الإحصاء في المدونات القاموسية
- من أجل إعداد قاموس تعليمي

الدكتور:عثمان لخضر
المدرسة العليا للأساتذة ببوسعادة
athmanlakhdar@gmail.com

الملخّص:

تستند مداخلتي إلى مفهوم الإحصاء المعجمي statistique lexicale، حيث أحاول المساهمة في عرض بعض معايير جمع المادة المعجمية وانتقاء المفردات - المداخل- اعتماداً على التكرار المطلق والنسبي كمعيارين مهمين في إحصاء المفردات الموجّهة لإعداد قاموس تعليمي متخصص، سيتوجّه بحثنا إلى إحصاء الوحدات المتخصصة المستعملة في مادة التربية العلمية والتكنولوجية في مستوى الابتدائي.

الكلمات المفتاحية: المدونة- الإحصاء - المعجم المدرسي- المفردات .

البحث في المدونات القاموسية المتخصصة:

ترتبط عملية البحث في وقتنا الحالي بعملية الحوسبة والمعالجة الآلية، إذن نحن إزاء المدونّة المحوسبة، وتعتبر هذه المدونات سندا أساسيا في معالجة المفردات، أي أنّها لم تعد تستخرج من قوائم اصطلاحية أو من خلال اجتهادات فردية بل من مدوّنة يتقاسم فيها المتخصص في الإعلام الآلي الدور مع اللساني في وضع البرامج الكفيلة بمعالجة المفردات اللغوية.

يمكن أن نرسم من الآن الخطوات العملية التي سنباشرها في عملية دراسة المدونة القاموسية، يتعلق الأمر بمرحلتين:

تتعلق الأولى بمواصفات المدوّنة ومعايير انتقائها والشهادات التي تقدمها لنا لمعالجة المحتوى.

أما الثانية فتتعلق بالبحث في النصوص قصد وضع بعض المعايير لتحديد المفردات وانتقائها قصد تعجيمها.

الإطار المنهجي:

أسسنا بحثنا على أسس نظرية معنى- نص(1) وهذا يجعلنا ننطلق في تناول هذه القضية من الملاحظتين الآتيتين:

- نقبل بالترادف الاصطلاحي ونعدّه معطيات يجب وصفها في إطار النظرية. .
- الترادف منشؤه في هذه القضية هو الإقليمية وهذا يعني أن النصوص التي سنختارها ستعكس لنا كل الاستعمالات
- محاولة لوضع برنامج معالجة مدونة لسانية آلية :

سنسعى إلى إبراز ملامح بعض المعايير والشروط التي ستسهّل على المتخصّص في الحوسبة معالجة المصطلحات بمختلف أنواعها: الأسماء والأفعال والمركبات وذلك بالتعرف على طرق استخراجها ووصفها تمهيدا لتعريفها ومن ثمّ إمكانية تعجيمها وتيسير التداول بها بين المتخصصين.

-1 - تعريف المدونة المتخصصة :

نقبل مبدئيا بالشرط المنهجي الذي يؤطر وضع التعريف، يشمل هذا الشرط كل المصطلحات، فتعريف مصطلح ما يخضع إلى إطار مرجعي محدد، تعريف المدوّنة لا يشذ عن هذا الأمر فتعريف المدوّنة المتخصصة في إطار مقاربة الدلالة المعجمية يختلف عن إطار المنظور المفهومي، ينظر للمدونة في إطار المنظور الثاني كمعطيات يتم تخزينها آليا كبنوك للمصطلحات مع وضع المواصفات القياسية لكل مصطلح، لكن دون اللجوء إلى الذكاء الاصطناعي للقيام بمختلف العمليات التي من شأنها أن تعالج المصطلح من جوانب مختلفة وضعا واستعمالا. يختلف الأمر من منظور الدلالة المعجمية فالمدونة المتخصصة الآلية أصبحت حسب "ساجيه" Sager- سند أساسي في معالجة المصطلحات، أي أنها لم تعد مستخرجة من قوائم اصطلاحية سابقة أو من خلال اجتهادات فردية بل من مدوّنة معلومات. تتمثل في الأصل في مجموعة نصوص متخصصة أي ليست مصطلحات معزولة، فالنص المتخصص هو «خزان كبير من المصطلحات، يستخدمه المصطلحي للاستدلال على الوحدات التي يمكن أن تكون جزءا من المعجم الذي يحضره » **(2).**

نقبل إذن بالخاصيتين الآتيين في تعريف المدونة المتخصصة:

خاصية 1 : النص المتخصص معطى أساسي في المدونة المتخصصة.

