

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي و البحث العلمي
جامعة وهران – السانبة

كلبة الآداب و اللغات و الفنون
قسم الترجمة
مدرسة الدكتوراه

بحث مقدم لنبل شهادة ماجستير في الترجمة
موسوم بـ:

استثمار المدونة في تدريب المترجم

إشـراف:

الدكتور: نصر الدين خليل

الدكتورة: خيرة حمر العين

إعداد الطالبة:

فاطمة الزهراء توتاوي

السنة الجامعية

2008 * 2009

كلمة لا بد منها! ح

قد يجيء بعض من لهم حق القعود لي بالمرصاد، زبانية العلم الشداد، يقبضون علي متلبسة بجرمي فيلقون بالأصفاد، و يعلنون على رؤوس الأشهاد: خذوها و عبرة لغيرها اجعلوها، و بحرّ نار النقد و الانتقاد صلوها، إنها قد اختلست ألفاظا من هنا، و استرقت عبارات مصادرها هاهنا! ثم أتتنا مدعية أن هذا من صنع أنا!

على رسلكم! لا تقذفوني بالعار و الشنار، فليس ثمة انتحال و لا اجتزار، إنما ثمة هضم و تمثيل للأفكار، أو قل هي رسكلة لما ألفيته في الآثار... و في الرسكلة إعادة لنظم ما تنافر، و جمع ما تناثر.. و فيها شيء من التنسيق و الترميق، بل و فيها شيء من الصهر و السهر، و المعالجة و الاعتلاج، حتى بلغت بها مدارج الإنضاج و الإنتاج!

و ما أظنني قد بلغت...

و إذا كنت حقا قد "سرقت"، فقد "سرق" قبلي هومير و شكسبير و مولير ذوو الأدب الرشيق، فهل شككتكم يوما في سناهم البراق؟! و إذا كنت حقا قد "سرقت"، فلا يزال "يسرق" النحل من الزهر الرقيق، فهل شككتكم يوما في أن العسل الرقيق من صنع "سراق"؟! ليس الخلق من العدم إلا من صنع البارئ الخلاق...

و الخلق من الخلق ليس من قبيل الاختلاق، و لا يضير في شيء مادامت تجمله مكارم الأخلاق...

أما الإهداء...

فإلى الظل الأقرب إلي من ظلي،

و إلى كل من لهم فضل علي سأظل أدين لهم به إلى أن يوافيني أجلي.

مقدمة

مقدمة

شهدت الصناعة اللغوية – شأنها شأن كل النشاطات البشرية – تحولاً ملحوظاً ونقلة نوعية، و يعود الفضل في ذلك إلى ما بلغته المبتكرات التكنولوجية من تطورات وما تمخض عنها من ثورة معلوماتية. و تجلّى ذلك بخاصة في ظهور معطيات نصية محفوظة الكترونياً، قد اصطلح أهل الاختصاص على تسميتها بـ "المدونات"؛ و تلكم معطيات فتحت أمام البحث آفاقاً بلا حدود: ففي مجال اللسانيات – مثلاً – سمحت إمكانية استغلالها بظهور اختصاص جديد يدعى "اللسانيات المدونة"، استثمرت في كثير من التطبيقات اللغوية، نحو: تدريس اللغات و تعلمها، البحث التوثيقي، و هلم جرا.

و الترجمة – بحكم سنة التقاطع بين الحقول المعرفية و تداخلها – قد استفادت هي الأخرى من هذه الموجة الجديدة في البحث، بحيث يتجلى ذلكم الاستثمار و تتبدى آثاره سواء فيما تعلق بالبحث في مجال الترجمة (الترجمات باعتماد المدونات) أو التدريب عليها أو ممارستها، فكان ميلاد "المقاربة باعتماد المدونات". وهو الأمر الذي جعل منها (أي المدونات) أداة رئيسة من أدوات المترجم و وسيلة مثلى لتدريبه، يعتمد عليها الأساتذة في كبرى أقسام الترجمة ومعاهدها في العالم.

و طالما أن هذه الموضة المحمودّة قد عرفها الغرب منذ ما يربو عن عشر سنوات، نرى أنه بات لزاماً على أقسام الترجمة في العالم العربي أن تحذو حذو نظيراتها، و أن تنتهج خطة تعليمية تسعى من خلالها إلى تكوين المكونين والمتعلمين على حد سواء في هذا المجال.

و على هذا الأساس، عقدنا العزم على تناول الموضوع بالفحص و الاستقصاء – ما استطعنا في ذلك سبيلاً – قصد إبراز أهمية المدونات و فعاليتها في تدريب المترجم، ومن ثم الممارسة الترجمية مستقبلاً. و الحال هذه، جاء بحثنا موسوماً بـ "استثمار المدونة في تدريب المترجم"، منطلقين من إشكالية قوامها مجموع التساؤلات الآتية:

ما هي المدونة و ما أهميتها في مجال تدريب المترجم و الممارسة الترجمية مستقبلا؟ مامدى فعاليتها في إعانة المترجم؟ هل يمكن الاستغناء عن الأدوات التقليدية في العمل الترجمي و الاكتفاء باستخدام مدونة؟ و ما هي صور تطبيقها داخل صف الترجمة؟ الواقع أن اختيارنا لهذا الموضوع لم يكن اعتباطا، بل إنه وليد قناعتنا بالأهمية التي يكتسبها، و سعي منا لوضع لبنة في صرح تعليمية الترجمة العربية، تضمن مواكبتها لكل ما جد في المجال بما يخدمها و يضمن تقدمها على شاكلة كبرى الأقسام في العالم الغربي. من هذا المنطلق، عكفنا على الموضوع بالدرس و التمهيص طويلا، وتقليب النظر و إمعانه في جميع أوجهه و جوانبه، فتوسعت لدينا أنظار و بدت لنا أفكار لم تكن قد جرت في خاطر أول وهلة، فازدنا حبا في الاستزادة و الاستفادة، و قلنا "رب زدني علما". و الحق أننا قد أكرهنا أيما إكراه على توزيع الهيكل الخارجي الذي يؤطر الإشكالية على نحو مغاير تماما للمخطط الذي رسمنا خطوطه الأولى و الأولوية في بادئ الأمر، و لا نرى في ذلك منقصة أو عيبا، طالما أنه "لا يكتب أحد في يومه كتابا إلا قال في غده: لو غير هذا لكان أحسن، و لو زيد هذا لكان يستحسن، و لو قُدم هذا لكان أفضل، و لو ترك هذا لكان أجمل (...)" – على حد قول العماد الأصفهاني.

و الحال هذه، عمدنا إلى تقسيم البحث إلى ثلاثة فصول:

الفصل الأول: المدونة.

الفصل الثاني: واقع البحث باعتماد المدونة عند الغرب و العرب.

الفصل الثالث: المدونة في الترجمة تنظيرا و تطبيقا.

أما الفصل الأول، فقد جعلناه عتبة البحث التي لا مناص من اجتيازها، بحيث حاولنا طرح جملة من المفاهيم و تقريبها إلى ذهن القارئ – قدر المستطاع – وهي مفاهيم تتعلق أساسا بماهية المدونة، و كيفية صناعتها و أنواعها، لنقف بعدها عند "الساكنات المدونة" من حيث المصطلح و المنطلقات التاريخية، و كذا المنهج والأدوات الإجرائية.

أما في الفصل الثاني، فقد سلطنا الضوء على واقع استثمار المدونات في بعض الأبحاث اللغوية عند الغرب و العرب. و نحن و إن كنا نسلم عن يقين بأن هذا الفصل يستوجب بحثا مستقلا، إلا أننا نرجو ألا يقذف في روع القارئ أن ذلكم الزخم من الأمثلة

التي ضربناها، إنما كان استرسالا عابثا و حشوا زائدا أملاه انسياقنا وراء البحث في الموضوع و اندفاعنا في المطالعة؛ بل إننا نبتغي من وراء ذلك دعوة القارئ إلى النظر فيما بلغته تلك الأبحاث عند الغرب و مقارنتها بما يعد غثا هزيلا عند نظرائهم العرب، ذلك أنهم ما قدروا المدونات حق قدرها من حيث فوائدها و خدمتها للغتنا؛ ثم إننا نسعى إلى إبراز كيفية اكتساح لسانيات المدونة شتى الحقول المعرفية كتمهيد للفصل الموالي.

في الفصل الثالث، يلقى القارئ عنصرين: أولهما جعلناه وقفا على أصول البحث الترجمي باستخدام المدونات و أهم ما توصل إليه الدارسون بهذا الصدد، والعنصر الثاني خصصناه لما نعهده صلب بحثنا كله و مدار الرحي فيه، فوسمناه بـ "تجربة الترجمة باعتماد المدونة"، إذ تناولنا فيه بالدرس و التحليل نتائج ورشات اختبار الطلبة من خلال تطبيقات بعض صور استخدام المدونة في الترجمة.

إننا، ونحن نأتي على ختام هذه الأسطار، على يقين جازم من أن البحث تكاد لا تسلم كل صفحة فيه من هفوة أو زلة تنم عن قصورنا و قصر باعنا في الموضوع، و لا نخال ذلك عيبا نحاسب عليه أو جرما نؤاخذ به، بل إنه سمة كل "باحث" "غُرّ" يدرج أولى درجاته في عالم البحث. وفوق ذلك، إن ما نرجو أن يغفر زلتنا و يصفح عن قصورنا هو طبيعة دراستنا التي تتسم بالأصالة و الجدة، من حيث إنها دراسة لم يطرقها أي دارس قبلنا بقسم الترجمة على مستوى الدول العربية؛ لذلك نأمل من خلال بحثنا أن يكون للجامعة الجزائرية قصب السبق و الريادة في تناول الموضوع، و إن كان للجامعة الكويتية سابق تجربة في تجسيده على أرض الواقع دون غيرها، بحيث تُعتمد مدونة متوازية عربية-انجليزية في تدريس الترجمة و المعجماتية، نصوصها مأخوذة عن منشورات المجلس الكويتي الوطني.

أما الحقيقة التي تجعل الشبان شيئا فمفادها أننا نكاد لا نعثر على مؤلف واحد باللغة العربية يتناول المقاربة باعتماد المدونات في أي مجال من المجالات، اللهم إلا ما بلغنا عن دراسة لغوية في تأصيل الألفاظ لماهر عيسى حبيب؛ إذ و في بحث له مقدم إلى ندوة تطوير المناهج و الاختصاصات الجامعية (جامعة حلب، 30-31 أيار 2007) موسوم بـ: "الارتقاء بتعليم العربية التقليدي إلى المستوى الإلكتروني"، نجد في الصفحة الخامسة منه

ضمن الإحالة رقم 16 ما نصه بالحرف الواحد: "قمنا بإجراء بحث أصيل اعتمادا على المدونة الحاسوبية و إمكانات الحاسب الآلي، ينظر بحثنا في الأصيل و الدخيل- دراسة لغوية تأصيلية في مجلة التعريب (الصادرة عن المركز العربي للتعريب و الترجمة والتأليف و النشر بدمشق عدد 31 كانون الأول 2006)، ص 29-79؛" أو ما وقع بين أيدينا من مقال لسلوى حمادة، عنوانه: "المدونات النصية، و دور اللغة العربية في التعامل معها"، و قد ذيلته بملاحظة قالت فيها ما نصه: "طُعّم هذا البحث بالكثير من الأجزاء المترجمة لعدم وجود كتابات باللغة العربية في المجال. و قد حاولت أن أبسط المعاني المقصودة و أدمج بأمثلة من اللغة العربية و إضافات لتوضيح المعاني و إجلالها مما قد يخل بصحة التعريفات الأصلية، و لكنني وجدت من الأمانة الإشارة للمصدر. أي أن الرجوع للمصدر لا يعني وجود نفس المعنى، و إنما معاني متقاربة أو متشابهة." و الواقع أننا ما كنا لنفقه حرفا مما ورد فيه، لو لم نرجع إلى المصادر و عيون الأخبار في هذا الموضوع باللغة الإنجليزية.

ختام هذه الأسطر أن الحمد لله الذي علم الإنسان ما لم يكن يعلم، و الشكر لأستاذي المشرقيّن: الأستاذ نصر الدين خليل و الأستاذة خيرة حمر العين.

و ما توفيقني إلا بالله العلي العظيم.

فاطمة الزهراء توتاي

الفصل الأول:

المدونة

- I- ماهية المدونة.
- II- صناعة المدونة.
- III- أنواع المدونة.
- IV- لسانيات المدونة:

المصطلح و المنطقات التاريخية.

المنهج و الأدوات الإجرائية.

إن المتأمل في واقع البحث في شتى الحقول المعرفية يدرك إدراكا لا يشوبه شك، أنه يسير وفق حركية لم يعهدها إلا منذ عهد غير بعيد؛ أي منذ ظهور الحواسيب الالكترونية و نظم الذكاء الاصطناعي، و اقتحامها لجل المعارف على اختلاف فروعها و تخصصاتها.

و واقع الدراسات اللغوية لا يشذ عن هذه الحركية الجديدة؛ إذ إن استثمار الحواسيب قد أعطى لتلك الدراسات بعدا يتماشى و المعطيات المعاصرة، من حيث توظيف التكنولوجيا توظيفا يسهم بحظ وافر في تطوير كثير من جوانبها و إثرائها، سواء الجوانب المنهجية منها أو المفاهيمية أو الإجرائية. و أضحت الأبحاث في مجال العلوم اللغوية – بكافة أنواعها – غثة هزيلة ما لم تتخذ التقنيات الحاسوبية سندا لها في أعمالها. و الراجح أن أكثر التطبيقات اللغوية تأثرا بالتكنولوجيا الحاسوبية هو ما كان على شاكلة: تعلم اللغات الأجنبية و الترجمة الآلية و كذا الصناعة المعجمية. و باتت الحاجة ملحة إلى توفير قواعد بيانات نصية *Textual databases* تمثل الواقع اللغوي، أي الاستخدام الفعلي للغة *Real language use*، – على حد تعبير بوب غدوين (Bob Godwin) –¹ و تخدم تلك الأعمال و تدعمها؛ مما حدا بالعاملين في المجال إلى إنشاء متون أو مدونات سرعان ما أثبتت فعاليتها من حيث إنها تخدم مبتغاهم، و ترقى بالأبحاث اللسانية رقيا لا يستهان به، لاسيما فيما يتعلق بـ "الدقة و السرعة و حجم النتائج و شمولها في الاستخدام الفعلي للغة".²

I - ماهية المدونة:

كلمة *Corpus* كلمة انجليزية مشتقة من الكلمة اللاتينية *Corpse*، بمعنى *Body* (أي "الجسم")، و جمعها في اللغتين *Corpora*، و يمكن أن تجمع في الانجليزية بصيغة *Corpuses*. و المدونة في أبسط تعريف لها حسب معجم ديفيد كريستال (*David Crystal*) عبارة عن "مجموعة من المعطيات اللسانية، سواء كانت نصوصا مكتوبة أم تحويلا من تسجيلات صوتية إلى مادة مكتوبة، و التي يمكن استخدامها كمنطلق لوصف اللغة أو كطريقة للتحقق من الفرضيات اللغوية و إثباتها." [الترجمة لنا]

"A collection of linguistic data, either written texts or a transcription of recorded speech, which can be used as starting-point of linguistic description or as a means of verifying hypotheses about a language".³

الظاهر أن هذا التعريف يحدد المدونة في مجموعة من النصوص ذات الحامل اللساني المكتوب أو المنطوق، تُجمَع بغرض دراسة الظواهر اللغوية. بيد أننا نجد تعريفا آخر في معجم أكسفورد (*Oxford*) للغة الانجليزية، نصه:

¹ - Bob, Godwin-Jones (2001) Emerging Technologies: Tools and Trends in Corpora Use for Teaching and Learning. *Language Learning and Technology* 5(3), September 2001, p.7.

² - سلوى حمادة، المدونات النصية و دور اللغة العربية في التعامل معها. مجلة وانا للترجمة و اللغات، السنة الأولى، العدد4، خريف 2007 . <http://www.watajournal.com/mag4/index.html>

³ - David, Crystal (1991) A Dictionary of Linguistics and Phonetics. Blackwell, 3rd edition. Viewable online at : <http://www.engl.polyu.edu.hk/corpuslinguist/corpus.htm>

"[...]a body of texts, utterances, or other specimen considered more or less representative of language, and usually stored as an electronic database."¹

"[المدونة] مجموع النصوص أو التلغظات أو غيرها من العينات التي تعتبر ممثلة للغة ما – بقدر دق أو جل –، و غالبا ما تحفظ ضمن شكل الكتروني." [الترجمة لنا]

و هذا تعريف يتوافق إلى درجة ما مع التعريف الذي اصطلح عليه العاملون في مجال معالجة اللغات الطبيعية TAL/NLP،* و نصه:

"[المدونة] بناء كبير من النصوص الالكترونية يستخدم في التحليل

الإحصائي اللغوي، و يتحقق من تكرار أو صحة القواعد اللغوية."²

بمعنى أن المدونة كم هائل من النصوص المقروءة آليا، سواء في شكلها المكتوب أو المنطوق، تتجلى من خلالها كيفية استعمال اللغة في سياقات طبيعية.³

و جدير بالذكر أن ما يميز المدونات المقروءة آليا (Machine readable corpora) أو التي في شكلها الالكتروني عن المدونات الورقية – التي كانت موضوعة الأبحاث لسنوات قد خلت – هو سهولة التعامل معها، و الاستفادة منها بسرعة كبيرة و دقة متناهية بفضل برمجيات خاصة (Software).*

"Machine readable corpora have a number of advantages over other forms of storage. Firstly, and most importantly, machine readable corpora may be searched and manipulated in ways which are simply not possible with the other formats. Secondly, machine readable corpora can be swiftly and easily enriched with additional information".⁴

"تمتاز المدونات المقروءة آليا عن غيرها ببضع فوائد، من حيث إنها تمكن الباحث من فحصها و التعامل معها وفق طرق بسيطة و في متناول يده، بخلاف المدونات المحفوظة ضمن أشكال أخرى. ثم

¹ - McArthur & McArthur (1992) *The Oxford Companion to the English Language*. Viewable online at: <http://www.engl.polyu.edu.hk/corpuslinguist/corpus.htm>

* - Natural Language Processing (NLP)/Le Traitement Automatique des Langues (TAL) est l'application de programmes et techniques informatiques à tous les aspects du langage humain. Parmi les applications les plus connues, on peut citer : La traduction automatique (historiquement la première application dès les années 1950) ; la correction orthographique ; la recherche d'information ; le résumé de texte ; la synthèse de la parole... Voir ATALA Pedie : Association pour le Traitement Automatique des Langues. <http://www.atala.org/AtalaPedie/>

² - سلوى، حمادة، مر. س، ص. 2.

³ - المدونة كما أثبتتها المعجم الموحد لمصطلحات اللسانيات (مكتب التنسيق و التعريب- الرباط 2002) هي: "ما يشكل الرصيد اللغوي، أو مجموع المعطيات اللغوية يخضعها الباحث للتحليل و الدرس". (ماهر عيسى حبيب، الارتقاء بتعليم العربية التقليدي إلى المستوى الإلكتروني. بحث مقدم إلى ندوة تطوير المناهج و الاختصاصات الجامعية – جامعة حلب 30-31 أيار 2007، ص. 5، الإحالة 15)؛ و هو تعريف نراه غير شامل.

* - تلفت انتباه القارئ إلى أن ما نعنيه بمصطلح "المدونة" في ما سيلي هو المدونة الحاسوبية أي الإلكترونية (Computer corpora)، ما لم نشر إلى خلاف ذلك.

⁴ - Tony, McEnery & Andrew Wilson (1993) *Corpora and Translation: Uses and future prospects*. UCREL, Technical Papers vol.2, March 1993.

إنها [أي المدونات الإلكترونية] قابلة للإثراء ببيانات إضافية بصورة سريعة و سهلة. " [الترجمة لنا]

و في هذا المقام، يستوقفنا سؤال لا بد منه: هل يمكن اعتبار جميع النصوص الإلكترونية مدونة؟

الواقع أن بعض المجموعات من النصوص الإلكترونية لا تمثل مدونة بخلاف الاعتقاد السائد بخصوصها؛ و تفسير ذلك، أن دراسة اللغة هي الدافع الرئيس إلى صناعة المدونات، أما سواها (أي باقي المجموعات) فتصمم لأغراض و أهداف أخرى.¹ و مثال ذلك:

أ- النص الواحد:

من المغالطة اعتبار النص الواحد مدونة² لفرق جوهري بينهما؛ إذ عندما يقوم الباحث بتفحص مقطع لغوي صغير قد استنبطه من نص ما، فإنه يسعى إلى معرفة إسهامه الخاص في معنى النص ككل، أما أن يتفحص المقطع اللغوي نفسه مستنبطاً من مدونة، فإنه يسعى إلى إدراك مدى إسهامه في التعميم (Generalization) الذي يراد من ورائه توضيح طبيعة اللغة ككل و بنيتها بعيداً عن فردية أو وحدانية التلفظ.

ب- الأرشفة:

يكمن وجه الفرق بين الأرشفة و المدونة في الغرض الذي جمعت لأجله المادة النصية؛ و هو فرق ينجر عنه اختلاف أو تباين في أولويات جمع المعلومات بخصوص كل نص. و كما يقول جون سينكلير (John Sinclair):

"[...] archive refers to sets of texts that do not need to be selected, or do not need to be ordered, or the selection and/or ordering do not need to be on linguistic criteria."³

"[إن تجميع النصوص في شكل أرشفة] لا يقتضي انتقاء تلك النصوص أو تنظيمها؛* أضف إلى ذلك، لا يخضع انتقاؤها أو تنظيمها لمعايير لسانية". [الترجمة لنا]

ج- الشواهد اللغوية (Citations):

يمثل الشاهد اللغوي اقتباساً نصياً وجيزاً يتضمن الكلمة أو العبارة التي تعد سبب انتقائه دون غيره؛ و تلك الشواهد لا تعد مدونة نظراً لغياب الاتصال النصي الذي يميز الأمثلة المأخوذة عن المدونة.

¹ - Cf. John, Sinclair (2005) Corpus and Text-Basic Principles. In: Martin, Wynne (ed.) *Developing Linguistic Corpora: a Guide to Good Practice*. Oxbow Books. Available online at: <http://www.ahds.ac.uk/creating/guides/linguistic-corpora/chapter1.htm>

² - نظرياً، يجوز أن يطلق مصطلح "مدونة" على أكثر من نص، ينظر: سلوى، حمادة، مر. س.، ص 2.

³ - John, Sinclair (1996a) EAGLES Preliminary Recommendations on Corpus Typology. Technical report, EAGLES (Expert Advisory Group on Language Engineering Standards).

* و في ذلك يقول بيل دود : Bill Dodd

"A random collection of texts is sometimes referred to as a text archive".

Cf. Bill, Dodd (2000) Introduction. The Relevance of Corpora to German Studies. In: Bill, Dodd (ed.) *Working with German Corpora*. Birmingham: Birmingham University Press, pp. 69-95.

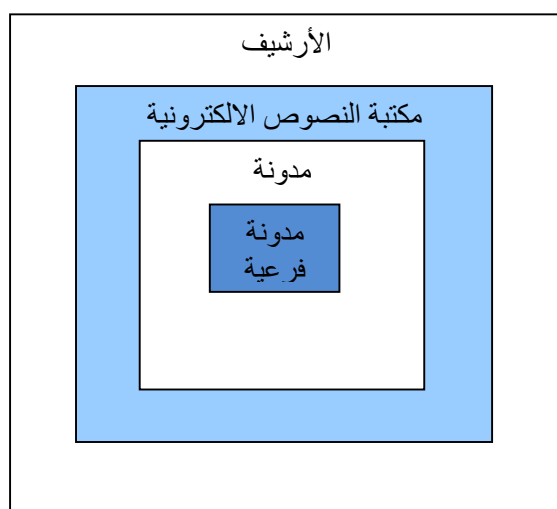
د- المقتبسات (Quotations):

يتم انتقاؤها تبعاً لمعايير داخلية (Internal Criteria) دونما تدخل الآلة، و لا تعد مدونة للأسباب نفسها التي لا تجعل من الشواهد مدونة.¹

هـ- نصوص الشبكة العنكبوتية (The World Wide Web):

بالنظر إلى عدم محدوديتها و تغيرها باستمرار، و أيضاً لأنها لم تصمم وفق منظور لسانی، إذ تعج بكثير من الصور و الأشكال و ما شاكلها. بيد أن تلك النصوص تمثل مصدراً رئيساً، و مورداً خصباً يعتمد عليه الباحثون في بناء مدوناتهم، و سنتعرض لذلك بشيء من التفصيل لاحقاً.²

و يمثل المخطط أدناه توزيعاً لكبرى المجموعات النصية، من اقتراح أتكين و كلير و أسلر (Atkins, Clear & Osler).³



الشكل (01): توزيع المجموعات النصية.

نخلص مما سلف إلى أن ما يصطلح على تسميته "مدونة" لا يعني البتة مجموعة نصية يتم جمعها بصفة عشوائية، بل إنها بناء كبير من العينات النصية الحقيقية والالكترونية، تخضع لجملة من المعايير التي ينبغي مراعاتها أثناء التجميع (Compilation).

و تلكم الخصائص الأربع هي ما يميز المدونات عن بقية المجموعات النصية الالكترونية. وإذا أمعنا النظر و أرجعنا البصر، ألفينا أن ما كان من خاصتي الحجم و الشكل، فهذان أمران لا ينتطح فيهما عنزان؛ و أما ما يقصد بالنص الحقيقي، فمعناه أن النص مثال عن اللغة و هي تتجسد في الواقع الملموس و تتكون في سياق التواصل (أي الكلام). و بقيت واحدة و هي المعايير نرى أنه لا بد من أن نقف عندها فيما يلي وقفة متأنية.

¹ - جدير بالملاحظة أن الشواهد اللغوية و المقتبسات تحاكي نوعاً ما الفهارس الآلية (Concordances) لا المدونة؛ إذ هي عبارة عن اقتباسات وجيزة من المدونة ينتقيها البرنامج الحاسوبي. ينظر العنصر IV من هذا الفصل.

² - ينظر الفقرة 2.2.II من هذا الفصل.

³ - W. John, MacMullen (2003) Requirements Definition and Design Criteria for Text Corpora in Information Science. SILS Technical Report TR 2003-03. Chapel Hill, NC: University of North Carolina, School of Information and Library Science, Technical Report Series.

"The corpus is a remarkable thing, not so much because it is a collection of language text, but because of the properties that it acquires if it is well-designed and carefully-constructed." John, Sinclair (2005) Op.Cit.

II- صناعة المدونة:

يقول فدير يكو زانتان (Federico Zanettin) في تعريفه للمدونة:
"A 'corpus' is a collection of **electronic texts** assembled according to explicit **design criteria** which usually aim at **representing** a larger textual population." ¹ [Emphasis added]

"المدونة كمّ من النصوص الالكترونية تجمع وفق معايير واضحة، و تهدف غالبا إلى تمثيل مجموعة نصية أوسع." [الترجمة لنا]
إن أدنى تأمل في هذه المقولة، يجعلنا نلمس تأكيدا على أن ما يصنع من النصوص المخزنة الكترونيا "مدونة" هو جملة المعايير و الأسس التي تحكم مادتها و تجعلها عينة ممثلة لمجموعة نصية أكبر منها حجما و أكثر تشعبا.
و علاوة على ذلك، فإن تلكم الأسس و المعايير هي الأخرى تحددها الغاية المرجوة من وراء إنشاء المدونة:

"[...] One of the key factors in design and use, [...], is the purpose for which the corpus has been constructed ."²

كأن تكون المدونة منطلقا للصناعة المعجمية، أو لصنع جدول بتكرارات الكلمات، أو لدراسة بعض الظواهر اللغوية، أو لدراسة اللغة عند كاتب/متكلم ما، أو خلال فترة زمنية معينة، أو لتدريب نظام لمعالجة اللغة الطبيعية (to train an NLP system)، أو كأداة في تدريس لغة لغير الناطقين بها، أو لدراسة ميكانيزمات اكتساب اللغة،... و على هذا النحو، فإن مادة المدونة التي يعتمدها الباحث في صناعة معجم لغوي – على سبيل المثال – هي بخلاف مادة مدونة قد يصنعها لغرض حصر مجموعة من الأنماط التركيبية أو البنيوية للغة. و في هذا الصدد، تقول فيرونيك ماغري- مورغ (Véronique Magri- Mourgues):

"Le corpus est un objet empirique et structuré selon les enjeux et les objectifs de la recherche. Il est par conséquent toujours

¹ - Federico, Zanettin (2002a) Corpora in translation practice. *Proceedings from First International Workshop on Language Resources (LR) for Translation Work and Research*, Gran Canaria, 28 May 2002.

* - بالنسبة لخاصية التمثيل، ينظر الفقرة 1.2.II من هذا الفصل.

² - W. J. MacMullen (2003) Op. Cit.

*contingent, déterminé par l'application que l'on veut en faire [...].*¹

"إن المدونة مادة تجريبية يتوافق هيكلها و رهانات البحث و كذا أهدافه؛ لذا تراها دائما محدودة بما يتفق و نوع التطبيق المراد إنجازه باعتمادها [...]." [الترجمة لنا]

إن ما سلف من معطيات يزيدنا قناعة من أن صناعة المدونة صَنَعَةٌ ليس يقدم عليها إلا من كان ملما بأبجدياتها؛ إذ حري بالباحث أن يضع خطة محكمة، و يجيب عن بعض التساؤلات بخصوص حجم المدونة، و نوع مادتها النصية، و خاصية التمثيل و التوازن (أي التنوع على مستوى النصوص و مصادرها، و مؤلفيها،...) دونما أن يغيب عنه أن تقرير ما تتضمنه المدونة وليد الأهداف المقصودة و الغايات المنشودة من وراء إنشائها. ثم إن جملة التساؤلات تلك، هي التي توطر الخطوات الأولية الثلاث في الصناعة المدونانية: تصميم المدونة و تجميع النصوص و إعدادها.

II. 1- تصميم المدونة (Designing the corpus):

تتضمن الخطوة الأولى عددا من الخيارات التي تتاح للباحث قصد تيسير عملية تصميم المدونة. و تتجلى تلكم الخيارات في مجمل الأسئلة التالية:

أ- ما نوع المدونة؟

- هل ينبغي أن تكون المدونة سانكرونية (آنية)، تمثل اللغة خلال فترة زمنية محددة (على غرار مدونة العربية الفصحى 2004 بجامعة مانشستر، محددة من الفترة قبل الإسلام حتى القرن الحادي عشر)؟

أم أنها مدونة دياكرونية شاملة، تمثل اللغة ضمن فترات مختلفة؟

- هل ينبغي أن تكون المدونة متوازنة (Balanced)، بمعنى أنها تمثل اللغة ككل؟

أم أنها مدونة مختصة، كأن تتناول خاصية لغوية معينة منظومة/مكتوبة، أو لهجة خاصة أو سجلا معيناً، أو ضرباً من النصوص كالنص الصحفي مثلا، أو لغة خاصة بمتكلم/كاتب ما؟

- هل ينبغي أن تكون المدونة مفتوحة (Ouvvert)، بحيث تضاف إليها نصوص جديدة في كل مرة؟

أم أنها مغلقة (Fermé)، أي تبقى ثابتة؟²

ب- ما هو حجم المدونة المعتمدة؟

الواقع أنه ليس ثمة اتفاق حول الحجم على وجه التحديد، لكن الملاحظ في كثير من كبرى المدونات التي صممت حتى الآن، أن عيناتها النصية غالبا ما تتضمن زهاء 2000-5000 كلمة أو 20000 كلمة، بغية الحصول على نتائج أدق و أصدق. ثم إنه من المفترض أن تكون المدونة العامة أكبر حجما من المدونة المتخصصة؛ كما أن الحجم

¹ - Véronique, Magri –Mourgues (2006) Corpus et Stylistique. Corpus n°5, Corpus et Stylistique – décembre 2006.

² - For more explanation, cf. Elizabeth, Marshman (2003) Construction et gestion des corpus: Résumé et essai d'uniformisations du processus de la terminologie. Defining the norms for the ELECTIK research group. [En ligne] <http://www.olst.umontreal.ca/pdf/terminotique/corpusenttermino.pdf> (page consultée le 24/07/2007). p2

غالبا ما يكون محددًا بالنظر إلى وفرة المادة النصية، و إمكانية الحصول عليها (حقوق التأليف).¹

فضلا على ذلك، يتوقف الحجم على الظاهرة المراد تناولها بالدرس والتحليل؛ إذ ينبغي أن تشتمل المدونة على عينة تمثل الظاهرة، و تفي بالغرض – على حد تعبير ستيفانوفيتش (Stefanowitch) –:

"[...] for corpus must contain a sufficient number of examples of the phenomenon you are planning to investigate; how many examples are sufficient depends on the phenomenon".²

فإن كنت تنوي دراسة البنى المورفولوجية أو التركيبية العامة، أمكنك ذلك باعتماد مدونة بحجم مليون كلمة؛ أما إذا كانت دراستك تنصب حول الوحدات المعجمية و البنى القواعدية الخاصة و كذا التعابير الاصطلاحية (Idioms)، فحري بك إنشاء مدونة أكبر. و جدير بالذكر أن هذين الخيارين أو المعيارين (نوع المدونة و الحجم) يرتبطان أساسا ارتباطا وثيقا بطبيعة مدونة الباحث؛ أي أن اعتماده خيارا دون آخر ليس اعتباطا، بل إنه يتوقف لا محالة على طبيعة مشروع بحثه.

ج- ما مدى ملائمة المدونة للمشروع ؟

في أوائل عمر البحث باعتماد المدونة، كان الحجم أشد ما يعنى به الباحثون، ثم سرعان ما تحول مدار اهتمامهم من حجم المدونة (Corpus Size) إلى مدى ملائمتها (Corpus Adequacy) لغرض البحث.³ و قد طرأ هذا التحول لاسيما بعد الفيض الواسع الذي شهدته النصوص الالكترونية.

و لما كانت نوعية نتائج البحث المُتَحَصَّل عليها باستخدام المدونة تتوقف أساسا على نوعية المدونة نفسها، فإنه لزام على الباحث أن يراعي تحقيق مجموعة من المسلمات، تحصرها مارشمان في ثلاث نقاط، هي:

- أن يكون مجال نصوص المدونة مُحدَّدا و محدودا.
- أن تكون النصوص على قدر كاف من التمثيل لدعم النتائج و تأكيدها.
- أن تكون المدونة طيعة عند استثمارها فتخدم غرض البحث، و لا يتأتى ذلك إلا إذا كان بناؤها محكما، و محتواها ملائما.⁴

د- هل يمكن استعمال تلك المدونة مجددا؟

كثيرا ما يؤكد أهل الاختصاص على أهمية، بل و ضرورة أن تكون المدونة قابلة للاستعمال ضمن مشاريع أخرى (La réutilisabilité)؛ سواء كانت تلك المشاريع شبيهة بالمشروع الأصل الذي صممت لأجله المدونة أول مرة، أم أنها تختلف عنه. و في ذلك تقول إليزابيث مارشمان:

¹ - Cf. N., Oostdijk (1988) A Corpus for studying linguistic variation. ICAME Journal 12, p.4.

² - Anatol, Stefanowitsch (2003a) Corpus compilation. Corpus linguistics handouts. Viewable online at : <http://www-user.uni-bremen.de/~anatol/materials.html>

³ - Krista, Varantola (2002) Disposable corpora as intelligent tools in translation. In Tagnin, S.E.O. (Org.) Cadernos de tradução: Corpora e Tradução. Florianópolis: NUT, 2002, vol.1, n°9, p.174.

⁴ - Elizabeth, Marshman (2003) Op. Cit., p. 3.

" Il doit être possible d'adapter le corpus à des besoins de projets autres que celui pour lequel le corpus a été créé. Cette nouvelle application pourrait être similaire ou complètement différent [sic] de l'application originale."¹

و لا شك في أن ذلك يتوافق و المنطق، لاسيما بالنظر إلى استثمار الوقت و الجهد الضروريين لبناء مدونة أخرى.

II. 2- تجميع النصوص (Obtaining the texts):

II. 1-2 الانتقاء:

بعد أن يتبنى الباحث ما يتسنى له من الخيارات أعلاه – بحكم طبيعة العمل–، يتجه إلى ثاني خطوة يسعى من خلالها إلى تجميع المعطيات النصية، التي تتوافق و تلك الخيارات، واضعا نصب عينيه أن المدونة ليست مجرد أرشيف من النصوص، بل إنها مجموعة نصية تخضع لمعيار الانتقائية. تماما كما أن عملية الانتقاء – في حد ذاتها – تقوم على معايير تتعلق أساسا بخاصة التمثيل و التنوع و الثراء و التوازن و كذا الحجم؛ و يفصلها جون سنكلير (John Sinclair) في مجمل النقاط التالية:

■ مراعاة الوظيفة التواصلية: على الباحث أن يتأكد من أن مجموعة النصوص التي انتقاها هي من صنع كاتب/متكلم أصلي (Native speaker/writer)، عالم بمجال اختصاصه؛ مما يعني أن انتخاب العينات النصية لا يتم بالنظر إلى لغتها في حد ذاتها، بل وفق وظيفتها التواصلية، في حدود المجتمع اللغوي الذي تترعرع في كنفه:

" The contents of a corpus should be selected without regard for the language they contain, but according to their **communicative function** in the community in which they arise."² [Emphasis added]

و على هذا الأساس، يصنف جون سنكلير المعايير المعتمدة في انتقاء المادة النصية إلى صنفين:

أ- صنف أول يتعلق بالمعايير الخارجية (External criteria)، و هي وليدة تفحص الوظيفة التواصلية.

ب- و صنف ثان يتعلق بالمعايير الداخلية (Internal criteria)، و هي تعكس تفاصيل لغة العينات النصية؛ هذا من جهة.

