها النحوية.

- تخلو اللغة العربية من الحروف الكبيرة capital letters التي تستخدم في بعض اللغات وفقًا لقواعد محددة، وتؤدي وظائف مهمة في بعض تطبيقات المعالجة الآلية كتَعرُّف الأسماء آليًّا named entity recognition.

المعجم ونظم المعالجة الآلية:

في سياق الدراسات اللغوية الحاسوبية يميَّز بين نوعين من المعاجم الآلية:

أ– **المعاجم الإلكترونية** electronic dictionaries ، وهي معاجم آلية مصممة للاستخدام البشري؛ والهدف منها تزويد المستخدم البشري بالمعلومات المعجمية، بحسب نوعية المعجم.

ب– **المعاجم الحاسوبية** computational lexicons ، وهي معاجم مصممة لاستخدامها في نظم معالجة اللغات الطبيعية. والهدف الأساسي من بناء هذه المعاجم هو توصيل المعلومات إلى النظم التي تقوم بتحليل اللغة الطبيعية– نصوصا أو كلاما– وإنتاجها.

يبدو واضحًا أن الاختلاف بين النوعين قائم على الوظيفة التي يختص بها كل منهما. وقد ترتب على اختلاف الوظيفة– أو الهدف من بناء المعجم– سؤال بدهي يتعلق بمدى كفاية المعلومات في المعاجم الإلكترونية لتلبية متطلبات نظم المعالجة الآلية.

إن المعجم الإلكتروني لا يختلف– في جوهره– عن المعجم الورقي المطبوع، "بيد أن المعلومات تكون موجودة في ملفات نصية text files"[(15)] ؛ لأن المستهدَف في الحالتين واحد، بل إن بعض المعاجم الإلكترونية لا تعدو– بالفعل– كونها نسخًا إلكترونية من معاجم ورقية؛ مع عدم إغفال مجموعة المميزات التي يضيفها المعجم الإلكتروني للمستخدم البشري.

والحقيقة أن هذا الأمر يختلف عن بناء معاجم تُستخدَم في تطبيقات المعالجة الآلية؛ لأنه على الرغم من اشتمال المعاجم الإلكترونية على قدر من المعلومات القيمة فإن هذه المعلومات لا تصلح للمعاجم الحاسوبية؛ إذ تحتاج المعاجم الحاسوبية إلى معلومات تفصيلية على المستويات الصرفية، والنحوية، والدلالية؛ والسياقية؛ لأن المستخدم في هذه الحالة هو الآلة، وليس الإنسان؛ ولهذا يحتاج هذا النوع من المعاجم إلى أن يكون أكثر شموليةً، وتنظيمًا، وتصريحًا، وتماسكًا، وشكليةً من معاجم الاستخدام البشري.

لقد اعتبرت النظرةُ التقليدية المعاجمَ الحاسوبية مجرد مستودعاتٍ للمعلومات المعجمية لمهام محددة، كالإعراب parsing، والتوليد generation، والترجمة. ومن هذه الوجهة وجب أن يحتوي

(15) Sterkenburg, P. (2003). P. 5.

المعجم على نوعين من المعرفة: معرفة لازمة للتحليل والتوليد النحوي، ومعرفة لازمة للتفسير الدلالي. لكن هذا التعريف خضع لمراجعة كبيرة – بعد ذلك – نتيجة للنضج الذي حدث في مجالي علم الدلالة وعلم اللغة الحاسوبي.

دور المعجم في نظم المعالجة الآلية:

تتطلب نظم المعالجة الآلية للغات الطبيعية– من أكثرها بساطة إلى أكثرها تعقيدًا– معرفة بالمفردات؛ ففي معظم الأحوال تحتاج هذه النظم– لكي تكون أكثر فاعلية وعملية– إلى تزويدها بمعجم أساسي يغطي الاستخدام الفعلي لمفردات اللغة، وكذلك المعلومات اللغوية المتعلقة بها. وحتى بالنسبة للتطبيقات البسيطة كالمدققات الإملائية spelling checkers قد لا يخلو الأمر من حاجة إلى معلومات صرفية، وإملائية، وصوتية، ونحوية، ودلالية لإزالة اللبس. ومن هنا أصبح تصميم المعجم وتطويره من القضايا الرئيسية في المعالجة الآلية للغات الطبيعية[16].