خاصية 2 : المصطلح وحدة معجمية كامنة غير قابلة للانفصال عن النص في المدونة المتخصصة.

حسب الخاصية2 : تتنوع المصطلحات حسب أقسام الكلام، فهناك من المصطلحات ما يرد أسماء أو أفعالا أو تراكيب مختلفة.

اختيار المدوّنة :

تعد مرحلة جمع المدوّنة وإعدادها مرحلة جد دقيقة، لأن نوعية النتائج التي سنحصل عليها هي رهن القرارات التي نتبناها، يتعلق الأمر بمدى أهمية النصوص وقدرتها التمثيلية représentativité أي قدرة المعطيات النصية المنتمية إلى مجال دراستنا – اللسانيات - على تقديم كل الخصائص التي تتضمنها المعاجم والقواميس اللسانية، فالنص اللساني الذي يتم اختياره في دراسة مصطلحات هذا التخصص

يجب أن يستجيب إلى شرط التمثيلية **(3)** لا يقتصر الأمر على اختيار مجال التخصص فقط بل بالتطبيق المستهدف من الدراسة. سنتناول في هذا الجانب معايير انتقاء النصوص.

معايير الانتقاء :

لانتقاء المدونة وتحديدها تقترح علينا "ماري كلود لوم" مجموعة من الضوابط اخترنا منها **(4)** : مجال التخصص – حجم المدونة.

نحاول في هذا البحث وضع نماذج لعملية استخراج المصطلحات من المدونات اللسانية قصد إعدادها لعملية التعجيم أي إمكانية إدراجها كمداخل للقاموس اللساني .

مجال التخصص:

المقصود هو تحديد مجال التخصص بدقّة، يمكن أن يندرج داخل التخصص الواحد مجموعة تخصصات، وهذا ما يسمى بتداخل الاختصاص أي داخل تخصص ما هناك تخصصات، كما يمكن أن يفهم بوجود مصطلحات تنتمي إلى مجموعة تقاطع بين تخصص ما وتخصصات مختلفة، إذ أن فهم المصطلح قد ينغلق أحيانا على المتلقي لعدم

تمثل «الأبعاد المرجعية والمعرفية والمنطقية أو الأبعاد المنهجية أو تصور المعطيات والوقائع أو العلاقة بين هذه الأبعاد مجتمعة (5)».

حجم المدوّنة :

سوف نباشر المعالجة الآلية لعينة بواسطة برنامج استخراج الوحدات المتخصصة وملاحظة تكرار الوحدات المرشحة التي تستجيب لمعايير انتقاء المصطلحات، عملية الكشف هذه ستسمح لنا بالحكم على صلاحية اختيار المدونة في البداية وبتحديد حجمها في أحسن شكل لتحقيق أهداف مشروعنا في صناعة المصطلح، سيكون هدفنا وصف عيّنة نصية، تتمثل في مجموعة نصوص تمّ انتقاؤها من كتاب التربية العلمية والتكنولوجية للسنة الخامسة ابتدائي. حيث سنرسم ملامح برنامج لمعالجة المصطلحات.

صغر حجم المدوّنة يعود إلى أسباب مختلفة :

– عملية الوصف ولائحة الاستبعاد في المدونات النصية الكبيرة تتطلبان إمكانيات في البرمجيات لا تتوفر للباحث ولا حتى لمن تقربنا منهم من المتخصصين في البرمجيات لأن الأمر يتعلق بالحوسبة اللغوية.

- البرمجيات المتوفرة في المعالجة الآلية للمدونات المتخصصة سواء أحادية اللغة أو متعددة اللغات هي برمجيات تتعامل باللغتين الفرنسية والانجليزية كبرنامج(6) termostst.