■ أما من جهة أخرى، فلا ينبغي على الباحث أن يختزل المادة المجموعة في أعمال كاتب/متكلم بعينه، بل حري به أن يوزعها على عدد من الكتاب/المتكلمين، كي لا يقع العمل في شرك الأسلوب "الشخصي" للكاتب/المتكلم الواحد؛ و هو بذلك (أي الباحث)

¹ - Ibid.

² - Cf. J. Sinclair (2005) Op. Cit.

يتحرى سمة الموضوعية و طابع الشمولية ما أمكنه إلى ذلك سبيلا، حرصا منه على إضفاء صبغة الدقة على نتائج بحثه.¹

■ فضلا على ذلك، ينبغي أن تتضمن مادة التحليل نصوصا بأكملها، بدل إدراج نصوص مبتورة، مخافة أن تسقط بعض الأساليب اللغوية و أدوات التعبير، أو المصطلحات والمفاهيم التي لا تظهر إلا في الجزء المبتور؛ ففي المطويات السياحية –على سبيل المثال – غالبا ما نلغى لغة التأثير و الإقناع في مطلع المطويات، ثم نجد الوظيفة الإخبارية فيما تلاه.

■ و في سياق آخر، ليس للباحث أن يدرج ضمن مدونته نصوصا مستقاة عن الكتب والمجلات و المذكرات و ما شاكلها من الكتابات، أو عن البرامج الحوارية و الثقافية والنشرات الإخبارية...، ما لم يطلب إنذا من الناشر، و ذلك مراعاة لحقوق التأليف.

■ و لما كان التطور سمة طبيعية و مطردة تغلب على اللغات، كان لزاما على الباحث أن يراعي شرط "الجدة" أثناء عملية انتقاء النصوص؛ و في ذلك سعي إلى رصد جل المستجدات اللغوية و المفاهيمية التي يمكن أن تطرأ.

■ و نتيجة للخاصة ذاتها (أي التطور المطرد)، تبقى اللغات الطبيعية ظاهرة من الظواهر اللامحدودة؛ و بناءً عليه، فإنه لا توجد مدونة مهما كان حجمها، و درجة الحرص و الدقة في تصميمها و الإحكام في بنائها، يمكن أن تتضمن الخصائص نفسها التي تتضمنها اللغة ذاتها.

و لما كانت الدراسات حول تلكم الظواهر اللامحدودة تتم بانتخاب عينات، يرى الدارسون في حقل اللغات أن لا ضرر و لا ضرار في اعتماد الطريقة ذاتها، شريطة أن تكون تلك العينات ممثلة قدر المستطاع للغة قيد الفحص – على حد تعبير سنكلير:-

" *Corpus builders should strive to make their corpus as representative as possible of the language from which it is chosen*".² [Emphasis added]

و عليه، يتضح أن العنصر الرئيس الذي يتم في حدوده تجميع المدونات على اختلاف صنوفها، هو أن تمثل مادتها النصية المنتخبة كافة الجوانب اللسانية للغة المراد التعامل معها، فتغطي قدرا من المجالات و السجلات و الوضعيات التواصلية.³ و في ذلك يقول الأنصاري ما نصه:

" *The appropriate design for corpus [...] depends upon what it is meant to represent. The representativeness of the corpus, in return, determines the kinds of research questions that can be*

¹ - ينظر: سلوى، حمادة (2007) مر. س.

² - J. Sinclair (2005) Op. Cit.

³ - ينظر العنصر VI من هذا الفصل، بحيث تمثل الأشكال تصاميم المدونات الثلاث الشهيرة في اللغة الانجليزية، و قد اعتمدت عدة متغيرات في ذلك، بحكم أنها عوامل تحكم التغير السياقي للغة، و هي من قبيل: المجال (كالسياسة، الأعمال...)، وضعية الكلام (رسمي، غير رسمي...)، العلاقة بين (المتحدث / الكاتب) و (المستمع / القارئ)، واسطة التواصل (مكتوب، منطوق...)، المنطقة (أي اللهجات)، المتحدث/الكاتب (جنسه، سنه، مستواه الاجتماعي، الثقافي...)، و قس على ذلك.

addressed and the generalizability of the results of the research.”¹

"إن التصميم الأمثل للمدونة [...] وَقَفَّ على ما يراد تمثيله من خلالها. و خاصَّة التمثيل تلك بدورها، تحدد نوع الأسئلة محط البحث، وكذا قابلية النتائج للتعميم." [الترجمة لنا]
بيد أن خاصة التمثيل لا تتحقق إلا باعتبار الآتي:
أ- تحديد أيِّ المعايير – أدناه – سَتُوظَّف في بناء المدونة، و جعلها منطلقاً لتصميم هيكل عناصرها الرئيسية:

- 1- طبيعة النص (Text mode): و يعنى به الشكل الذي يتمظهر من خلاله النص، أو نمط الحامل اللساني (كأن يكون مكتوباً أو منطوقاً).*
 - 2- مصدر النص (Text source): ** كأن يكون مأخوذاً عن كتاب، جريدة، رسالة، أو رواية...، إن كان الحامل مكتوباً.
 - 3- حقل النص (Text domain) أي مجاله، كأن يكون أكاديمياً أو عاماً.
 - 4- اللغة أو ضروب اللغة التي تتضمنها المدونة.
 - 5- موقعة النصوص (Text location)، و يتجلى ذلك مثلاً مع اللغة الانجليزية: لغة المملكة المتحدة، أستراليا.
 - 6- الفترة الزمنية للنصوص أو تاريخها.²
- و لتقريب هذه النقط إلى ذهن القارئ، لا ضير في أن نضرب مثلاً لذلك من خلال ميكاز (MICASE)، مدونة ميتشيغن للغة الانجليزية الأكاديمية المنطوقة (Michigan Corpus of Academic Spoken English)،³ وهي مدونة قوامها تسجيلات صوتية أجريت في مجمع آن أربور (Ann Arbor Campus) بجامعة ميتشيغن سنة 1997؛ و عليه، فإن نوع لغة المدونة هي اللغة الانجليزية الأمريكية (American English).⁴
- و يؤكد سنكلير في سياق آخر على ضرورة أن يختار المصمم من تلك المعايير ما تيسر له، و ما ييسر عليه عمله (أي انتقاءه للنصوص). كما يلح على ضرورة أن تتسم تلك المعايير بالبساطة، قصد ضمان هامش أصغر من الأخطاء في حال وقوعها، لاسيما إذا كانت تلك المعايير على درجة من التعقيد أو التفاوت فيما بينها، إذ يجب تجاوزها.

¹ - Sameh, Al-Ansary et al. (2007) Building an International Corpus of Arabic (ICA): Progress of Compilation Stage. 7th International Conference on Language Engineering, Cairo- Egypt, 5-6 December 2007, p. 1.

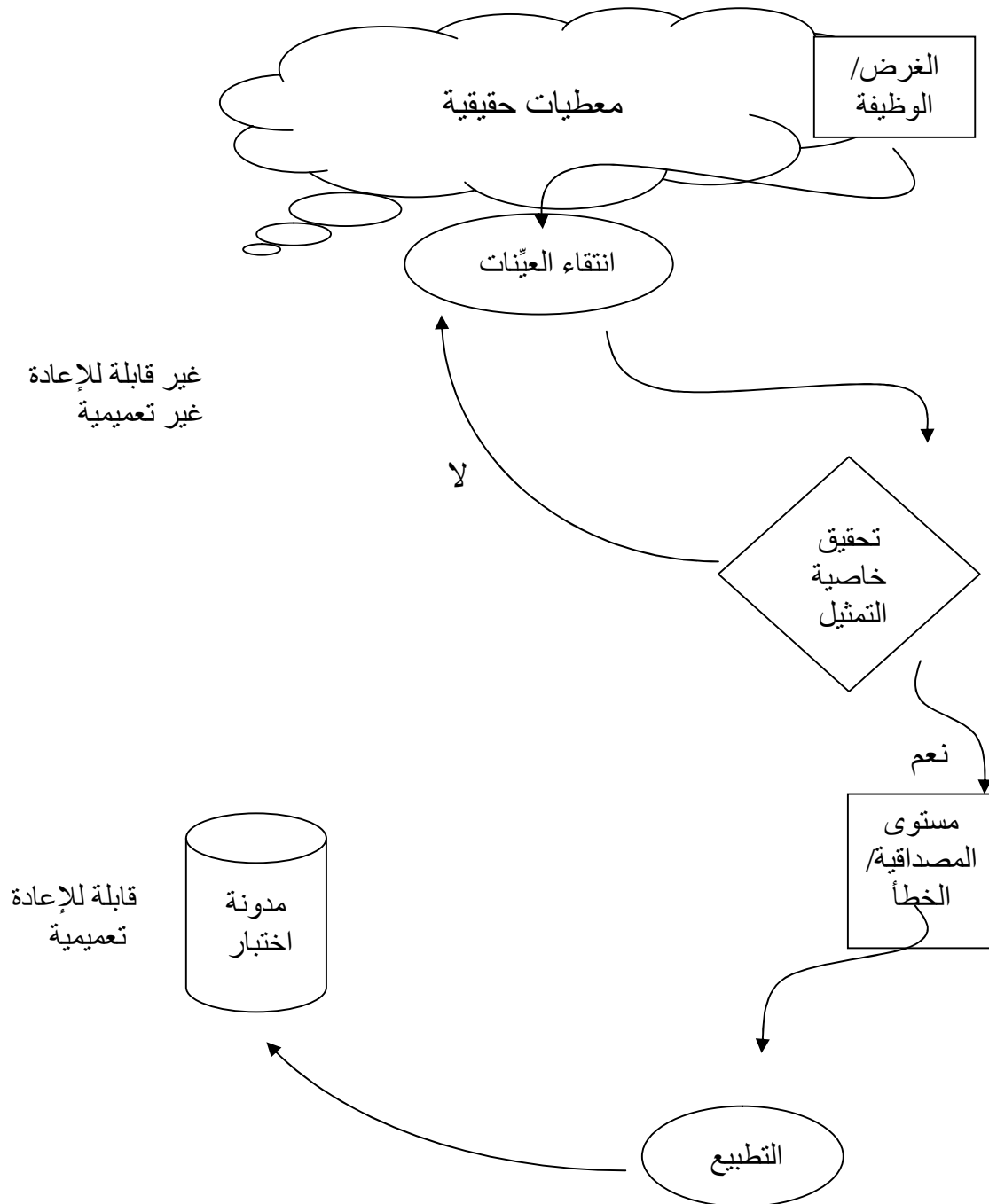
* - يضيف سنكلير نمطاً ثالثاً يدعو بالنمط الإلكتروني (Electronic mode)، لكننا نرى أنه يمكن اعتماد هذا النمط ضمن تصنيف آخر لطبيعة النصوص؛ إذ إنها تكون إما ورقية أو إلكترونية.

** - و إن كان سنكلير يشير إلى هذا المعيار بـ (Text type)، إلا أننا نحبذ "مصدر النص" (Text Source)، تفادياً للبس الذي يمكن أن يقع فيه القارئ بالنظر إلى أصناف النص المعهودة (Typologie textuelle).

² - Cf. J. Sinclair (2005) Op .Cit.

³ - See the MICASE website, <http://www.hti.umich.edu/m/micase>

⁴ - نلفت عناية القارئ إلى أن اسم تلك المدونة يحوي ضمنياً المعايير المشار إليها أعلاه - عدا التاريخ - (نمط الحامل اللساني: منطوق، حقل النص: أكاديمي، لغة المدونة: الانجليزية الأمريكية، موقعة النصوص: ميتشيغن).



الشكل (2) خاصة التمثيل في تصميم المدونات.¹

و لنا أن نلاحظ ذلك - على سبيل المثال - في "ميكاز"، بحيث نجد أن المعيار الثاني (مصدر النصوص) قد أسقط عند تصميم المدونة، نظرا لعدم توافقه و المعيار الأول، إذ إنه يخص النصوص ذات الحامل المكتوب فقط.

ب- وضع جرد كامل بأنواع النصوص التي عُثِرَ عليها، بالاعتماد على معايير خارجية.

¹ - Cf. MacMullen (2003) Op. Cit., p. 10.

ج- ترتيب تلك الأنواع من النصوص وفق مبدأ الأولوية – على أن تؤخذ بعين الاعتبار كل العوامل التي يمكن أن تنقص أو تزيد من أهمية نوع النص – على شاكلة العوامل المطروحة أعلاه.

د- تقدير حجم معين لكل نوع نصي، و ذلك بالنسبة إلى:

(*) الحجم الكلي للعنصر. (*) عدد أنواع النصوص. (*) أهمية كل نوع.¹

هـ- ينبغي أن تكون المدونة متوازنة، من حيث إنها تتضمن تشكيلة متنوعة من النصوص و مصادرها و كُتابها، و وضعيات التواصل ومستويات التخصص – على سبيل المثال – ما لم يقصد الباحث أن يفرد بالدراسة عاملا من هذه العوامل؛² كأن يتناول بالتحليل مثلا أسلوب كاتب ما.

و هكذا، ليس للباحث أن يصل إلى نتائج أدق و أشمل، إلا إذا كانت مدونته محكمة البناء، دقيقة التصميم من وجهين:

- أن تتوافق مادتها و غرض البحث.

- أن تمثل تلك المادة اللغة قدر الإمكان.

"[...] the contents of the corpus should be chosen to support the **purpose**, and therefore in some sense represent the language from which they are chosen"³. [Emphasis added]

و في السياق نفسه، تتحدث إليزابيث مارشمان – هي الأخرى – عن جملة من المعايير التي تحكم عملية انتقاء النصوص، كي تدرج ضمن مدونة خاصة بالمصطلحية؛ كما تؤكد على أن تلك المعايير ذاتها هي بمثابة عناصر رئيسة تساعد على تسيير المدونة ككل، و استثمارها لاحقا ضمن مشروع البحث. و الباحثة تقترح إحداث نظام خاص، يسمح بانتخاب النصوص لاحقا وفق عدة معايير، على شاكلة: اللغة، الميدان والفروع، أنواع النصوص، التاريخ، مستوى التخصص، و ما إلى ذلك. و تذهب إلى ضرورة أن يكون النظام محددا بواسطة تلك المعايير نفسها.⁴

أ- اللغة:

يشترط في الأغلب الأعم، أن تكون النصوص أصلية، و محررة في اللغة الأم للكاتب، إلا إذا كان متمكنا من اللغة الثانية مثله كمثل الناطق بها. و قد جاء التأكيد على أن تكون النصوص أصلية، و ليست ترجمات مخافة أن تكون تلك قد حررها كاتب غير ضليع في المجال (لاسيما أن المدونة – و كما نبهنا – خاصة بالمصطلحية).

ب- الميدان والفروع :

ينبغي أن يكون الميدان محددا و محدودا، قبل أن يشرع الباحث في صناعة المدونة؛ و من ثم، يفصل في ما إذا كان النص (س) ينتمي إلى الميدان محط البحث. مثال ذلك: بإمكان الباحث أن يضيف إلى مدونته القانونية أي نص يتحدث عن قرار وزاري مثلا أو يفسره، لكن لا يمكن أن يضيف مقالا افتتاحيا يقيم أثر هذا القرار في المجتمع.

¹ - Ibid.

² - Cf. E. Marshman (2003) Op. Cit., p. 4.

³ - Cf. J. Sinclair (2005) Op. Cit.

⁴ - Cf. E. Marshman (2003) Ibid., pp. 6-15.

فضلا على ذلك، يستطيع الباحث أن يجعل فروعاً للميدان الواحد، مما يسمح بالحصول على مدونات فرعية (Sous-corpus) انطلاقاً من تلك الفروع. مثال:

قانون الأسرة.

قانون العمل.

الخ...

في مدونة قانونية

Microsoft Access - [Gestion des corpus]

Fichier Edition Affichage Insertion Format Enregistrements Outils Fenêtre ?

Référence

Nom de fichier	Langue	Domaine
modèle_eti.txt	anglais	Informatique
Sous-domaine		
traitement de texte		
Code de la recherche	Genre de document	Niveau de spécialisation
2003-01-01-01	Article de journal	Vulgarisation
Auteur	Destinataire	
Autre	Non-initié	
Référence		
Smith, Bob. "Word for Dummies", Montreal Gazette, 23 janvier 2003. [http://www.gazette.com/janvier/smith/word.htm]. (Visité le 1 janvier 2003). [modèle_eti.txt]		
Date de parution	Méthode de saisie	Date de saisie
2003	Électronique	2003-01-01
Nombre de mots	Format du fichier	
403	Texte ASCII	
Annotation		
Étiqueté (WINBRIL)		
Commentaires		
Auteur : journaliste. Pagination non-respecté.		

Enr : 1 sur 2

La référence complète, en format bibliographique. (Modèle à

الشكل (03) نظام انتخاب نصوص المدونة.

ت- جنس النص (Genre du texte):

يحدد مستوى التخصص و مؤهلات الكاتب و خبراته في المجال؛ مثلاً: مقال في مجلة مختصة، مقال في صحيفة، مقال في مجلة تبسيطية، كتيب تقني، كتاب دروس، و هلم جرا.

ث- مستوى التخصص (Niveau de spécialisation):

في مدونة مختصة، ينبغي تجميع نصوص قد حررها متخصصون في الميدان/الفرع. كما ينبغي إدراج عدة مستويات قصد الحصول على متغيرات المصطلح الواحد، أو مرادفاته، كل حسب المستوى الذي وظف فيه؛ على غرار: تقني جدا، تقني، مختص، تبسيطي، تكوين، عام. و لمعرفة مستوى تخصص ما، ينبغي أن يأخذ الباحث بعين الاعتبار كلا من: نوع النص و كاتبه و المتلقي.

ج- مؤهلات الكاتب و خبراته:

ينبغي أن يكون الكاتب خبيراً بمجال اختصاصه، و أن يكون النص المنتقى صادراً عن هيئة في ذات الاختصاص. و يمكن أن نميز بين: كاتب مختص (Expert)، و آخر شبه مختص (Relative- Expert) و الأستاذ (Teacher).

ح- المتلقي:

تصنف النصوص أيضاً حسب المتلقي، و هو عنصر لا تخفى مكانته في الوضعية التواصلية على أحد. و يمكن أن نميز بين متلق مختص (Expert)، و آخر مطلع (Initié) و ثالث غير مطلع (Non- initié) و الطالب،... مثال ذلك: درس في الكيمياء العضوية موجه لطلبة السنة الجامعية الأولى.

و جدير بالإشارة، أنه يمكن تحديد نوع المتلقي انطلاقاً من نوع النص، أو باعتماد معايير بين نصية (كالكثافة المصطلحية، و شرح المصطلحات).

خ- المرجع:

ينبغي توثيق النص و تدوين المرجع الذي استقي منه تدويننا كاملاً، مما يسمح بتحقيق النسخة المستخدمة منه بالعودة إلى الأصل، و كذلك للحيلولة دون المشاكل التي يمكن أن تنجر عن حقوق التأليف.

د- التاريخ أو الفترة الزمنية للنصوص:

يرتبط هذا المعيار بمشروع البحث و بنوع المدونة المراد صناعتها (سانكرونية، دياكرونية،...). مثلاً: جل الأعمال في مجال المصطلحاتية تكون نصوصها حديثة أو مواضيعها المطروقة معاصرة؛ أما إذا كان المفهوم قد مضت على ظهوره حقبة من الزمن، فإنه يحبذ العودة إلى النصوص التي ظهر فيها أول مرة و الاستعانة بها، طالما أنها تحوي سياقات تصف ذلك المفهوم أو تشرحه. و بالمثل، فإن تجميع نصوص "قديمة" في مجال الترجمة، قد يفيد في شرح بعض المفاهيم لإنجاز ترجمات معاصرة لنصوص أقدم زمنياً.

ذ- شكل الملف (Format de fichier):

تحفظ النصوص عموماً ضمن الشكل txt. قصد تسهيل معالجتها و تحليلها بأدوات عدة (المفهرسات الآلية concordanciers، المرمزات الآلية étiqueurs، الخ). كما أنه ثمة أشكال أخرى، على غرار: doc، rtf، XML، HTML، SGML.

II. 2-2- مصدر النصوص (Corpora Sources):

يتعين على الباحث أن يجد مصادر مناسبة يستقي منها نصوصاً تتناول الموضوع محط اهتمامه ليدرجها ضمن مدونته بعد أن يحولها إلى الصورة النصية (Plain text format). و لا يشكو الباحثون اليوم ندرة تلك المصادر أو انعدامها، بل إنها تعرف فيضاً

منقطع النظير، لاسيما بفعل حركية الاتصال و التواصل الزاخرة التي يشهدها العالم هذه الأيام.

1- استخدام الماسح الضوئي:

إن أبسط مصادر المدونة، يتمثل في المادة المطبوعة على اختلاف ألوانها و تباين أشكالها، من كتب و مجلات و موسوعات و جرائد و مطويات، و ما شاكلها من منشورات و مطبوعات؛ يقوم الباحث بتحويلها إلى شكل الكتروني باستخدام الماسح الضوئي (Scanner)؛ بيد أن ما يعاب على هذه الطريقة هو استغراق وقت أكبر مقارنة بطرق آخر.

2- طلبات الترجمة (Assignments as corpora):

إذا كنت مترجما - من العربية إلى الفرنسية/الفرنسية إلى العربية مثلا - فلا شك أنك تتلقى عددا لا يستهان به من طلبات الترجمة (لاسيما إذا كانت في شكل الكتروني). و تعد تلك الطلبات مصدرا لصنع مدونتك الخاصة بعد إعادة حفظ نصوصها ضمن الصورة النصية.

و في هذا الصدد، يتحدث ويلكنسن (Wilkinson) عن تجربته الخاصة؛ إذ تمكن من صنع مدونة سياحية باللغة الفنلندية بحجم سبعين ألف كلمة، مصدرها المطويات السياحية التي حصل عليها عن طريق زوجته، و هي مترجمة محترفة من الفنلندية إلى الانجليزية. و لم تستغرق من وقته عملية الصنع سوى سوياعات، طالما أن طلبات الترجمة تلك كانت تصل ضمن شكل الكتروني.¹

و بالقياس، يمكن لأستاذ الترجمة - هو الآخر - أن يستغل وفق الطريقة نفسها مجموع النصوص التي يقدمها لطلبته.

3- شبكة الويب:

تزخر الشبكة العنكبوتية بنصوص متباينة اللغات؛ و هي تشكل مصدرا هاما لصناعة المدونات. بيد أن تلك النصوص على كثرتها الكثرة و وفرتها السافرة، تفرض على الباحث توخي الكثير من الحذر؛ إذ ينبغي عليه أن ينتقي منها ما يراه مناسباً و ملائماً و موثوقاً منه.* و الحال هذه، لا مناص من أن يلجأ إلى العدد الكبير من الجرائد و المجلات و مختلف الدوريات المتخصصة على مستوى الشبكة؛² سواء كانت منشورات متاحة للاستعمال، أم أنها تتطلب اشتراكا للحصول على بعضها، و بعضها الآخر يعرض مادته للبيع، و مثال ذلك: إصدارات المنظمات الدولية (على غرار هيئة الأمم المتحدة وتنظيماتها المختلفة)، أو الإصدارات الحكومية (كالجريدة الرسمية)، الوكالات للأنباء (رويتر، ب ب سي،...)، أو الدوريات نصف الأسبوعية و الشهرية (على شاكلة: الخبر، التايمز، لوموند ديبلوماتيك)، و نحوها.

¹ - Cf. M. Wilkinson (2006) *Op. Cit.*

* - على الباحث أن يحذر مختلف المدونات المعروفة باسم (Blog)، أو المقالات التي ينشرها الهواة لا المتخصصون.

² - لمعرفة مختلف المقاربات المنتهجة في البحث عبر الويب، ينظر:

W. H. Fletcher (2007) *Concordancing the Web: Promise and Problems, Tools and Techniques*. In: Marianne Hundt, Nadja Nesselhauf & Carolin Biewer (eds.) *Corpus Linguistics and the Web*. Amsterdam: Rodopi, pp. 25-45.

و بعد الانتقاء، تنتقل النصوص و تُلصق (Copy-Paste) ضمن صفحة من نظام معالجة النصوص العادي (Microsoft Office Word)، ثم تحفظ ضمن الصورة النصية (Plain text).

و الواقع أن الحصول على مدونة باعتماد صفحات الويب ليس بالأمر الهين؛ إما لأن المادة النصية يخضع بعضها لحقوق التأليف، أو أنها ليست متاحة، و إما لأن عملية تحويلها إلى مادة لغوية محض تتطلب تدخل الباحث، و تفرض عليه أن يُعْمَل مقصده، ويستأصل كل ما كان على شاكلة الصور أو الجداول أو الرسوم التخطيطية. و كلما كانت المواقع الالكترونية أخاذة و أكثر تعقيدا (كأن تكون النصوص تشعبية Hypertexts)، كانت عملية الحصول على المادة النصية أصعب و أشق على الباحث - كما تقول لين باوكر Lynne Bowker :-

"[...].good web design is not conducive to easy corpus building !".¹

غير أن ما يبعث على الطمأنينة، و يدفع إلى المضي قدما في عملية صنع المدونة وفق هذه الطريقة، هو أن الوقت المستغرق و الجهد المبذول يمثلان "عبئا بسيطا إذا ما [قورنا] بالإمكانات الهائلة التي ستخدم بها هذه المتون مجال التطبيقات اللغوية".² ثم إن الويب يتميز عن بقية المصادر المستخدمة في صناعة المدونة ببعض الخصائص:

أ- الاكتمال و سعة المجال: يمكن للمدونة أن تُغفل جنسا كتابيا أو مجالا ما، كما يمكن ألا تورد ما يكفي من الأمثلة بخصوص بعض التعابير/التراكيب، و تلك تسهل موقعتها على مستوى الويب.

ب- الطبيعية و الجدة: سرعان ما يصيب الهرم نصوص المدونة؛ في المقابل، تزخر صفحات الويب بنصوص تتعلق بمواضيع معاصرة و أمثلة حقيقية عن الاستعمال اللغوي، سواء ما كان سائرا حاليا أو ما كان قد ظهر لتوه.

ج- التنوع اللغوي: ثمة بعض اللغات أو اللهجات على صفحات الويب، لم تحظ بفرصة جمع أي مدونة تتناولها.

د- خاصية التمثيل: بالنظر إلى تزايد نسبة المعلومات و التواصل الموزعة عبر النت، لا يخفى على أحد أن لغة الويب تنعكس باستمرار على لغتنا و تثريها إثراء فاحشا. هـ- الملاءمة و عدم الكلفة: إن الاستبحار في صفحات الويب غير مكلف، كما أن كثيرا من برمجيات تصفح الويب و استنباط المعلومات منه في متناول الباحثين و الطلبة على حد سواء (هذا عدا محركات البحث).³

و- تيسير عملية تجميع النصوص: و من خصوصيات الويب أيضا أنه المصدر الأكثر احتواء للنصوص في شكلها الالكتروني، مما يسهل إلى حد ما عملية التجميع.

¹ - Lynne, Bowker (2002) Working Together: A Collaborative Approach to DIY Corpora. Paper presented at the First International Workshop on Language Resources for Translation Work and Research, Gran Canaria, 28/05/2002. Viewable online at:

<http://www.ifi.unizh.ch/cl/yuste/postworkshop/repository/lbowker.pdf>

² - سلوى، حمادة (2007) مر. س.

³ - Cf. W. H. Fletcher (2007) Op. Cit., pp. 28-29.

فضلا على ذلك، يمكن أن تعالج المدونة و تحلل بشكل مباشر بفضل بعض البرمجيات المتوفرة على الشبكة.

ز- الدينامية و المعاصرة: ثم إن ما يميز مدونة مصنوعة باعتماد الويب هو ديناميته و معاصرتها، من حيث إنها تحوي آخر المستجدات المصطلحية والاصطلاحية، و التي يمكن أن يستغرق البحث عنها في مصادر آخر وقتا أكبر؛ و ما على الباحث إلا أن يتأكد من نوعية المعلومة و من موثوقية مصدرها - كما أسلفنا¹.

4- ثمة أشكال أخرى من المصادر المعتمدة في صناعة المدونات؛ على شاكلة: المنشورات ضمن القرص الليزري CD-ROM، البرامج الإذاعية/التلفزيونية و المكالمات الهاتفية، و هذه يمكن أن تفيد في صناعة المدونات المنطوقة وفق شكلين:
- مادة صوتية مسجلة غير مكتوبة كمدونات مؤسسة (إل دي سي)*.
- مادة منطوقة تم تحويلها إلى مادة مكتوبة باستخدام المحولات الصوتية (كما في المدونات المنطوقة لسلسلة كولنز- كوبيلد (Collins-Cobuild)).

II-3- إعداد النصوص:

بعد أن يجمع الباحث النصوص التي تخدم بحثه، يتجه إلى ثالث خطوة في عمله، ألا و هي إعداد تلك النصوص إعدادا يتسنى له من خلاله إدراجها ضمن مدونته؛ إذ يجب تحويلها إلى شكل الكتروني كي يسهل التعامل معها آليا (و غالبا ما تحول إلى الصورة النصية Plain text format)، كما يتعلق الأمر أيضا بتزويد تلك النصوص و إثرائها ببعض البيانات الإضافية، و هي عملية يصطلح على تسميتها بـ "تحشية المدونة" (Enriching corpora). أما تلكم البيانات فتصنف إلى:

II-3-1- بيانات الوثيقة (Document information):²

و تتعلق بمصدر النص المراد إدراجه ضمن المدونة، نحو: اسم المؤلف/المتكلم، اسم الناشر (ضمانا لحقوق التأليف)، و كذا بيانات سياقية (Extralinguistic/contextual information) نحو: مكان النشر و تاريخه (أو التسجيل في حال مدونة منطوقة)، سن المؤلف/المتكلم و جنسه، المستوى الاجتماعي و الثقافي، الخ.

II-3-2- التنقيح (Markup):

كثيرا ما تتطلب بعض المدونات التنقيح و التشذيب، لاسيما في ما يتعلق ببيانات حول بنية الوثيقة الأصل، على شاكلة: العناوين الرئيسية، الفقرات، بداية و نهاية كل صفحة من الوثيقة، و إعداد الصفحة (الحروف الطباعية، قياسها، نمطها،...) و ما إلى ذلك من خصوصيات الوثيقة.³

¹ - Cf. K. Varantola (2002) Op. Cit., p. 186 ; Cf. F. Zanettin (2002) Op. Cit.

* - The Linguistic Data Consortium (LDC) is a consortium compiling and distributing data banks of mostly the English language but also other languages (University of Pennsylvania). Available at: <http://www ldc.upenn.edu>

* المصطلح مأخوذ عن: سلوى، حمادة، مر.س. ص 7. و ننبه إلى أن هذه المعالجة واحدة من اهتمامات اللسانيات الحوسبية.
² - قد يجد القارئ أنه لمن المفارقة أن نتحدث عن النص في موضع، و عن الوثيقة في موضع آخر، لكننا نذهب إلى ما ذهب إليه أحمد الديداوي (الترجمة و التواصل: ص 20) من أن النصوص في أغلبها الأعم "تتخذ في واقع الأمر أشكالا هجينة"، و أن الوثيقة خليط من النصوص تتعاضد لمعالجة موضوع ما (نفسه: ص 23).

³ - Cf. A. Stefanowitsch (2003a) Op. Cit.

II-3-3- الترميز (Annotation):

و يعنى به تحليل مادة المدونة الخام (Raw corpora) تحليلًا لغويًا بأية صورة من الصور، أو على أي مستوى من المستويات اللغوية، بحيث يزودها الباحث ببيانات ميتالغوية عدة وفق الحاجة:

- أ- على المستوى الفونولوجي (Phonetic annotation)، و يتم يدويًا على:
 - مستوى الفونيمات/الأصوات، كأن تضاف بيانات حول كيفية نطق كلمة ما في مدونة منطوقة؛ و النسخ الصوتي (Phonetic transcription) ضرب من الترميز الصوتي يستخدم في تحويل المدونات المنطوقة إلى مدونات مكتوبة.¹
 - مستوى التنغيم (Prosodic annotation)، و ذلك بإضافة بيانات خاصة بالتشديد (Stress) و النبر (Intonation) و الوقف (Pauses) في مدونة منطوقة.
- ب- على المستوى الصرفي: أي الترميز على مستوى السوابق و اللواحق و جذوع الكلمات.

ج- على المستوى المعجمي: و هو الترميز على مستوى أقسام الكلام، أو متغيرات اللميمة (Lemma).

د- على المستوى الدلالي (Semantic annotation): و يقصد به تزويد نصوص المدونة ببيانات حول الحقول الدلالية، و تبيان العلاقات الدلالية بين وحدات النص الواحد، أو الخصوصيات الدلالية لكلمات النص. مثال ذلك الشكل رقم 4.

AT1	The	Z5
MC	one	N1
NN1	disadvantage	A5.1-
IO	of	Z5
JJ	woolen	O1.1
NN2	clothes	B5
VBZ	is	A3+
CST	that	Z5
PPHS2	they	Z8
VM	can	A7+
VV0	become	A2.1
JJ	uncomfortable	O4.2-
II	in	Z5
RG	very	A13.3
JJ	hot	O4.6+
NN1	weather	W4

الشكل (4) الترميز على المستوى الدلالي.²

¹ - تعد مدونة London- Lund أول مدونة تخضع لهذا النوع من الترميز، ينظر:

Frank van Eynde (2008) Corpus construction and annotation. Centre for Computational Linguistics. Katholieke Universiteit Leuven. Viewable online at:

<http://www.ccl.kuleuven.be/Courses/LEA/corpusannotatie.ppt>

Cf. also: T. McEnery & A. Wilson (1996) Corpus Linguistics, Chapter 2. Edinburgh.

² - Cf. T. McEnery & A. Wilson (1996) *Ibid*.

يتضمن المثال ثلاث خانات: تشير الخانة الأولى من الرموز إلى أقسام الكلام، و الخانة الثانية تتضمن كلمات النص، أما ثالثها فتتضمن الشارات الدلالية *Semantic tags*؛ بمعنى أن قراءة النص تتم شاقولياً. تتكون كل شارة دلالية من:

* حرف يدل على حقل الخطاب العام.

* رقم يدل على فرع أول للحقل.

* فاصلة تشير إلى فرع آخر ضمن الحقل ذاته (إن وجد).

* علامتا (-) أو (+) تحيلان على التضاد (إن وجد) أو تفاوت في الدرجة على مستوى السلم الدلالي.

و لنأخذ مثال الشارة الدلالية +O4.6، و هي شارة تحيل على الكلمة (أو الصفة) *Hot* تنتمي إلى الحقل "المواد و الخصائص" *O*، ضمن الفرع "الخصائص الفيزيائية" *O4*، وكذا الفرع الأدنى "الحرارة" *O4.6*، درجتها "عالية" (+) بدل "منخفضة" (-).
هــ على المستوى التركيبي: و يتعلق بترميز المركبات و الجمل (*phrases and sentences*)، و هو ما يعرف بالإعراب (*Parsing*).

و- على مستوى الخطاب (*Discoursal annotation*)¹، و نميز بين:

- بيانات خاصة بالعلاقات العائدية (*Anaphoric annotation*)، لاسيما الإحالة إلى المرجع بواسطة الضمائر (*Pronoun reference*)، مثال ذلك:

(1 *Feodor Baumenk* 1), a former Nazi death campguard, has asked the U.S Supreme Court to allow <REF=1 him to retain <REF= 1his American citizenship. (2 *The Hartford Courant* 2) said...

المفتاح:

(1), (2) ...= antecedent

<= anaphoric

>= cataphoric

REF= central pronoun²

- بيانات تداولية (حول أفعال الكلام مثلاً).
- بيانات أسلوبية (الخطاب المباشر، الخطاب غير المباشر،...الخ).³

و نكتفي فيما يلي بإيراد صورتين من صور الترميز، هما: الترميز بأقسام الكلام (*POS tagging*)، و إعراب البنى التركيبية (*Syntactic structure parsing*).

¹ - مثال ذلك: مدونة (London -Lund) المنطوقة، و تتضمن 16 شارة خطابية (*Discourse tags*)، نذكر منها : "apologies" (sorry, excuse me, ...), "greeting" (hello, hi..), "politeness" (please,...), "responses" (really, that's right...), etc.

² - شمة أمثلة قليلة عن المدونات التي تحمل ترميزاً للعلاقات العائدية، و لعل أشهرها (The Lancaster / IBM anaphoric Treebank)، وهي تعتمد على ما جاء به هاليداي و حسن في مؤلفها:

M.A.K. Halliday & R. Hasan (1976) Cohesion in English. London: Longman.

³ - Cf. Geoffrey, Leech (2005) Adding Linguistic Annotation. In: Martin, Wynne (ed.) *Op. Cit.*, viewable at: http://www.ahds.ac.uk/creating_guide/linguistic-corpora/chapter_2.htm

1- الترميز بأقسام الكلام (*Etiquetage des catégories grammaticales*):

و يعرف كذلك بالترميز النحوي (*Grammatical Tagging*)، و هو أن تحلل مادة المدونة تحليلًا لغويًا على مستوى أقسام الكلام، بحيث تُدَيَّل كل وحدة معجمية من المدونة بشارة (*Tag/Label*) تبرز قسم الكلام الذي تنتمي إليه تلك الوحدة (فعل، اسم، حرف، صفة، ظرف،...)، على أن يتم الربط بينهما (أي الوحدة المعجمية و الشارة) بواسطة عروة " _ " ¹. مثالًا: في مدونة مرزمة، نجد المقطع: *Machines_ NNS* الوحدة المعجمية: *Machines*. العروة: " _ " .