إن الأداء النهائي لنظم المعالجة الآلية يعتمد على كفاءة المعجم المرتبط بها؛ إذ تؤدي المعلومات الصرفية، والنحوية، والدلالية، والسياقية– التي يزوِّد المعجمُ النظامَ بها– دورًا مهمًّا في هذا الأداء، ومع ذلك قد تتنوع مستويات التفصيل ونوعية المعلومات التي يحتويها المعجم باختلاف نظام المعالجة.

ويبدو هذا التنوع واضحًا عندما نقارن بين بناء معجم يقتصر استخدامه– مثلًا – على التصنيف النحوي POS tagging، وآخر يستخدم في الإعراب الآلي parsing، أو الترجمة الآلية machine translation، أو غير ذلك من نظم المعالجة. فعند بناء مصنف نحوي tagger يجب أن يكون النظام قادرًا على تعيين كل الصيغ الممكنة للكلمة في اللغة، وهو ما يمكن إنجازه عن طريق عمل قائمة بكل الصيغ الصرفية الممكنة للكلمة في المعجم، لكن هذا الأمر لا يحدث فعليًّا؛ لأن معظم النظم تحتوي على محلل صرفي morphological analyzer، أو مُجَذِّع stemmer يقوم باختزال هذه الصيغ إلى وحداتها المعجمية الفرعية lexemes ، ومن ثم يُبحث عن هذه الوحدات في المعجم.[17]

وعلى هذا يحتوي المعجم المرتبط بالمُصَنِّف النحوي على الوحدات المعجمية الفرعية مع معلومات حول التصنيف النحوي لها parts of speech ، ولا يحتاج هذا النظام إلى معلومات دلالية، أو تصنيفات فرعية subcategorizations، إلا في حالة استخدام نتائج هذا المُصَنِّف مداخلَ للمعرب parser، أو لتطبيقات أخرى أكثر تعقيدًا، كالترجمة الآلية[18].

([16]) Brisco, T. (1991). Lexical issues in Natural Language Processing. In Klein, E. & Veltman, F. (1991). Natural Language and Speech. Springer, Verlag.

([17]) Megerdoomian, K. (2003). In Farghaly, A. (ed). Handbook for Language Engineers. Center for the Study of language and Information. Page 25.

([18]) Ibid. P. 25.

ويُقصَد بالتصنيفات الفرعية للأفعال الأقسام النحوية الفرعية التي تُعبِّر بها الأفعال عن المواضيع الدلالية semantic arguments الخاصة بها عن طريق وسائل نحوية مختلفة. ويطلق على كل قسم من هذه الأقسام الفرعية– التي يمكن أن يظهر فيها الفعل– اسم "إطار تصنيف فرعي" sub categorization frame. وتبدو معرفة هذه الإطارات من الأهمية بمكان في الإعراب parsing [19]. ونستطيع أن نتبيَّن أهمية هذه التصنيفات لمهام الإعراب إذا نظرنا إلى إعراب المثالين التاليين:

أخبرت زوجها أن الطفل في المكان المتفق عليه.

وجدت الزوجة المكان المتفق عليه.

فإذا علمنا أن للفعل "أخبر" الإطارَ (NP NP S)، الذي يعني (فاعل subject، ومفعول object، وجُمَيْلَة clause)، وأن الفعل "وجد" ليس له هذا الإطار، ولكن له الإطار (NP NP) – مثلًا – أمكننا أن نلحق عبارة "المتفق عليه" إلى الفعل "أخبر" في الجملة الأولى، وإلى "المكان" في الجملة الثانية، وذلك على النحو التالي:

أخبرت [الزوج] [أن الطفل في المكان المتفق عليه].