وصف البرنامج المقترح في تحليل النصوص :

قمنا استنادا إلى المعايير السابقة بتوجيه الحاسوبي إلى وضع برنامج لاستخراج المصطلحات فاعتمد البرنامج في وضع على جانبين:

أحدهما لساني وقد اختص الباحث بضبط معاييره اللغوية، والثاني حاسوبي حيث قام الحاسوبي "رقيق يونس" بوضع البرنامج وفق متطلبات البرمجة.

الخصائص التقنية للبرنامج:(7)

قاعدة المعطيات Data base XML: - لغة البرمجة: C#

Logicielle:Visual studio 2015 يتم تشغيل هذا البرنامج على نظام التشغيل ويندوز فقط.

Target frame work 4.5: .NET.

معايير في معالجة النصوص المتخصصة:

نشير في البداية إلى أن فكرة الاستعانة بالحاسوب كانت مطروحة منذ البدايات الأولى لظهور علم المصطلح مع التوجه الفوستري غير أن الأمر اقتصر آنذاك على عملية التخزين فقط ووصف المصطلحات في جذاذات ويتم اللجوء إليها عند الحاجة باعتبارها خزانا للمصطلحات من أجل عملية الاستقراء تمهيدا لتوحيدها،

جاء في مقدمة مقال "لمونيك سلودزيان" عنوانه بروز علم مصطلح نصي ما نصه: « إن الإسهام الذي قدمته المعلوماتية لعلم المصطلح منذ مطلع السبعينيات يكاد ينحصر في أنظمة قواعد البيانات العلائقية الهادفة إلى تخزين المصطلحات ومعالجتها ونشرها ... وبفعلهم هذا- أي المتخصصون في علم المصطلح- هم يتبعون تعاليم "فوستر" الذي كان ينظر إلى المعلوماتية بوصفها أحد الميادين المؤلفة لعلم المصطلح (8)».

أمّا في التوجّه الحديث تجاوز فكرة التخزين إلى المعالجة الآلية حيث تخضع المدونات النّصية المتخصصة إلى مختلف العمليات الآلية في عمل جماعي يتقاسمه اللغوي والحاسوبي والمتخصص(9).

استندنا في وضع البرنامج إلى أسس لسانية يؤطرها التوجه الدلالي المعجمي(10)، حيث يتم النظر إلى المصطلحات كوحدات معجمية وليس كوحدات مفهومية، وبالتالي فمعالجة المصطلحات وفق هذا المنظور تتم في إطار النصوص وليس كوحدات معزولة، يتّم الاعتداد هنا بمختلف أقسام الكلم واعتبارها مصطلحات ولا يتم الاقتصار على صنف الأسماء فقط.

ماهي المعايير التي يتم الاستناد إليها في معالجة النصوص المتخصصة لاستخراج المصطلحات آليا ؟

المعايير اللسانية التي استندنا إليها في وضع البرنامج والتي استند إليها الحاسوبي فيما بعد في وضع البرنامج أخذناها من تجربة الباحثة الكندية "ماري كلود لوم" في من خلال كتابها: علم المصطلح، مبادئ وتقنيات، واقتصرنا على المعايير التي استطعنا التعامل معها في اعداد برنامج تحليل النصوص، وهي:

1- **التكرار** :تتم المعالجة الآلية من الناحية النظرية بالاعتماد على الجانب الإحصائي، واستعنا في هذا الجانب بالإحصاء المعجمي كما وضع أسسه مولر(11)، يقسم التكرار في الإحصاء عموما إلى نوعين؛ تكرار مطلق وتكرار نسبي

أ- **التكرار المطلق** FA : يتم حساب تكرارات الكلمات، والتكرار مؤشر أساسي على تخصص الكلمة في النص، حيث غالبا ما يكون ارتفاع تكرار الكلمة دليلا على تخصص الكلمة، واعتبارها كلمات مفتاحيه للنص . يسمى هذا النوع من التكرار بالتكرار المطلق حيث يقتصر فيه الأمر على عدّ مرات ظهور الكلمة في النص.