شارة قسم الكلام: *NNS* (اسم جمع عام *Plural Common Noun*). و تستعمل تقنية الترميز بأقسام الكلام بصورة أكبر مقارنة بباقي التقنيات المشار إليها سابقًا، و مرد ذلك أنها تعد القاعدة الأساس التي يتم الانطلاق منها في الترميز على المستويين الدلالي و التركيبي، و لأنها تتم على قدر عال من الدقة باستخدام الحاسوب. فضلًا على ذلك، يولي الباحثون في مجال المعالجة الآلية للغات هذه العملية أهمية كبرى نظرًا لفائدتها في كثير من التطبيقات، لعل أهمها: ²

- رفع اللبس و استجلاء الغموض، مثالًا: كلمة "deal" الانجليزية إما أن تكون اسماً (*NN*)، أو فعلاً (*VB*).
- كما أنها تساعد في النطق، مثالًا كلمة: *INSult* و تتكون من مقطعين، يلفظ بنبرة مشددة أولها و يلفظ آخرها بنبرة مخففة إذا كانت اسماً، و *inSULT* بالتشديد على المقطع الثاني إذا كانت فعلاً. و كذا هي الحال مع *OBject/obJECT*، *CONtent/content*، و هكذا دواليك.
- تحديد معنى الكلمات التي تتشابه كتابة و تختلف معنى (*Homographs*)، مثال ذلك: "booted" إذا كانت صفة فهي تعني "ينتعل جزمة"، و إذا كانت فعلاً فتحمل معنى "رفس، ركل".
- تحديد جذوع الكلمات (*Stemming*)، مثالًا: حرف "s" الانجليزي إما أن يكون علامة الجمع بالنسبة للأسماء (أسلحة *arms*)، و إما أن يكون علامة المفرد الغائب بالنسبة للأفعال (تسلح/ يسلح *arms*).
- تساعد في الترجمة الآلية، و مثال ذلك استنباط القاعدتين بخصوص كلمة *Content* الانجليزية:

¹ - Cf. T. McEnergy & A. Wilson (1993) *Op. Cit.*, p. 2.

² - عن الترميز بأقسام الكلام و أهميته، ينظر:

F. van Eynde, J. Zavret & W. Daelemans (2000) Part-of-Speech Tagging and Lemmatization for the Spoken Dutch Corpus. In: M. Gravididou et al.(eds.) *Proceedings of the Second International Conference on Language Resources and Evaluation* (pp. 1427-1433). Athens.
Dan Jurafsky (2007) Lecture 6: Part of Speech Tagging (I). *LINGUIST 180: Introduction to Computational Linguistics*.

- و عن المرزمات الآلية، ينظر الملحق (1).

(E) content + N → (F) contenu + n.

(E) content + Adj. → (F) content + Adj.

satisfait + Adj.

■ تساعد في البحث المصطلحي و التوثيقي.

و من أمثلة المدونات المرمزة بأقسام الكلام، نذكر: المدونة الوطنية البريطانية (BNC)، مدونة براون (The Brown Corpus)، و مدونة لوب (The LOB Corpus).
و المثال التالي يبرز هذا النوع من التحشية، و هو مثال مستنبط من مدونة "لوب" (Lancaster-Oslo /Bergen Corpus)¹، مرفقا بمفتاح لأهم الشارات الواردة فيه:

^Colour_NN is_BEZ highly_RB important_JJ in_IN
choosing_VGB autumn_NN footwear_NN._.
^the_ATI autumn_NN range_NN of_IN shades_NNS
is_BEZ almost_RB bewildering_JJ, and_CC
there_EX are_BER some_DTI exciting_JJ new-
comers_NNS [...]

مفتاح الشارات : Key

ATI article neutral for number. **BER** are (present plural form of the verb BE).

BEZ is, 's (-s form of the verb BE). **CC** coordinating conjunction. **DT** singular determiner.

DTI determiner neutral for number. **EX** existential there. **IN** preposition. **JJ** general adjective

NN singular common noun. **NNS** plural common noun. **RB** general adverb.

VBG present participle of lexical verb.

و جدير بالإشارة، أن الترميز بأقسام الكلام يتم بواسطة برمجيات خاصة هي الرموزات الآلية² Taggers، تعتمد كل واحدة منها على سلسلة خاصة من الشارات الشارحة للمكونات اللغوية Tagsets، أشهرها تلك التي استخدمت في ترميز مدونة براون (87 شارة)، و Lancaster UCREL C5 سلسلة تحوي 61 شارة استخدمت في ترميز المدونة الوطنية البريطانية، و كذا Lancaster C7 التي تحتوي 145 شارة.³

¹ - Cf. T McEnery & A. Wilson (1993) Op. Cit., pp. 2-3.

² - بعض تلك البرامجيات متوفر مجانا على شبكة النت، وأشهرها: CLAWS tagger

<http://www.comp.lancs.ac.uk/computing/research/ucrel/claws/tagsservice.html>

³ - Brown corpus tagset viewable online at: <http://www.scs.leeds.ac.uk/amalgam/tagsets/brown.html>

C7 tagset at <http://www.comp.lancs.ac.uk/ucrel/claws7tags.html>

2- الإعراب (تحليل الجمل) (Parsing):

يعنى به إسناد البنية السطحية لنص ما بواسطة التحليل الشجري - حسب الشكل (5). - و نسبة إلى هذا النوع من التحليل، تعرف المدونات المرمزة على هذا النحو بالمُشجَّرات *Treebanks.

و كما سلف و ذكرنا، يتم الإعراب انطلاقاً من الترميز بأقسام الكلام - كما هو موضح في المثال أدناه - باستخدام برمجيات خاصة، هي "المعرّبات الآلية" *Parsers*؛ بيد أن نتائج العملية ليست بنفس دقة النتائج المحصل عليها عند الترميز بأقسام الكلام، لذا تتم العملية إما يدوياً، أو شبه آلياً (أي بتدخل الإنسان).

و تعتبر مدونة بان *Penn Treebank* أول مدونة تخضع لهذا النوع من الترميز باعتماد سلسلة من الشارات تعدادها 45 شارة¹، هذا فضلاً عن مشجرات *Nijmegen* و مدونة سوزان *Susanne corpus*.
المثال:

S[NP Claudia_NP1 NP][VP sat_VVD[PP on_P[NP a_ATI stool_NN1 NP]PP] VP]S]

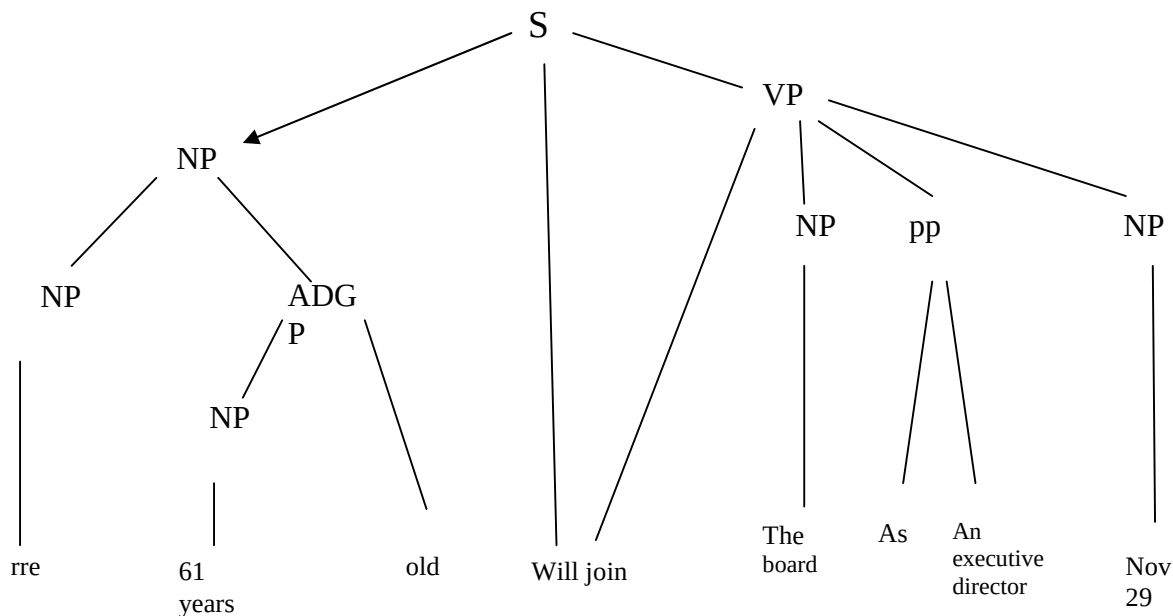
مفتاح الشارات (Key):

S = sentence.

NP = noun phrase, *VP* = verb phrase.

PP = prepositional phrase.

N = Noun, *V* = verb, *AT* = article, *P* = preposition.



الشكل (5) التحليل الشجري.

* - المُشجَّر: هو الثوب المزين على هيئة الشجر، وتلقب بالمشجر كتابة أهل الصين التي هي مشبكة كالشجر. ومنه استوحينا هذا المصطلح كمقابل للإنجليزي *Treebanks* (البنوك الشجرية).

¹ - Cf. annex (2) Penn Treebank Tagset from:
<http://www.cs.colorado.edu/~martin/SLP/Figures/>

و نلاحظ في المثال أن ما يجمع بين الوحدة المعجمية و الشارة (في الترميز بأقسام الكلام) هو العروة " _"، أما المكونات التركيبية فتحدد (في النوع الثاني من الترميز) بواسطة معقوفين. و الترميز التركيبي يفيد في:

- تدريب المعربات الآلية في مجال اللسانيات الحوسبية.
- استخلاص البيانات و المعلومات (كتأليف رصيد لغوي).

و بصرف النظر عن أهمية ترميز المدونات بمختلف أشكاله، غالبا ما يوصي المشتغلون في المجال بضرورة الاحتفاظ بنسخة عن المدونة على هيئتها الخام.

III - أنواع المدونة:

تصنف المدونات إلى أصناف عدة تبعا للغة، والغرض الذي صنعت لأجله.

1- من حيث نمط الحامل اللساني:

أ- مدونات اللغة المكتوبة *Written corpora*، مثل: مدونة بكوالتير العربية Buckwalter (1986-2003)، صنعها تيم بكوالتير Tim Buckwalter، بحجم 2.5-3 بليون كلمة، باعتماد المصادر المنشورة على شبكة الويب، و لغرض معجمي.

ب- مدونات اللغة المنطوقة *Spoken corpora*، مثل مدونة الأخبار الإذاعية (2000) من إنتاج جامعة بنسلفانيا، و تتضمن أكثر من مائة و عشرة (110) برامج إذاعية، عن إذاعة "صوت أمريكا" *Voice of America* بغرض التعرف إلى الكلام *Speech recognition*¹.

2- من حيث اللغة الفصحى/اللغة العامية:

أ- مدونات اللهجات *Dialect corpora*، مثل: مدونة كولهوم *CALL HOME* (1997)، من إنتاج جامعة بنسلفانيا؛ و تتكون من مائة و عشرين (120) محادثة هاتفية، مصدرها متحدثو اللهجة المصرية؛ و لقد صممت بغرض التعرف على الكلام عبر خطوط الهاتف.

ب- مدونات اللغة الفصحى، مثل: مدونة العربية الفصحى *Classical Arabic Corpus* (CAC) من صنع UMIST بالمملكة المتحدة (2004)، مادتها منطوقة بحجم 5 ملايين كلمة، و لقد صممت لغرض التحليل المعجمي باعتماد موقعي: "المحدث" و "الوراق"².

3- من حيث العموم:

أ- مدونات عامة *General Corpora*، تتضمن نصوصا متنوعة، سواء في شكلها المكتوب أو المنطوق، و تتناول مواضيع شتى؛ و غالبا ما تدعى بالمدونة المرجع *Reference Corpus*، نظرا لتعدد أغراضها، و لأهميتها كمادة مرجعية في تعلم اللغة، الترجمة،... مثل: مدونة براون (لصناعة المعاجم، تصميم برامج معالجة

¹ - للتعرف على أنواع المدونات، ينظر كل من: سلوى، حمادة (2007) م.ر.س.

Latifa, Al-Sulaiti (2005) Arabic Corpora. Viewable online at :

http://www.comp.leeds.ac.uk/eric/latifa/arabic_corpora.htm

Maryam, Mohammadi Dehcheshmeh (2007) Specialized monolingual corpora in translation. In: Gabe, Boker (ed.) *Translation Journal* 11(2), April 2007.

² - Cf. www.alwaraq.com
www.muhammadith.org

النصوص، أغراض تعليمية،...)، و كذا المدونة الوطنية البريطانية بحجم مائة مليون كلمة تقريبا.

ب- المدونات المختصة *Specialized Corpora*¹، تقتصر على نوع نصي معين، أو جنس كتابي محدد، أو مجال دون آخر؛ بعضها يعرف كذلك بمدونات لغة الاختصاص *LSP Corpora*، على غرار كنكود *Cambridge and Nottingham Corpus of Discourse in English*، بحجم خمس ملايين كلمة. و من المدونات المختصة نجد أيضا: مدونة ميتشيغن للغة الانجليزية الأكاديمية المنطوقة *MICASE* بحجم 1.7 مليون كلمة.

4- من حيث التعدد اللغوي:

أ- مدونة أحادية اللغة *Monolingual Corpus*، يتم تدوين مادتها النصية ضمن لغة واحدة، على شاكلة: مدونة النهار (2001) من صنع إلرا * *ELRA*، و هي مدونة مكتوبة بحجم 140 مليون كلمة، صممت لأغراض بحثية عامة باعتماد صحيفة النهار اللبنانية.

ب- مدونة ثنائية اللغة *Bilingual Corpus*، على شاكلة: المدونة الانجليزية النرويجية المتوازنة *ENPC*، بحجم 2.6 مليون كلمة، تتوزع على 100 نص أصل و 100 نص ترجمة، تمتد فترتها التاريخية من العام 1975 إلى غاية العام 1995.²

ج - مدونة متعددة اللغات *Multilingual corpus*، مادتها النصية مدونة في لغات عدة. على غرار: المدونة متعددة اللغات من إنتاج *UMIST* بالمملكة المتحدة، و هي مدونة مكتوبة بحجم 11.5 مليون كلمة، صممت لغرض الترجمة.

5- من حيث زمن اللغة:

أ- مدونة سانكرونية، محدودة بفترة زمنية ما، على شاكلة: مدونة العربية الفصحى (2004) من صنع جامعة مانشستر؛ و هي مدونة الشعر العربي من فترة ما قبل الإسلام إلى غاية القرن الحادي عشر.

ب- مدونة تاريخية أو دياكرونية شاملة، مختلفة الفترات، بغية دراسة تطور اللغة على طول تلك الفترات؛ و أشهر تلك المدونات:

The Helsinki Diachronic Corpus of English Texts، بحجم 1.5 مليون كلمة، نصوصها محررة بين سنوات 700 - 1700، و كذا مدونة آرشر *The ARCHER A Representative Corpus of Historical English Registers* بحجم 1.7 مليون كلمة، و تشمل الفترة الممتدة من 1650 إلى 1990.

6- أصل اللغة المدونة:

أ- لغة المدونة أصلية *Original*.

ب- لغة المدونة عبارة عن ترجمة *Translation*، مثل:

¹ - Cf. Guy, Aston (1999) Corpus use and learning to translate. *Textus* 12, pp. 289-314.

* - ELRA: European Language Resources Association. See:

<http://www.elra.info/>

² - English Norwegian Parallel Corpus at:

<http://www.hf.uio.no/ilos/forskning/forskningsprosjecker/enpc>

مدونة "تيك" *Translational English Corpus – TEC* (مركز الترجمات بمانشستر) مادتها نصوص ترجمها ناطقون باللغة الانجليزية من عدة لغات مصدر، و الغرض منها دراسة النصوص الترجمة و البحث في خصوصياتها.¹

7- من حيث مستخدم اللغة:

أ- مدونة خاصة بابن اللغة *Native speaker*، مثل مدونة:

Louvain Corpus of Native English Essays (LOC-NESS)

ب- مدونة خاصة بمتعلم اللغة *Learner corpus* مادتها النصية ناتجة عن متعلمي لغة أجنبية، مثال ذلك: *ICLE-The International Corpus of Learner English* ، بحجم عشرين ألف كلمة، مصدرها محاولات الطلبة متعلمي اللغة الانجليزية كلغة أجنبية.

و تفيد هاتان المدونتان في تحديد أوجه الاختلاف بين نصوص يحررها المتعلم في لغة ما، و أخرى يحررها ابن تلك اللغة.² فضلا على ذلك، يمكن ترميز المدونة الخاصة بالمتعلم بواسطة "شارات الأغلاط و الأخطاء" *Error Tags* ، لتحديد موضع الخطأ الذي ارتكبه المتعلم و كذا نوعه؛ و تفيد هذه المدونة المرمزة في الصناعة المعجمية.³

8- من حيث تحشية المدونة/التحليل اللغوي:

أ- مدونات خام *Unannotated/Raw corpora*، توجد مادتها النصية على صورتها الخام دون إضافة أي معلومات.

ب- مدونات مرمزة *Annotated corpora* : و هي مدونات حولت من صورتها الأولية الخام، إلى مدونات محللة تحليلًا لغويًا بإضافة البيانات اللغوية المختلفة – كما بيناه سلفًا –.

9- من حيث العلاقة بين لغات المدونة:

أ- المدونات المتوازية *Parallel corpora*: تتكون من مادة نصية في لغة (أ) مرفقة بترجمات في لغة (ب) أو لغات عدة. على غرار: مدونة مصر (1999) من صنع جامعة جون هوكينز، و هي مدونة مكتوبة تتضمن النسخة العربية للقرآن الكريم، مرفقة بنسخة انجليزية لترجمة معانيه؛ و منها أيضا مدونة لوب (*Lancaster-Oslo/ Bergen Corpus*).

ب- المدونات المقارنة⁴ *Comparable Corpora*: تتكون هي الأخرى من مادتين نصيتين متماثلتين من حيث معايير التصميم في لغتين (أ/أ)، دون أن تكون إحدهما ترجمة للأخرى. و من أمثلتها:

المدونة الانجليزية العالمية *ICE The International Corpus of English* (1991) وسنفصل في هاتين المدونتين لاحقا.

¹ - Cf. Maeve, Olohan (2002a) Comparable Corpora in Translation Research: Overview of Recent Analyses Using the Translational English Corpus, *LREC Language Resources in Translation Work and Research Workshop Proceedings*, pp. 5-9.

² - Cf. M. Mohammadi Dehcheshmeh (2007) *Op. Cit.*

³ - مثال ذلك، ينظر الشكل (14) ص 79، و أيضا:

Geoffrey, Leech (2004) *Op. Cit.*

⁴ - Cf. John, Sinclair (1996) *Op. Cit.*

10- المدونة البيداغوجية *Pedagogic Corpus* :

و هي مدونة تتضمن جميع النصوص التي تعرض لها المتعلم (على شاكلة النصوص ضمن الكتب المدرسية، المطالعة،...الخ)، و كذا جميع الأشرطة التي استمع إليها. و تفيد هذه المدونة في التعرف على جميع ضروب استعمال الكلمة/العبرة الواحدة ضمن سياقات عدة، مما يدعم التحسين اللغوي لدى المتعلم.

11- مدونة الضبط و المراقبة *Monitor corpus* :

تتضمن النوع نفسه من النصوص، و تساعد في ضبط التغيرات التي تطرأ على اللغة و مراقبتها، من خلال تزويدها بجديد تلك النصوص بصفة منتظمة و مطردة (يومية، شهرية، سنوية). و الحال هذه، فإن كل نسخة جديدة من النصوص يمكن أن تقارن بسابقتها من النسخ. و أشهر هذه المدونات: بنك اللغة الانجليزية *The Bank of English* بجامعة برمنغهام.¹

¹ - Cf. M. Mohammadi Dehcheshmeh (2007) *Op. Cit.*

David, Evans (2006) Unit 1: Introduction. Corpus Building and Investigation for the Humanities (*Online introduction to Corpus Investigative Techniques*), AHRC. Viewable online at: <http://www.humcorp.bham.ac.uk/>

‘La linguistique de corpus existe, elle est récente et sa méthodologie et son épistémologie se forment, pour la forger, il faut simplement la reconnaître’. G. Williams (2006:157).

IV- لسانيات المدونة: المصطلح و المنطلقات التاريخية، المنهج و الأدوات الإجرائية.

1- المصطلح:

لا يخالّن القارئ – و هو يطالعه هذا العنوان – أن محط اهتمامنا، و مريض إشكالنا، يكمن في إيجاد تعريف لمفهوم "لسانيات المدونة"، بل إن ما نريد إثارتة أبعد من ذلك، طالما أنه لا ريب في أن إيجاد تعريف لهذا المفهوم لا يخلق إشكالا كبيرا، كما لا يثير جدلا عظيما بالقدر الذي يثيره المصطلح ذاته. بيد أنه، و عملا بالتقليد أو ما تقتضيه الأعراف (أي تحديد المفاهيم و ضبطها)، وتمهيدا للطرح الذي سيلي، نرى أن لا ضير في أن نسوق تعريفا بسيطا لهذا المصطلح، يتفق عليه جل المشتغلين في المجال، مفاده أن لسانيات المدونة هي تلك الدراسة التي تتناول اللغة و ظواهرها انطلاقا من المدونات.

"Corpus linguistics is simply the study of language through corpus-based research[...]"¹

نعود إلى حيث توقفنا آنفا، أي إلى الجدل القائم حول المصطلح ذاته. لسانيات المدونة – و كما سيأتي بيانه – قد تطورت في رحاب بيئة أنجلوساكسونية، لذا لا نلحظ طرعا لمثل هذا الإشكال في اللغة الانجليزية؛ إذ إن مصطلح *Corpus linguistics* محل إجماع بين أهله. و الحال نفسها بالنسبة للعربية، حيث إن المقابل العربي "لسانيات المدونة"² قد صيغ صياغة تستوفي دقة المفهوم.

أما فرنسيا، فلم يرس البحث بعد على مصطلح واحد و موحد؛ و لما كانت اللغة الفرنسية لا تجيز رصف كلمتين دونما رابط لتشكيل مصطلح مركب (على شاكلة الانجليزية أو العربية)، فإن علة ذلكم الشقاق حروف الجر (De/Des/Sur)، إذ بين قائل بـ *Linguistique de corpus*، و آخر يدعوها بـ *Linguistique des corpus* و ثالث يسميها *Linguistique sur corpus*؛ لكن جيوفري ويليامز له القول الفصل و الفيصل في هذه المسألة، من خلال مقال طريف له، موسوم بـ:

‘La linguistique et le corpus: une affaire prépositionnelle’.³

و قبل فك معضلة هذه الحروف، نرى فريقا آخر – و لسنا ندري إن كان الأخير فيكفينا شرّاً هذا التعدد المصطلحي – يزعم أن مصطلح *Les linguistiques de corpus* هو

¹ - Cf. <http://www.engl.polyu.edu.hk/corpuslinguist/corpus.htm>

² - تمت صياغة المصطلح العربي بمحاكاة عن المصطلح الانجليزي على غرار غالبية المصطلحات في مجال اللسانيات. غير أننا نشير إلى أن سلوى حمادة تعتمد مصطلحا آخر (التدوين اللغوي)، ويبدو أنها قد اشتقت "التدوين" من لفظ "مدونة"! وفي اعتقادنا أن هذا المصطلح لا يستوفي دقة المفهوم. ينظر مقالها السابق.

³ - Cf. Geoffrey C., Williams (2006) La linguistique et le corpus: Une affaire prépositionnelle. In: F. Rastier & M. Ballabriga (Dir.) *Corpus en Lettres et Sciences Sociales – Des documents numériques à l'interprétation, actes du XXVII colloque d'Albi, Langues et Signification*, p. 151.

المعادل الأنسب، أي بالجمع،¹ على غرار الكلمة الانجليزية *Linguistics*. لكن ما لا يجب أن يغيب عن الأذهان، هو أن هذه اللفظة الانجليزية و غيرها كثير (على شاكلة: *Physics Mathematics*) و إن كانت تبدو جمعا، إلا أن تركيبها - في الواقع - لا يتغير *Les mots invariables*، أو ما يعرف في اللغة العربية بـ "اسم الجمع"، و هو ما تضمن اسم الجمع و ليس له مفرد من لفظه، نحو: "خيل، جمع، قوم، جيش"، فكلها جمع لا مفرد له.*

و للعلة نفسها، فإن استخدام الحرف *de* يعني حضور اختصاص واحد، أما استخدام *des* فيعني تضافر عدة اختصاصات بدل عدة مقاربات ضمن الاختصاص الواحد. أما أن يوظف الحرف *Sur*، فمعناه أن فروعا أخرى عن الأصل (أي اللسانيات العامة) يمكن أن تستثمر المدونات، من غير أن يحيل استخدام هذا الحرف على لسانيات المدونة في حد ذاتها، و يسوقنا هذا إلى طرح مسألة طبيعة المدونات.

"[...]'de' signifiera la présence d'une discipline unique, 'des', que plusieurs disciplines - au lieu de plusieurs approches de la même discipline- sont en jeu. 'Sur' est une interprétation supplémentaire impliquant que d'autre [sic] domaines de la linguistique peuvent utiliser les corpus sans faire de la linguistique de corpus perse, ce qui soulève la question de la nature des corpus." ²

من هذا المنظور، ليس من اللغو في شيء الجزم أن ما يُعنى بمصطلح "المدونة" في "لسانيات المدونة" هو ما اصطلحنا عليه منذ أول محطة في هذا البحث،* فضلا على أن بناء تلك المدونة يشكل جزءاً لا يتجزأ من الأسس و المفاهيم الإجرائية التي تقوم عليها تلك اللسانيات. و الحال هذه، فإننا نقر ضمناً بوجود مدونات أخرى يمكن دراستها دراسة لسانية؛ مما يعني أيضا إقرارنا بوجود ما يدعوه أولئك بـ "*Linguistique sur corpus*".³ فإن نحن سلمنا بذلك، فلا حاجة بنا إلى التملص من حقيقة مفادها أن لسانيات المدونة حقل قائم بذاته، استطاع اقتحام كثير من الحقول المعرفية، بحكم سنة الاحتكاك والتقاطع التي تحكمها.

و تتبني لسانيات المدونة - شأنها شأن مثيلاتها من المعارف - على جملة من الأسس المفاهيمية و المصطلحية، و المعطيات المنهجية و الإجرائية؛ بيد أنه، و قبل أن نمضي قُدماً في الحديث عن هذه و تلك و نسترسل في سرد تفاصيلها، نرى أنه لا ضير في أن نشير - و لو عرضاً - إلى بعض المحطات في تاريخ الفكر اللساني التي أفرزت

¹ - قد ورد استعمال هذا المصطلح عند بنديكت بنسمان، في مذكرة لها لنيل درجة الدكتوراه.

Pincemin, Bommier-Bénédicte (1999) Diffusion ciblée automatique d'informations: Conception et mise en œuvre d'une linguistique textuelle pour la caractérisation des destinataires et des documents. Chapitre IV, A, pp. 415-427. Université Paris IV, Sorbonnes.

* - أما "شبه الجمع"، فهو ما تضمن معنى الجمع، وُفرّق واحده بالتاء، نحو: "ورق"، و"ثمر"، فإن المفرد "ورقة" و"ثمرة".

² - Geoffrey, Williams (2006) *Op. Cit.*, p. 151.

* - أي أنها مجموعة نصية الكترونية، تمثل الواقع اللغوي، وتجمع بغرض دراسة اللغة وظواهرها.

³ - *Ibid.*, p. 157.

هذا الحقل المعرفي، بعد أن فرشت له الأرضية على كثير من الأصعدة و الجوانب، سواءً تعلق الأمر بالجانب المفاهيمي أم المنهجي أم الإجرائي.

و نحن إذ نقف عندها، لسنا نرجو من وراء ذلك الدخول في معمعة التاريخ لذلك الفكر، بل نسعى إلى بسط جملة المعطيات التي كان لها الفضل في تأطير حقل لسانيات المدونة، و تشكيل ذلك الرصيد المعرفي و الإجرائي الذي تنفرد به، و يجعل منها اليوم واسطة العقد، و همزة الوصل بين جل الحقول المعرفية؛ أو كما يقول ج. ويليامز:

"Je ne retrace pas des origines pour faire de l'histoire, mais pour expliquer des paradigmes de recherche actuels".¹

2- المنطلقات التاريخية:

انبثقت لسانيات المدونة كاتجاه في البحث اللغوي عن اللسانيات التطبيقية، وعرفت تطوراً واسعاً في رحاب البيئة الأنجلوساكسونية* انطلاقاً من تقليدين كبيرين هما: تدريس اللغة الانجليزية كلغة ثانية (لاسيما في فترة ما بعد الحرب العالمية الأولى)، و المقاربة السياقية في اللسانيات التي تنسب إلى فيرث Firth، الذي لا ينكر أحد مدى إسهاماته في الميدان، لاسيما بعد أن طور مفهومي السياق الثقافي و سياق المقام، اللذين اقترضهما عن مالينوسكي Malinowski الأنثروبولوجي الشهير؛ فكان أن برزت على يده إلى الوجود "السياقية" Contextualisme معارضة المقاربة العقلانية.

و لعل أبرز علمين من أعلام لسانيات فيرث هما تلميذاه: هاليداي (M. A. K. Halliday) و سنكلير (J. Sinclair)، اللذين حملا المشعل من بعده، و صنعا في ميدان اللسانيات أعاجيب لا يزال يسير على هديها كثيرٌ من الباحثين التابعين؛ فهاليداي رائد النحو النسقي الوظيفي Grammaire systémique et fonctionnelle، و هو نحو وصفي تعتمد لسانيات المدونة السياقية la linguistique de corpus contextualiste، بمعنى أن الرجل قد اهتم بالجانب النحوي. أما سنكلير فقد تولى أمر العناية بالجانب المعجمي،* لذا يعد رائد تحليل المدونة السياقية.²

أ- هاليداي و النحو النسقي الوظيفي:

النحو النسقي الوظيفي نحو وصفي نصي، يولي اهتمام بالغا بفكرة "السياق"، وهو ما نتوسمه في مقدمة مؤلف هاليداي الموسوم بـ: "مدخل إلى النحو الوظيفي".

"Just as each text has its environment, the 'context of situation' in Malinowski's terms, so the overall language system has its environment, Malinowski's 'context of culture'. The context of culture determines the nature of the code. As a

¹ - Geoffrey, Williams (2006) Ibid., p. 156.

* - ليس يجدي نفعاً البحث في من كان له السبق إلى استخدام المدونات FRANTEXT, Brown, OSTI... لا سيما بالنظر إلى العزلة النسبية التي كان يعيشها الباحثون زمانئذ في غياب البريد الإلكتروني. ثم إننا حين نتحدث عن تطورها في رحاب البيئة الأنجلوساكسونية، نشير ضمناً إلى أن أوائل اللسانيين قد جمعوا مدوناتهم الورقية الخاصة، إلا أن ذلك كان في حدود قدرات الباحثين قبل ظهور الحواسيب وانتشارها.

* - لا يتبادرن إلى ذهن القارئ، أن ما تعنى به لسانيات المدونة هو وصف المستوى المعجمي، بل إنها تتجاوز ذلك.

² - نحيل القارئ إلى مقالتهما القيمتين كل في مجال اختصاصه:

M. A. K. Halliday (1966) Lexis as a linguistic level. In: C. E. Bazell, J. C. Catford, M. A. K. Halliday & R. H. Robins (eds.) In Memory of JR FIRTH (pp. 148-162). London: Longman.

J. Sinclair (1966) Beginning the study of lexis. In: C. E. Bazell, J. C. Catford, M. A. K. Halliday & R. H. Robins (eds.) Op. Cit. (pp. 410-430). London: Longman.

language is manifested through its texts, a culture is manifested through its situations; so by attending to text-in-situation a child construes the code, and by using the code to interpret text he construes the culture".¹

"بما أن كل نص ذو فضاء خاص به، أو ما يعرف بـ 'سياق المقام' - حسب مالينوسكي -، فإن النظام اللساني بأكمله ذو فضاء هو الآخر، يعرف بـ 'السياق الثقافي' - وفق ما يذهب إليه مالينوسكي دائماً- و ذلكم سياق يحدد طبيعة السنن. و كما أن اللغة تتبدى من خلال نصوصها، فإن الثقافة بدورها تتمظهر من خلال مقاماتها؛ و هكذا فإن الطفل حين ينتبه إلى النص ضمن سياقه، يفسر السنن؛ و بتوظيف هذه في تأويل النص، يفسر الثقافة." [الترجمة لنا]

و التحليل ضمن هذا النموذج ينطلق من النص إلى الجملة فالكلمة، آخذا بعين الاعتبار التأثير المتبادل - لا محالة - بين النحو و المعجم *La lexis*؛ لذلك فإن النص - وفق ما يذهب إليه هاليداي- وحدة معنوية و ليست نحوية. و المعاني تتجسد من خلال الكلمات و التعابير؛ و في غياب نظرية لتلك الكلمات و التعابير - أي النحو- يستحيل تأويل معنى أي نص.

"A text is a semantic unit, not a grammatical one. But meanings are realized through wordings and without a theory of wordings -that is, a grammar - there is no way of making one's interpretation of the meaning of a text".²

و هذا النموذج نسقي متعدد المستويات، و لا شك أن ثمة تأثير متبادل بين مكونات النص جميعها، و بين النص و فضائه. و الحال هذه، يضطلع الاتساق النصي بدور رئيس في النحو. و باعتباره نحواً وصفياً، يحتل النحو النسقي الوظيفي مكانة هامة في درس المدونات.

ب- سنكلير و تقرير أستي *OSTI Report*:

كان سنكلير قد أكد في المقال ذاته و المشار إليه أعلاه (1966) أن البحث السياقي لا يزال في أول عهده.³ و سرعان ما اهتدى إلى أن استخدام الحاسوب قد يدفع بعجلته إلى الماضي قديماً؛ و في حديث بينه و بين ويليامز جيوفري، صرح الباحث بأن امتلاك جهاز حاسوب يومئذ و تخصيصه للبحث اللساني، كان أعجوبة من الأعاجيب، لاسيما إذا سلمنا بأن تلك التقنية كانت حكراً على نخبة من الباحثين في العلوم التجريبية.

"A l'époque, le fait d'avoir un ordinateur dédié à un projet linguistique était quelque chose d'extraordinaire dans un

¹ - M.A.K. Halliday (1985) *An Introduction to Functional Grammar*. London: Baltimore, Md/UAS: Edward Arnold. Preface: xxxi.

² - Ibid., Preface: xvii.

³ - Cf. J. Sinclair (1966) *Op. Cit.*

*monde où uniquement les élites des sciences dures y avaient accès."*¹

و هكذا، قام بصناعة مدونة إلكترونية، أجرى عليها سلسلة من الدراسات و نشر نتائجها إبان السبعينات ضمن تقرير – و إن لم يلق رواجاً زمانئذ لتتم إعادة نشره حديثاً (2004) – كان قد وضع أسس لسانيات المدونة السياقية؛ و هو يحمل عنوان: تقرير أستي حول مشروع بحث المعجم، عن مكتب المعلومة العلمية و التقنية *Office for Scientific and Technical Information*.²

ج- مدرسة برمنغهام و مشروع كوبيلد:

في إطار التعاون بينها و بين معاجم كولينز *Collins*، سعت جامعة برمنغهام من خلال مشروع كوبيلد (*COBUILD*) إلى بناء مدونة ضخمة، بغرض استخدامها كمرجع في اللغة الانجليزية، و استثمارها في صناعة معجم للمتعلمين قوامه المعطيات نتاج تحليل المدونة. و تلك نزعة بخلاف النزعة التي كانت سائدة آنئذ، حيث كان يُعتمدُ فيها كثيراً على حدس المعجماتي. و علاوة على ذلك، يعود الفضل إلى كوبيلد في تخريج أول معجم يُنزل المتلازمات اللفظية منزلتها الحق.

و في رحاب مدرسة برمنغهام³ نفسها، تحقق الحلم الذي طالما راود فيرث، بأن توشجت العلائق بين اللسانيات و المعجماتية و توحدتا فيما بينها.

و ما من شك في أن ما تحقق بفضل كوبيلد، قد دفع بدور النشر الأخرى إلى إعادة جدولة أعمالها وفق الخطة نفسها؛ فكان أن ظهرت إلى الوجود المدونة البريطانية الوطنية *BNC* في إطار اتحاد آخر، و دأب الباحثون يطورون مشروع كوبيلد، إلى أن خرجوه في زي "بنك اللغة الانجليزية" (*Bank of English*).

و جدير بالإشارة، أنه في الوقت الذي كانت مدونة كوبيلد أكبر المدونات حجماً آنذاك، تمت صناعة مدونات أخرى أصغر، لغرض الأبحاث في لغات الاختصاص بمدرسة برمنغهام.

د- و المدونة هذه الأيام:

إن اعتماد تقنية الحاسوب قد أعطى للسانيات المدونة نفساً جديداً و دفعا فريداً، لاسيما بعد ظهور المفهرسات الآلية *Les concordanciers*. و ما تلك إلا أداة أو وسيلة لبلوغ الغاية و المرام: تحليل اللغة قيد الاستعمال أو في سياق التواصل. و لا شك أن لويليامز جيوفري عبارة طريفة في هذا الصدد نقترضها لما تحمله من دلالات و إحياءات، إذ يقول:

*" En linguistique de corpus contextualiste, l'outil informatique n'est qu'une loupe pour mieux voir".*⁴

¹ - G. Williams (2006) *Op. Cit.*, p. 154.

² - Cf. J. McH. Sinclair, S. Jones & R. Daley (2004) *English Collocation Studies: The OSTI Report*. London/ New York: Continuum.

J. McH. Sinclair *et al.* (1970) *English Lexical Studies: Report to OSTI on Project C/LP/08*. Department of English, University of Birmingham.

³ - نشأت المدرسة في الستينات مع سنكلير و كولثارد (*Coulthard*)، وكانت تعنى بتطبيقات اللسانيات التطبيقية في التدريس، كما أنها كانت تهتم بتحليل الخطاب (الاسيما العلمي) بنية تدريس لغات الاختصاص.

⁴ - G. Williams (2006) *Op. Cit.*, p. 155.