وجدت [الزوجة] [المكان [المتفق عليه]].

وتلتقي فكرة الإطار التصنيفي مع تقسيم الفعل عند النحاة العرب من جهة المفعول؛ فهناك الأفعال اللازمة intransitive verbs ، وهي الأفعال التي تكتفي بفاعلها، ولا تقع على مفعول به، وهناك أفعال متعدية transitive verb ، وهي على أقسام : منها ما يتعدى بنفسه إلى المفعول دون حاجة إلى حرف جر، ومنها ما يحتاج إلى حرف جر. والأفعال المتعدية تنقسم من حيث عدد مفاعيلها إلى ثلاثة أقسام: أفعال تتعدى إلى مفعول واحد، وثانية تتعدى إلى مفعولين، وهي على ضربين: فمنها أفعال تتعدى إلى مفعولين أصلهما المبتدأ والخبر، وأخرى تتعدى إلى مفعولين ليس أصلهما المبتدأ والخبر، وثالثة تتعدى إلى ثلاثة مفاعيل. وبتطبيق فكرة الإطارات التصنيفة على هذا التقسيم يمكن ملاحظة بعض الإطارات التصنيفية:

نوع الفعل	الإطار	الوظيفة	الفعل	مثال
لازم	NP	فاعل	عاد	عاد المسافر.
لازم	NP PP	فاعل – فضلة	مات	مات المريض في المستشفى.
متعدٍ لمفعول واحد	NP NP	فاعل – مفعول	قاتل	قاتلُ الجنديُّ العدوَّ.
متعدٍ لمفعول واحد	NP Clause	فاعل – مكمل	علِمَ	يَعلَمُ اللهُ (أنني صادقٌ).
متعدٍ لمفعولين أصلهما	NP NP NP	فاعل – مفعول	عَلِمَ	علمتُ الطالبَ مجتهدًا.

(19) Foundations of Statistical NLP. Pages 271 – 272.

المبتدأ والخبر		– مفعول		
متعدٍ لمفعولين ليس أصلهما المبتدأ والخبر	NP NP S	فاعل– مفعول– مفعول	أَخْبَرَ أعطى كسا	أخبر المعلمُ التلميذَ أن إجابته سديدة. أعطى الغنيُّ الفقيرَ إحسانا. كسا المحسنُ البائسَ ثوبا.
متعدٍ لثلاثة مفاعيل: الثاني والثالث أصلهما مبتدأ وخبر	NP NP NP NP	فاعل – مفعول1 – مفعول2 – مفعول3	أَنْبأَ أَعلَمَ	أنبأتُكَ الفوزَ أكيدًا. أعلمَ الوالدُ ابنَه الصدقَ منجاةً.

جدول يوضح بعض إطارات التصنيفات الفرعية في إطار تقسيم الفعل بحسب المفعول به

ويكون إعراب الأمثلة– وفقًا لهذه الإطارات التصنيفية– على النحو التالي:

عاد [المسافرون].

مات [المريضُ] [في المستشفى].

قاتل [الجنديُّ] [العدوَ] [اللدودَ]].

يعلم [اللهُ] [أنني صادق].

علم[ت] [الطالبَ] [مجتهدًا].

أخبر [المعلمُ] [التلميذَ] [أن إجابتَه سديدةٌ].

أعطى[الغنيُّ] [الفقيرَ][إحسانًا].

كسا [المحسنُ] [البائسَ] [ثوبًا]،

أنبأ[تُ]-[كَ] [الفوزَ] [أكيدًا].

أَعلمَ [الوالدُ] [ابنَه] [الصدقَ] [منجاةً].