ب- **التكرار النسبي** F.R : يتم هنا الاعتداد بالنسبة المئوية لتكرارات الكلمة في النص أي نسبة التكرار المطلق للكلمة إلى التكرارات الإجمالية لكل كلمات النص. نعبر عنها بالنسبة الآتية :

$$\frac{\text{تكرار الكلمة}}{\text{تكرار الإجمالي}}$$

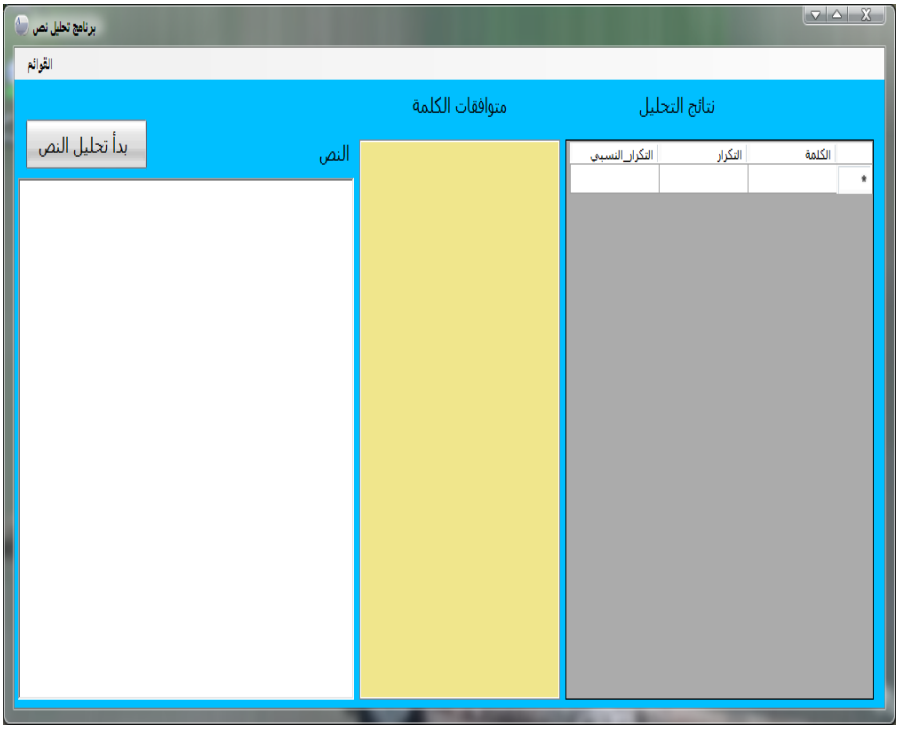
.

ويعبر عن النتيجة بعدد محصور بين 0 -1 مع الإشارة إلى أن النتائج المتحصل عليها يجب أن تظهر بأكبر عدد ممكن من الأرقام بعد الفاصلة لتقريب النسبة أكثر.

2- المتوافقات :

بعد أن يتم تحديد تكرار الكلمات يستعين الباحث بمعيار آخر في استخراج المصطلحات، فالتكرار وإن كان ضروريا في تحديد تفاوت مفردات النص في الأهمية، إلاّ أنّه لا يعني بالضرورة تخصص مفردة ما، بل لا بد من اللجوء إلى معيار التوافق أي لائحة تحتوي على السلسلة التي يتم البحث عنها مرفقة بسياقاتها، ومن التجارب الرائدة في هذا المجال برنامج KWIC وهو مختصر لعبارة كلمة مفتاح في السياق :

Key Word in Context وقد استعنا بفكرة هذا البرنامج في وضع معيار المتوافقات في برنامجنا .



واجهة برنامج تحليل النصوص:

تحليل النص: بعد أن يتّم إدراج النص المخصص للتحليل في الجزء الخاص بذلك وهي المساحة البيضاء التي تحتل الحيز الأكبر من الواجهة، يتم التعامل مع معطيات النص بواسطة الضغط على تحليل النص لتظهر مباشرة لائحة تمثل العنوان التالي وهو الكلمة وتكرارها.