و على صعيد آخر، ننبه إلى أن البحث في ميدان لسانيات المدونة، يتوزع اليوم عبر خمسة مراكز كبرى:

- جامعة برمنغهام مع فريق سنكلير، ثم خليفته توبير Teubert.
- جامعة أستون Aston ببرمنغهام مع روي Roe.
- جامعة ليفربول مع هوي Hoey و سكوت Scott.
- مركز لنكستر Lancaster (يتبنى أعمال Leech في مقاربته، التي تعتمد كثيرا على المعالجة الآلية للغات (TAL).
- و أكسفورد.

أما المراكز الثلاثة الأولى فسياقية إلى "حد النخاع"، و قلما تتدخل على مستوى المدونات، تأسيسًا بسنكلير الذي يعرضّ بالنواجز على فكرة عدم ترميزها (l'idée de zéro annotation).¹ و يبقى الهدف الذي تسعى من أجله: تدريس اللغات، لاسيما لغات الاختصاص منها، و كذا تطوير المعجماتية. أما مركز "لنكستر" فينزع إلى مقاربات أوسع في صناعة أدوات الترميز، و كذا التطبيقات في المعالجة الآلية للغات. إن هذه الرؤى الاستشرافية، و الإسهامات المعرفية، هي بمثابة أولى الإرهاصات في العمل اللساني باعتماد المدونات الالكترونية، و بواكير نشأة الاتجاه اللساني المدوناتي؛ لذا كان لزاما علينا الوقوف عندها - و لو لبرهة - وفق ما يقتضيه المقام المنهجي.

3- المنهج و الأدوات الإجرائية:

1-3 المنهج:

يعد "المعيار" و "الاستعمال" من المفاهيم التي فصل فيها اللسانيون و رسموا حدودها منذ أمد طويل. و ذهبوا إلى أن ما يميز المعيار عن الاستعمال، هو نفسه ما يميز اللغة عن الكلام؛ إذ يحصل لدينا من جهة ما هو متصل بالنظري، أي نتاج الدراسات الوصفية ممثلا في "المعيار - اللغة"؛ و النتاج الفعلي المنجز والملاحظ ممثلا في "الاستعمال - الكلام"، من جهة أخرى.²

و لسانيات المدونة - كما أشرنا غير ذي مرة - تدرس اللغة كما هي في عالمها الحقيقي، أي في سياقاتها الطبيعية.

"La linguistique de corpus est une linguistique appliquée, la théorie est issue de la pratique, et non l'inverse. La langue est atteinte à travers la parole [...] et n'a pas d'existence propre en dehors du contexte".³

¹ - Cf. J. McH., Sinclair (ed.) (2005b) *How to use corpora in language teaching*. Amsterdam: John Benjamins.

² - Cf. Christian, Vicente (2006) *Linguistique de corpus et traduction spécialisée: Eléments pour approche traductologique du lexique spécialisé*. In: D. Gouadec (Dir.) *Traduire pour le Web*. Paris : La Maison du Dictionnaire, pp. 147-156.

³ - G. Williams (2006) *Op. Cit.*, p. 156.

بمعنى أنها تتناول الظاهرة اللغوية، و تعكف على دراسة الأداء اللغوي *Performance* دون العمليات الذهنية؛ و هنا تتجلى المفارقة المنهجية في البحث التي انتهجها كل من النموذج اللساني المدوناتي و النموذج التوليدي التحويلي.¹ فكثيرا ما نسمع عن الثورة التي أحدثها نعوم تشومسكي في الأوساط الألسنية، حيث راح يهاجم تلك المبادئ التي تركز عليها المدرسة التجريبية في تحليل الوقائع اللغوية، كما هاجم أولئك اللسانيين الذين كانوا يعتمدون على مدونات صغيرة من الأمثلة أو النصوص.²

و هو بذلك يذهب – في ما يذهب إليه – إلى دحض فكرة التحليل اللغوي باعتماد الاتجاه الوصفي و التجريبي، طالما أنه يرى أن الهدف ليس وصف اللغة في حد ذاتها؛ بل حري بالباحث أن يحلل اللغة بناءً على حدس المتكلم بها/المستمع، و كذا القدرة أو الآلية الضمنية التي يمتلكها فطريا و تتيح له إنتاج عدد لا حد له من الجمل في لغته.³ و تشومسكي إذ يُعوّل كثيرا على الكفاءة اللغوية دون الأداء، إنما يعوّل على متكلم/مستمع نموذجي، ليقصي بذلك كل شذوذ يحتمل في الوقائع اللغوية داخل المجتمع الإنساني.

و فضلا على ذلك، فإنه يؤمن –في ما يؤمن– بمقولة الكليات النحوية *Universals*، و هي جملة من المبادئ الثابتة و القواعد العامة يراها تحكم لغات العالم أجمع، و طفق يبحث عن "نحو كلي" قوامه جملة تلك الكليات.

و لا ريب في أن التوليديين قد أشاحوا عن لسانيات المدونة أسوةً بفكرة لبالي،⁴ مفادها:

"La linguistique a pour unique et véritable objet la langue envisagée en elle-même et pour elle-même."

و عليه، عكف أولئك يبحثون في اللغة و النظام، و نَسِوا إلى درس الكفاءة اللغوية وفحصها، و ما قدّروا المدونات حق قدرها و لو أنها تمثل أحسن تمثيل الملفوظات الحقيقية، و الأداءات الفردية المنجزة في إطار واقع اجتماعي- تاريخي مُحدّد و مُحدّد – على حد تعبير دامون مايافر *Damon Mayaffre*:-

"[...] le corpus, en tant que recueil d'énoncés effectivement émis, en tant que recueil de performances individuelles produites par des locuteurs debout dans une réalité sociale et

¹ - Cf. Tamás, Váradi (2001) The linguistic relevance of corpus linguistics. In: P. Rayson, A. Wilson, T. McEnery, A. Hardie & S. Khoja (eds.) *Proceedings of the Corpus Linguistics 2001 Conference* (pp. 587-594). Lancaster: Lancaster University.

² - Cf. Maia, Belinda (2002) Do-it yourself disposable, specialised mini-corpora — Where next? ...In: Tagnin, S.E.O (Org.) *Cadernos de tradução: corpora e tradução*. Florianópolis: NUT, 2002, 1(9), p. 220.

³ - ينظر: أحمد حساني، دراسات في اللسانيات التطبيقية، ص 75.

⁴ - بخلاف الاعتقاد السائد بأن الفكرة لدي سوسير. ينظر:

S. Bouquet (2005) Après un siècle, les manuscrits de Saussure reviennent bouleverser la linguistique. *Texto!* http://www.revue-texto.net/saussure/Sur_Saussure/Bouquet_Après.html

historique déterminée (et déterminante), était non seulement négligé mais d'une certaine manière banni".¹ [Emphasis added]

كأنما لشد ما كان يخشاه أولئك، أن تلطخ المدونات نحوهم الكلي بالثقافة والمجتمع، أو أن تنقل إليه بعض عدوى الشذاذ، أو أن تشوبه شائبة الخيار والانتقاء، و في ذلك تشويش لـ "النظام" (اللساني طبعاً)، و إخلال بما تعاقدا عليه.² ذلك أن المدونات تتعلق بلسانيات الكلام، و تلكم لا تُمْتُ إلى لسانياتهم بصلة، إنما هي تأصر لسانيات الاجتماع والنفس، بل و تحليل الخطاب و التداوليات والأسلوبيات، و قس على ذلك؛ و هذا نوام تشومسكي بعينه قد صرح ذات عهد، لسنا نراه ببعيد(1999)، قائلاً:

*"Corpus linguistics does not exist!"*³

إنها "حرب باردة" – على حد تعبير مايافر – تلك التي اشتملت بين أنصار النظرية و دعاة التجريب، بين أشياح الاستنتاج و أتباع الاستقراء. لكن ما ذهب إليه الرجل (أي تشومسكي) في كثير من السياقات، و هو يسعى جاهداً إلى إرساء قواعد "نظريته النموذجية" – كما يحلو له أن يدعوها في كثير من المقامات –، هي تحصيل حاصل لا اعتناقه النزعة اللسانية العقلانية، تأثراً بالفلسفة الديكارتية.⁴

و حقيق بنا أن نعترف بأن نزعته تلك ليست وليدة رغبة إحياء تقاليد الماضي أو تشبثاً بها، بل قد فرضها واقع الدراسات اللسانية زمانئذ. ثم إن اعتراضه على البحث اللساني باعتماد المدونات – أيام الناس تلك – لمشروع من وجهين: أولهما، صغر حجم تلك المدونات؛ و ثانيهما، أن ذلكم البحث كان معظمه يتم وفق طرق تقليدية. أما و اليوم، وقد تواسجت القرابة بين تكنولوجيا الحواسيب و المدونات النصية، تغيرت النظرة إلى لسانيات المدونة، و غدت تستقطب اهتمام كثير من الباحثين.

¹ - Damon, Mayaffre (2005) Rôle et place des corpus en linguistique: Réflexions inductives. *Texto!* [En ligne], décembre 2005, vol. x, n. 4. Consultée le 10/11/2006.

http://www.revue-texto.net/Reperes/Themes/Mayaffre_Corpus.html

² - Cf. *Ibid.*

³ - Chomsky, entretien avec Baas Aarts, cité par F. Rastier (2005a:40) Enjeux épistémologiques de la linguistique de corpus. In : G. Williams (éd.) *La linguistique de corpus*, Rennes, PUR, pp. 31-45. [En ligne sur] *Texto!*

http://www.revue-texto.net/Inedits/Rastier_Enjeux.html

⁴ - من الطبيعي والمنطقي أن يتأثر كثير من اللسانيين بالمدارس الفلسفية نظراً للتداخل أو التقاطع المعرفي، فاعتقاد تشومسكي الراسخ – مثلاً – بأن الحدس هو المصدر المناسب في تحليل الظواهر اللغوية، لا تفسير له إلا تأثره الواضح بمذهب العقلانيين حين يقولون بأن أصل العلم الإنساني ومصدره الأول هو العقل. و هذا المذهب يتعارض ومذهب التجريبيين الذين يأخذون بمسألة أن التجربة هي المصدر الجوهري والأسبق لكل معرفة، وهو يدعى أيضاً بمذهب الحسيين، إذا اعتبر الإدراك عن طريق الحواس أصل كل علم.

و الحق أن اختبار مسلمات كل مذهب قد أكد صحتها وصدقها بالنظر إلى سياقاتها ونسقاتها؛ مما يعني أن تبني أحدهما دون الآخر ليس إلا موقفاً فلسفياً شخصياً نابعا عن اقتناع براهيه. بيد أن الواقع يثبت أن مشكلة التعارض/التقابل بين هاتين النزعتين التقليدية، يتعين تجاوزها. وذلك ما توصل إليه الفيلسوف الألماني إيمانويل كانط *Emmanuel Kant* (1724-1804) صاحب المذهب النقدي، حين استطاع التوفيق بين المذهبين و إصلاح ذات بينهما، وقد برهن على تكاملهما بجميع متناقضاتها، بقوله أن "العلم نتيجة اجتماع عاملين أحدهما بصوري [أي شكلي، يتعلق بالقواعد المنطقية والقوالب الفكرية] يرجع إلى طبيعة العقل ذاته، والآخر مادي يتكون من الإحساسات الداخلة في الإدراكات الحسية. فإذا لم يوجد أحد هذين العاملين، استحال وجود علم حقيقي [...]". و لتقريب ذلك إلى الذهن، نضرب مثلاً: ليست الملاحظة مجرد مشاهدة بل تتضمن أيضاً المراقبة الذهنية، أي أن الباحث يلاحظ بواسطة حاسة البصر مثلاً، ثم يتدخل فكره يوحى إليه بفكرة يفترضها من أجل تفسير ما يبحث عن فهمه؛ وهنا ننتبين تكامل الحس والفكر وتقاربهما.

لمزيد من التفاصيل، نحيل القارئ على مؤلف قيم للباحث الألماني أرفلد كولبي *Oswald Kulpe* (1915-1982) في: المدخل إلى الفلسفة، تر: أبو العلا عفيفي، ط5، مكتبة النهضة المصرية القاهرة.

و الواقع أن ما جاءت به النظرية التوليدية التحويلية لم يكن يوماً ليقف حجر عثرة أمام تطور لسانيات المدونة (السياقية خصوصاً) - وفق ما يشير إليه ليون Léon¹، اللهم إلا بعض الألق الذي فقدته لبرهة، و تلكم طبيعة كل بحث يهفو و يكبو حتى يستقيم عوده.

و بخلاف لسانيات تشومسكي، لا تستند مقاربة الوقائع اللغوية باعتماد المدونة إلى الأحكام النابعة من الحدس، تدعمها الأمثلة الموضوعة (Invented examples)؛ بل إنها تتم بناءً على قاعدة نصية معتبرة، تتضمن أمثلة حقيقية (Authentic examples) من الاستخدام الفعلي لتلك اللغة.

و على هذا النحو، بمقدور الباحث أن يقارب اللغة من وجهة لسانية بحث، أو من منظور الخطاب، باعتبارها ظاهرة اجتماعية ثقافية - على حد تعبير إيفان Evans:-

"The language in a corpus can be studied from both a purely linguistic point of view and from the perspective of discourse as a social and cultural phenomenon."²

و يمكن أن نجل ضمن جدول تلخيصي - أدناه - أهم الفروقات أو أوجه الاختلاف بين الاتجاهين: التوليدي التحويلي، و اللساني المدوناتي؛ و هي تشكل في الوقت نفسه، المرتكزات أو المبادئ التي يقوم عليها كل اتجاه.³

نظرية النحو التوليدي - التحويلي	لسانيات المدونة
- اللغة (الكفاءة).	- الكلام (الأداء).
- مستمع / متكلم مثالي .	- التنوع .
- اللغة : ملكة ذهنية فطرية.	- اللغة : ظاهرة اجتماعية.
- الحدس .	- التجريب .
- الكليات .	- الاختلافات.
- النحو (الكلي).	- المعنى.
← نزعة عقلانية .	← نزعة تجريبية.

الجدول (1) النظرية التوليدية التحويلية ولسانيات المدونة.

و الواقع أنه ليس من الغلوّ في شيء القول إن اعتماد المدونات في التحليل اللغوي - لا محالة- قد فرضته الحاجة الماسة إلى منهج قوامه التجريب، بعد أن تأكد لدى الباحثين (على غرار سنكلير) أنه مهما بلغت قوة حدس المتكلم أو قدرته الضمنية، فإنها لا تكفي كمصدر لمعالجة الوقائع اللغوية، أو كدليل يُهتدى من خلاله إلى حقائقها.

¹ - Cf. J. Léon (2005) Claimed and unclaimed sources of Corpus Linguistics. *Henry Sweet Society Bulletin*, n° 44, pp. 36-50.

² - D. Evans (2006) *Op. Cit.*

³ - الجدول مقترح عن شارمن لي مع بعض التصرف:

- Sherman, Lee (without a year) *Corpus Linguistics: Developing a PolyU Language Bank*. (Power Point, p. 3.)

• - و التجريب L'expérimentation غير التجربة L'expérience ؛ إذ إن التجربة تستغرق مفهوم "التجريب" على اعتبار أنه العمل المحدد بشروط و وسائل معينة، تخضع فيه الفروض للتحقق؛ فضلاً على مفهوم الاستدلال التجريبي ويعنى به حوصلة ما أسفر عنه التجريب من نتائج قصد صياغة قانون أو نظرية.

"[...] Probably the most frequent tenet found in the literature is **the need for empiricism** on the grounds that native speaker intuition about one's language are generally a poor guide to linguistic reality[...]." ¹ [Emphasis added]

بناءً على ما سلف من معطيات، و دائماً في سياق المقارنة بين الاتجاهين، نتبين أن لسانيات المدونة لا ترمي إلى التنظير حول ما هو ممكن في اللغة، كما هي الحال بالنسبة لنحو تشومسكي الذي يمكن أن يولد عددا لا حصر له من الجمل؛ بيد أنه – و كما يقول Krieger – نحو لا يأخذ في الحسبان الخيارات المحتملة التي يتبناها فعليا الناطقون بتلك اللغة.²

فهي – أي لسانيات المدونة – تعنى باكتشاف تنميطات الاستعمال اللغوية (المعجمية، التركيبية، البنوية،...) التي توفرها المعطيات التجريبية؛ كما تعنى بما يمكن استخلاصه بخصوص السلوك اللغوي، كأن نتبين إلى أي مدى تستعمل تلك التتميطات، أو نهتدي إلى العوامل السياقية التي تحكم تغيراتها، و ما إلى ذلك. و في هذا الصدد، يقول مايف أولوهان (Maeve Olohan) ما نصه:

"According to Michael Stubbs [...], corpus linguistics is concerned with 'what frequently and typically occurs' as opposed to isolated, unique instances of language [...]."³

و في السياق نفسه، يميز سنكلير بين صنفين من الاهتمامات أو الأهداف التي يعنى بها المشتغلون في لسانيات المدونة:⁴

- أهداف خاصة، يعنى فيها بالنتيجة المتحصل عليها؛ مثال ذلك: صنع مراقب الإملاء (Spell checker).

- أهداف عامة، و هي تشمل مهمات و وظائف أكثر تعقيدا؛ على شاكلة: بحث شروط الترجمة الآلية، أو التواصل بين الإنسان و الآلة. فضلا على دراسة "اللغة لذاتها" – على حد تعبير سنكلير –، و كذا "تطبيق النظريات السابقة على المدونات بغية التحقق منها وتطويرها".

و الحال هذه، ليس من المغالاة الجزم بأن استثمار المدونات قد غيّر كثيرا من نظرتنا إزاء اللغات، كما رسم لنا آفاقا غير التي عهدناها في مقاربتنا للنحو و المعجماتية و دراسة النصوص، بل و مكانة اللغة في المجتمع.

و للسائل أن يتساءل: ما حدود ذلك الاستثمار أو الآفاق حتى؟ لعل الإجابة عن هذا السؤال نستشفها في ثنايا سؤال آخر، صيغته: ما حاجتنا إلى رسم الحدود في استثمار المدونات،

¹ - Bill, Dodd (1999) Op. Cit.

² - Daniel, Krieger (2003) Corpus Linguistics: What it is and How it can be applied to teaching. *The International TESL Journal*, vol. IX, n° 3, March 2003. Viewable online at: <http://iteslj.org/Articles/Krieger-Corpus.html>

³ - Maeve, Olohan (2002a) Op. Cit., p. 5.

⁴ - Cf. Sandra, Halverson (1998) Translation studies and representative corpora: Establishing links between translation corpora, theoretical/descriptive categories and a conception of the object of study. *Meta* 43(4), p. 407.

طالما أن تلك المدونات ذاتها لا حدود لها، شأنها شأن اللغة لا تنفك تتطور – على حد قول سنكلير:-

"[...] the corpus 'has no final extent because, like the language itself, it keeps on developing' (Sinclair 1991:25)."¹

ثم إن اقتحام لسانيات المدونة لكثير من الحقول المعرفية – كما سنتبينه في محطة لاحقة من هذا البحث – يدحض فكرة الحدود تلك، ويتجاوزها لي طرح قضايا أخرى أهم. و رغم كل ما جادت به لسانيات المدونة – و لا تزال – خدمة للغة و علوم أخرى، إلا أن طائفة من الباحثين – و طبيعي أن يكونوا أكثر عددا و أعز نفرا – يشككون في ما إذا كانت علما مستقلا من العلوم اللغوية، و يعتبرونها في مقابل ذلك مجرد طريقة لدراسة اللغة.²

و نحسب هذا التصور زعما بل وهما، بالنظر إلى كل ما أسلفنا، أو ما سنأتي على ذكره من معطيات، تفند ما ذهب إليه أولئك، و تترر بلا شك حقيقة مفادها أن: لسانيات المدونة تمتاز برصيد مفاهيمي و إجرائي لا يستهان به، مما يجعل منها حقلا معرفيا قائما بذاته، مثله كمثل الحقول المعرفية الأخرى.

و لو افترضنا أن لسانيات المدونة هي على النحو الذي يزعمون، و سلمنا مبدئيا بما يقولون، و أمعنا النظر فرأينا أنه لا ينكر عاقل إمكانية استخدام المدونات كأداة ضمن مقاربة إمبريقية في الدراسات اللغوية، لعدنا أدراجنا و قلنا بما سبق و أقررنا: إن تلكم المقاربة لا تنتظم أو تستقيم إلا في رحاب البحث اللساني باعتماد المدونة؛ بحيث إنها تنبني أولا على ملاحظة الواقعة المشاهدة، أي الظاهرة اللغوية، فصياغة الفرضيات بشأنها (الافتراض)، و هي بمثابة تفسير مؤقت للظاهرة المبحوثة، يحتل الصحة فيصير قانونا يحكم مجرى الظاهرة و ما شابهها، أو البطلان فيُعدّل أو يُعَدّل عنه إلى غيره.³ و للثبوت من ذلك، لا بد من إخضاع تلك الفرضيات للتجريب بالنظر ضمن المعطيات الحقيقية بواسطة الأدوات الخاصة (المفهرسات الآلية)*؛ فإذا أسفر هذا التحقق عن مطابقة بينهما (أي بين الفرضيات و المعطيات)، و رأينا أنها تتسم بالعموم و تُسحب على المجموعة ككل، تحولت من مجرد افتراضات مبدئية إلى نظرية أو على الأقل حكم نظري.

و الحقيقة أن هذا الحكم النظري ذاته، يمكن أن يستند إليه الباحث في وضع فروض أخرى، و هكذا دواليك. أضف إلى ذلك، فإن ما برّره التجريب و ثبّته التعميم، يمكن أن يسهم في الحصول على تنبؤات جديدة Prediction. و الشكل التبسيطي الموالي يوضح هذه الدورة و عناصرها:⁴

¹ - C. Tribble (1997) Improvising corpora for ELT: Quick-and-dirty ways of developing corpora for language teaching. In: J. Melia & B. Lewandowska-Tomaszczyk (ed.) PALC' 97 Proceedings. Lodz: Lodz University Press. Online at:

<http://www.tribble.co.uk/text/Palc.htm>

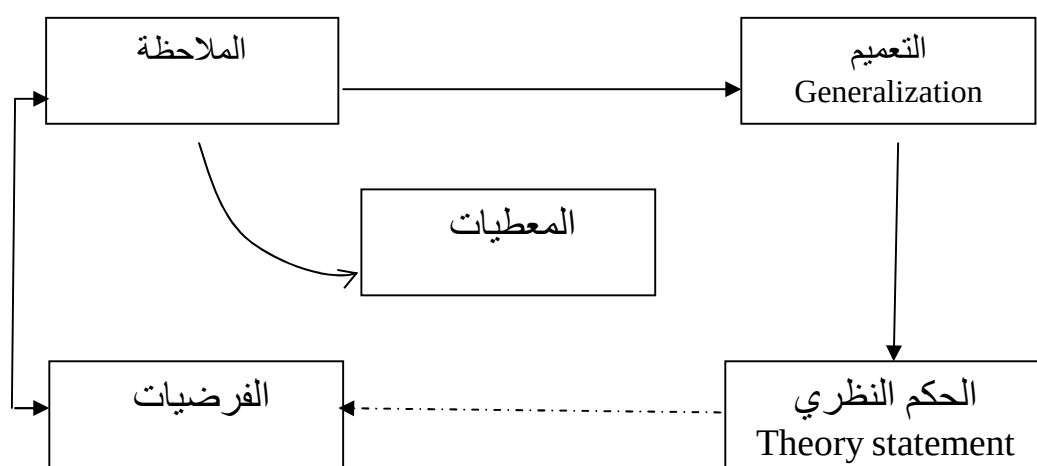
² - Cf. W. John, MacMullen (2003) Op. Cit.

³ - يعرف ماخ (ت. 1916) الفرضية بأنها " تفسير مؤقت لوقائع معينة بمعزل عن امتحان الوقائع، حتى إذا ما امتحن في الوقائع أصبح من بعد، إما فرضا فاشلا يجب العدول عنه إلى غيره، وإما قانونا يفسر مجرى الظواهر. " نقلا عن: عبد الرحمن بدوي، مناهج البحث العلمي، وكالة المطبوعات، ط3، الكويت، 1977، ص 145.

* - يشترط أن ينوع المجرب عمله لتأكيد طابع الشمولية.

⁴ - المخطط عام يمكن تطبيقه في دراسات لسانية و غير لسانية. ينظر:

W. John, MacMullen (2003) Op. Cit.



الشكل (6) مخطط عام للنموذج الإمبريقي من اقتراح Tognini-Bonelli.

إن هذا المخطط عام يعكس جليا خطوات المنهج التجريبي، و هو منهج ينبني أساسا على الاستدلال الاستقرائي،* بمعنى انتقال الفكر من حكم جزئي (أي على كل فرد من أفراد مجموعة معينة) إلى حكم كلي يتناول كل أفراد تلك المجموعة.¹ و الاستقراء في الفكر القديم، أي عند أرسطو بصفته أول من وظف المصطلح، معناه:

"إقامة قضية عامة ليس عن طريق الاستنباط، و إنما بالالتجاء [كذا وردت] إلى الأمثلة الجزئية التي يكمن فيها صدق تلك القضية العامة، أو البرهنة على أن قضية ما صادقة صدقا كليا بإثبات أنها صادقة في كل حالة جزئية إثباتا تجريبييا."² أما في الفكر الحديث، فيعرفه فرنسيس بيكون أنه:

"[...] استخلاص القواعد العامة من الأحكام الجزئية."³

هذا عن طبيعة الاستقراء من حيث المفهوم؛ أما عن طبيعته من الناحية الإجرائية، فإن المنهج الاستقرائي – و كما رأينا – ينبني على ثلاث خطوات رئيسة هي: ملاحظة الوقائع، فوضع الفروض، فالتجريب لاستخلاص القوانين. و بعضهم يجعل استخلاص القوانين أو النظريات خطوة رابعة، إلا أن من يتبنى التقسيم الثلاثي، يعدها لصيقة بخطوة التجريب و تابعة لها.*

* - ثمة ثلاثة أنواع متميزة من الاستدلالات (البراهين):

- استدلال استنتاجي: ينتقل فيه من العام إلى الخاص، وهو استدلال تقوم عليه المنهجية التي يتبعها الرياضي في البرهنة على استنتاجاته.

- استدلال تمثيلي: ينتقل فيه من الخاص إلى الخاص.

- استدلال استقرائي: ينتقل فيه من الخاص إلى العام؛ وهذا النوع هو الذي يقوم عليه العلم من خلال المنهج التجريبي الذي يوظفه، لأنه استدلال يتسم بالخصوصية و الإنتاج. ويطلق اسم المنهج الاستقرائي على المنهج التجريبي. ينظر: جمال الدين، بوقلي حسن، إشكاليات فلسفية، ط1، الديوان الوطني للمطبوعات المدرسية، الجزائر، 2007.

¹ - ينظر: جمال الدين، بوقلي حسن، نفسه، ص ص 68-70.

² - أرسطو، منطق أرسطو، تحقيق عبد الرحمن بدوي، ج3، مكتبة النهضة المصرية، 1952، ص 713.

³ - مهدي، فضل الله، مدخل إلى علم المنطق، ط3، دار الطليعة، بيروت، 1985، ص 244.

* - قد يستغني الباحث عن بعض خطوات هذا المنهج كالملاحظة مثلا، ويستبدلها بغيرها تماشيا مع طبيعة الموضوع الذي يدرسه، وخصوصيات العلم الذي ينتمي إليه؛ فيبتغي عليه اللجوء إلى الاستدلال عندما لا يتمكن من المشاهدة الحسية المباشرة، كما هي الحال في رؤية مركز الأرض لمعرفة هل هو صلب، سائل، غاز، [...] و يستطيع أن يعرف ذلك عن طريق الآثار التي تتركها هذه الظواهر. إلا أن

و ثمة ميزة هامة في النتيجة التي يؤول إليها المنهج الاستقرائي، بوصفها تتضمن عنصري التعميم و التنبؤ.

أما التعميم فيعني به "قابلية انتظام الظواهر في المكان"، أي أنه في دراستنا لظاهرة في مكان (أ)، فإن تفسيرها ينطبق على كل الظواهر المتشابهة في أي مكان و ضمن الشروط المتماثلة. أما التنبؤ فيعني به "قابلية انتظام الأشياء في الزمان"، أي أن المستقرئ ليس في حاجة إلى دراسة الظاهرة في كل لحظات الزمان، بل حسب أنه يلاحظ ظاهرة ما في زمن ما، ليتوصل إلى أن القوانين التي تسير وفقها، ستكون دائماً على هذا النحو على مدى الزمان.¹

فضلا على ذلك، يقترض المنهج الاستقرائي من الرياضيات لغتها الدقيقة، ليصب في قوالبها قوانينه. و يتجلى ذلك خاصة من خلال توظيف الإحصاء، الذي يضيف مسحة خاصة على الدراسات التي تعتمد، بحيث تُسجل الحقائق المتعلقة بالظواهر المبحوثة في صورة قياسية عددية، و تُلخص وفق طريقة تسمح بالإلمام باتجاهات هذه الظواهر والعلاقات القائمة بينها،² و كذا بضبط مقاديرها و تغييرها كمياً، تكشف دلالاتها عن درجة من الموضوعية و اليقين.

تلکم هي أهم مقومات المنهج التجريبي. و بإسقاط تصوراته على الظاهرة اللغوية، و باعتماد المخطط السابق (الشكل-6)، نتحصل على جملة من المسلمات بخصوص البحث اللساني المدوناتي:

أ- تتسم المعطيات التي تخضع للملاحظة و التجريب بالدقة، و هي ممثلة للمجموعة موضوع الدراسة.

ب- ترد المعطيات ضمن سياقاتها؛ و من ثمة، فهي تسهم في إثبات أو نفي الفرضية/مجموع الفرضيات بخصوص الظاهرة اللغوية.

"Empirical data provide **context** and the ability to **confirm** or **deny** what until then may have been only hypothesized." ³ [Emphasis added]

ج- تتسم التجارب المنجزة بالخصوبة و الإنتاج و تلك سمة البحث العلمي 'Experimental reproducibility'، وفق ما يذهب إليه أليفي كرايف (Olivier Kraif):

"Toute démarche expérimentale se fonde sur un principe fondamental: la reproductibilité des faits observés lorsqu'un certain nombre de conditions sont satisfaites [...]."⁴

الاستخلاص يكون بالتجريب في معناه الواسع. وعلى هذا الأساس، يمكن القول بأن مفهوم الملاحظة لا يبتعد اليوم عن مفهوم التجريب [...] لأن الملاحظة قد تلعب دور التجريب، ولأن التجريب في واقع أمره ليس سوى نوعاً آخر من الملاحظة. ينظر: جمال الدين، بوقلي حسن، مر. ن، ص 167.

¹ - عبد الرحمن بدوي، مر. س، ص 172.

² - توفيق، الطويل، أسس الفلسفة، ط5، دار النهضة العربية، 1967، ص 174.

³ - W. John, MacMullen (2003) Op.Cit., p. 4.

⁴ - Olivier, Kraif (2001) Constitution et exploitation de bi-textes pour l'aide à la traduction. Thèse—Sciences du Langage. Université de Nice Sophia Antipolis.

هذا فضلا عن إمكانية إعادة تلك التجارب و تكرارها (سواء كانت لك أم لغيرك).
و في ذلك فائدة، ليس من وجه أنها تسهم في إثبات النتائج و تأكيدها فحسب، بل إنها تسهم في دفع عجلة البحث و ترقيته.
د- يستعين الباحث في عمله ببرمجيات خاصة؛ و هو بذلك لا يهدف إلى استخراج البيانات آليا، بل إلى دراستها و تحليلها.
هـ- علاوة على ذلك، يستند البحث المدوناتي إلى التحليل الكمي *Quantitative analysis*، من خلال توظيف بعض المقاربات الإحصائية و القياسات الكمية* في عملية بناء المدونة مثلا، أو في حساب التوارد (*Occurrences*)؛ و تحليل المدونات على هذا النحو يسمح بتعميم الحكم و جعله صالحا على المجموعة ككل.¹
و- لا يكشف التحليل عما هو متداول فحسب؛ بل يمنحنا القدرة على التنبؤ.
و على هذا النحو، نخلص إلى القول: إن المقاربة اللسانية المدوناتية مقاربة استقرائية (*Approche inductive*)، يعتمد فيها الباحث على الضابط التجريبي، بغية فهم الظاهرة اللغوية فهما موضوعيا، و توخيا للدقة العلمية.

2.3 آليات البحث و أدواته:

ما نعنيه بـ " آليات البحث"، مختلف الطرق التي تتيح للباحث المعالجة الآلية للظواهر اللغوية ضمن المدونات، سواء من حيث استخراج البيانات و ضبطها و عرضها في لغة الأرقام، و تنظيمها أو دراستها و تحليلها بدرجة أعلى من الدقة و السرعة و حجم النتائج، مقارنة بآليات البحث التقليدية، مستعينا ببرمجيات خاصة صممت على نحو يخدم ذلك الغرض؛ بحيث تتولى مهمتين رئيسيتين، هما:
- تنظيم مكونات المدونة و تهيئتها على نحو يُيسّر على الباحث معاينتها و دراستها.
- ضبط تلك المكونات في شكل معطيات إحصائية.²
و من ثَمَّ، تنقسم آليات البحث باعتماد تلك البرمجيات إلى:

1.2.3- الفهرسة الآلية (*Concordance*):³

و يراد بها وضع قائمة لكلمات المدونة؛ أي إحصاء جميع تواتراتها (*Occurrences*) و عرضها على الواجهة البحثية ضمن سياقاتها التي وردت فيها.
و الفهرست الآلي الأكثر تداولا حاليا يدعى بـ: "كويك" الكلمة المفتاحية ضمن السياق (*Key Word in Context /KWIC*)، بحيث يتكون من عدة أسطر (*lines*) تستقل عن بعضها بعضا،* تتوسط كل سطر منها الكلمة المفتاحية أو العقدة (*Node*)،** يحيط بها عدد من الكلمات يمينا (مرافق العقدة اليمين *right co-text*) أي الصدر، و شمالا (مرافق

* - لاسميا تلك المستعملة في القياسات المعجمية (*Lexicométrie*).

1 - Cf. Sandra, Halverson (1998) *Op. Cit.*, note 2, p. 512.

2 - Cf. D. Evans (2006) *Op. Cit.*

3 - Cf. A. Stefanowitsch (2003c) *Corpus Linguistic Methods. Corpus Linguistics Handouts*, p. 1. Viewable online at: <http://www-user.uni-bremen.de/~anatol/materials.html>

N. Catherine, Ball (1997) Tutorial: Concordances and Corpora.

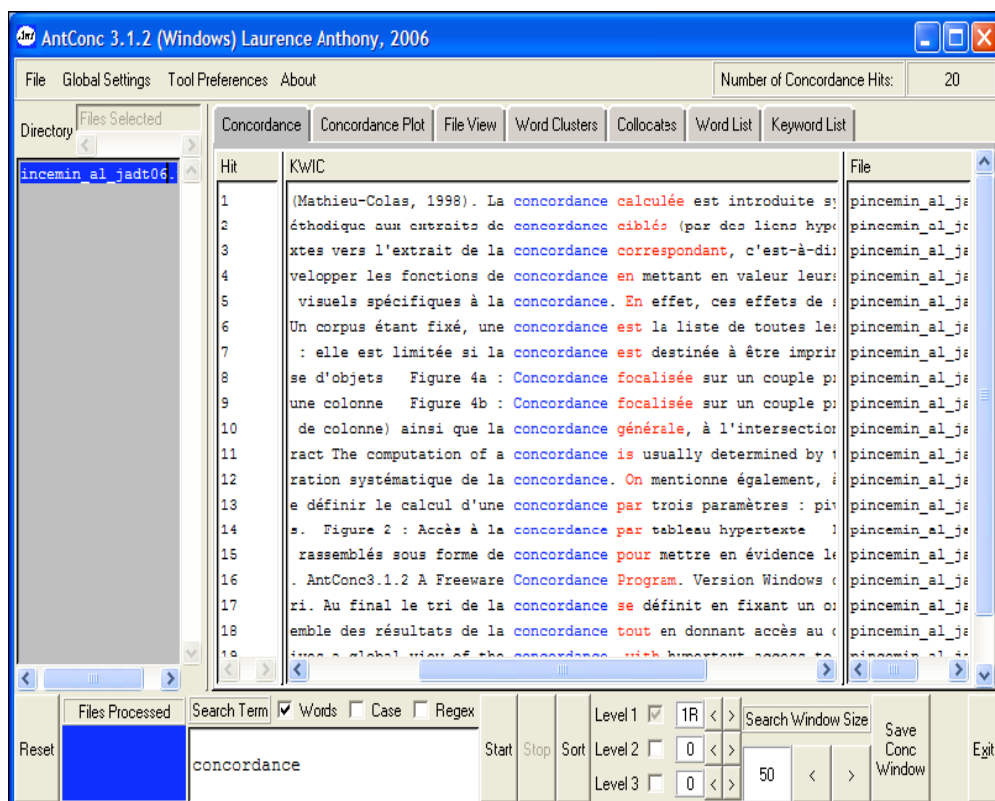
Pierre, Isabelle (1992) *La bi-textualité : Vers une nouvelle génération d'aides à la traduction et terminologie.*

Meta 37 (4), p. 727.

* - بمعنى أن كل سطر (أو أكثر) مأخوذ من نص واحد مثلا أو مدونة (في حالة المدونات الفرعية).

** - يمكن أن تكون وحدة البحث أي الكلمة المفتاحية مفردة، أو مركبا أو جملة.

العقدة اليسار (left co-text) أي العجز،* على حد سواء، فتشكل السياق ككل (Context)، و الذي غالبا ما يحدد طوله بسبع أو ثماني (7-8) كلمات، أو ما يعادل أربعين (40) حرفا في الاتجاهين كليهما. بيد أن تحديد هذا الطول يتوقف على نوع المفهرس المستخدم؛ إذ إن بعض المفهرسات الآلية¹ تعرض الجملة التي وردت فيها العقدة كاملة، بدل تحديد عدد الكلمات و بتر الجملة، كما في المثال أدناه.