أما المعلومات الدلالية التي يجب أن يحتويها المعجم الحاسوبي فليس هناك اتفاق كبير حول مقدارها، ولعل التوافقَ أقلُّ حول طريقة تمثيل هذه المعلومات داخل المعجم

وعلى الجانب الآخر تؤدي محاولة التمثيل الكامل لمعنى الكلمة على قضايا تمثيل معرفية معقدة، كما أنها أقل مرونة على المستوى النظري، وأقل مطاوعة على المستوى الحاسوبي، حتى المعلومات شديدة البساطة يمكن أن تنطوي على مشكلات؛ مثال ذلك ما نلحظه من أن فاعل الفعل "تحمَّل" (experiencer) يجب أن يمتلك صفة الإحساس، لكن هذا الفعل قد يأتي مع النباتات، أو مع الجماد، كما في المثالين التاليين:

لا تتحمل النباتات الصحراوية التربة القلوية.

يتحمل الحديد ما لا يتحمله الخشب.

ومن ثم يؤدي اعتبار "صفة الإحساس" قيدًا ملزمًا إلى فشل إعراب الجملتين السابقتين. ومن ثم تستخدم هذه المحددات لتوجيه عملية المعالجة– بشكل عام– عن طريق التفضيلات preferences بدلًا من التحكم التام في عملية الإعراب.

وتحتاج بعض نظم الفهم الآلي للغات الطبيعية– كاسترجاع المعلومات information retrieval، وإجابة الأسئلة question answering – إلى معلومات تتعلق بمعاني الكلمات؛ حتى تقوم بعمل الاستنتاجات المناسبة. وعلى هذا يحتاج المعجم المرتبط بهذه النظم إلى أن يعكس العلاقات الدلالية كالترادف، والتضاد، والعلاقات الهرمية hierarchical relations بين الكلمات كي تقوم هذه النظم بتحليل "معنى" الجملة[20].

مقارنة بين الدراسات الحاسوبية للمعجم والمناهج اللغوية التقليدية:

تتمثل مظاهر الاختلاف بين الدراسات الحاسوبية للمعجم والمناهج اللغوية التقليدية في أربع نقاط[21]:

أولا: وضوح التمثيل المعجمي: lexical representation

يعد وضوح التمثيل المعجمي من أهم مبادئ بناء المعجم الحاسوبي؛ فالمعرفة التي يحتويها المعجم يجب أن تكون مفصلة – بشكل كافٍ – بحيث تدعم نظم المعالجة، ولاسيما النظم التي تعتمد على القواعد rule-based systems، وهي نظم تعتمد على نظريات لغوية، مثل النحو الوظيفي المعجمي[22] lexical-functional grammar، ونحو الاعتماد[23] dependency syntax، والنحو التصنيفي[24] categorical grammar ... إلخ.

---

(20) Ibid. 26.

(21) Ibid. P. 161.

(22) النحو الوظيفي المعجمي lexical-functional grammar نظرية عن بنية اللغة الطبيعية، وعن كيفية ترابط عناصر البنية اللغوية. ويعكس اسم النظرية اختلافها عن النظريات الأخرى من جهتين: فهي نظرية معجمية تقوم على أن العلاقات بين الصيغ اللغوية – مثل العلاقة بين صيغتي بناء الفعل للمعلوم والمجهول - هي عموميات حول بنية المعجم، وليست عمليات تحويلية تشتق صيغة اعتمادًا على الأخرى، كما يحدث في النحو التحويلي. وهي نظرية وظيفية تقوم على العلاقات النحوية – مثل العلاقة بين الفاعل والمفعول- تعد أبنية أساسية، وبدهية، ولا تُعَيَّن باعتبار المفاهيم الدلالية (العامل agent، والمعمول patient)، أو تعريفات بنية العبارة phrase structure configuration. انظر: Ibid. P. 464.