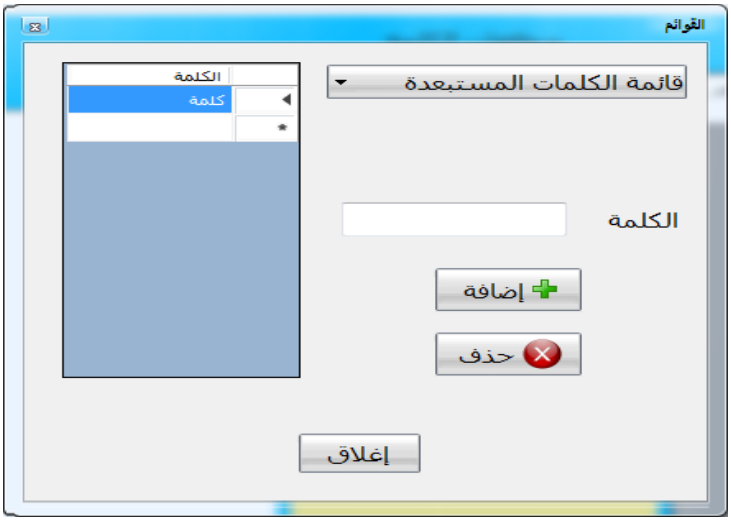
الكلمة: يتم تقطيع النص إلى كلمات ، يستند مفهوم الكلمة هنا إلى ما تم تحديده للبرنامج في قاعدة المعطيات.

التكرار المطلق : يتم حساب تكرارات الكلمات، ويتم ترتيبها على هذا الأساس.

التكرار النسبي: يبرز نسبة التكرار المطلق إلى عدد الحالات الممكنة.

المتوافقات C: حيث يظهر المعالج في المساحة ذات اللون الأصفر متوافقات الكلمة أي المحيط اللغوي للكلمة ما يسبقها وما يليها مباشرة . وهذا ما يسهل بطبيعة الحال الحكم على تخصص النص المعالج تعديل القوائم: يتيح هذا العنوان مجموعة خيارات بمجرد الضغط عليه وتتمثل في قائمتين الأولى خاصة بالحروف والثانية بالكلمات وكلاهما

مخصصتان لعملية التعديل إما بالإضافة وإما بالحذف. حيث يساهم هذا العنوان بتعديل عملية التحليل ووضع بالأساس للتعامل مع النصوص المتخصصة.



واجهة خاصة بنافذة القوائم حيث يتم فيها تحديد لائحة الاستبعاد.

النص المقترح :

نقوم بتحليل النصوص الواردة في فقرة (ما يجب الاحتفاظ به مع التمرين الثاني) من كتاب التربية العلمية والتكنولوجية الخامسة ابتدائي، في **الوحدة 1 :** مصونية كتلة المادة عند انحلالها في الماء.

النصوص هي :

- بعد تحريك محتوى الكأس الذي يحتوي على الماء والسكر تختفي مادة السكر.
- عند اختفاء مادة بعد انحلالها في الماء فإن اختفاءها يكون ظاهريا لأن كتلتها تبقى محفوظة .
- بالتسخين أو التبخر يمكننا استرجاع مادة منحلة في الماء .
- الزيت ينتحل في الماء
- السكر ينحل في الماء الفاتر أكثر مما ينحل في الماء البارد .
- الملح لا ينحل في الماء.

نتائج التحليل كما ظهرت على واجهة البرنامج

وصف النتائج وتحليلها:

نشير في البداية إلى أن لائحة الاستبعاد المدرجة في عنوان القوائم قد زوّدت بمجموعة من الكلمات والحروف، قصد عزلها أثناء التحليل باعتبارها وحدات نحوية لا يمكن التعامل معها في النصوص المتخصصة باعتبارها وحدات معجمية يمكن أن تتخصص، فكانت لائحة الاستبعاد كما وردت في الصورة الآتية :

نتائج التكرار:

تظهر نتائج التحليل أن التكرار المطلق لا يتأثر بقوائم الاستبعاد، فمهما عمدنا إلى تعديل قائمة مفردات النص بالزيادة أو الحذف فإن تكرار مفردة معينة لا يتغير، وفي المقابل فإن التكرار النسبي يتغير بالتعديلات التي تطرأ على لائحة الاستبعاد لأن الأمر يتعلق بعلاقة التكرار المطلق بمجموع التكرارات الكلية لكل مفردات النص، أظهر البرنامج النسب المؤوية لتكرار كل مفردة غير أن الأمر لا يزال في حاجة إلى تدقيق أكثر حيث نحتاج إلى ضبط القيم بالفاصلة.