الشكل (7) فهرست آلي.

فضلا على ذلك، يمكن أن تتنازد تلك الأسطر أو السياقات وفق طرق عدة: ألفائيا تبعا للكلمات التي تسبق الكلمة المفتاحية أو تعقبها؛ أو حسب ترتيب كل سطر ضمن المدونة ذاتها، و ما إلى ذلك.²

* - "الصدر" و "العجز" اجتهد شخصي.

¹ - Concordancers يقصد بها البرمجيات التي تتولى عملية الفهرسة (أو المعالجة الآلية عموما). وفي الأصل يطلق اسم Concordancer على الشخص الذي كانت تناط به مهمة قراءة النص، ووضع فهرست لجميع كلماته وسياقاتها؛ وفي معجم الفرنسية Le Petit Larousse Illustré (2007) نجد تعريفا لمصطلح La concordance biblique، نصه : 'Index alphabétique systématique des mots employés dans les livres bibliques avec l'indication, pour chaque terme cité, des passages où on le rencontre.'

لذلك رأينا في مصطلحي الفهرست/ المفهرس الآلي أفضل معادلين، في حين نجد أن سلوى حمادة والد يودان لفظ "مطابقة" كمعادل، والواقع أننا لا نراه مناسباً. ينظر:

سلوى حمادة، مر. س.

محمد الدداوي، منهاج المترجم: بين الكتابة والاصطلاح والهوية والاحتراف، ط.1، المركز الثقافي العربي، الدار البيضاء - المغرب/بيروت- لبنان، 2005، ص 125.

² - Cf. Lynne, Bowker (1999) Exploring the potential of corpora for raising language awareness in student translators. *Language Awareness* 8(3/4), p.162.

و لا يعتقدنّ - بأي حال - أن استخدام المفهرسات الآلية في معالجة المدونات يقتصر على الاستخراج الآلي للبيانات؛ بل إنه يمكّن من ضبط القوالب المتكررة (Recurring Patterns)، و عرض الأجزاء النصية التي تنتمي إليها. و لا يقف الباحث عند هذه المرحلة الكمية البحث، بل يتعداها إلى التحليل، و بالتالي اكتشاف الكثير حول الاستخدام الفعلي للكلمات (أو وحدات البحث عموماً) و سلوكها ضمن مختلف سياقاتها؛ و في ذلك استكمال في كثير من الأحيان، أو تغيير في أحيان أخرى لما تنطوي عليه الأدوات المرجعية المعيارية كالمعاجم و كتب النحو (...) وفق ما يذهب إليه فيديريكو زانتان:

"The analysis of these **recurring patterns** highlights the **behaviour of actual language in context**, and complements and sometimes challenges the information provided by standard reference tools such as dictionaries and grammars."¹
[Emphasis added]

و الحال هذه، يمكن مقارنة شتى الاستعمالات للكلمة الواحدة، و تحليل العبارات و المركبات و الجمل، و إيجاد التعابير الاصطلاحية و المصطلحية، لاسيما في حال الاستعانة بمفهرسات ثنائية اللغة أو متعددة اللغات في تحليل المدونات المتوازية، نذكر منها:

Multilingual Concordancer for Windows, Text Aligner, WordSmith Tools, Paraconc.

بحيث تقوم تلك البرنامجيات باستخراج جميع تواردات أي وحدة بحث (عجم *lexème*، كلمة، عبارة، جملة، ...) في لغة (أ)، ثم يتم تنظيرها مع ترجمتها/ ترجماتها في لغة (ب) أخرى.²

بالإضافة إلى ذلك، ثمة ميزة أخرى تنفرد بها المدونة، من حيث إن استثمارها في البحث اللغوي يعطيه بعداً آخر غير معهود؛ فالباحث حين يتفحص الفهرست الآلي في شكله المتداول، أي الكلمة المفتاحية ضمن السياق *KWIC*، و هو الشكل المعياري لتقرير الطبيعة التكرارية*، يجد أن البعد الأفقي هو البعد النصي، الذي يكتشف من خلاله (بعد القراءة طبعاً) تطور النص و المعنى الذي يصنعه باعتباره سلسلة خطية. بينما يبدي البعد الشاقولي أوجه الائتلاف و الاختلاف بين كل سطر و ما يحيط به من أسطر أخرى، و هنا

¹ - Federico, Zanettin (1998) Bilingual comparable corpora and the training of translators. *Meta* 43(4), p. 616. Cf. also: Sylviane, Granger (1996) From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In: Aijmer K., Altenberg B. & Johansson M. (eds.) *Languages in Contrast: Papers from a Symposium on Text-based cross-linguistic Studies*, Lund, 4-5 March 1994. Lund: Lund University Press, pp. 37-51.

² - Cf. Sylviane, Granger (2003) The corpus approach: a common way forward for Contrastive Linguistics and Translation Studies. In: S. Granger, Jacques Lerot & Stephanie Petch-Tyson (eds.) *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam/ New York: Rodopi, p. 24.

* يشير سنكلير إلى أن الطبيعة التكرارية *recurrence* - والتي يعتبرها قوام كل شيء - هي بخلاف الطبيعة الوحيدة أو الفريدة *uniqueness*

"All rely on recurrence as their starting point" (Sinclair:2005).

ممكن "القيمة المضافة" الهامة التي تتأتى من خلال استثمار المدونة، و هي قيمة تمكن الباحث من التعميم انطلاقاً من التكرارات (Recurrences).¹ و ليست تخفى على أحد أهمية استخدام المفهرسات في تعلم اللغات، بعد أن كانت حكراً على الدراسات الأدبية و الأسلوبية؛ إذ - و كما أسلفنا- تزرع المدونات بنفيس الأمثلة الحقيقية عن الاستخدام اللغوي، و بمقدور المتعلم أن ينقب عنها بفضل المفهرسات، فيغتني و يثري رصيده اللغوي، لا بل و يجدده - وفق ما يذهب إليه بوب غدوين-:

"An extensive linguistic corpus is a gold mine of authentic language use and mining that through KWIC concordances can provide students with multiple contexts from which to learn new vocabulary."²

و تلكم البرمجيات منها ما هو مجاني يمكن تنزيله أو تحميله انطلاقاً من الويب، على غرار MonoConc و Conc 1.80؛ و بعضها غير متاح على مستوى الشبكة، إنما يتطلب تسجيلاً أي اشتراكاً أو شراء نسخة عنه، و من أمثلتها:

Multiconcord, Paraconcord, Glossa Net, Corpus Eye *

و لا يخالّن القارئ أن الذاكرة الترجمية هي نفسها المفهرسات الآلية، بل ثمة فرق جوهري بينهما؛ فبرمجيات الذاكرة الترجمية MT - أو كما يدعوها اليداوي ببرنامج الذاكرة الحاسوبية- بمثابة "مخزون من النصوص المتوازية (الأصل و ترجمته) يعطي للمترجم اقتراحات للترجمة استناداً إلى النصوص المسجلة آنفاً، إذ يتعرف على الجمل و المقاطع المتشابهة أو المتقاربة فيبرزها و يشير إليها [...]"، و أبرزها: ملتيترانس Multitran، ترادوس Trados، ميتاريد Metaread، وردفاست Word Fast...،

أما المفهرسات الآلية (و هو يدعوها ببرنامج المطابقة)، فتستخدم "للتنبية لا إلى الجمل" ** و إنما إلى المصطلحات و التعابير المصطلحية المتكررة و يوردها في سياقاتها [...]".³ و في معرض حديثها عن أهمية ما حققه الباحثون في مجال المدونات، تشير سيلفيا برنارديني⁴ Silvia Bernardini إلى أن المدونات - عموماً - لم تلق الترحيب في بادئ الأمر، و عزف عنها المحترفون المشتغلون في شتى الميادين، لا لسبب إلا لأنها كانت تتطلب وقتاً و فيراً و جهداً كبيراً، سواء في تصميمها أو تجميعها و استثمارها، دونما تحسن يذكر على مستوى الإنتاجية. و على العكس من ذلك، لاقت الذاكرة الترجمية رواجاً واسعاً، لاسيما بالنظر إلى سهولة عملية صنعائها، و يسر استخدامها؛ و ما كانت الحال لتتغير أو النظرة لتتبدل لولا ظهور المفهرسات الآلية.

¹ - Cf. John Sinclair (2005) Op. Cit.

² - Bob, Godwin-Jones (2001) Op. Cit., p. 9.

* - ينظر الملحق (3) من هذا البحث (برنامجيات معالجة المدونات).

** - الواقع أنه ثمة مفهرسات تستخدم لاستخراج الجمل أيضاً.

³ - ينظر: محمد، اليداوي، مر. س، ص ص 125-126.

⁴ - Cf. Silvia, Bernardini (2006) Corpora in translator education and translation practice: Achievements and challenges. *Proceedings of LREC 2006*. Viewable online at : http://mellange.eila.jussieu.fr/bernardini_lrec06.pdf

و نفس الملحوظة تنطبق على محركات البحث على مستوى الشبكة، على شاكلة: جوجل Google أو كوبرنيك Copernic؛ بحيث تقدم تلك المحركات واجهة بحثية تعرض من خلالها شبه فهرست Quasi-concordance بالنصوص التي ترد فيها الكلمة المفتاحية، غير أنها تفتقر إلى الخصائص التي تميز المفهرسات الآلية، لذلك عمد الباحثون في المجال إلى استحداث برمجيات جديدة للاستبحار في الشبكة (باعتبارها مدونة في حد ذاتها)، نذكر منها: Boot CaT من تصميم باروني -برنارديني Baroni-Bernardini، TextStat (Matthias Hüning) و KwicFinder لـ "فلتشر" (Fletcher) و WebConc لـ "كليغارييف" (Kilgariff).¹

2.2.3- قائمة الوصيلات Collocate list²

و يقصد بها قائمة جميع الكلمات التي تنزع إلى التوارد مقترنة بالعقدة، لذلك تدعى بـ "وصيلات العقدة" (Collocates)، و تتخذ رتبة معينة بالنسبة إليها، كأن تمتد مثلا من الكلمة الرابعة يسارا إلى الكلمة الرابعة يمينا (في فهرست إنجليزي)، و ترفق بتكرارات ورودها في تلك الرتبة، و هو ما يوضحه الشكل أدناه.

L3	L2	L1	R1	R2	R3
24 the	8 the	15 and	36 the	27 to	29 to
7 my	7 hand	8 to	23 by	12 the	16 of
6 to	5 which	7 would	16 a	7 seal	7 in
6 that	4 it	7 which	8 him	6 a	5 and
5 of	4 of	7 had	6 them	5 great	4 that
5 in	4 and	6 may	5 to	4 of	4 the
4 a	4 or	5 have	5 it	3 much	2 than
3 and	3 their	4 has	5 us	2 system	2 rise
2 it	3 This	4 is	3 all	2 entire	2 build
2 with	2 conditions	3 be	2 so	2 lot	2 deal
2 this	2 signals	3 or	2 more	2 his	2 trouble
2 many	2 was	3 that	2 widespread	2 be	2 concern
2 other	2 this	3 might	2 such	2 damage	2 shear
	2 no	3 can	2 any	2 particular	
	2 process	2 it	2 her	2 more	
	2 must	2 but	2 an	2 aerator	
	2 in	2 will	2 increased	2 trouble	
	2 has	2 are			
		2 could			
		2 being			
		2 fats			
		2 been			
		2 hymen			
		2 was			
		2 thus			

الشكل (8) قائمة الوصيلات.

¹ - Cf. S. Bernardini (2006) Op. Cit.

- F. Zanettin (2002a) Op. Cit.

² - Cf. Anatol, Stefanowitsch (2003) Op. Cit., p. 1.

مصطلح 'الوصيلات' اجتهد شخصي، تفاديا للبس الذي يمكن أن يقع بالنسبة لمصطلح 'المتلازمات' Collocations.

و بالمثل، تتخذ قائمة الوصيات أشكالاً عدة، كأن تنتظم من الوصيات الأكثر وروداً فالأقل ثم الأقل، أي وفق ترتيب تنازلي – كما في المثال أعلاه. و تمكن الوصيات مقترنة بالعقدة من دراسة المتلازمات اللفظية (Collocations)، أي "توارد كلمتين أو أكثر في نطاق حيز محدد يجمع بينهما داخل النص" – على حد تعبير سنكلير (1991: 170).¹

3- جدول التكرار (Frequency list):

تسمح برامجيات التحليل بالكشف عن عدد الكلمات المتباينة ضمن مدونة، أي الأنواع Types، من جهة؛ و عن عدد مرات ظهور كل كلمة ضمن المدونة نفسها، أي النماذج Tokens من جهة أخرى.²

مثال ذلك، جملة: *A lion is a dangerous animal* تتكون من ست كلمات، أي ستة نماذج؛ بيد أن حرف *a* يظهر مرتين، لذلك فإن هذه الجملة تتكون من خمس كلمات متباينة فقط (أي خمسة أنواع).

انطلاقاً من هذه المعطيات، يمكن تشكيل جدول التكرار، و هو جدول يتضمن جميع كلمات المدونة المتباينة (الأنواع)، مشفوعة بتكراراتها (النماذج)، حسب ستيفانوفيتش: "A frequency list is a list of all words in a corpus together with their frequency."³

و يُرتَّب جدول التكرار إما ألفبائياً، و إما وفق ترتيب تنازلي أو تصاعدي (ينظر المثال).

69377	the	6688	on	3541	which	2235	who	1749	can
36146	of	6325	be	3270	were	2219	will	1725	only
28708	and	5351	at	3259	one	2204	more	1689	other
25926	to	5272	by	3221	you	2178	no	1609	some
23321	a	5135	i	2997	her	2162	if	1604	new
21222	in	5114	had	2947	all	2071	out	1587	could
10517	that	5110	this	2847	she	1967	so	1575	time
9993	is	4553	not	2706	there	1967	said	1565	these
9755	was	4354	from	2705	would	1875	up	1401	two
9491	he	4345	are	2653	their	1858	what	1390	may
9435	for	4350	but	2653	we	1845	its	1359	then
8655	it	4194	or	2567	him	1806	about	1346	first
7244	with	3909	have	2464	been	1784	into	1336	any
7211	as	3711	an	2421	has	178	than	1318	do
6955	his	3592	they	2316	when	1751	them	1296	such

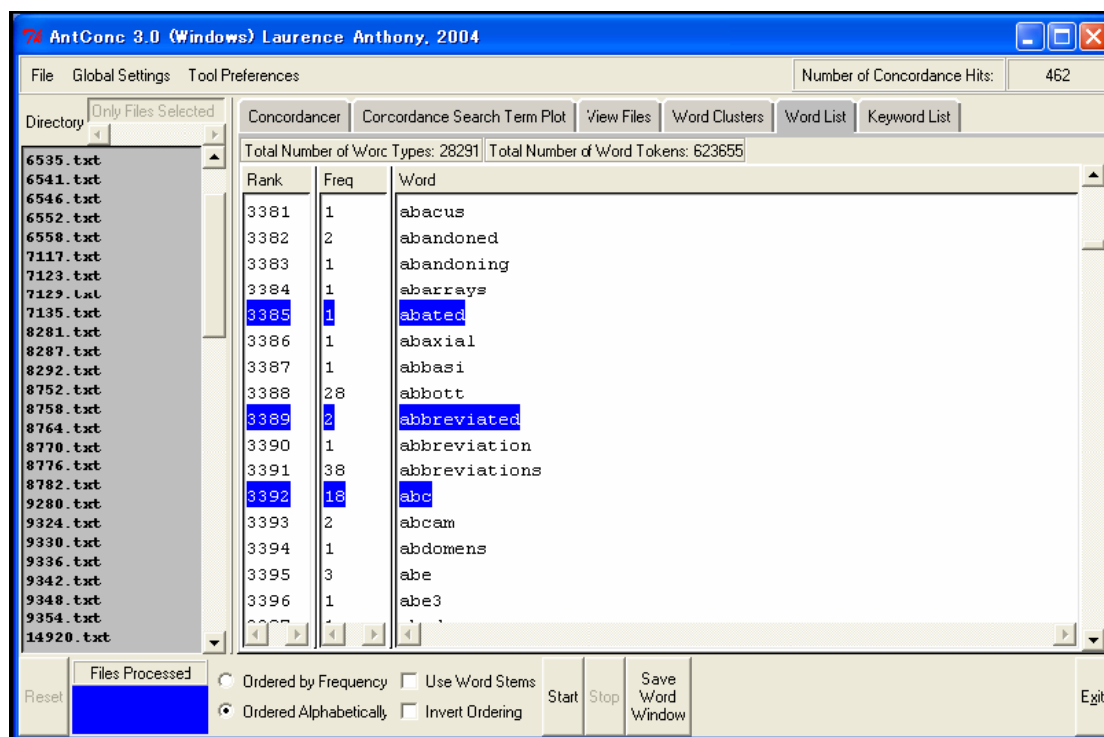
الجدول (2) جدول التكرارات.

¹ - John, Sinclair (1991) qt. in: Dodd Bill (2000) Op. Cit.

² - Cf. Lynne, Bowker (1999) Op. Cit., p. 162.

³ - Cf. A. Stefanowitsch (2003c) Op. Cit., p. 2.

و ثمة ضرب خاص من جداول التكرار، يدعى بجدول توزيع التكرارات ¹ *Distribution Frequency List*، و هو يتضمن التكرارات الصّرف، و كذا توزيع الكلمة ضمن المدونات الفرعية أو الملفات – الجدول رقم 3-.



الشكل (9) جدول التكرارات عن AntConc.²

و المثال يتناول الكلمات العشر الأولى من جدول التكرار ضمن مدونة براون. يمثل العدد الأول من كل سطر التكرار الصّرف للكلمة، و العدد الثاني (15) يمثل عدد الأجناس الكتابية التي تتوارد فيها الكلمة؛ أما العدد الأخير، فيتعلق بعدد النصوص التي وردت فيها الكلمة.

الجدول (3): جدول توزيع التكرارات.³

69971-15-500 THE
36411-15-500 OF
28852-15-500 AND
26149-15-500 TO
23237-15-500 A
21341-15-500 IN
10595-15-500 THAT
10099-15-485 IS
9816-15-466 WAS
9543-15-428 HE

¹ - Cf. *Ibid.*, p.3.

² - Cf. Laurence, Anthony (2004) AntConc: A Learner and Classroom Friendly Multi-Platform Corpus Analysis Toolkit. In: *IWLeL 2004, an Interactive Workshop on language e-Learning*, pp. 7-13.

³ - *Ibid.*

الفصل الثاني:

واقع البحث باعتماد المدونة عند الغرب و العرب

- I- الواقع عند الغرب .
 - I-1- المدونة في اللسانيات و الأبحاث اللغوية.
 - I-2- المدونة و تعليم اللغات.
 - I-3- المدونة و الترجمة الآلية.
- II- الواقع عند العرب.
 - II-1- المدونة في خدمة العربية.
 - II-2- المدونات العربية.

"Pour les disciplines qui font leurs premières pas en linguistique de corpus, il s'agira donc plutôt de déterminer quels sont les types de questionnements pour lesquels le recours aux corpus semble utile et d'en fixer les règles. Pour les autres, il s'agira de mettre à distance des pratiques déjà anciennes, d'en évaluer la pertinence, d'en mesurer les acquis et les limites".

Sylvie, Mellet (2002) Corpus et recherches linguistiques. Corpus, n°1, Corpus et recherches linguistiques – nov. 2002.

لقد كان الجدل لا يزال قائماً - و إلى حين - عما إذا كان اعتماد المدونات في البحث ضرورة ملحة، أم أنه من قبيل الترف الفكري و البذخ الذي لا طائل من ورائه. و ذريعة هؤلاء أنه ليس ثمة أي مدونة - مهما بلغت سعة و عظمة - بمقدورها أن تتضمن جميع عناصر اللغة، أو أن تحيط بخصائصها و دقائقها؛ لذلك فلا جدوى من اعتمادها. أما أدعياء الضرورة، فيقولون أن كل مدونة - مهما دقت حجماً - قد بينت كثيراً من الحقائق التي لا يمكن للمرء أن يدركها أو يجدها في أي مصدر آخر.¹ و الحق أن الاحتكام إلى الواقع قد بت في الأمر بما لا خيار و لا عودَ فيه؛ فبيّن أن ما يذهب إليه أشياع البحث باعتماد المدونة حقيقية ناصعة لا غبار عليها. و هو الواقع نفسه الذي سوف نحتكم إليه في هذه المحطة، و عبر إطلالة سريعة؛ إذ نقف في البداية عند أهم ما بلغته الدراسات في هذا المجال عند الغرب، من حيث تنوع صور تطبيقات استخدام المدونات في اللسانيات و الأبحاث اللغوية، على غرار المعجماتية و المصطلحية و النحو و الدلالات و الأسلوبيات و المعالجة الآلية للغات، و قس على ذلك. بعدها نحاول الكشف عن حظ العربية من ذلك كله.

I- الواقع عند الغرب:

حقيق بنا أن نذكر، قبل الاستفاضة في تفاصيل هذه المحطة، بأن اعتماد المدونات في البحث ليس في حد ذاته بدعاً و لا مُحَدَّثاً؛ بل إن محل الجدة يكمن في استثمار البرمجيات التي غيرت من "كفاءة و نوعية الأبحاث التي كانت تعتمد على دراسة التراكيب اللغوية، و أصبحت الآن تعتمد على الاستعمال اللغوي و التطبيق الفعلي".² ثم إن القطيعة التامة بين العهدين في تاريخ لسانيات المدونة، و أولى أمارات العصر الجديد، يحددها المؤرخون - لاسيما مكنري و ويلسن³ - انطلاقاً من فترة الستينات، حين بدأ راندولف كويرك Quirk Randolph في تصميم خطة مشروعه الفذ الطموح (SEU) SURVEY OF ENGLISH USAGE، أي مسح للاستعمال في اللغة الانجليزية، و شرع في تجسيد أولى لبناته على أرض الواقع. و هو من أوائل المدونات المعاصرة، يحوي مائتي (200) عينة، يبلغ حجم كل واحدة منها زهاء 5000 كلمة؛ مصدرها النصوص المكتوبة (100) والمنطوقة (100)، يمتد تاريخها من سنة 1953 إلى غاية سنة 1987 (ينظر الشكل - 10 - أدناه).⁴

و في الوقت نفسه، طفق نلسن فرانسر (Nelson Francis) و هنري كوشيرا (Henry Kucera) بجامعة براون، يجمعان مدونة براون (Brown) الانجليزية الأمريكية المكتوبة ضمن 500 عينة، بحجم 2000 كلمة تقريباً في كل واحدة.

¹ - Cf. Sameh, Alansary et al. (2007) Op. cit., p. 1.

² - سلوى، حمادة (2007) مر. س. ، ص 4.

³ - Cf. T. McEnery & A. Wilson (1996) Corpus Linguistics. Edinburgh University Press.

⁴ - و أول ما تطرق إلى الحديث عن المدونة عام 1960، حين نشر: "Towards a Description of English Usage." للمزيد من التفاصيل بخصوص المدونة، ينظر:

<http://www.ucl.ac.uk/english-usage>

و اتخذت "براون" نموذجا قاعديا في تصميم المدونات لعدة سنوات عقب نشرها؛¹ إذ صممت على شاكلتها و مقاساتها مدونات أخرى، تصميماتا تقريبا أو مطابقا لها مطابقة تامة، على غرار: مدونة كولابور (Kolhâpur Corpus) (الانجليزية الهندية)، و مدونة ولينغتن (Wellington Corpus) (الانجليزية النيوزيلندية)، و المدونة الاسترالية للغة الانجليزية (Australian Corpus of English). و لعل أهمها مدونة لوب (LOB) للغة الانجليزية البريطانية Lancaster–Oslo/Bergen Corpus من تصميم ستينج جوهانسن (Stig Johansson) و جيوفري ليتش² (Geoffrey Leech)، باعتبارها نسخة طبق الأصل عن "براون"، إذ يبلغ حجم كل واحدة منها مليون كلمة ضمن 500 عينة نصية، بحجم 2000 كلمة في كل نص، هذا فضلا على عدد الأصناف النصية (15).

أ- نصوص ذات أصل مكتوب (100)

- أ- مطبوعة (46)
 - إعلامية (28): صحافة (8) – أكاديمية (فنون وعلوم) (13)
 - إدارية (4) – قانونية (3)
 - توجيهية (6)
 - تأثيرية (5)
 - خيالية (7)
- ب- غير مطبوعة (36)
 - مراسلات (21): اجتماعية (13) – غيرها (8)
 - يوميات شخصية (4)
 - كتابات دائمة (11): خيالية (5) – إعلامية (6)
- ج- محررة للإلقاء (18)
 - محادثات (6)
 - مسرحيات (4)
 - أخبار (3)
 - خطابات محررة (3)
 - قصص (2)

¹ - Cf. W. Francis & H. Kucera (1979) *Brown Corpus Manual: Manual of Information to Accompany A Standard Corpus of Present-Day edited American English, for use with Digital Computers* (3rd ed.) Providence: Brown University.

² - Cf. S. Johansson et. al. (1978) *Manual of Information to Accompany the Lancaster-Oslo/ Bergen Corpus of British English, for use with Digital Computers*. Department of English, University of Oslo.

II- نصوص ذات أصل منطوق (100)

أ- مونولوج (24)

تلقائي (18)

خطابات (10)

تعاليق (8): رياضة (4) – غيرها (4)

خطب محضرة لكنها غير محررة (6)

ب- حوار (76)

محادثة وجه لوجه (60): مسجلة سرا (34) – غيرها (26)

محادثات هاتفية (16)

الشكل (10): تصميم مدونة SEU

و لا شك أن السر في ما حققته المدونتان معا من نجاح، مرده جودة التصميم تلك، يضاف إليها ما تضمنته من أنواع نصية متعددة تعددا يستقطب كثيرا من المستخدمين كل حسب حاجته، إذ شكلت المدونتان المادة الخام لعدة دراسات وصفية في علم الصرف، والمعجم (LEXIS) و النحو و الخطاب؛ فأفرزت رصيذا معرفيا حول الاستخدام اللغوي أكثر دقة و وثوقا مقارنة بالدراسات التقليدية التي كانت تعتمد على الحدس.¹ و نشير إلى أنه و لأكثر من أربعين سنة، كان البحث لا يزال جاريا باعتماد تلكما المدونتين في الدراسات المقارنة، إلى جانب ما استنسخ على غرارهما من مدونات أخرى: فراون (Frown) و فلوب (FLOB) أوائل التسعينات. و في العام 1975، قام (Jan Svartvic) بصناعة مدونة لندن- لوند (London-Lund Corpus) استنادا إلى عملي كويرك و فرانسنز- كوشيرا، و تلك هي أول مدونة قوامها اللغة المنطوقة. ثم كان ميلاد مدونة "كونكود" عام 1994 عن جامعة كامبريدج،² Cambridge and (Cancode) Nottingham Corpus of Discourse in English.

أما (BNC) المدونة البريطانية الوطنية، فقوامها 4124 عيّنة، و حجم كل واحدة أكثر من 40.000 كلمة. مادتها النصية مكتوبة و منطوقة، تمتد فترتها التاريخية من سنة 1960 إلى سنة 1993. و لقد انتهج مصمموها مقاربة أكثر حذقا في صناعتها، مقارنة بالجيل الأول من المدونات (Brown و SEU)، من حيث تحديد التصنيفات النصية المعتمدة في بناء المدونة؛ إذ تصنف كل عيّنة وفق سلسلة واسعة من المتغيرات أو العوامل التي تؤثر في اللغة (كالبينة و المستوى الاجتماعي و الجنس) – ينظر الشكل (11) أدناه.

¹ - Cf. Sylviane, Granger & Viana, Vander (2007) Corpus Linguistics, Language Learning and ELT: Interviewing Sylviane Granger, Mindbite, 2007, 1, p.11.

² - Cf. S. Alansary et al. (2007) Op. Cit., p. 1.

I- المنطوق (915)

نوع التدخل: حوار (672)، مونولوج (218)، غير مصنف (25).

المنطقة: الجنوب (269)، المنطقة الوسطى (208)، الشمال (334)، غير مصنفة (77).

سياق محدد (762)

المجال: تعليمي (144)، أعمال (136)، مؤسساتي (241)، تسليية (187)، غير مصنف (54).

ديمغرافي (153)

السن: 0-14 (26)، 15-24 (36)، 25-35 (22)، 45-59 (20)، 60+ (20).

القسم: AB (59)، C1 (36)، C2 (31)، DE (20)، غير مصنف (7).

الجنس: ذكر (73)، أنثى (75)، غير مصنف (5).

II- المكتوب (3209)

الفترة الزمنية : 1960-74 (53)، 1975-1993 (2596)، غير مصنفة (560).

الوسط : كتب (1488)، دوريات (1167)، متفرقات منشورة (181)، متفرقات غير منشورة (245)، للإلقاء (49)، غير مصنفة (79).

المجال : خيالي (625)، العلوم الطبيعية (144)، العلوم التطبيقية (364)، علوم اجتماعية (510)، شؤون دولية (453)، تجارة (284)، فنون (259)، الأفكار و المعتقدات (146)، تسليية (374)، غير مصنفة (50).

الجمهور: السن (أطفال (45)، مراهق (74)، بالغ (3086)، غير مصنف (4))، الجنس (ذكر (63)، أنثى (167)، مختلط (2034)، غير مصنف (945))، المستوى (ضعيف (702)، متوسط (1674)، عال (824)، غير (9)).

المؤلف: النوع (مشترك (397)، متعددة (1357)، فردي (1331)، مجهول (122)، غير مصنف (2))، الجنس (ذكر (948)، أنثى (445)، مختلط (208)، مجهول (117)، غير مصنف (1491))، السن (0-14 (22)، 15-24 (15)، 25-35 (38)، 45-59 (80)، 60+ (70)، غير مصنف (2899)).

إضافات: مكان النشر، نوع العينة (نصوص كاملة (267)، البداية (599)، الوسط (555)، النهاية (127)، مؤلفة (18)، غير مصنفة (1643))، وضع التلقي (ضعيف (801)، متوسط (903)، عال (1059)، غير مصنف (446)).

الشكل (11): تصميم مدونة BNC.

و يتضح من خلال الشكل أن الجزء المنطوق ينقسم إلى قسمين :

- قسم محدد بسياق ما، يتضمن تحويلات لتسجيلات تمت خلال مناسبات أو لقاءات معينة.
 - و قسم ثان ديمغرافي، يتضمن تحويلات لمحادثات تلقائية عفوية بين أفراد المجتمع.
- في حين أن انتقاء النصوص المكتوبة تم وفق ثلاثة معايير، هي: المجال و الفترة الزمنية و الوسط.¹

- المجال *Domain*: 75% من النصوص المدرجة (المكتوبة) تنتمي إلى النصوص الإخبارية (*Informative*)، و هي نصوص مستقاة من مجالات العلوم التطبيقية، الفنون، الاقتصاد و المالية، العلوم الطبيعية، الاجتماعية، الشؤون الدولية، الخ؛ أما النسبة المتبقية (أي 25%) فهي نصوص من الخيال الأدبي (*Imaginative*).
- الوسط *Medium*: 60% من النصوص مأخوذة عن الكتب، 25% من الدوريات (كالصحف و المجلات...)، 5-10% متفرقات من المنشورات (المطبوعات السياحية، النشرات الدعائية،...)، 5-10% مادة نصية غير منشورة (رسائل شخصية، يوميات و مذكرات)، و نسبة ضئيلة تقل عن 5% مادة نصية مكتوبة لغرض الإلقاء (كالخطب السياسية مثلا).
- الفترة الزمنية *Time*: بما أن المدونة سانكرونية، فإن مادتها تنتمي إلى الحقبة الزمنية نفسها تقريبا.

و ما حجم المدونة و تشكيلة المتغيرات تلك، إلا دليلا قاطعا على العناية الفائقة التي أولاهها المصممون للمدونة؛ بيد أنها تكاد تقتصر إلى مطلق التمثيلية أو التوازن، لاسيما إذا لاحظنا البون الشاسع بين حجمي النصوص المكتوبة (90%) و النصوص المنطوقة (10%) التي تتشكل منها.

و مع ذلك، لا ننكر أن المدونة البريطانية الوطنية – و إلى حد الساعة – هي بمثابة محاولة طيبة، و سعي محمود لإنشاء مدونة شاملة، جامعة للانجليزية الخاصة بقطر وطني معين. كما أن جسامه حجمها نقيية قد انفردت بها دون غيرها، مكنت الباحثين من جمع مدونات متوازنة أصغر، انطلاقا من مادة المدونة ذاتها، بغرض مشاريع أبحاث خاصة.²

أما المدونة العالمية للغة الانجليزية ³ *International Corpus of English* فيهدف مصمموها إلى مقارنة المدونة البريطانية الوطنية حجما و أهمية؛ بيد أن هذه المدونة تتكون من مدونات فرعية بحجم 1 مليون كلمة، تتباين لغاتها بتباين الأقطار

¹ - Cf. BNC (2003) *British National Corpus*.

<http://www.natcorp.ox.ac.uk/>

² - Cf. BNC site :

<http://www.hcu.ox.ac.uk/BNC>

Demo access is available at :

<http://sara.natcorp.ox.ac.uk/lookup.html>

³ - Cf. *International Corpus of English*, University of Hong Kong.

<http://www.ucl.ac.uk/english-usage/ice/>

<http://www.hku.hk/english/research/ice/index.htm>

- Cf. also: S. Greenbaum (ed.) (1996) *Comparing English Worldwide. The International Corpus of English*. Oxford: Clarendon Press.

الانجليزية. و قد شرع في تصميمها (سنة 1990) خمس عشرة فريقا من الباحثين في مختلف أرجاء العالم (بريطانيا، أستراليا، سنغافورة، الهند، كندا، جنوب إفريقيا، نيوزيلندا...). و بعد الجمع، سيبلغ حجم المدونة 15 مليون كلمة، بحيث إن قوام كل مدونة فرعية منها 500 عيّنة، بحجم 2000 كلمة تقريبا، تتضمن نصوصا منطوقة (300) وأخرى مكتوبة (200)، تتوزع على عدد من الأنماط، كما يوضحها الشكل (12) أدناه.

محادثات (90)	خاص (100)	حوارات (180)	الجزء المنطوق (300)
مكالمات هاتفية (10)			
دروس الصف (20)	عام (80)		
نقاشات مذاعة (20)			
لقاءات مذاعة (10)		مونولوج (120)	
جلسات برلمانية (10)			
استجابات مضادة (10)			
صفقات (أعمال) (10)			
تعاليق (20)	غير محرر (70)	محرر (50)	
خطابات غير محررة (30)			
تظاهرات (10)			
عروض قانونية (10)			
أخبار مذاعة (20)		محررات طلابية (20)	الجزء المكتوب (200)
محادثات مذاعة (20)			
محادثات غير مذاعة (10)			
محاولات الطلبة (10)			
امتحانات (10)		مراسلات (30)	
مراسلات اجتماعية (15)			
مراسلات الأعمال (15)			
إنسانيات (10)			
علوم اجتماعية (10)		أكاديميات (40)	مطبوع (150)
علوم طبيعية (10)			
تكنولوجيا (10)			
تقارير صحفية (20)			
وثائق إدارية (10)		رورتاج (20)	
مهارات/هوايات (10)		توجيهي (20)	
افتتاحيات (10)		تأثيري (10)	
روايات (10)		إبداعي (10)	

الشكل (12): تصميم مدونة ICE.

تلکم إذن قائمة بأهم المدونات التي شكلت محطات رئيسة في الصناعة المدوناتية الغربية، أو لنقل الانجليزية بصفتها اللغة الحظية، بالنظر إلى ما نالته من موفور العناية وبالع الاهتمام، و ما استأثرت به من حصة الأسد في مجال لسانیات المدونة.¹ و بعد هذه

¹ - يمكن التعرف إلى مدونات أخرى، و ذلك بزيارة موقع البوابة "Gateway".

GATE: General Architecture for Text Engineering. <http://www.gate.ac.uk>

الوقفة العجلى، ننتقل في ما يلي إلى تفحص أهم المجالات التي استثمرت فيها تلك المدونات.

I. 1- المدونات في اللسانيات و الأبحاث اللغوية:

الواقع أن الحديث عن أهمية استثمار المدونات في مجال اللسانيات، و مختلف الأبحاث اللغوية، هو حديث عن أهمية المعطيات التجريبية. تلك المعطيات تثير العجب، بل العجاب*، حيث تمنح الباحث قدرة إصدار أحكام لا تشوبها شائبة الذاتية، و لا هي تتبني على ما يدعوه بعضهم بإدراك الفرد للغته حسياً، – على حد تعبير الأنصاري؛ بل إنها أحكام تتأسس على أرضية صلب من الموضوعية و الدقة.

*"Empirical data enable the linguist to make objective statements, rather than these which are subjective, or based upon the individual's own internalized cognitive perception of language."*¹

ذلك أن اللغة ليست تتكشف أمام مستكشفها، فتفصح له عن جميع خباياها ومكوناتها، و هو ليس له من العدة و العتاد سوى حدسه و بضع أمثلة لغوية؛ بل إنها تفرض عليه امتلاك ترسانة هائلة من النصوص كالتى توفرها المدونات.