(23) النحو الاعتمادي dependency syntax هو نوع من النحو الشكلي يقيم علاقات بين عناصر التركيب، تمثل في صورة شجرة اعتماد dependency tree، وهي عبارة عن مجموعة من العقد nodes يحدد الترابط بينها العلاقات التركيبية. وتحتوي كل شجرة على رئيس (رأس) governor وأتباع dependents، وذلك على النحو التالي: الفعل هو الرأس في الجملة، والتركيب الاسمي يكون تابعًا. وفي تركيب الجار والمجرور يكون حرف الجر هو الرأس، في حين يكون الاسم المجرور تابعًا. وفي

ثانيا: البنية الشاملة للمعجم The global structure of the lexicon:

يجب أن تكون للبنى التي تربط الكلمات كلها في المعجم نفس الأهمية الخاصة بالبنى التي تربط عناصر الكلمات. كما يجب اعتبار التجمعات اللفظية– كالتصاحبات اللفظية collocations، والتعابير الاصطلاحية idioms، والكلمات المركبة compound words وغيرها– مداخل معجمية يتم تمثيلها داخل المعجم.

ثالثا: التغطية الكافية للمجال sufficient coverage of its domain

يجب أن يكون المعجم شاملا بمجال المعجم؛ فالمعجم الحقيقي يمكن أن يشتمل– مثلًا – على حوالي أربعمائة ألف (400,000) مدخل.

رابعا: قابلية التقييم:

تُقَيَّم المعاجم الحاسوبية باعتبارات ثلاثة هي:

- **التغطية coverage:** وتكون التغطية– أفقيًّا– على مستوى المعجم، ورأسيًّا على مستوى المعلومات المعجمية.
- **القابلية للتوسيع** extensibility: ما مدى سهولة إضافة المعلومات إلى المدخل المعجمي؟ وإلى أي مدى يمكن أن تتسق المعلومات الجديدة مع البنى المعجمية الأخرى؟
- **الاستخدام** utility: ما مدى فائدة المداخل المعجمية لتطبيقات المعالجة الآلية؟

خاتمة

تركيب الإضافة يكون المضاف هو الرأس والمضاف إليه يكون تابعًا. وفي التركيب الإسنادي يكون المبتدأ رأسا والخبر يكون تابعًا في الجملة الاسمية، ويكون الفعل رأسًا والفاعل تابعًا في الجملة الفعلية. انظر: Crystal, D. (2008). P. 137.

(24) النحو التصنيفي categorical grammar نوع من النحو الشكلي، ابتكره المناطقة logicians في العشرينيات والثلاثينيات من القرن العشرين، وطوره اللغويون في الخمسينيات. ويقوم على فكرة تقسيم التعبير النحوي إلى: مجموعات أو أقسام أساسية primitive categories يشتق منها مجموعات اشتقاقية derivative categories فإذا افترضنا مجموعتين أساسيتين هما: المجموعة (N) للاسم و(S) للجمل فإن كلمة مثل (go) تعين لها المجموعة (N/S) ، التي تعني أن كلمة go يمكن أن تتحد مع الاسم (N) لتكوين الجملة (S). انظر: Ibid, P. 68.

المعجم ونظم المعالجة الآلية

أ.د. وفاء كامل فايد

مؤتمر جمعية هندسة اللغة 2020

(ملخص البحث)

- **مقدمة**

يتناول البحث علاقة المعجم بالنظم المختلفة التي تتعامل مع المعالجة الآلية للغات الطبيعية، فيتحدث عن مدى الإفادة من الحاسب الآلي في صناعة المعاجم. وعن بدايات استخدام الحاسوب في الدراسات اللغوية العربية.

- **مقارنة بين المعاجم الورقية والإلكترونية.**
- **الإمكانات التي تتيحها التكنولوجيا للمعجمي.**
- **مفهوم المعالجة الآلية للغات الطبيعية، والهدف منها.**
- **اللغة العربية وتحديات المعالجة الآلية.**
- **المعجم ودوره في نظم المعالجة الآلية.**
- **مقارنة بين الدراسات الحاسوبية للمعجم والمناهج اللغوية التقليدية.**
- **خاتمة.**