يمكن أن نعبّر عن ذلك في الجدول الآتي :

الكلمة	التكرار	التكرار النسبي	القيم العشرية
الماء	7	17	0.17
مادة	3	7.31	0.0731
ينحل	3	7.31	0.0731
السكر	2	4.87	0.0487
انحلالها	1	2.43	0.0243
منحلة	1	2.43	0.0243
مج التكرارات	41	100	0-1

تُظهر القيّم الواردة في التحليل أن هناك عيّنة من المفردات ذات تكرار مرتفع نسبيا مقارنة بالمفردات ذات القيم التكرارية المنخفضة، يمكن لهذه المفردات أن تكون مرشحة للتخصص كمصطلحات، غير أن هذا المعيار يبقى غير كاف بل لابد من مراعاة متوافقات كل كلمة، والتي تسهم في تحديد مجال التخصص.

المتوافقات :

بعد تحديد التكرار نقوم بالبحث عن المتوافقات الخاصة بكل مفردة أي سياق الكلمة.

تظهر النتائج الخاصة بالكلمات الواردة في مقدمة التحليل ككلمة (الماء) النتائج الآتية الخاصة بالمتوافقات :

يحتوي على الماء والسكر

انحلت في الماء وتحصلنا

مادة بعد انحلالها في الماء فان اختفاءها

منحلة في الماء

كنتيجة أولية يتاح لنا الحكم بأن:

التركيب انحل في الماء هو متلازمة لفظية

يمكن أن نستعين مستقبلا بهذا المفهوم أي التلازم اللفظي كمؤشر نحيل الحاسوبي إليه لإدراجه في برنامج المعالجة ويتم ذلك بحساب تكرار هذه التركيبات وتصنيفها في لائحة خاصة.

بالنسبة للأفعال الواردة بعد الوصف هي كلمات يمكن عدها متخصصة، وذلك باعتبار متوافقاتها، فهي تحيل إلى المركب انحل في الماء لو وسعنا سياق المفردة في البحث وهذا ما يسمى في نظرية معنى- نص بالفواعل الدلالية Actants sémantiques حيث يتحدد الفاعل دلاليا قبل ورود الكلمة في السياق.

نشير في هذا السياق إلى تعدد أشكال المصطلح فلم يعد المصطلح وحدة إسمية بل يمكن أن يتخذ المصطلح أقساما مختلفة من الكلم ، وهذا ما تظهره نتائج التحليل.

ينظر الشكل الآتي: واجهةالبرنامج أدرجنا فيها جذر/ نحل / في البحث

يكشف التحليل أن مادة "نحل" الخاصة بعنصر الماء كمتلازمة لفظية ظهرت بالصيغ الآتية: ينحل – انحلال – منحلة حيث جاء المصطلح فعلا ومصدرا واسم فاعل.

سبب هذا التنوع يعود إلى الاعتبار المنهجي الذي انطلقنا منه وهو اعتبار النص الحاضن الأساسي للمصطلح وليس التصنيف المسبق في شكل قوائم جاهزة.