و علاوة على ذلك، ليس من الهين أن يمحس الباحث اللغة، و ينساب بين دقائقها و ذراتها، إلا بفضل البر(نا)مجيات الخاصة، فتحصل لديه بذلك رؤى جديدة حول بنية تلك اللغة، فضلاً على ما يتكشف له من خصائصها الكمية، على حد قول تاماس فارادي (Tamás Váradi):

*"[...] the technological facility to sift through large amounts of data leads to new insight into the structure of the language. One key aspect of language use that CL is particularly well suited to reveal is its quantitative characteristics."*²

و تلكم ليست إلا واحدة أو اثنتين من إسهامات لسانيات المدونة في دفع عجلة البحث اللساني للمضي قدماً. فضلاً على ما أحدثته من ثورة جعلت كثيراً من اللسانيين على اختلاف مللهم و نحلهم، بما فيهم أولئك الذين صعدوا لها الخد أول الأمر، يرتدون ويدخلون في معسكرها فوجاً بعد فوج، و ينضوون تحت لوائها و قد جاؤوا من كل فج؛ لأنهم رأوا فيها موضة جديدة في التفكير اللغوي – وفق ما يذهب إليه ليتش³ Leech، اجتثت بعضاً من رواسخ المعتقدات المتجذرة لديهم بشأن اللغة. و لا يُعْتَقَدَنَّ – بأي حال – أنهم قد اعتنقوها، لأن كل موضة تسلب عقولهم و تخلق ألبابهم، فينسلخون من معتقداتهم، كما النهار ينسلخ من الليل؛ بل لأن ما كانوا ما يدينون به من أحكام و معتقدات لسانية بدا

* - والعُجَاب ما جاوز حد العَجَب.

¹ - Cf. S. Alansary et al. (2007) Op. Cit., p.2.

² - Tamás, Váradi (2001) The Linguistics Relevance of Corpus Linguistics .In: Paul Rayson, Andrew Hardie & Shereen Khoja (eds.) *Proceedings of the Corpus Linguistics 2001 conference*. Lancaster University (UK), 29 March - 2 April 2001. University Centre for Computer Corpus Research on Language. Technical Papers, vol. 13, Special Issue, p. 587.

³ - Cf. S. Granger & v. Viana (2007) Op. Cit., p. 11.

لهم من زخرف القول و سفساف العمل، ما لم يستند إلى نصوص من رحم الواقع، تسمح لهم بـ "تسييق" تحاليلهم للغة. و في ذلك يحضرنا قول الأنصاري :

"Because corpora consist of texts, they enable the linguists to **contextualize** their analyses of language; corpora are very well suited to more functionally based discussion of language and linguistics."¹ [Emphasis added]

ثم إن ما تطمح إليه اللسانيات من بلوغ مراتب العلوم، يحتم عليها – بلا شك – أن تتزيا بزيتها، و تأخذ بأخذها، و تدرج مدارجها؛ أذناها درجة أن تتوفر على قدر من الفرضيات و النظريات تخضعها للاختبار و الكشف، و تتيح للباحثين فرصة إجراء تلك الاختبارات و إحداثها مرارا (Experiments must be reproducible). و ذلكم المطمح ضرب من المحال، أو فوق الحيازة و فوق المنال، ما لم تستعن اللسانيات بالمدونات التي تخدمها من أوجه ثلاثة:

- أنها السبيل إلى دراسة اللغة قيد الاستعمال.
- أنها أدوات اختبار لدراسة نتائج اللغة الطبيعية.
- أنها المورد الأساسي المعتمد في تطوير البرامج (يات) اللغوية.²

1.1.I المعجمية، المعجماتية و المصطلحية:

يعنى المعجماتيون ببحث معاني الكلمات، و مرادفاتها و كذا استخداماتها. و قد توسعوا في أبحاثهم تلك لاسيما بعد اعتمادهم المدونات كمادة خام في الصنعة المعجماتية. و ذلكم التحول وليد قناعة، مفادها أن معرفة المعاني الحقيقية للكلمة، و استعمالاتها الدقيقة، ليست تتأتى إلا بفحص مختلف سياقات ورود تلك الكلمة،³ و في ذلك عميم الفائدة فضلا على واقعية المسعى.

و لما كانت المادة النصية تحمل في طياتها كثيرا من البيانات، سواء تعلق الأمر بالمؤلف أو الجنس الكتابي أو عامل المكان أو الترميز بأقسام الكلام، و ما إلى ذلك، فإنه لمن السهل تحديد انتماء استعمالات بعض الكلمات أو التعابير، على أساس أنها كلمات أو تعابير نمطية تختص بها منطقة دون أخرى، أو ينفرد بها جنس كتابي دون آخر، و قس على ذلك.

و يعتمد في الصنعة المعجماتية – أيام الناس هذه – على المدونات بمختلف أشكالها، لما توفره من جداول التكرار و الفهارس الآلية و المتلازمات اللفظية، و كذا الأمثلة التوضيحية و البيانات النحوية.⁴ و لا غرو إذا جزمنا وفقا لذلك، بأن ما توصل إليه الباحثون في مجال لسانيات المدونة، سواء من الأولين أو التابعين، قد غيّر كثيرا من نظرة اللسانيين إلى اللغة.

¹ - S. Alansary et al. (2007) Op. Cit., p.2.

² - Cf. José, Soler (1993) Text Corpora: Meeting the Challenge of Information Excess. *Library & Linguistic Computing* 8 (4), pp. 224-226.

³ - Cf. Salem, Ghazali & Abdelfattah, Braham (2001) Dictionary Definitions and Corpus_ Based Evidence in Modern Standard Arabic. *Arabic NLP Workshop at ACL/EACL*. Toulouse, France.

<http://www.elsnet.org/arabic2001/ghazali.pdf>

⁴ - Cf. S. Alansary et al. (2007) Op. Cit., p. 2.

و لناخذ مثال المدونة الممتدة* *Open-ended corpus* (كما تدعوها سلوى حمادة)، أو مدونة المراقبة *Monitor corpus* (كما تعرف لدى المشتغلين في المجال)، و هي – تبعا للمصطلحين – مدونة في دينامية دائمة، من حيث إنها مفتوحة تضاف إليها نصوص جديدة بصفة دورية و منتظمة، تجمع بغرض مراقبة اللغة و سيرورة تطورها. و يعتمد المعجمانيون بنية ضبط جميع الكلمات الجديدة التي تدخل المعجم، أو إحصاء الكلمات التي تكتسب معان جديدة بفعل التطور اللغوي.

و هكذا يمكن تخريج معاجم ضمن طبقات مزيدة و منقحة بأسرع ما يمكن تصوره، تنتظم مداخلها بناء على التكرار بخلاف التقليد المتعارف عليه، أي الاشتقاق أو التأثيل (*Etymology*). و كذا هي الحال بالنسبة للتعاريف، إذ ترد جامعة مانعة، طالما أنه اعتمدت في وضعها أمثلة حقيقية، حسب أوليفي:

"[les corpus] permettent de donner une assise empirique aux définitions en liant le sens à l'usage, l'intuition du lexicographe n'étant plus, et ne devant plus être, le seul guide."¹

و كثيرا ما ينادي القائمون في المجال، لاسميا هانكز² (*Hanks*)، بجملة من المسلمات مفادها:

1. أن تعاريف المعاني الحرفية للكلمات التي تتوفر لدينا بناء على المعاجم و الحدس، ليست على درجة أعلى من الدقة و الموثوقية، بقدر ما هي عليه تلك التي تستنبط انطلاقا من البحث في المدونة.
2. ثم إن التباينات في المعنى – وفق ما تشير إليه المعطيات المدوناتية – هي وليدة التباينات في سياقات التراكم و التعبيرات الاصطلاحية.
3. و آخرها، إن ما نعتبره بحكم الحدس معنى مجازيا، غالبا ما يطلع الأصل أو المعنى النموذجي في سياق الاستعمال.

و يعج تاريخ المعجمانية بكثير من الأمثلة، التي نستشف من خلالها أهمية استثمار المدونات في تلك الصناعة؛ و طالما أن المقام يضيق بنا عن الإلمام بها جميعا، نكتفي بذكر أهمها و أشهرها، باعتبارها محطات رئيسة بل إحداثيات أساس في معلم ذلك التاريخ. يشير جيوفري ويليامز إلى أن الصناعة المعجمانية قد أدركت أهمية المدونات مع معجم جونصون (*Johnson*) منذ العام 1755 كفكرة، أما مبدأ العمل فيختلف اختلافا جذريا؛ بحيث يعد المعجم المذكور أول معجم يصنف باعتماد مدونة في شكل جذاذات عمل مرفقة بشواهد حقيقية. و يقول ويليامز ما نصه بالحرف الواحد:

" Une telle affirmation est peut-être un peu osée, mais pas infondée puisqu'avec Johnson débute une tradition

* - أما النوع الآخر، فهي المدونة الثابتة *Finite corpora*.

¹ - Kraif, Olivier (2001) *Op. Cit.*, pp. 214-215.

² - P. Hanks (2002) *Literal and Metaphorical Word Meaning*. Toscan Word Center Document; qt. in: s. Ghazali & A. Braham (2001) *Op. Cit.*, pp. 1-2.

*lexicographique plus normative que prescriptive mais basée sur des textes authentiques [...].*¹

إن ما يذهب إليه الكاتب، و الحق يقال، ليس من الزيف أو التطاول في شيء؛ إذ إن ما أحدثه جونصون من تقليد، أضحي أساس معجم أكسفورد للغة الانجليزية. بعدها شرع المعجمانيون في تصنيف شكل معجمي آخر موجه لمتعلمي اللغة الانجليزية، وفق تقاليد أكسفورد، ألا و هو معجم أكسفورد للطلاب المتقدمين (*Oxford Advanced Learner's Dictionary*)، و هو سليل معجم هورنبي (*Learner's Dictionary of Current English*) الصادر عام 1948. بيد أن تلك المعاجم و إن كانت في الأساس قد صنفست استنادا إلى جذايات، فإنها حوت أمثلة استقيت من اللغة العامة.²

و ساد ذلك التقليد ردحا طويلا من الزمن، إلى غاية أوائل التسعينات، حيث غدت معاجم أكسفورد في الطليعة،³ تهتم باللغة الانجليزية (لغة اليوم)، معتمدة على أحدث و أعظم برنامج في البحث اللغوي قوامه مدونة أكسفورد بحجم 800 مليون كلمة.

و المتصفح لإحدى طبعات المعجم (ط4، 2002)، يجد فيها كثيرا من المزايا من حيث إخراج المعجم إخراجا طيبا، مما يجعل منه السفر الحقيقي، و المرجع الغني الذي لا غنى عنه للمتمدرس و المتمرس على حد سواء. و واحدة من تلك المزايا، أنه يضم كثيرا من الكلمات قد أمعن مؤلفوه في تدقيقها و ضبطها، بما يتيح تمييز الفوارق بين معانيها و إن كانت تبدو مترادفات؛ كما بينوا مواطن كثير من الأخطاء الشائعة، فصححوها بعد أن عللوها، و مثال ذلك كلمة (*Steal*) في الشكل أدناه.

و الواقع أن ما عرفته معاجم أكسفورد من تحولات جذرية و نقلة نوعية، يعود الفضل فيها إلى معجم كوبيلد أحادي اللغة لطلاب اللغة الانجليزية (1987) *Collins Cobuild English Language Dictionary*، الذي يعد منعطف حاسما في تاريخ المعجمانية، إذ تم تصنيفه استنادا إلى أكبر مدونة مرجع آنئذ؛ ليصير بذلك تقليدا معهودا و عرفا محمودا في تصنيف معاجم المتعلمين. و البداية كانت أوائل الثمانينات، حيث تأسس كوبيلد (*Collins Birmingham University International Language Database*)، كمجمع بحث مقره بجامعة برمنغهام برئاسة جون سنكلير، في حين تتولى عملية تمويله هاربر كولينز للنشر *Harper Collins*. و لعل أهم إنجاز حققه مشروع كوبيلد، يتمثل في إنشاء مدونة بنك اللغة الانجليزية (*The Bank of English*)، بحجم 450 مليون كلمة، مادتها النصية مكتوبة و منطوقة، تجدد بانتظام.⁵

¹ - Cf. Geoffrey, Williams (2006) *Op. Cit.*, p. 152.

² - Cf. *Ibid.*

³ - للمزيد، ينظر الموقع:

<http://www.askoxford.com/oec/mainpage>

⁴ - Cf. COBUILD. University of Birmingham & Harper Collins Publishers, available at:

<http://titania.cobuild.collins.co.uk/>

⁵ - وغالبا ما تدعى بمدونة كوبيلد، ويمكن تحميل نسخة *Demo* من مواقع كوبيلد، بغرض استخدامها في توليد الفهارس الآلية بخصوص الكلمات أو المتلازمات اللفظية على حد سوى.

WHICH WORD? steal / rob / burgle / thief / robber / burglar			
Verbs			
Object of verb	steal	rob	burgle
money, property	<i>My wallet was stolen.</i>	<i>My wallet was robbed.</i>	<i>My wallet was bur- gled.</i>
person	<i>I've been stolen.</i>	<i>I've been robbed.</i>	<i>I've been burgled (= my house).</i>
bank, public building	<i>Thieves stole a bank last night.</i>	<i>Thieves robbed a bank last night.</i>	<i>Thieves burgled a bank last night.</i>
office, house, etc.	<i>Our house was stolen.</i>	<i>Our house was robbed.</i>	<i>Our house was bur- gled.</i>

Steal. You can also **steal from** somebody: *He had been stealing from his employer for years.*

Rob. You can also **be robbed of** something: *She was robbed of all her money and clothes.*

Burgle. This is BrE. AmE is **burglarize**.
Look also at the verb MUG.

Nouns
A **thief** is a person who steals things, especially secretly and without using violence. A **burglar** is a person who gets into a house, office, etc., especially by force, in order to steal things. A **robber** is a person who steals property from a person or public place, especially using violence.
Look also at MUGGER.

الشكل (13): مثال عن معجم أكسفورد (Steal).

يمكن استثمار المدونة في عدة أبحاث، بيد أنها قد صممت خصيصاً لغرض استخدامها في تخريج معاجم جديدة للغة الانجليزية، لعل أشهرها معجم (ب ب سي) للانجليزية BBC English Dictionary ، و هو معجم يعكس بدقة اللغة التي تذاق وفقها أخبار BBC¹. يضاف إلى ذلك معجم كولنز- كوبيلد للغة الانجليزية Collins Cobuild English Dictionary، إذ نجد في مقدمته ما نصه:

"All the statements that Cobuild makes are based on examination of the English **language in use**, as represented in the Bank of English – our present collection [...] The examples are all taken from this corpus, and **represent real language in use**; that is to say, each example is a chunk of language that served a **purpose** in a real text or conversation, and is not made up [...] The corpus is a rich store of information about the language. It enables dictionary writers to examine how words have been used, and to write dictionary entries which reflect the central and typical uses of the language. Every

¹ - Cf. S. Alansary et al. (2007) Op. Cit., p. 4.

entry in this dictionary is based on the **evidence** of the corpus".¹ [Emphasis added]

"إن ما خلصنا إليه في هذا المعجم من نتائج، قوامها ما لاحظناه من استعمالات اللغة الانجليزية تماما كما وردت في المدونة التي اعتمدناها – بنك اللغة الانجليزية [...] و لقد استقينا الأمثلة كلها من المدونة ذاتها، و هي تمثل الاستعمال الحقيقي للغة؛ بمعنى أن كل مثال عبارة عن مقطعة لغوية سبق و استعملت لغرض ما في نص أو محادثة حقيقيين، و لم تكن من قبيل الموضوعات [...] * . و المدونة بمثابة خزانة هامة تزخر بالمعطيات اللغوية؛ بحيث تتيح للمعجمائين إمعان النظر في كيفية استعمال الكلمات، و وضع مداخل معجمية تصور تصويرا دقيقا الاستعمالات اللغوية الأساس منها و النموذجية. و لقد وضعنا كل مدخل من هذا المعجم استنادا إلى ما ثبت لدينا من خلال المدونة ". [الترجمة لنا]

و جدير بالملاحظة أن ما يمتاز به معجم كوبيلد من المزايا – كما وردت على لسان مؤلفيه –، قد دفعت بدور نشر أخرى إلى الحذو حذوه، و السير على نهجه، كما هي الحال بالنسبة لمعاجم أكسفورد مثلا.

أما كامبريدج للدراسات اللغوية (Cambridge Language Survey)، فقد طورت مدونتين هما: مدونة كامبريدج العالمية (Cambridge International Corpus)، و مدونة كامبريدج للطلاب (Cambridge Learner's Corpus)؛ و ذلك بغية الإسهام في تخريج عدد من المعاجم اللغوية، على شاكلة Cambridge International Dictionary of English. أما معجم كامبريدج للطلاب المتقدمين Cambridge Advanced Learner's Dictionary، فقد اعتمد مؤلفوه على المدونة الثانية في إخراجهم وفق صورة حديثة، عمدوا فيها إلى إضافة بعض الملحوظات درءاً للأخطاء الشائعة التي يقع فيها المتدربون² – على غرار المثال أدناه.



الشكل (14): مثال عن معجم كامبريدج.

¹ - Collins Cobuild English Dictionary (1996) London: Harper Collins Publishers.

² - Cf. S. Alansary et al. (2007) Op. Cit., p. 4.

* - أي المصطنعة.

أما لونغمان للنشر Longman Publishers ، فقد عمدت إلى بناء مدونة كبيرة قوامها اللغة الانجليزية الأمريكية، في نمطها المنطوق و المكتوب، لأجل صناعة معجم لونغمان للغة الانجليزية الأمريكية Longman Dictionary of American English، كما استعانت بالمدونة البريطانية الوطنية (BNC) في تخريج معجم لونغمان للانجليزية المعاصرة¹ Longman Dictionary of Contemporary English

تلكم هي – إذن- أهم المدونات التي اعتمدت في صناعة أشهر المعاجم الانجليزية. و تبقى القائمة طويلة لا نستطيع أن نوفىها حقها لعنتين؛ أولاها أن المقام لا يسمح لنا بالاستفاضة أكثر مما استفضنا، و ثانيتهما أننا لا نسعى إلى إحصاء جميع تلك المعاجم أو جردها، بل إننا نتوسم من خلال ما ضربناه من أمثلة تبليغ فكرة، لا مرأى في أنها على قدر بالغ من الأهمية، مفادها أن المعجماتية قد استفادت كثيرا من لسانيات المدونة؛ إذ بفضلها عرفت كيف تزواج بين المقاربتين المعيارية و الوصفية، أي بين كيف يجب أن تستعمل اللغة، و كيف تستعمل اللغة فعلا، ليتحقق بذلك ما تنبأ به كثيرون من ثورة مدوناتية.

و لنا أن نتصور كيف استطاعت لسانيات المدونة اقتحام المصطلحية، أو الأخرى كيف اضطرت المصطلحية إلى إقحام لسانيات المدونة في عقر أبحاثها، من خلال استثمارها للمدونات الثنائية أو المتعددة اللغات، لاسيما المتوازية منها و المتخصصة، إذ بتنظيرها* يمكن الحصول على ثنائيات أو قوائم من المصطلحات، تسهم بحظ وافر في بناء المعاجم المتخصصة.

2.1.I المدونات، النحو و الصرف:

كان النحاة – في بادئ الأمر- يعتمدون أساسا على ما يتمتعون به من حدس في وضع قواعد اللغة، أو يستعينون بنطاق تلك اللغة للحكم على سلامة الجمل نحويا Grammaticality، أو مجانبتها للصواب؛ و في ذلك، تغيب بل إقصاء تام لكل طريقة من الطرق التجريبية في دراسة جانب الاستخدام الفعلي. و من ثم فإنهم لا يعيرون التنوع في الاستعمال اللغوي أهمية تذكر.²

و النحوي، و هو يشتغل على المستويات الثلاثة: مستوى الكلمة و مستوى الجملة و مستوى الخطاب، يسعى إلى فهم البنية اللغوية واستبيان كنهها؛ إلا أن استثمار المدونات في ذلك يوفر عليه الكثير، بحيث يتيح له إدراك البنى التي تتخذها مختلف التراكييب النحوية، و يمكنه من الإلمام بشتى ضروب استعمالاتها، و التعرف على مختلف العوامل التي تحكم الخيارات بين البنى رغم تنوعها، مما يسمح له بتوظيف ما تسنى له الحصول عليه في وضع قواعد اللغة أيا كانت.

أضف إلى ذلك، إن ما يصنع من المدونات أداة هامة و لازمة من لوازم البحث في النحو، هو إمكانية التحليل الكمي للعينات الممثلة لضرب لغوي بأكمله. فضلا على دورها

¹ - Cf. Ibid.

*- ينظر الفصل الثالث.

² - Voir: J. Aarts (1991) Intuition-based and observation based grammars. In j. Aijmer & B. Altenberg (eds.) Directions in English Corpus Linguistics. London /New York: Longman

(أي المدونات) كمعطيات تجريبية يتسنى من خلالها اختبار الفرضيات التي يصنعها المنظرون في النحو.¹

و تاريخ النحو الإنجليزي – هو الآخر – حافل بالأمثلة التي تعبر عن السيادة التي يتمتع بها تقليد اعتماد المدونات في مثل هذه الأبحاث و الدراسات. و نذكر في هذا الصدد مؤلفين هامين لراندولف كويرك (Randolph Quirk) و سيدني غرينبوم (Sidney Greenbaum) و جيوفري ليتش (Geoffrey Leech) و يان سبارتفيك (Ian Svartvic)، هما:

• نحو الإنجليزية المعاصرة (1972)

A Grammar of Contemporary English

• النحو الشامل في اللغة الإنجليزية

(1985) *A Comprehensive Grammar of the English Language (CGEL)*

و في المؤلفين كليهما، يلقى القارئ كثيرا من النقاش بخصوص البنى النحوية، و الذي يستند إلى تحاليل عن مدونة لندن. و في معرض الحديث عن الجملة الاسمية في المؤلف الثاني، يورد المؤلفون جدول توزيع تكرارات الجمل الاسمية البسيطة و المركبة في مختلف الأجناس الكتابية بالمدونة ذاتها.²

و ثمة دراسات نحوية أخرى تستثمر التحاليل الكمية للمعطيات، على غرار دراسة *Schmied* (1993) لصلة الموصول. كما يولي الباحثون اهتماما بالغاً بدراسة التكرارات النحوية، كدراسة *Oostdijk* و *Haan* (1994) التي تنبني على تحليل تكرار مختلف أنواع الجملة في اللغة الإنجليزية.³

كما أن مؤلف *Greenbaum* (1996) و الموسوم بـ : (*Oxford English Grammar*)، يعتمد كلية على الجزء البريطاني من المدونة العالمية للغة الإنجليزية (*ICE-GB*).

و استنادا إلى ما توفره مدونة بنك اللغة الإنجليزية (*Bank of English*) من معطيات، أصدرت كوبيلد سلسلة في النحو تعد مرجعا بحق يتزود منه طلاب الإنجليزية.⁴ أما مدونة لونغمان للغة الإنجليزية المنطوقة و المكتوبة، الأمريكية و البريطانية (بحجم 40 مليون كلمة)، فقد اعتمدها بيبر *Biber* و رفاقه في تأليف

Longman Grammar of Spoken and Written English (1999)

و في ذلك يقول:

" The LGSW adopts a corpus-based approach, which means that the grammatical descriptions are based on the patterns of structure and use found in a large collection of spoken and

¹ - Cf. S. Alansary (2007) *Op. Cit.*, p. 5.

² - Cf. *Ibid*, p.6.

³ - Voir: J. Schmied (1993) Qualitative and quantitative research approaches to English relative constructions. In: C. Souter & E. Atwell (eds.) *Corpus-based computational linguistics* (85-86). Amsterdam: Rodopi. N. Oostdijk & P. de Haan (1994) Clause patterns in Modern British English. A Corpus-based (quantitative) study, *ICAME Journal* 18, pp. 41-80.

⁴ - Cf. J. M. Sinclair (1987) (ed.) *Looking Up—An account of the COBUILD Project in lexical computing*. London: Collins ELT.

*written texts, stored electronically, and searchable by computer."*¹

" إن هذا المؤلف يتبنى المقاربة المدونانية، بمعنى أن ما يتضمنه من وصف نحوي قوامه قوالب البنى و الاستعمالات المبنوثة في ثنايا مجموعة كبيرة من النصوص المنطوقة و المكتوبة، المحفوظة الكترونياً، و التي يمكن بحثها باعتماد الحاسوب". [الترجمة لنا] و لا مشاحة في أن المؤلف يتسم بطابع الجدة و الابتكار، من حيث اهتمامه بالمعطيات المدونانية، و اعتماده في درس الفروق بين اللغتين المنطوقة و المكتوبة. و لقد كان لمؤلف كويرك "النحو الشامل في اللغة الإنجليزية (CGEL)" أثراً على المؤلف المذكور، بوصفه سابقة في النحو الإنجليزي، و منه استلهم بيير و رفاقه فكرة تخريج "نحو لونغمان في الإنجليزية المنطوقة و المكتوبة" (LGSWE)، و هو القائل في هذا الصدد:

*" [CGEL is] a previous large-scale grammar of English from which we have taken inspiration for a project of similar scope."*²

و في الضفة الأخرى، نلفى كثيراً من الباحثين (و على رأسهم Jan Aarts) يرافعون بحماسة في دعوى "الجيل الجديد من مؤلفات النحو الإنجليزي باعتماد المدونة" -على حد تعبير يان آرتس-³، إذ عمد إلى وضع أربعة ضوابط يشترط توفرها في كل مؤلف نحوي يتصدى للغة المجسدة في الواقع الملموس. و قد صادف حدوث ذلك عشية ظهور "نحو لونغمان الإنجليزية المنطوقة و المكتوبة" (LGSWE)، و كانت المصادفة أكبر حين تبين أن المؤلف يوافق تلك الضوابط، و يجسدها على الوجه الأكمل، فكان أشمل و أفضل مرجع في النحو الإنجليزي بوصفه:

1- يتناول أربعة سجلات جوهرية في اللغة الإنجليزية، و يقارن بين المنطوق منها و المكتوب.

2- و هو يجمع بين الوصفين الكمي و الكيفي للمعطيات ضمن المدونة.

3- كما أنه يأخذ بعين الاعتبار أثر مختلف العوامل الخطابية على الخيارات اللغوية؛ بمعنى أنه يحدث علاقة بين الظواهر التي تقع خارج النظام اللغوي و الظواهر التي تنتمي إليه.

4- فضلاً على ذلك، يأخذ بعين الاعتبار التعالق *Interdependence* بين المعجم و النحو، بحيث يحدد الوحدات المعجمية التي تنزع إلى التوارد مع بعض البنائات التركيبية.¹

¹ - D. Biber, S. Johansson, G. Leech, S. Conrad & E. Finegan (1999) Longman Grammar of Spoken and Written English. Harlow: Pearson Education, p. 4.

² - Ibid, Preface: viii.

³ - Jan, Aarts (2000) Towards a new generation of corpus-based English grammars. In: B. Lewandowska-Tomaszczyk & P.J. Melia (eds.) PALC'99: Practical applications in language corpora. Frankfurt am Main: Peter Lang, pp. 17-36.

* - تشمل السجلات الجوهرية الأربعة (Core registres)، كلا من: المحادثة، الخيال، لغة الصحافة، النثر الأكاديمي.

3.1.I المدونات، الدلائيات و الأسلوبيات:

إنها لمقاربة جديدة تلك التي يعتمد فيها الباحث على المدونة لاشتقاق معاني الكلمات، و معالجة المسائل المتعلقة بها على اختلاف صنوفها (كالترايف و الاشتراك اللفظي *Hyponymy*، و قس على ذلك)، معالجة تتسم بكثير من الموضوعية في الطرح والتناول، و الدقة في النتائج، نظرا لاستنادها إلى معطيات تجريبية، بعيدا عن مظان الاستبطان. تلك المقاربة هي نتاج استثمار لسانيات المدونة في حقل الدلائيات. و وجه الاختلاف بين هذه و تلك، أن الأمثلة المأخوذة عن المدونات أمثلة في سياق التواصل – كما أشرنا غير ذي مرة-؛ أما تلك التي يوظفها الاستبطان فلا سياق لها، و في ذلك تقول أن كوندامين (Anne Condamines):

" En effet, les textes qui nous parviennent sont forcément des textes *situés* alors que les exemples produits par introspection sont censés être totalement *décontextualisés*".² [Emphasis added]

و في بحث له حول السمات الدلالية (*Semantic distinctions*) في اللغة الإنجليزية، بيّن Mindt إمكانية استخدام مدونة في وضع معايير موضوعية لربط الوحدات اللسانية بمعانيها، كما أنه يتحدث عن التوجهات التقليدية في المجال التي كانت تعتمد على الحدس كمرجع في وصف تلك المعاني.³ و يؤكد Mindt في سياق متصل، بأن السمات الدلالية ترتبط في النصوص بسياقات مميزة و ملحوظة، منها ما يتعلق بالتركيب و الصرف و التنعيم. و باعتبار الفضاءات التي تتموقع فيها الوحدات اللسانية، يمكن للباحث أن يتوصل إلى مؤشر لسمة دلالية معينة، على نحو يعتمد على التجريب و يتصف بالموضوعية.⁴ أما في حقل الأسلوبيات، فإنه لا يخفى على أحد مدى إسهام لسانيات المدونة في دراسة الأعمال الأدبية، بل الخاصة "الأدبية" بعامة (*Literariness*). و يؤكد إيفان (Evans)⁵، على أن التحليل الإحصائي للنصوص الأدبية (أو ما يعرف بالقياسية الأسلوبية (*Stylometrics*) كثيرا ما وظفه الباحثون للفصل في النصوص المتنازع فيها بإثبات أصحابها الشرعيين.

و في سياق آخر، يتحدث عن أهمية مقارنة مدونة صغيرة من النصوص الأدبية بمدونة أخرى مرجع، و ذلك للكشف عن "الأدوات و الإجراءات الأدبية" أو المهارات، ولمعرفة كيف أنها تختلف عن تلك المعهودة في اللغة الإنجليزية اليومية.

¹ - Cf. Joybrato, Mukherjee (2006) Corpus linguistics and English reference grammars. In: Antoinette Renouf & Andrew Kehoe (eds.) *Language and Computers. The Changing Face of Corpus Linguistics*, vol. 55 (pp. 337- 354). Amsterdam/ New York: Rodopi.

² - Anne, Condamines (2006) Modes de construction du sens en corpus spécialisé. *Cahiers de Grammaire* 30, 'Spécial Anniversaire', p. 75.

³ - Cf. Dieter, Mindt (1991) Syntactic evidence for semantic distinctions in English. In: J. Aijmer & B. Altenberg (eds.) *Directions in English Corpus Linguistics*. Amsterdam: Rodopi.

⁴ - Cf. T. McEnery & A. Wilson (2001) *Corpus Linguistics*. Edinburgh University Press.

⁵ - Cf. D. Evans (2006) *Op. Cit.*

و في إحدى دراساته ذات الصلة، يتحدث "لو" ¹ (Louw) عن أهمية المعطيات المدوناتية في إثبات أحكام الطلبة بخصوص النصوص الأدبية و هي أحكام نابغة من الحدس. مثال ذلك:

حين راح الطلبة يدرسون عبارة "wielding a..."، قصد إثبات فرضية أن استعمالها في البيت: "and crawling sideburns, wielding a guitar" استعمال ساخر تهكمي. و كما كان متوقعا، وجد الطلبة أن هذه العبارة تستعمل – في الغالب الأعم – مع ضرب من السلاح، أو أداة تستخدم لذات الغرض، كأن يقال مثلا:

He was wielding a large knife. أي: يلوح/ يهدد بسكين كبيرة.

و قد اكتشف الطلبة – علاوة على ذلك – من خلال جدول التكرار، أن تلك العبارة تستعمل استعمالا ساخرا بدرجة أعلى بخلاف ما توقعوه.²

4.1.I- المدونات و المعالجة الآلية للغات:

إن لكثير من المشتغلين في حقل لسانيات المدونة، اهتمامات و اعتبارات لسانية حوسبية،³ أكثر من أنها لسانية صرف؛ حيث يعتمد هؤلاء اللسانيون إلى بناء مدونات بغرض استخدامها في بعض الأبحاث التي تتعلق باللسانيات الحوسبية، على غرار: الترميز بأقسام الكلام، الإعراب، استخلاص البيانات و المعطيات، تطوير آليات التعرف إلى الكلام، و نحوها من الأبحاث التي تدخل في إطار ما يعرف بالمعالجة الآلية للغات الطبيعية.

و لما كانت تلكم هي بعض اهتمامات الباحثين في المجال المذكور، فإن ما يصنعونه من مدونات تختلف – هي الأخرى – من حيث التصميم،⁴ عن المدونات المستخدمة في أبحاث لسانية بحث؛ فبينما يحرص الباحث في لسانيات المدونة الوصفية (*Descriptive Corpus Linguistics*) أيما حرص على تصميم مدونات ممثلة، نجد نظيره في مجال المعالجة الآلية للغات لا يكثر لذلك، طالما أن تدريب نظام في الإعراب أو التعرف إلى الكلام – مثلا – لا يستدعي بالضرورة اعتماد مدونة متوازنة و ممثلة.⁵

بيد أن تباين الاهتمامات و اختلاف التوجهات، لا يعني بتاتا الانزواء أو استقلال الحقلين، و أن بينهما برزخ لا يبغيان؛ بل إن بينهما لتواشجا و تمازجا واضحين للعيان، بما يكفي لإحداث تقاطع معرفي و إجرائي بين الاختصاصين أو الفرعين، مما يخدم الأصل – أي لسانيات المدونة عامة. و لنا أن نتوهم مثال ذلك في الأدوات التي تعمل اللسانيات الحوسبية على تطويرها – كآليات الترميز خصوصا –، و هي تعد أداة رئيسة يتوقف عليها إجراء هام في لسانيات المدونة الوصفية، أي الترميز بأقسام الكلام أو الإعراب على حد سواء.

¹ - Cf. B. Louw (1997) The Role of Corpora in Critical Literary Appreciation. In: A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (eds.) *Teaching and Language Corpora*. London: Longman.

* - البيت من قصيدة Elvis Presley، للشاعر الإنجليزي الأمريكي Thom Gunn.

² - Cf. David, Evans (2006) *Op. Cit.*

³ - تعنى اللسانيات الحوسبية *Computational Linguistics* بتطوير أدوات تحليل المدونات، و خوارزميات معالجة اللغات الطبيعية ودراساتها صرفيا، نحويا، دلاليا،... وكذا تحليل مختلف الظواهر اللغوية.

⁴ - و قد أشرنا – غير ذي مرة – إلى أن تصميم المدونة يحدده غرض التطبيق المراد إنجازه باعتماده.

⁵ - Cf. S. Alansary et al. (2007) *Op. Cit.*, pp. 7-10.

و صور ذلك التقارب و التعاون كثيرة، كما أن ثمارهما واضحة؛ و من ذلك مَثَلُ المدونة البريطانية الوطنية، التي تعكس بصدق معنى أن يلتقي اللساني بالحوسبي، ويستلهم كل واحد من الآخر. فالمدونة قام بإنشائها فريق ¹UCREL بجامعة لنكستر، و هو فريق يتكون من باحثين في لسانيات المدونة (الوصفية) و آخرين في مجال اللسانيات الحوسبية. و فوق ذلك، قام الفريق بتدوين نظام الترميز ²(CLAWS) الذي استخدم في ترميز المدونة عينها.

و مثال ذلك أيضا، ما توصل إليه الباحثون في مجال اللسانيات الحوسبية من استحداث نظام نيمجن (Nijmegen) للترميز و الإعراب، و استخدامه في معالجة الجزء البريطاني من المدونة الإنجليزية العالمية (ICE-GB)، مما وفر على اللسانيين كثيرا من الجهد و الوقت المستغرقين قبلا في تحليل المدونة يدويا.³

I.1.5- المدونات واللسانيات الاجتماعية:

تُعنى اللسانيات الاجتماعية بكشف تلك العلاقات القائمة بين الظواهر اللغوية والاجتماعية، فتدرس الاختلاف اللغوي باعتباره أمانة من أمارات الانتماء إلى طبقة اجتماعية معينة، أو جماعة محددة ... أي أنها تتقصى تأثير مختلف العوامل الاجتماعية (كالسن و الجنس و المستوى الاجتماعي و الثقافي...) على استعمال الفرد للغة، من منطلق أن تلك المحددات تسهم في فهم الملفوظ و إدراكه، وفق ما يذهب إليه يودا (YODA):

" Pour la sociolinguistique, la compréhension d'un énoncé dépasse le cadre linguistique et englobe les facteurs sociaux." ⁴

و كالعادة، عرفت لسانيات المدونة كيف تتوغل إلى الحقل، و تستقطب اهتمام السوسيولسانيين؛ بحيث عمدوا إلى تبنيتها في أبحاثهم، لاسيما تلك التي تتعلق بالسجل. و لا شك أن فكرة اعتماد المدونات في الدراسات المتعلقة بالسجل، نابعة من صميم الصعوبات التي تواجه الباحث ذي العدة التقليدية، من جهة؛ و بالنظر إلى متطلبات مثل هاته الدراسات من جهة أخرى. و يحددها الأنصاري في مجمل النقاط التالية:⁵

- إدراج كم هائل من النصوص؛ ذلك أن الأبحاث التي تقوم على بضعة متبضعة من النصوص، غالبا ما تفتقر إلى الدقة.

- اعتبار مجموعة واسعة من الخصائص اللسانية؛ ذلك أن الأبحاث التي تعتمد على قلة قليلة من تلك الخصائص، لا تعطي توضيحات شاملة بخصوص السجل، ثم إن التعميم انطلاقا منها نادرا ما يحتمل الصواب.
- المقارنة بين السجلات؛ و ذلك لضرورة اعتماد حد فاصل *baseline* للمقارنة، من أجل معرفة ما إذا كان توظيف خاصية لسانية معينة ضمن سجل ما، شائعا أم نادرا.

¹ - UCREL: University Centre for Computer Corpus Research on Language, at Lancaster University.

<http://www.comp.lancs.ac.uk/computing/research/ucrel>

² - CLAWS (the Constituent Likelihood Automatic Word-tagging System) is a POS tagging software.

<http://www.comp.lancs.ac.uk/computing/research/ucrel/claws/>

³ - Cf. Alansary et al. (2007) *Ibid.*, pp. 7-10.

⁴ - Lalbila Aristide, Yoda (2005) La traduction médicale du français vers le mooré et le bias. Thèse, p. 104.