نشير أيضا في هذا السياق إلى أن تحليل المتوافقات الخاصة بكل كلمة يأخذ بعين الاعتبار قائمة المستبعدات، ففي تحليل البرنامج لكلمة الماء نجده يورد لنا في المتوافقة الأولى النتيجة الآتية :

- انحلالها في الماء فإن اختفاءها

حيث اقتصرت النتيجة على الكلمة السابقة واللاحقة بمرتبتين وهذا يعود بطبيعة الحال إلى محاولتنا إدراج كلمة في وفإن في قائمة الاستبعاد لذا أقصيت من نتائج التحليل . مواصفات مثلى للمعالجة :

بقي أن نشير إلى أننا اتفقنا مع الحاسوبي على استكمال بعض المكونات التي نراها ضرورية في برنامج مستخرج المصطلحات حيث اقترحنا عليه إضافة النتيجة الخاصة بعمليات البحث السابقة لتحديد الكلمات المرشحة كمصطلحات ثم وضعها في معجم أولي للمصطلحات وذلك بإتباع الخطوات الآتية:

توصية : العمل الجماعي المشترك بين اللساني والحاسوبي اقتصاد في الجهد والوقت ويحقق نتائج عملية مفيدة.

تؤخذ النسب التكرارية الأعلى في حدود 50% من النتائج المتحصل عليها ثم تتم عملية البحث في هذه العينة المحدودة بالكشف عن المتلازمات والمتوافقات لتحديد أيها يقبل تلازما لفظيا بالكلمات المختارة فقط أي تبقي عملية البحث محصورة بين الكلمات المختارة فقط، والنتائج التي يتم الحصول عليها ستكشف عن المصطلحات الفعلية في المجال المتخصص للنص.

الإحالات:

(1) ينظر:

Igor Melcuk :la lexicologie explicative et combinatoire

في هذا الكتاب جانب تطبيقي مهم في شرح أسس نظرية معنى- نص .

(2)-Marie -Claude L'Homme: La terminologie principes et techniques, presses de l université de Montréal, Canada,2012,P119

(3) التمثيلية représentativité : (قدرة المعطيات المنتمية إلى عينة محددة على تقديم كل الخصائص التي تتضمنها العينة الواسعة فالمثال المعتمد في دراسة ظاهرة لغوية ما يجب أن يستجيب إلى شرط التمثيلية) المعجم الموحد لمصطلحات اللسانيات مدخل رقم 1376

(4)-Marie -Claude L'Homme: La terminologie principes et techniques, presses de l université de Montréal Canada,2012,P126

(5) عبد القادر الفاسي الفهري: اللسانيات واللغة العربية: منشورات عويدات، بيروت ،1986، ص12

(6)-Marie -Claude L'Homme: La terminologie principes et techniques, P126

(7) البرنامج من وضع رقيق يونس: حاسوبي، خريج قسم الرياضيات والإعلام الآلي، جامعة زيان عاشور بالجلفة، تخصص نظم المعلومات .

(8) هنري بيجوان وفيليب تواورن : المعنى في علم المصطلحات، تر: ريتا خاطر، بيروت، 2009، ص105

(9) المرجع نفسه، ص 106

(10) ينظر في هذا الاتجاه:Alain Polgaire: La lexicologie et sémantique lexicale-

-Marie -Claude L'Homme: La terminologie principes et techniques

(11) Charles Muller: Principes et Méthodes de statistique lexicale , Edi Champion, Paris ,1992, P47.

المصادر والمراجع:

(1) عبد القادر الفاسي الفهري: اللسانيات واللغة العربية، منشورات عويدات، بيروت ،1986.

(2) هنري بيجوان وفيليب تواورن : المعنى في علم المصطلحات، تر: ريتا خاطر المنظمة العربية للترجمة، ط1، بيروت، ديسمبر 2009 .

(3) Alain Polgaire: La lexicologie et sémantique lexicale, presses de l université de Montréal Canada,

(4) Charles Muller: Principes et Méthodes de statistique lexicale , Edi Champion, Paris ,1992,.

(5)-Marie -Claude L'Homme: La terminologie principes et techniques presses de l université de Montréal Canada,2012

المعاجم والقواميس:

(1)المنظمة العربية للتربية والثقافة والعلوم، مكتب تنسيق التعريب: المعجم الموحد لمصطلحات اللسانيات انجليزي – فرنسي – عربي، مطبعة النجاح الجديدة، المغرب، 2002

الوثائق الرسمية :

وزارة التربية الوطنية : كتاب التربية العلمية والتكنولوجية للسنة الخامسة ابتدائي،الجزائر، 2016-2017