⁵ - Cf. Alansary et al. (2007) *Op. Cit.*, p. 8.

و على هذا النحو، نرى أن تلك الصعوبات تجعل من استخدام المدونات ضرورة لا بد منها؛ و كثيرة هي تلك الأبحاث التي تناولت الموضوع بعمق، فبينت أن اللغة غالبا ما تسلك سلوكات مختلفة باختلاف السجلات، و لكل قوالبه و قواعده الخاصة، أي أن الخصوصيات اللغوية تتغير بتغير السجلات.¹

و من أمثلة الدراسات السوسiolسانية باعتماد المدونات، نذكر دراسة Kjellmer²، تتناول التحيز لدى الرجال في اللغتين الإنجليزية البريطانية والأمريكية، حيث عمد إلى ملاحظة توارد ضمائر المذكر و المؤنث، و كذا توارد الثنائيات (رجل/ رجل) و(امرأة/نساء) في مدونتي "براون" و "لوب".

و لقد رأينا - في ما سبق - كيف أن تلك المتغيرات أو المحددات تحكم توزيع النصوص ضمن المدونة البريطانية الوطنية (BNC)، لاسيما في الجزء المنطوق منها؛ إذ نجد أن كل محادثة تشتمل على رأسية الملف و بيان بتلك المتغيرات (كس كل متحدث و جنسه) في بداية العينة.

و ثمة برنامجيات (على غرار SARA) صممت بغرض قراءة الرأسيات، و القيام بتحليل المدونات انطلاقا من جملة تلك المتغيرات.³ أما بالنسبة للمدونة العالمية للغة الإنجليزية، لاسيما الجزء البريطاني منها (ICE-GB)، فإن البيانات الإثنوغرافية التي تتعلق بالكاتب/المتكلم، محفوظة ضمن قاعدة خصصت لذلك؛ و صمم برنامج خاص لتحليل المدونة يدعى ICE CUP⁴.

I.2- المدونات و تعليم اللغات:

إن ما حظيت به اللغة الإنجليزية من عناية، و ما استأثرت به من اهتمام في مجال لسانيات المدونة، قد صاحبتة خطوة أخرى في حقل تعليمية اللغات باعتماد نفس الأدوات، أي المدونات. و قد ظهر ذلك جليا أوائل التسعينات، من خلال مؤلفين لبعض الباحثين الرواد في هذا المجال:

- J. Johns (1986) *Micro – concord: A language learner's research tool.* System 14 (2), pp.151-162.
- C. Tribble & G. Jones (1990) *Concordances in the classroom: A resource book for teachers.* Harlow, England: Longman.

و لقد تلاهما زخم هائل من الدراسات، خصصت لمختلف أوجه استخدام المدونات في تعليمية اللغات، من أبرزها تلك المنشورات الصادرة عن مؤتمرات TALC حول التعليمية و المدونات اللغوية (Teaching and Language Corpora).⁵

¹ - Cf. D. Krieger (2003) *Op. Cit.*

² - Cf. G. Kjellmer (1986) *The lesser man: Observations on the role of woman in modern English writings.* In: J. Aarts & W. Meijas (eds.) *Corpus Linguistics II* (163- 176). Amsterdam: Rodopi.

* - رأسية (Header): بمعنى السطر أو النصيص الذي يضاف آليا إلى أعلى كل صفحة نقوم بطباعتها من الحاسوب؛ كما يمكن أن يحمل البيانات الخاصة بالنص، كمؤلفه، العنوان، المصدر، ... و يمكن أن يرد السطر أو النصيص أسفل الصفحة فيسمى حاشية (Footer)؛ كان نجد مثلا في صفحة:

<file:///post01/mes documents/Guy Aston Corpus use and learning to translate.htm>

³ - SARA web site : <http://www.natcorp.ox.ac.uk/tools/sara/>

⁴ - Cf. Alansary et al. (2007) *Op. Cit.*, p. 9.

⁵ - Voir: L. Burnard & T. McEnery (eds.) (2000) *Rethinking Language pedagogy from a corpus perspective: Papers from the Third International Conference on Teaching and Language Corpora.* Frankfurt: Peter Lang. B. Kettemann & G. Marko (eds.) (2002) *Teaching and learning by doing corpus analysis.* Amsterdam/ New York: Rodopi.

و الواقع أن الاهتمام بهذا الموضوع نابع أساسا عن الوعي بأهمية الأمثلة المستقاة من المدونات، إذ تخول للمتعلّم النظر في ضروب شتى من الجمل التي سوف يصادفها - بلا شك- أثناء التطبيق الفعلي للغة، أي استعمالها في مواقف حياتية حقيقية، وفق ما يذهب إليه الأنصاري في قوله:

" [...] Corpus examples are important in language learning as they expose students to the kinds of sentences that they will encounter when using the language in real life situations."¹

ثم إن هذه الممارسة في إطار التعليمية، أي اعتماد المدونات، سواء في تدريس اللغة الأم، أو اكتساب اللغة الأجنبية، ممارسة تعكس - هي الأخرى - فصلا بين المقارنتين التجريبية و العقلانية، و انفصالا عن الثانية، خاصة بعد أن تقطن المشتغلون في المجال إلى أن التدريس القائم على الوسائل التي لا تمت إلى التجريب بصلة، أو حتى الطريقة في حد ذاتها، لن يؤتي ثمارا طالما أنه لا يتوخى الرؤية العلمية الموضوعية. وعلى هذا النحو، عرفت لسانيات المدونة كيف تشق طريقها داخل صف تعليم اللغات. و يحدد الأنصاري في ثلاث نقاط أهمية المقاربة باعتماد المدونة في مجال اكتساب اللغة:²

- أولها: إن ما تتضمنه المدونات من مادة نصية هائلة، تسمح - بعد تحليلها- بإجراء التعميم، و ذلك عبر مراحل مختلفة من العملية التعليمية.
 - و ثانيها: يفيد تحليل المدونات في اكتشاف كثير من الخصائص اللسانية، على شاكلة القوالب الاعتيادية و المألوفة، مما يعني إمكانية الإلمام بالاستعمالات اللغوية، و ذلك عبر مراحل العملية التعليمية.
 - و آخرها: إن المادة النصية تنتمي إلى تشكيلة متنوعة من السجلات، و تحليلها (في إطار المدونة) يعطي أفقا أوسع بخصوص تطور اللغة.
- و الحال هذه، جنح كثيرون إلى البحث في سبل تعزيز هذا التوجه الجديد، و توطيد دعائمه، فركزوا - في بادئ الأمر- على بعض الجوانب التي تتعلق خاصة بحجم المدونات التي ينبغي إدراجها (كبيرة أم صغيرة)، و نوعها (عامة أم متخصصة، مرمرزة أم خام).³

و من أمثلة تلك الدراسات، دراسة كل من أستون (1997) و روي (2000) حيث يوصيان بضرورة استخدام مدونات صغيرة الحجم تهيأ وفق حاجات المتعلّم، بينما يعول آخرون كثيرا على فهرسة مدونات كبيرة، أمثال: برنارديني (2000) Bernardini، و تشانغ، فارن و كسان - فانغ (2003) Cheng, Warren & Xun-feng.

¹ - Cf. S. Alansary et al. (2007) Op. Cit., p. 9.

² - Cf. Ibid.

³ - Cf. Angela, Chambers (2005) Integrating Corpus Consultation in Language Studies. LIT 9(2), p. 111.

⁴ - Voir: G. Aston (1997) Small and large corpora in language learning. In: B. Lewandowska-Tomaszczyk & P.J. Melia (eds.) Practical applications in language corpora (51-62). Lodz, Poland: Lodz University Press. P. Roe (2000) The ASTCOVEA German Grammar in conText Project. In: B. Dodd (ed.) Working with German Corpora (199-216). Birmingham, England: University of Birmingham Press.

S. Bernardini (2000b) Systematising serendipity: Proposals for concordancing large corpora with language learners. In: L. Burnard & T. McEnery (eds.) Op. Cit., pp. 225-234.

W. Cheng, M. Warren & X. Xun-feng (2003) The language learner as language researcher: Putting corpus linguistics on the time-table. System 31(2), pp. 173-186.

و ثمة مسائل بيداغوجية أخرى جديرة بالدرس و التمحيص، قد تناول بعضها كثير من الباحثين، و بعضها الآخر يستحق المزيد من البحث و التعمق في التناول، على غرار: أهمية الاستخدام الفوري للمدونات، أي إدماج المتعلم مباشرة في الوسط المدوناتي (Direct access) دونما وساطة الأستاذ؛ و من أمثلة الأبحاث التي تناولت هذه المسألة، نذكر دراسة كل من برنارديني (2002)، كنيدي و ميشلي (2001) Kennedy & Miceli، أما جونز Johns (1997:113)، و إن كان يؤكد مبدئياً الفكرة، إلا أنه يوصي باستخدام النتائج المحصل عليها من المدونات مع تدخل الأستاذ كمرحلة أولى في العمل.¹ و من أمثلة تلك المسائل أيضاً: ضبط الإستراتيجيات التي تتيح للمتعلم الاستفادة من عملية الاستخدام المباشر؛ يضاف إلى ذلك كله، تحديد الأدوات أو الوسائل التي تسهم في إدماج المدونات ضمن العملية التعليمية، و تضمن نجاحها على أكمل وجه.² و تؤكد أنجيلا تشامبرز في سياق متصل على ضرورة إيلاء فائق العناية لمسألة إعداد الدرس و تصميمه؛ و ذلك برسم خطة محكمة تؤطر العملية التعليمية باعتماد المدونات في مجال اللغات، ضمن الدراسات العليا.

"[...] there is still considerable scope for developments, particularly in the area of course design and structure, concerning how one can successfully integrate corpus consultation into a program of language study in higher education".³

و تحضرنا بهذا الصدد، تجربة رائدة للثلاثي ماكنري – بايكر – ويلسن (1995) بجامعة لنكستر، حيث حاول الباحثون تحديد أهمية سياتر (Cytos) و فعاليتها في تدريس أقسام الكلام طيلة فصل دراسي كامل؛ و ذلك بمقارنة عمل فوجين من الطلبة، اعتمد في تدريس الفوج الأول على سياتر، أما الفوج الثاني فقد تلقى الدروس وفق الطريقة التقليدية المعهودة، فكان أن حقق الفوج الأول نتائج طيبة و أفضل بكثير من الفوج الثاني.⁴ و تحقيقاً لتلك الأهداف المقصودة و الغايات المنشودة، ثمة ضربان رئيسان من المدونات يعتمدان داخل صف تعليم اللغات:

- مدونة الناطق باللغة الأصلي (The Native-speaker corpus).

- مدونة المتعلم (The Learner corpus).

أما المدونة اللغوية الأولى فهي معيارية، لأن مادتها النصية (منطوقة أم مكتوبة) صادرة عن متكلمين أصليين بتلك اللغة، على غرار: المدونة البريطانية الوطنية مثلاً. و بناءً هذا النوع من المدونات يأتي لتأكيد مسلمتين من المسلمات التي تقوم عليها لسانيات المدونة:

¹ - Voir: S. Bernardini (2002a) Exploring new directions for discovery learning. In: B. Kettemann & G. Marko (eds.) Op. Cit., pp. 165-182.

A. Chambers & I. O'Sullivan (2004) Corpus consultation and advanced learners' writing skills in French. ReCALL 16(1), pp. 158-172.

C. Kennedy & T. Miceli (2001) An evaluation of intermediate students' approaches to corpus investigation, LLT 5(3), pp. 77-90.

T. Johns (1997) Contexts: The background development, and trialing of a concordance-based CALL program. In: A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (eds.) Op. Cit., pp. 100-115.

² - Cf. A. Chambers (2005) Op. Cit., p. 111.

³ - Ibid., pp. 111-112.

⁴ - Cf. T. McEnery & A. Wilson (2001) Op. Cit.

1- محدودية "المعجم الذهني الخاص"؛ إذ - و الحق يقال - إن ما يمتلكه الفرد من معرفة حدسية بلغته التي يتكلم بها، يتيح له تحليل المعطيات اللغوية، بيد أن ذلكم الحدس لا يمكن أن يعتمد عليه دائماً في معرفة كيف تستعمل اللغة فعلاً، ثم إن الاستبطان وحده لا يكفي بأي حال لإدراك معاني الكلمات و سلوكاتها.

2- إن دراسة المعطيات المدونانية توحى بالكثير عن الطبيعة التكرارية التي تتسم بها السلوكات اللغوية في أغلبها الأعم، كما تكشف عما هو متواضع عليه من تلك السلوكات. و هي تبين، فضلاً على ذلك، ما ينطوي عليه كل من السياق (Context) و مرافقات النص (Co-Text) من أثر على كيفية استعمالنا للكلمات، و على طريقة فهمنا.¹

أما النوع الثاني، أي مدونة المتعلم، فقوامها نصوص صادرة عن متعلمي اللغة الأجنبية، على غرار المدونة العالمية لمتعلمي اللغة الإنجليزية International Corpus of Learner English (ICLE) بحجم مليوني كلمة، و مدونة لونغمان للمتعلمين (Longman's Learner Corpus) و مدونة جامعة هونغ كونغ للعلوم والتكنولوجيا² Hong Kong University of Science and Technology (HKUST) Learner Corpus.

و بالمثل، فإن الدافع إلى صناعة مثل هاته المدونات هو محدودية القدرة الحدسية، إذ لا شك أن الأستاذ المتمرس يدرك جيداً مواطن ضعف طلبته، و مواقع الزلل لديهم، سواء على مستوى المفردات أو البنى؛ بيد أن هذا الإدراك يبقى محدوداً، و لا يلم إلا ببعض جزئيات المشاكل و العوائق التي تعترض المتدرس عبر مختلف الأطوار التعليمية.³

و في تجربة لها بهذا الصدد، حاولت⁴ (S. Granger) دراسة الاستعمالات اللغوية لدى طلاب اللغة الإنجليزية المتقدمين، من جهة؛ و لدى الناطقين بها من جهة أخرى، بغية المقارنة بينها و تحديد أوجه الاختلاف، و لقد استعانت في ذلك بمفهرس آلي، فتوصلت بعد التحليل إلى نتائج هامة، نذكر منها تلك التي تتعلق باستعمالات الفعل (discuss): إذ وجدت أن توظيف هذا الفعل يكون في صيغة المبني للمجهول عند الناطقين الأصليين باللغة الإنجليزية، كأن يقال مثلاً :

- It will be discussed more fully below...
- As discussed in Chapter 3,...

بينما ينزع المتعلم إلى توظيفه ضمن صيغة المبني للمعلوم (مع ضمير المتكلم)، على شاكلة:

- I would like to discuss....
- I will discuss...

علاوة على ذلك، قد تأكد لديها من خلال الدراسة أن العديد من الطلبة لا يدركون بأن هذا الفعل فعل متعد، بحيث أنهم يميلون - مثلاً - إلى قول:

- Politicians discuss about this problem.

و الأصح أن يقال:

¹ - Cf. Michael, Rundell & S. Granger (2007) From corpora to confidence. MED Magazine, Issue 46, August 2007. Retrieved: 10/11/2007.

² - See: ICLE at <http://cecl.fltr.ucl.ac.be/research.html>

³ - Cf. M. Rundell & S. Granger (2007) Ibid.

⁴ - CF. S. Granger & Vander Viana (2007) Op. Cit., p. 12.

- *Politicians discuss this problem.*

و تذهب الباحثة إلى أبعد من ذلك، حين تلج على أستاذ اللغة الإنجليزية اعتماد المدونات في ممارسته التعليمية، و استعمالها تماما كاستعمال المعجم أو كتاب النحو؛ بحيث ما إن يطرأ إشكال، من قبيل: هل نستعمل *increase* أم *increase of*؟ في أي سياق نستعمل كلمة *person* بالجمع؟ و هلم جرا، يطلب من أحد طلبته النظر في المدونة، فتتحقق بذلك فائدتان: يحفز الطالب - من ناحية- و تتاح له فرصة الاستكشاف والاكتشاف، فتتولد لديه الحماسة لذلك؛ و من ناحية أخرى، تعين المدونة الأستاذ في عمله كثيرا، و تسعفه من حيث لا يدري، لاسيما إذا لم يكن ناطقا أصليا باللغة الإنجليزية، إذ - و الحق يقال- قد يعجز عن الإلمام بجميع جوانب الاستعمال، و ليس غريبا أن يغيب عنه جلها إن هو أدرك بعضها.¹

و على هذا النحو، نجد *Michael Hoey* يحصي قائمة بأهم الأسئلة التي يمكن أن يطرحها الأستاذ/ المتعلم بخصوص كلمة ما:

- 1- إلى أي القوالب المعجمية تنتمي هذه الكلمة؟
 - 2- هل تقترن الكلمة على نحو منتظم بمعان أخرى؟
 - 3- أي البنى النموذجية التي تظهر فيها هذه الكلمة؟
 - 4- هل ثمة علاقات بين استعمالات الكلمة/ معانيها و البنى التي تظهر فيها؟
 - 5- هل ترتبط الكلمة بموقع معين ضمن هيكل النص؟²
- و الإجابة عن هذه الأسئلة لا تتوفر إلا بتوفر مدونة و مفهرس آلي؛ ذلكم أنها أسئلة تعبر عن واقع استعمال الكلمات، و هو واقع لا يتأتى و لا يُستبين إلا من خلال النظر في سياقات حقيقية كالتّي تتضمنها المدونات، و يستعان بمفهرسات آلية في استكشافها. و جدير بالإشارة أن التواصل مع تلك المفهرسات و استخدامها في التدريس يتراوح بين مستويين:

- أ- تواصل خفيض المستوى (*Low Contact*): و فيه يعتمد الأستاذ/ المتعلم على المفهرس الآلي كأداة مرجع، إلى جانب أدوات أخرى، كالمعجم مثلا.
 - ب- تواصل عالي المستوى (*High Contact*): بمعنى أن الفهرست الآلي يستخدم داخل الصف بصفته أهم مورد لغوي، و يعتمد عليه اعتمادا كلياً.
- و الملاحظ في أغلب الأحيان، أن جميع النشاطات أو التمارين التي تحضر باعتماد الفهرست الآلي، تتمحور حول حذف الكلمة المفتاحية أو تركها. فإذا كان الأستاذ يميل إلى المقاربة الاستقرائية في التدريس، قدم لطلبته الفهرست دونما حذف العقدة، على أساس أنها أمثلة تخدم درسه. أما إذا كان ممن يحبذون الاستنباط، فإنه يقدم لطلبته الفهرست على أساس أنه معطيات للتحليل، كأن يطلب منهم استخراج المتلازمات اللفظية، أو استنتاج "القواعد" بأنفسهم، بدل الاعتماد على القواعد التقنينية (*Prescriptive rules*) المبنوثة هنا و هناك في كتب النحو.³

¹ - Cf. *Ibid.*, p. 14.

² - Michael, Hoey (1997) From concordance to text structure: New uses for computer corpora. Paper presented at the 1997 PALC Conference. Poland: Lodz.

³ - Cf. <http://www.nsknet.or.jp/~peterr-s/concordancing/usingconcs.html>

و في حال حذف العقدة، يعتمد الطالب على معرفته المعجمية و النحوية كي يتوصل إلى الكلمة المطلوبة (أي ملأ الفراغ)؛ و هذه الطريقة مفيدة قصد تقييم مدى استيعاب الطالب، و معرفة مدى تمكنه من وحدة معجمية معينة أو بنية ما.¹

تلكم - إذن - بعض نماذج توظيف الفهرست في تدريس اللغات، و هو توظيف يعكس بجلاء تحولاً و تطوراً في الأدوار بالنسبة للأستاذ أو الطالب على حد سواء؛ فصورة الأستاذ التقليدية التي كانت فيما مضى ترتسم أمامنا، قد تغيرت بتغير مهمته من إلقاء الدروس في شكلها الروتيني إلى مهمة تسهيل سبل البحث باعتماد المدونات والمفهرسات الآلية. و كذا هي الحال بالنسبة لصورة الطالب، من مجرد طرف مستقبل، إلى عنصر فاعل في العملية التعليمية لاسيما بجعل أدوات البحث في متناوله، كما لو أننا بذلك نزيح بعض ثقل المسؤولية عن كاهل الأستاذ و نلقيناها على عاتق الطالب - على حد تعبير غدوين جونز-:

*"This kind of activity provides a means of putting research tools into the hands of students, and working towards shifting some of the responsibility for learning on to their shoulders."*²

و قبل وضع نقطة آخرَ آخر السطر من هذا العنصر، لا ضير أن نجمل في شكل نقط أهم الأوجه التي يمكن أن تخدم وفقها المدونة (أو لسانيات المدونة عامة) تعليمية اللغات.

1- التعرف إلى مختلف معاني الكلمات و استعمالاتها، مثلاً: الأفعال التي تتعلق بالحواس الخمس *touch, see, feel, hear, smell* في صيغة الحاضر تستخدم مسبقة بالفعل المساعد (*can/can't*)؛ أما إذا لم تكن تعني الحواس الخمس و في باقي الاستعمالات، فنستعملها في صيغة الحاضر (*simple present tense*).

2- تحليل تلك الاستعمالات، مثال ذلك: متى يجدر بنا استعمال الفعل *Obtain* بدل الفعل *Get*؟

3- المقارنة بين الاستعمالات اللغوية (لاسيما في مدونة المتعلم/ مدونة المتكلم الأصلي باللغة)، أو بين الاستعمالات في النمطين المنطوق/ المكتوب (...).

4- التمكن من مداراة اللبس و استجلاء الغموض الذي تخلقه الكلمات منفردة، و بالتالي تجاوز الأخطاء المعجمية التي يرتكبها الطالب، طالما أنه من خلال المدونات يكتشف الكلمات ضمن سياقاتها الحقيقية.

5- تمييز الكلمات في استعمالاتها الأساسية و النموذجية، بدل توظيفات مصطنعة أو شعرية أو مغلوبة .

6- فحص ترتيب الوحدات اللغوية، أي موقعها بالنسبة لما يحيط بها من وحدات أخرى (بمعنى توزيعها).

7- استنباط التعابير الجزلة المفيدة، و المتلازمات اللفظية التي يمكن توظيفها لاحقاً في المواقف ذاتها، أو السياقات نفسها.

8- التعرف إلى بنية الخطاب في شكله المنطوق و المكتوب.

¹ - Cf. *Ibid.*

² - Bob, Godwin-Jones (2001) *Op. Cit.*, p. 10.

9- الإلمام بالخصوصيات اللغوية، التي يختص بها نوع من النصوص دون آخر.
10- بإمكان الطالب التحقق من القاعدة اللغوية، بالبحث عن الأمثلة ضمن المعطيات التجريبية التي توفرها المدونات.

11- تحسين المهارات الإنشائية لدى الطالب.

12- إعداد التمارين و النشاطات الموجهة لذلك.¹

3.I- المدونات و الترجمة الآلية:

إن كثيرا من الأبحاث اليوم في حقل الترجمة الآلية، لا تهدف إلى استحداث نظام حاسوبي يقوم مقام المترجم، و يعفيه إعفاء تاما من عمله؛ بل إنها تسعى جاهدة إلى تسخير الحواسيب لتكون من معينات المترجم في عمله، دون الاعتماد عليها اعتمادا كليا.* و تعتمد تلك الأبحاث على المدونات، لاسيما المتوازية منها؛ بحيث يتم تنظيرها على مستوى الكلمات، أو على مستوى الجمل باعتبار علامات الوقف.

و بالنظر إلى ما حققته طرق الاحتمال من نجاحات في حقول أخرى من المعالجة الآلية للغات الطبيعية، سعى براون و غيره إلى تطبيق تلك الطرق على الترجمة الآلية لحل إشكالاتها و فك اشتكالاتها؛ فقاموا باستحداث نظام في الترجمة الآلية قوامه الاحتمالات، تم تدريبه على مدونة فرنسية – انجليزية منظرة. و يقوم النظام المذكور باختيار الجملة-الترجمة الأكثر احتمالا في اللغة المستهدفة، و التي توافق الجملة في اللغة الأصل، وفق نموذجين من الاحتمالات:

- نموذج لغوي ثلاثي الأشكال (Trigram)، قوامه سلاسل ثلاثية الكلمات، و قد وضع في الأصل خصيصا لنظام التعرف إلى الكلام.
- نموذج الترجمة، و قد اشتق من تنظير المدونتين الفرعيتين الانجليزية و الفرنسية على مستوى الكلمة، و كذا بيانات مواقع الكلمة ضمن جمل المدونة.

أما سادلر فيكتور (Sadler Victor) فيقترح مقاربة بديلة، يعتمد فيها على قاعدة هامة و هي ثنائية اللغة، مع إعراب كل لغة على حدى، ثم تُنظَر فيما بينها الوحدات الناتجة. و تتم الترجمة عن طريق عزل الوحدات الممكنة في النص الأصل، و تستعاد تلك الوحدات إلى جانب ترجماتها ضمن القاعدة، ثم ينسق في ما بينها.² و تلکم تعرف بالترجمة الآلية باعتماد الأمثلة (La Traduction automatique basée sur l'exemple (TABE) ، تؤطرها خمس مراحل – وفق المخطط 15 أدناه:

¹ - See: Daniel, Krieger (2003) Op. Cit.

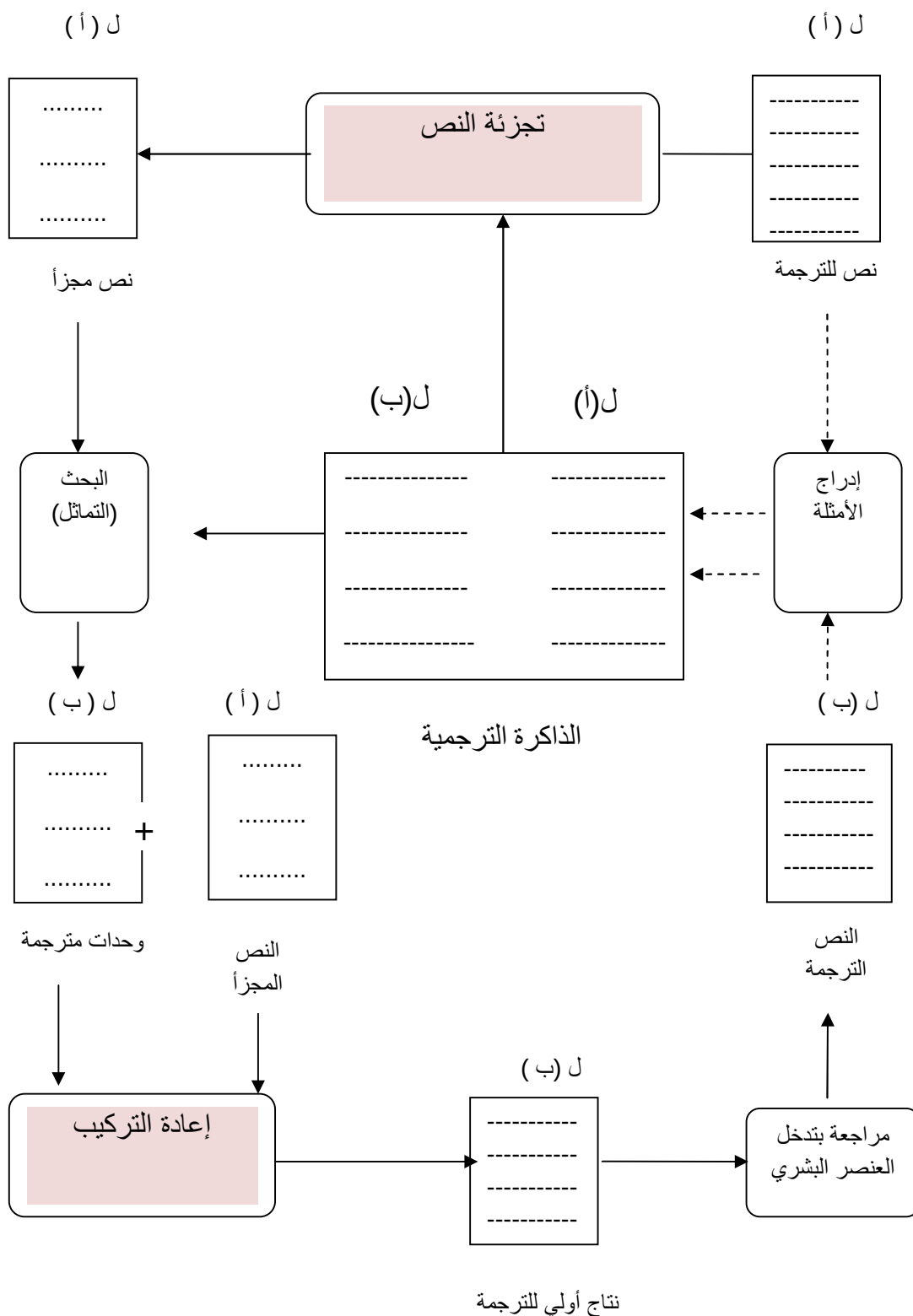
Jane, Willis (1998) Concordances in the classroom without a computer. In: Brian, Tomlinson (ed.) *Materials development in language teaching*. Cambridge.

Anthony, Laurence (2005) AntConc: Design and Development of a Freeware Corpus Analysis Toolkit for the Technical Writing Classroom. *IEEE International Professional Communication Conference Proceedings*, pp. 729-735.

* و هي ما يدعوه الديداوي بالترجمة الحاسوبية في الحالة الأولى، و الترجمة المحوسبة في الحالة الثانية. ينظر:

محمد، الديداوي (2000) مر. س، ص ص 131-135.

² - P. Brown, V. Della Pietra, S. Della Pietra, J. Lafferty & R. Marcer (1993) Analysis, Statistical Transfer and Synthesis in Machine Translation. Submitted to TMI-92, 4th International Conference on Theoretical and Methodological Issues in Machine Translation, 1992.



الشكل (15) مراحل الترجمة الآلية.

- تشطير نص الترجمة و تجزئته إلى وحدات.
- البحث عن الوحدات المتماثلة أو المتشابهة ضمن الذاكرة (MT).
- إعادة تركيب ترجمات الوحدات.
- مراجعة ناتج الترجمة (Post-édition). حفظ النص الأصل و ترجمته بعد مراجعتها ضمن الذاكرة، باعتبارهما مثالين جديدين.¹
- و تبعا للمخطط، فإن الذاكرة الترجمية عبارة عن قاعدة من الأمثلة، و بناؤها يتم وفق المراحل التالية:
- جمع المدونة النصية و انتقاؤها، بحيث أن مادتها نصوص مترجمة تشكل أمثلة تترجم على منوالها نصوص أخرى.
- تنظير النصوص ثنائية اللغة، أي تحديد مواطن التطابق بين شقي المدونة.
- حفظ الأمثلة ضمن الذاكرة .
- استخلاص المعطيات الترجمية انطلاقا من تلك الأمثلة.

باعتبار ما تقدم من معطيات، و هي غرفة من بحر و ذرة من رمل، يتضح أن التجربة الانجليزية في لسانيات المدونة تجربة رائدة و غنية، و أن الصنعة المدونانية بكل مكوناتها و مقوماتها، قد قطعت أشواطاً مرموقة منذ زمن بعيد، فبلغت الذروة عند أهل تلك اللغة. و الواقع أن كثيراً من اللغات الأوروبية (كالفرنسية و الألمانية)، و الآسيوية (كالصينية و اليابانية) راحت تسعى إلى مجازاة الانجليزية، و تتلمس طريقها إلى ذلك؛ والحق أنها قد بلغت بعض مرادها، لتضيف بذلك روافد أخرى إلى ذلك الصرح الذي لا ينفك يتعاضم. و لنا أن نتساءل: أين هي اللغة العربية من ذلك كله؟

¹ - Cf. Olivier, Kraif (2001) *Op. Cit.*, pp.

“What is required are experiments on a very large and representative datasets of common every day Arabic”.

A. Gowder & A. de Roeck (2001) Assessment of a Signification Arabic Corpus. Workshop Proceedings of Arabic Language Processing: Status and Prospects (pp.73-79).Toulouse, France.

II-الواقع عند العرب:

يسألونك عن حال العربية، و حظها من الصنعة المدونانية؛ قل إنها أشبه بأرض قفر و واد غير ذي زرع، قد بدأ يمر عليه بعض القطر فيهبز تربته، و يحي مواته، أو قل إنها قد بدأت تخرج من بوتقتها و تنتفض لتتحرر من قوقعتها، لاسيما خلال السنوات القليلة الماضية، بعد أن صارت محط اهتمام العديد من الباحثين في كثير من مجالات المعالجة الآلية للغات الطبيعية.

و بحكم ما تستدعيه الأبحاث في تلك المجالات من موارد لغوية هامة، وبالنظر إلى افتقار العربية لمثل تلك الموارد،¹ قامت ثلة جلة من الباحثين، و قد شمرت عن سواعدها، بصنع مدونات عربية خدمة لتلك الدراسات، و مجارة لبقية اللغات. و هي مدونات سوف نتبين شيئاً من أحوالها، و نستكشف بعض أصولها خلال هذه المحطة. لكن قبل هذا و ذاك، نرى أنه لا ضير في أن نقف وقفة، نستوضح من خلالها – في بادئ الأمر – بعض أوجه خدمة المدونات للغة العربية.

1.II- المدونات في خدمة العربية:

يقول جيوفري ليتش، و هو يتحدث عن الدلائيات، يوم كانت تخطو في عالم البحث أولى الخطوات:

“ Une discipline autonome commence non avec des réponses, mais avec des questions”.²

و الحق أن لسانيات المدونة – أو لنقل المدونات – في العالم العربي فتية حديثة العهد، لذا لا مناص من أن نبدأ بسؤال حقيق بالطرح قبل أي طرح: فيم يمكن أن تخدم المقاربة المدونانية الأبحاث العربية؟

إن نظرة واحدة إلى واقع تلك الأبحاث – على بكرة أبيها – تكفي للإجابة عن هذا السؤال، و تقطع قول كل خطيب؛ إذ ما نراه اليوم قد لاكته كثير من الأفواه، و أضحى له في كل مصر العديد من النظائر و الأشباه، حتى ملت تقليبه الأذهان، و برمت سماعه الأذان، لأنها لم تجد فيه إلا الغث الهزيل، لا يسمن و لا يغني من جوع. لذا رأى أصحابها أن لا مناص لهم من تحسس السبيل و البحث عن البديل؛ فكان أن وجدوا في المدونات ضالتهم.

و الحق أن السر يكمن في الإحصاء اللغوي الذي يطبق على مستوى المدونات، من حيث إنه يثري اللغة العربية من عدة جوانب، على شاكلة الجانب المعجماتي و الجانب النحوي الصرفي.¹

¹ - Cf. Mahtab, Nikkhou & Khalid Choukri (2005) Report on Survey on Arabic Language Resources and Tools in Mediterranean Countries. NEMLAR, Center for Sprogteknologi, University of Copenhagen, Denmark.

² - Geoffrey, Leech (1977) Semantics. Penguin Books.

فالمدونات تتيح للباحث حصر جميع سياقات ورود الكلمة الواحدة؛ و من ثم بمقدوره تحديد مختلف معانيها و أوجه استعمالها، سواء كانت استعمالات حقيقية أم مجازية أم مغلوبة أم أنها تتباين من قطر عربي إلى آخر، و ذلك من مسلمة أنه بتعدد النصوص تتعدد السياقات فالمعاني و الاستعمالات. و هذا بدوره يفك عقدة اللبس و يحل إشكالها، و لنا أن نتوهم مثال ذلك من خلال الجدول (4) أدناه.

كما تستخدم المدونات أيضا في تحليل تكرار الكلمات، و تحديد أيها الأكثر شيوعا؛ و كذلك هي الحال بالنسبة للمترادفات و الأضداد، و ما شاكلها من مختلف الظواهر اللغوية. و مثال ذلك يمكن أن نستوحيه من الجدول (5)، إذ يبين أن للسياق المكاني دورا في انتخاب مفردات/ تعابير دون غيرها، فكلمة "ملاحم" – مثلا – في عمان شائعة للدلالة على مجال الجزارة، في حين أن الدارج و المتداول في الجزائر هو "المجازر"، بينما لفظ "ملاحم" يطلق على الموقعة العظيمة أو الفن المعروف من الفنون الأدبية.

الجملة	معنى الكلمة (كبد)
- لقد أجرى عملية زرع الكبد.	العضو.
- كانت الشمس في كبد السماء.	وسط الشيء.
- كبد الأرض.	ما في باطنها من الذهب و الفضة.
- كبد القوس.	ما بين طرفي علاقتها.*
- لقد فقد فلذة كبده.	إبن(ت)ه.

الجدول (4) معاني الكلمة "كبد" (اسم) في استعمالاتها الحقيقية والمجازية.

البلد	الجملة	معنى "ملاحم"
الجزائر	ولكن افتقار الأدب العربي لهذا النوع من الشعر لا يعني عدم احتوائه على المعاني والمفردات كالبطولة والشجاعة والفخر والحماسة التي تتصف بها الملاحم المعروفة.	الموقعة العظيمة. فن من الفنون الأدبية.
عمان	كما ينفذ قسم المراقبة الصحية حملات تفتيش واسعة النطاق على محلات الاستهلاك [...] المتمثلة في بقالات بيع المواد الغذائية و الملاحم والمخابز والمطاعم ومحلات الحلاقة.	محل الجزار.

جدول (5) معاني كلمة "ملاحم" في الجزائر و عمان.

ثم إن اعتماد المدونات في الأبحاث، يبين الثراء المفرداتي و التنوع الدلالي الملحوظين في اللغة العربية، و تباينهما من بلد لآخر أو حتى في البلد الواحد، مثال ذلك: تعيين وزارة التربية و التعليم في مصر و العربية السعودية – مثلا – ب "وزارة

¹ - ينظر بالتفصيل كل من:

سلوى حمادة (2007) مرس.

S. Alansary et al. (2007) Op. Cit.

Ahmed Abdelali, Jim Cowie & Hamdy S. Soliman (2005) Building a Modern Standard Arabic Corpus. Workshop on Computational Modeling of Lexical Acquisition. The Split Meeting. Croatia, 25th to 28th of July 2005.

* - العلاقة: أي ما تعلق به القوس.

المعارف"، و في موريتانيا بـ "وزارة التهذيب"، أو أن نجد مثلاً كمعادل للفظ الانجليزي *Dormitory* لفظة "مرقد" ("الخبر" الجزائرية)، و لفظة "عنبر" ("الدستور" الأردنية)، و "إقامة" ("الحياة" اللندنية).¹

فضلاً على ذلك، يمكن دراسة تكرارات الكلمات باعتبارها وحدات معجمية تنتمي إلى أقسام مختلفة من أقسام الكلام؛ إذ الأرجح أن نجد أن الكلمة العربية الواحدة يمكن أن تكون اسماً، أو فعلاً، أو حرفاً حسب السياق، مثال ذلك "كلمة عين":

الجملة	قسم الكلام الذي تنتمي إليه كلمة (عين)	معناها
- جامعة عين شمس.	اسم علم مركب تركيب إضافة.	/
- عين الماء .	اسم.	ينبوع الماء.
- عين الإنسان .	اسم .	عضو البصر.
- عين وزيراً	فعل مبني للمجهول/ المعلوم.	نصب (ه) وزيراً.
للخارجية		
- رأيتُه هو بعينه.	اسم .	للتوكيد.

جدول (6) معاني كلمة "عين".

كما يمكن أن يعتمد الباحث إلى دراسة مدى تأثير غياب الحركات (من فتحة وكسرة و ضمة) و علامة السكون؛ إذ يتغير معنى الكلمة الواحدة بتغيير طفيف على مستوى الحركة أو التشديد، مثال ذلك "نفس":

نَفْس (بالشيء : ضن به، و على فلان بخير: حسده عليه).

نَفْس (كان نفيساً مرغوباً فيه).

نَفْس (عنه الكربة: لطفها و فرجها).

و كذلك هي الحال بالنسبة للأحرف المزدوجة، التي تغير من معاني الكلمات مع اعتبار السياق دائماً، نحو:

نافس (فلاناً في الأمر: باراه فيه).

تنافس (القوم في الأمر: تباروا فيه).

تنفّس (الصباح: انبلج، الصعداء: تنفّس تنفساً طويلاً من تعب أو كرب).

و ليس من الخوارق في شيء، إن فكر الباحثون في استحداث نظام التشكيل الآلي للنصوص العربية، أو إن هم اهتموا إلى تطبيق آليات الذكاء الاصطناعي في تفهم اللغة العربية غير المشكولة،² طالما أن المدونات تيسر لهم ذلك أيما تيسير. هذا ناهيك عن التحليل النحوي الذي يتم وفقه تشطير الجملة إلى عناصرها النحوية الأولية؛ أما التحليل و التوليد الصرفي، فيتم فيه البحث عن جذور العقد و مشتقاتها، بفضل مولد صرفي يولد جميع الكلمات الممكنة انطلاقاً من جذر العقدة.

و يمكن أن تكون البرمجيات ذاتها مزودة – علاوة على ذلك – بـ "قاعدة بيانات معجمية دلالية، حيث تخزن مع كل مشتقة معلوماتها المعجمية و قيودها الدلالية".³ و في

¹ - Cf. A. Abdelali, J. Cowie & H.S. Soliman (2005) Op. Cit.

² - ينظر: سلوى، حمادة (2007) مرس.

³ - نفسه.

ذلك عميم الفائدة، لاسيما في تصويب الأخطاء الصرفية منها، و النحوية و الدلالية. و لنا أن نتصور مدى الإسهام في الارتقاء بالأساليب التعليمية للغة العربية، سواء تعلق الأمر بأبنائها أم بغير الناطقين بها، من الأساليب الروتينية التقليدية إلى الأساليب التحفيزية الالكترونية. و صور ذلك كثيرة، كأن يعتمد الأستاذ/ المتعلم إلى تفصي استعمالات الكلمة الواحدة و تكراراتها، و معانيها ضمن مختلف الأنواع النصية؛ و له أن يتعرف إلى مختلف القوالب المعجمية و شتى البناءات التركيبية التي تنزع إلى الورد في نطاقها، فيصنع و ينسج ما شاء على منوالها؛ كما له أن ينظر في الجملة العربية و نظمها، و يتبين ما يحكم سبكها. هذا علاوة على أهمية دراسة اللغة ككل في شكلها المنطوق و المكتوب، و المقارنة بينهما على المستويين المعجمي و التركيبي.

أما الأساليب العربية، فلا شك أنها – هي الأخرى – ستظفر بكثير من الدقة والموضوعية في الطرح و التناول، لاسيما إذا عرفت كيف تستثمر المدونات في بحث أساليب الكتابة لدى الأوليين و المعاصرين – على سبيل المثال –، قصد النظر في تطورها، أو حصر النصيب الأوفر مما تزدها به و تزخر من دقة المعاني و إحكام المباني و ظلال و نبرات و ألوان و دلالات، و شرحها للناشئة و المتعلمين يولد لديهم الملكة الإنشائية و يعززها.¹

إن تلك المعطيات بأسرها ليست سوى بضع من بضعة أمثلة عن عربية المستقبل، بل مستقبل العربية، إذا ما نهض بها المنتطسون، و فطنوا إلى حاضر اللغات الذي هم عنه غافلون، حاضر كأي من أمة ارتقت فيه بلغتها، و تجاوزت زمن البردي* إلى كل ما هو تقني و الكتروني.

و لا مشاحة في أن الدراسات الغربية، سواء تلك التي سلف و أحدثنا منها ذكرا، أو تلك التي لم نستطع أن نحيط بها خبْرا، تدل على عظم "الفجوة الرقمية" بينها و بين الدراسات العربية التي تعد على رؤوس الأصابع؛ بل الأدهى و الأمر، أنها دراسات لم يُكْتَبَ لبعضها أن يرى النور إلا حديثا، أو مذ خمس سنوات تحديدا، مع "الملتقى السنوي حول اللسانيات العربية"، بحسب ما ورد على لسان باركنسن و فاروانه (Parkinson & Farwaneh):

*"[...] four of the papers deal with the area of corpus linguistics (new for this series), including papers from both a computational and variationist point of view."*²

و قد تطرقت المقالات الأربع إلى دراسة الأدوات و الحروف في اللغة العربية (مقال فان مول، و مقال باركنسن)، و بنية الجملة الاسمية و أنواعها (مقال الأنصاري)، و الإنتاج المعجمي (مقال تايلور).³

¹ - ينظر: سلوى، حمادة (2007) نفسه.

* - البردي: نبات كالقصب، كان قدماء المصريين يستخدمون قشره للكتابة.

² - D. Parkinson & S. Farwaneh (eds.) (2003a) *Perspectives on Arabic Linguistics xv: Papers from the Fifteenth Annual Symposium on Arabic Linguistics*. Salt Lake City.

³ - Cf. L. Al-Sulaiti & E. Atwell (2006) The design of a corpus of contemporary Arabic. *International Journal of Corpus Linguistics* 11(1), pp. 5-7.

و المقالات الأربع هي على التوالي:

و في بحث له حول استثمار المدونات في دراسة اللغة العربية الفصحى المعاصرة (MSA)، قام فان مول * بصناعة مدونة مرمزة قوامها مادة نصية بحجم 240000 كلمة، مأخوذة عن ثلاث محطات إذاعية عربية: الجزائرية و المصرية و العربية السعودية. عمد من خلالها إلى دراسة استعمال حرفي المستقبل "السين" و "سوف"، فكان من جملة ما توصل إليه أنه ليس ثمة ما يميز بين الحرفين، بخلاف ما تنص عليه القواعد العربية من استعمال حرف "السين" للدلالة على المستقبل القريب، أما "سوف" فحرف استقبال أطول زمانا من سابقه. و يذهب الباحث إلى القول إن تحليل المدونات يتيح فرصة اكتشاف مظاهر فريدة و رؤى جديدة بخصوص بناء اللغة العربية و استعمالاتها:

*"Indeed, we may obtain a completely new look, and new insights in the structure and the use of the Arabic language by analysing corpora."*¹

كما استخدم المدونة نفسها في ضبط متلازمات لفظية جديدة، و تدقيق معان حديثة؛ فضلا على تحديد بعض التنوع في الاستعمال اللغوي في الأقطار العربية الثلاثة، مثال ذلك: عبارتا "في الوقت نفسه" و "في نفس الوقت"، حيث إن الأولى متواردة بكثرة في المدونة المصرية، و لا أثر لها في المدونة الجزائرية، أما في المدونة السعودية فتحظى العبارتان بنفس قيمة التوزيع.²

و ثمة دراسة أخرى باعتماد مدونة قوامها مليون كلمة، يطرح فيها أمين المهنا مسألة التراكيب المصطلحية الجديدة في المحررات العربية العلمية، و يقارنها بما خرجته المجامع اللغوية و اصطلحت عليه.³

أما في مجال الصنعة المعجمية، فإن معظم المعاجم الحالية في اللغة العربية المعاصرة لم يستند إلى المدونات في تخريجها؛ لذلك، فإنها تقتقر فقرا مدقعا إلى المتلازمات اللفظية و المستجدات المصطلحية، كما أنها متحجرة يأخذ بعضها عن بعض، سواء من حيث المحتوى أو طريقة ترتيب مختلف المعاني و وصفها، و في ذلك تبعات قد تنعكس – لا محالة – سلبا على الوضع التحصيلي لمتعلمي اللغة العربية من أبنائها أو من غير الناطقين بها على حد سوى، وفق ما يذهب إليه سالم غزالي و عبد الفتاح براهيم:

"Arabic dictionaries are notorious for representing a fossilised version of our language as each one is to great extent

M. van Mol (2003) Evolution of MSA, the Case of some Complementary Particles. In: D. Parkinson & S. Farwaneh (eds.) *Op. Cit.*, pp. 135-147.

D. Parkinson (2003b) Future Variability: A Corpus Study of Arabic Future Particles. In: D. Parkinson & S. Farwaneh (eds.) *Op. Cit.*, pp. 191-211.

S. Alansary (2003) NP-Structure Types in Spoken and Written Modern Standard Arabic (MSA) Corpora. In: D. Parkinson & S. Farwaneh (eds.) *Op. Cit.*, pp. 149-180.

S. Taylor (2003) Comparing Frequencies of Lexical Productions in Arabic Words. In: D. Parkinson & S. Farwaneh (eds.) *Op. Cit.*, pp. 135-147.

* - الباحث من معهد اللغات الحية، الجامعة الكاثوليكية بلوفن Leuven البلجيكية.

¹ - Mark, van Mol (2000a) Exploring annotated Arabic corpora: Preliminary results.

http://www.ilt.kuleuven.ac.be/ilt/arabic_pdf/tunis.pdf.

² - *Ibid.*

³ - Cf. A. Al-Muhanna (2003) Scientific and technological terms transfer into Arabic: A corpus –based study of Arabic noun+ noun and noun+ adj. compounds. –Ph.D. thesis, UMIST, UK

*a reflex of the preceding ones in content, [...] and the way different meanings are ordered. This state of affairs is not likely to help the great majority of Arabic learners whether Arab School children or speakers of other languages."*¹

و قد ضربا لذلك مثلا، حين عمدا إلى تقصي معاني الفعل (أخذ) انطلاقا من مدونة عربية بحجم 1.500.000 كلمة، و ألقيا أنه إلى جانب المعاني "الحرفية" المبنوثة في ثنايا المعاجم "التقليدية"، يؤدي هذا الفعل معنيين آخرين عند وروده في بعض التراكيب، لا يتوفران في معجم "الوسيط" (1960)، و لا غيره من المعاجم الحديثة نسبيا. أما مارك فان مول، فقد توصل إلى تطوير قاعدة بيانات علائقية "عرالات" (AraLat) (Relational database) لصناعة معجم هولندي - عربي /عربي - هولندي. و تتوفر القاعدة على معجم للمتعلمين بحجم 19 ألف كلمة عربية مترجمة ضمن سياقاتها، بالإضافة إلى أكثر من 10 آلاف مثال إيضاحي، تم انتقاؤها جميعا من مدونة ممثلة بحجم ثلاث ملايين كلمة.²

أما هوغلند (Hoogland) فقد صمم مدونة قوامها أكثر من مليوني كلمة، لاستخدامها في صناعة معاجم هولندية- عربية /عربية -هولندية. في حين أن زمانك (Zemanek) عمد إلى تأليف معجم عربي - تشيكي، معتمدا على مدونة بحجم خمسين مليون كلمة.³

أما ماثيو جيدار (Mathieu Guidère) فيتحدث عن تنظير مدونة متوازية فرنسية - عربية، و استثمارها في مجال الترجمة الآلية و تطويرها في اللغة العربية المعاصرة، وهو مجال يفتقر حاليا إلى مدونة مرجع يعتمد عليها في تطبيقاته.

" In the current state of machine translation research from and into Arabic, no reference corpus is yet available".⁴

2.II- المدونات العربية:

غني عن البيان أن محل العقبة الكأداء في أي عمل ناشئ، هو الحصول على الأدوات اللازمة لأدائه و تجسيده على أرض الواقع. و البحث في المعالجة الآلية للغة العربية لا يزال في المهد، يحتاج إلى من يشمله بالرعاية حتى يشتد عوده، و يتعهده بالاهتمام حتى يقوى صلبه؛ و هو في كل ذلك يفتقر إلى الموارد الأولية التي ينهض بها،

¹ - S. Ghazali & A. Braham (2001) Dictionary Definitions and Corpus –based Evidence in Modern Standard Arabic. Arabic NLP Workshop at ACL /EACL. Toulouse, France. <http://www.elsnet.org/arabic/2001/ghazali.pdf>

² - Cf. M. van Mol & H. Paulussen (2001) AraLat: A relational database for the development of bilingual Arabic dictionaries. In: S. Lee (ed.) *Proceedings of Asialex 2001, Asian Bilingualism and the Dictionary* (pp. 206-211). Seoul, August 2001.

³ - Cf. J. Hoogland (1996) the use of OCR Software for Arabic in order to create a text corpus of Modern Standard Arabic for lexicographic purposes. In A. Ubaydy (ed.) *Proceedings of the International Conference and Exhibition on Multi- Lingual Computing* (pp. 2701-2716), Cambridge University.

- P. Zemanek (2001) CLARA (Corpus Linguae Arabic)/ An Overview in ELSNET (ed.) *Proceedings of ALC /EACL workshop on Arabic language processing*. Toulouse, France.

<http://www.elsnet.org/acl2001-arabic.html>

⁴ - Mathieu, Guidère (2002) Toward Corpus-Based Machine Translation for Standard Arabic. *Translation Journal* 6(1), January 2002.

بدءاً بالمدونات و وصولاً إلى البر(نا)مجيات الضرورية لمعالجتها. و لا تخفى على القارئ الآثار السلبية لذلك الافتقار، التي تنعكس على نوعية التطبيقات المنجزة في المعالجة الآلية للعربية.

و لولا تلك المحاولات العديدة، و المساعي الحميدة لصناعة مدونات عربية، رغم ما تحمله من مزايا و عيوب، لنفضنا الكف بالكف و قلنا أنه قد اتسع الخرق على الراقع بما لا حيلة له فيه، نظراً لعظم تلك "الفجوة الرقمية" بين البحث العربي و نظيره في الغرب.

1.2.II- مدونة كلارا (CLARA):

و هي مدونة مكتوبة بحجم 50 مليون كلمة، صنعت لأغراض معجماتية بجامعة تشارلز Charles- براغ Prague (1997). و المدونة و إن كانت لغتها في الظاهر العربية الفصحى المعاصرة، إلا أن زهاء 300 ألف كلمة مصدرها اللغات الغربية، لاسيما الانجليزية منها. و هي تغطي جغرافيا كلا من: الجزيرة العربية، سوريا، مصر، و بعض الأمثلة من دول أخرى على غرار: تونس و المغرب، بنسبة ضئيلة لا تتعدى 5%. و من ناحية المحتوى، فإن حوالي 50 % من النصوص مصدرها الدوريات العلمية، و 35 % من الكتب، و 15% متفرقات (أي مصادرها شتى)، و قد تم الحصول على معظمها عن طريق نسخها بواسطة الماسح الضوئي* (لاسيما الروايات التي تشكل 15%، و كذا الميادين المعرفية كالفلك و الطب، والقانون و التكنولوجيا، الخ).¹

2.2.II- مدونة نيوزوير العربية (Arabic Newswire Corpus):

ظهرت مدونة نيوزوير إلى الوجود عام 2001، و هي من صنع دايفيد غراف (David Graff) و كيفن ولكر (Kevin Walker) بجامعة بنسلفانيا يبلغ حجمها 80 مليون كلمة مكتوبة، و قد تم جمعها خلال الفترة من 13 ماي 1994 إلى غاية 20 ديسمبر 2000. و صممت لأغراض شتى، كالتعليم و تطوير أساليبه و تقنياته؛ بيد أن ما يعاب عليها أنها ليست ممثلة للغة العربية ككل، بحكم اعتمادها على المادة الصحفية فقط (عن وكالة فرنسا و وكالة أمة Ummah Press، و غيرهما).²

3.2.II- مدونة الحياة (2002):

تم استحداث المدونة في إطار مشروع بحث بجامعة اسكس (ELRA) Essex، و هي مدونة مكتوبة بحجم 18639264 كلمة ضمن 42591 مقالا، يشمل: عموميات، الأخبار، اقتصاد و أعمال، رياضة، الحواسيب و الانترنت، علوم و تكنولوجيا، سيارات؛ و هو التقسيم نفسه المعتمد ضمن صحيفة "الحياة" اللبنانية، إذ استقيت منها مادة المدونة، بعد ابتياع نسخة إلكترونية منها. و قد تم إنشاء المدونة بغرض الهندسة اللغوية و استعادة البيانات؛ بيد أن ما يعاب عليها اعتماد مصمميها على أرشيف سنة واحدة فقط من صحيفة "الحياة"، فضلا على تبني تصنيف لمادتها مطابق لتصنيف الزوايا ضمن الصحيفة ذاتها.³

4.2.II- مدونة النهار (2001):

* - وهو متوفر عن شركة صخر: Arabic OCR Program Automatic Reader

¹ - Cf. S. Alansary et al. (2007) Op. Cit., p11.

² - Cf. Latifa Al-Sulaiti's home page.

http://www.comp.leeds.ac.uk/eric/latifa/arabic_corpora.htm

³ - Cf. A. Goweder & A. de Roek (2001) Op. Cit.

- S. Alansary et al. (2007) Op. Cit., p. 12.

و هي مدونة مكتوبة من صنع إلرا لأغراض بحثية عامة، بحجم 140 مليون كلمة، مادتها مستقاة من صحيفة "النهار" اللبنانية من العام 1995 إلى العام 2000، بمعدل 45 ألف مقال و 24 مليون كلمة سنوياً. و يتضمن كل مقال بيانات عامة، من قبيل: العنوان، اسم الصحيفة، التاريخ، البلد، النوع، الصفحة، الخ.¹

II.2.5- مدونة العربية الفصحى CAC (Classical Arabic Corpus):

جمع هذه المدونة عبد الحميد عليوة بجامعة مانشستر- معهد العلوم والتقنيات UMIST (2004)، بغرض التحليل المعجمي. و هي مدونة مكتوبة بحجم 5 ملايين كلمة، تتضمن قصار النصوص الشعرية من الفترة قبل الإسلام إلى غاية القرن الحادي عشر، مأخوذة عن موقعي "المحدث" و "الوراق". تقسم فيها النصوص إلى أقسام أربعة: ما يتعلق بالمبادئ و المعتقدات، الأدب، اللسانيات، و العلوم.²

II.2.6- المدونة العامة للعربية العلمية GSAC (General Scientific Arabic Corpus):

و هي مدونة مكتوبة، بحجم 1.6 مليون كلمة؛ قام بجمعها "أمين المهنا" بجامعة مانشستر - معهد العلوم و التقنيات، بغرض دراسة كيفية صياغة المصطلحات التقنية والعلمية - لاسيما المركب منها- في اللغة العربية؛ فضلاً على ذلك، يقارن الباحث بين ما يتبعه الكتاب العرب من آليات في الوضع، و ما اقترحتة الجامعات العربية و أقرته. ما يعاب على المدونة أنها لا تمثل غالبية اللغة العربية، و لا تغطي التنوع اللغوي الذي تنطوي عليه، إذ إن نصوصها مستقاة عن موقع المجلة الكويتية "علوم وتقنيات". و جدير بالإشارة أن جزء منها، أو ما يعادل 1 مليون كلمة تم ترميزه، و يؤكد "المهنا" أن العملية كانت دقيقة بنسبة 92%.³

II.2.7- مدونة أرابيك جيغا وورد (Arabic Gigaword Corpus):

ظهرت المدونة في 22 جويلية 2003، و هي مدونة مكتوبة بحجم حوالي 400 مليون كلمة، من صنع دايفيد غراف بجامعة بنسلفانيا - إل دي سي (LDC). و المدونة بمثابة أرشيف شامل للمعطيات النصية المستقاة من الوكالات، على شاكلة: الوكالة الفرنسية للصحف، وكالة الحياة للأنباء، وكالة النهار للأنباء، وكالة اكسانوا Xinhua للأنباء. و قد صممت لغرض المعالجة الآلية للغة، و نموذجتها، و استعادة البيانات.⁴

II.2.8- المدونة المتوازية متعددة اللغات:

و هي أول مدونة تتضمن اللغات الثلاثة (العربية، الإسبانية، الانجليزية) من تصميم فريق بحث بمختبر اللسانيات الحوسبية - Autónoma بجامعة مدريد، بحجم زهاء 3 ملايين كلمة (العربية منها 901511 كلمة، الإسبانية 1343225 كلمة، الانجليزية 1073209 كلمة). و لقد عني المصممون بنوعية المدونة عناية فائقة، فحددوا بعض الضوابط تتعلق أساساً بطبيعة ترجمة النصوص المنتقاة و مصدرها، و خاصة التمثيل، وإتاحة المدونة ضمن شكل اليكتروني و استعمالها في حدود ما يسمح به قانوننا، لذلك حرص الفريق على:

- ألا تكون النصوص المدرجة نتاج ترجمة آلية.

¹ - Cf. <http://www.elda.org/catalogue/en/text/w0027.html>

² - Cf. Latifa Al-Sulaiti's home page.

³ - Cf. Latifa Al-Sulaiti's & Eric Atwell (2006) Op. Cit., p. 4.

⁴ - Ibid.

- أن تمثل الاستخدام المعاصر للغة.
 - أن تكون المصادر مجانية و متاحة ضمن شكل إلكتروني.
 - أن تحترم حقوق التأليف، و أن تستخدم النصوص وفق الوجه المشروع.¹
- و قد وقع اختيار الباحثين على الوثائق الصادرة عن الهيئة الأممية، للاعتبارات التالية:

- اللغات العربية و الاسبانية و الانجليزية تنتمي إلى اللغات الرسمية على مستوى الهيئة.
 - نوعية الترجمة مضمونة.
 - نصوصها تمثل الاستعمال اللغوي المعاصر.
 - وفرة الوثائق و إتاحتها، و كذا تنوعها بتنوع تنظيمات الهيئة.
- بعدها، عمد المصممون إلى معالجة المدونة و تنظيرها على مستوى الجملة، وفق الثنائيات (عربي - اسباني، اسباني - انجليزي، عربي - انجليزي)، و قد كانت النتائج المتحصل عليها مرضية بنسبة تتعدى 90% في كل ثنائية.² و في مرحلة أخيرة، ثم ترميز المدونات الفرعية الثلاث على مستوى أقسام الكلام.
- و ثمة ملاحظتان بخصوص المدونة: إحداهما أن حجمها ضئيل جداً، و أخراهما أن ما توصلت إليه من نتائج - بإقرار من مصمميها أنفسهم - يثير كثيراً من التساؤلات، يدفع بأصحابها إلى اتخاذ مسارات في البحث أكثر عمقا و تدقيقاً، كما تقتضي توسيع المدونة بما يكفي للإجابة عن تلك التساؤلات. بيد أنها - و الحق يقال - قد أشرعت أبواباً جديدة كانت من قبل موصدة، و شرعت للباحثين الطريق إلى صناعة المدونات (العربية) المتوازية متعددة اللغات.³

و تستوقفنا في هذا المقام بعض الملاحظات نخالها على قدر بالغ من الأهمية، مفادها أن ما ذكرناه - لحد كتابة هذه الأسطر - من مدونات عربية أو تلكم المقيدة ضمن جدول الملحق*، هي مدونات خام (عدا المدونة العامة العربية العلمية) لا يعتمد عليها لتصنيف للأجناس واضح أو متعارف عليه، كما هو الشأن في المدونات الغربية؛ كما أنها لا تعطي بيانات خاصة بعدد الكلمات في كل جنس كتابي - باستثناء مدونة "كلارا". ثم إن بعضها غير متاح للعامة، بسبب حقوق التأليف أو ما شاكلها من المسائل القانونية، وبعضها الآخر يمكن ابتياعه من "إل دي سي" (LDC) أو "إلرا" (ELRA).⁴

و يتحجج كل مصمم بقوله إن صناعة هذه المدونة أو تلك، لم تكن بنية إتاحتها للاستخدام العام؛ بل قد أملت عليها الظروف الخاصة ببحثه. و هي الذريعة نفسها التي يتخذها أولئك ليحاجوا بها في مسألة عدم اهتمامهم بالأجناس الكتابية أو حجمها، ودراستها

¹ - Cf. Doaa Samy, Sandoval Antonio Moreno, José M. Guirao & Enrique Alfonseca (2006) Building a Parallel Multilingual Corpus (Arabic-Spanish-English). LREC, pp. 2176-2181.

² - Ibid., p. 2177.

³ - جزء من المدونة متاح على مستوى الشبكة: <http://www.llf.uam.es/>

* - يمكن النظر في مدونات أخرى لخصناها في الجدول ضمن الملحق، نقلاً عن:

L. Al-Sulaiti & E. Atwell (2006) Op. Cit., pp. 3-4.

⁴ - LDC, The Linguistic Data Consortium. University of Pennsylvania.

<http://www.ldc.upenn.edu>

ELRA, European Language Resource Association.

<http://www.elra.inf/>

قبل إدراج النصوص ضمن المدونات؛¹ مما يعني أن مجال استخدامها محدود، في الوقت الذي تعاني العربية فاقة شديدة إلى مثل تلك الموارد. ثم إن ما يذهب إليه أولئك المصممون يتنافى مع القول بضرورة إنشاء مدونة تصلح لأغراض شتى، و تخدم مجالات عدة في الدراسات اللغوية؛² أو أن تكون متاحة للباحثين كافة، و ذلك أضعف الإيمان.

إننا لا نبيغي من وراء هذه الملاحظات انتقاص بعض شأن تلك المدونات، بل إنها – و الحق يقال – ذات فائدة جمة كمنطلق و موطئ لكثير من الباحثين، تفتنوا إلى تلك العيوب و أحصوها، محاولين تجاوزها في مسعاهم إلى صناعة مدونات أخرى، تمثل بدقة اللغة العربية الفصحى المعاصرة؛ و لنضرب لذلك مثلين: مثل مدونة اللغة العربية المعاصرة CCA، و مدونة اللغة العربية الدولية ICA.

9.2.II - مدونة اللغة العربية المعاصرة :Corpus of Contemporary Arabic

طرحت المدونة في نسختها الأولى عام 2004، و هي من تصميم لطيفة السليطي و إريك أطول من جامعة ليدز (Leeds-UK) بالمملكة المتحدة، تتضمن مادة نصية منطوقة و أخرى مكتوبة بحجم حوالي مليون كلمة.³ و قد صممت المدونة بالنظر إلى محدودية المدونات التي صنعت حتىئذ، من حيث مصادر النصوص و أنواعها، و كذا الحاجة الملحة إلى توفير مدونة معاصرة للغة العربية، تغطي نطاقا واسعا من الأنواع النصية و تتاح مجانا على مستوى الشبكة⁴ لكافة الباحثين/ الدارسين في مجال المعالجة الآلية للغة العربية و البحث اللساني عامة، و كذا تخصيصها لتعليم اللغة العربية لغير الناطقين بها (TAFL). و خدمة لذلك، قام المصممون بانتقاء المادة النصية (لاسيما المكتوبة منها) بناءً على رؤى و اقتراحات المستخدمين من أساتذة اللغة العربية، و المشتغلين في مجال الهندسة اللغوية لاستحداث برمجيات المعالجة الآلية للغة العربية و تطويرها، وفق ما ورد ضمن الاستبيان الذي وزع على الفريقين.⁵

و ينقسم الاستبيان إلى ثلاث مجموعات من الأسئلة، تتضمن المجموعة الأولى أسئلة عن المعلومات الشخصية (اسم المؤسسة/ الشركة التي يتنسب إليها الأستاذ/ الباحث، طبيعة العمل/ الخدمة المقدمة، و العنوان،...)، أما المجموعة الثانية فتحتوي قائمة بواحد و أربعين نوعا من أنواع النصوص أو الأجناس الكتابية، طلب من الفريقين تصنيفها من النصوص الأفيد إلى أدناها فائدة. و تلك الأنواع/ الأجناس قد انتخبها المصممون انطلاقا من مدونات سابقة، و بناءً على نتائج دراسة أخرى أجريت بجامعة ليدز هدفها تحديد الأنواع النصية الأكثر ترجمة، و المستخدمة في جميع مدونة متعددة اللغات تقيم من خلالها الترجمة الآلية. و هي تنتمي إلى:

- نصوص مكتوبة: الخيال، الفنون، العلوم، الأعمال، متفرقات.
- نصوص منطوقة: التلفزة، الإذاعة، محادثات و مكالمات.

¹ - Cf. L. Al-Sulaiti & E. Atwell (2006) Op. Cit., p. 8.

² - مثال ذلك ما تلفاه عند مصممي المدونات الغربية سواء التقليدية منها (براون و لوب، ...) أو المعاصرة (ب أن سي، أي سي، ...) .

³ - و قد أشار الباحثان إلى أن النسخة الأولى لا تتضمن كثيرا من النصوص المنطوقة، لأن معالجتها تتطلب وقتا أطول مقارنة بالمدة الزمنية للمشروع المقدر بسنة واحدة، و ينوي الباحثان التفرغ لها لاحقا.

⁴ - النسخة الأولى متوفرة على الموقع: <http://www.comp.leeds.ac.uk/latifa/>

⁵ - Cf. L. Al-Sulaiti & E. Atwell (2006) Op. Cit., pp. 12-20.

أما آخر تلك المجموعات، فيتضمن سؤالاً واحداً للمشتغلين في الهندسة اللغوية، وأربعة عشر سؤالاً للأساتذة؛ و تتعلق بالعوامل التي دفعت بهم إلى إتباع ذلك الترتيب أو التصنيف، وكذا اقتراحاتهم لنصوص أخرى يرون أنها مفيدة، قد أسقطها الاستبيان. و هكذا، كان للنصوص التي أحرزت أكبر عدد من الأصوات، الحظ الأوفر من عدد الكلمات، و حتى تلك التي صنفت ضمن خانة النصوص الأدنى فائدة أدرجت ضمن المدونة و لم تُقص، لأنها "مفيدة" في نظر بعض المشاركين في الاستبيان. ثم عمد الباحثان إلى جمع المادة النصية من مصادرها، بعد أن حصلوا على إذن بتوظيف تلك العيون و المصادر في تزويد مدونتهما بالعينات النصية المطلوبة، و هذا مراعاة لحقوق التأليف، و للحيلولة دون الوقوع في المشاكل التي تنشأ جراء خرق مثل تلك المسائل القانونية. و أغلب تلك المصادر – إن لم نقل كلها – و التي استمدت منها النصوص المكتوبة عبارة عن مواقع الكترونية تضمنت:

- 1- مجلة "العربي" (الكويت). 2- مجلة "أفق" (العربية السعودية).
- 3- موقع "القصة العربية". 4- مجلة "PC" العربية (الإمارات المتحدة).
- 5- موقع "BBC" العربية (المملكة المتحدة).
- 6- مجلة "سيدتي" (المملكة المتحدة). 7- مجلة "عالم الاقتصاد" (السعودية).
- 8- مجلة "نزوى" (عمان). 9- الدورية الطبية العربية (سوريا).
- 10- موقع أكالات (الإمارات). 11- موقع "إسلام أون لاين" (قطر).
- 12- صحيفة "الرأي العام" (الكويت). 13- مجلة "المعرفة" (السعودية).
- 14- مجلة "آخر ساعة" (مصر). 15- عالم الكمبيوتر (مصر).
- 16- الكمبيوتر في العالم العربي (المملكة المتحدة).
- 17- صحيفة "الرأي" (قطر). 18- علوم وتكنولوجيا (الكويت).

أما عن الجانب المنطوق، فمن المزمع أن تجمع تسجيلات تغطي كثيراً من المواضيع، و هذا بالتعاون مع بعض الوسائط الإعلامية العربية – على حد قول الباحثين¹. و هكذا بلغ حجم مدونة اللغة العربية المعاصرة – في نسختها الأولية – حوالي 843 ألف كلمة ضمن 416 ملفاً، و قد تم حفظ كل نص في شكل الوثيقة XML، بعد تزويده برأسية تتضمن أهم البيانات الخاصة بالمؤلف و المصدر (أي توثيق النصوص بمراجعها)، و كذا اعتبار العوامل غير اللسانية (كالسن، الجنس، البلد، المستوى الاجتماعي، الثقافي،...).

و فوق ذلك، يطمح المصممان إلى ترميز المدونة على مستوى أقسام الكلام، نزولاً عند رغبة كثير من الأساتذة و المهندسين في اللغة العربية، و تلبية لحاجاتهم. و غير خاف أن سير العمل في تصميم المدونة قد اعترضته بعض المشاكل، لاسيما في ما يخص تصنيف النصوص و عددها و حجم كل عينة و كذا خاصية التمثيل.²

أ - تصنيف النصوص:

ليس من الهين أن تقرر إلى أي نوع/ مجال ينتمي النص، و قد سبق لجون سنكلير أن نبه إلى هذه القضية الحساسة، و تناولها بشيء من التفصيل و التحليل، فخلص إلى

¹ - Cf. L. Al-Sulaiti & E. Atwell (2006) Op. Cit., p. 19.

² - Cf. Ibid., pp. 28-30.

وجوب مراعاة العوامل الداخلية و الخارجية لحل ذلك الإش(ت)كال. و بيّن أن كثيرا من التصنيفات النصية تستند إلى الموضوع الذي يعالجه النص، تماما كما هي الحال في التصنيفات التي تتبعها الصحف و المجلات.¹ و يبدو أن المدونات العربية السابقة قد احتكمت إلى هذا القول (أي تصنيف النصوص حسب الموضوع Topic)؛ بيد أن سنكلير و إن كان قد أشار قبلا إلى الأخذ بهذا المعيار، إلا أنه يعود إلى القول بأن تصنيفا كهذا يفقد كثيرا من العناية و التنسيق؛ لذلك عمد إلى تقديم اقتراح بـ 35 صنفا.² بينما يرى شاروف (Sharoff) أن هذه التصنيفات يستحيل تطبيقها عمليا، و يوصي بتصنيف آخر يتضمن 8 أصناف أساسية أو ميادين عامة، تشمل أنواعا نصية أخرى، و هي:

- ميدان العلوم الطبيعية (الرياضيات، البيولوجيا، الفيزياء، الكيمياء، الخ).
 - ميدان العلوم التطبيقية (الزراعة، الطب، الهندسة، الإعلام الآلي، علوم البيئة، الخ).
 - ميدان العلوم الاجتماعية (القانون، التاريخ، الفلسفة، علم النفس، اللغات، التربية، الخ).
 - ميدان السياسة (الداخلية و الدولية).
 - ميدان التجارة (المالية، الصناعة).
 - ميدان الحياة (ميادين عامة، كالإعلانات، المكالمات/ المحادثات، المراسلات، الخ).
 - ميدان الفنون (الشعر، المسرح، و السير الذاتية، الخ).
 - ميدان الترفيه (الرياضة، السفر و السياحة، الموضة، الخ).³
- و قد قام الباحث بتصنيف نصوص مدونات هائلة في اللغة الانجليزية و الروسية و الألمانية و الصينية وفقا لهذا المخطط، و هي التصنيفات ذاتها التي اعتمدت في تصميم مدونة اللغة العربية المعاصرة، كما ضُمّت نصوصا لأشهر أعمدة الأدب العربي، على غرار: طه حسين، نجيب محفوظ، توفيق الحكيم، جبران خليل جبران؛ بالإضافة إلى نصوص دينية مستنبطة من الذكر الحكيم و الأحاديث الشريفة. و قد زوّدت المدونة بقاعدة معطيات، حفظت على مستواها بيانات خاصة بكل نص: الرقم التعريفي، العنوان، المصدر، عدد الكلمات، سنة النشر، و اسم المؤلف.⁴
- ب - عدد النصوص:

لم يتمكن الباحثان من الحصول على جميع النصوص التي كان في نيتهما إدراجها ضمن المدونة، نظرا لاستحالة إيجاد مواردها على مستوى الشبكة، مثال ذلك: صعوبة العثور على مصادر للمسرحيات أو المراسلات التجارية. ثم إن بعض النصوص لم يتسنّ لهما اقتناء إلا النزر اليسير منها، على شاكلة "قصص الأطفال" التي أخذت من مجلة

¹ - Cf. J. Sinclair (1996b) Preliminary Recommendations on text-typology. EAGLES Document EAG-TCWG-TTYP/P.

<http://www.ilc.cnr.it/EAGLES96/texttype.textty.html>

² - Cf. Ibid.

³ - Cf. Serge, Sharoff (2004) Towards basic categories for describing properties of texts in a corpus. In: M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, R. Silva, C. Pereira, F. Carvalho, M. Lopes, M. Catarino & S. Barros (eds.) LREC04: Proceedings of Language Resources and Evaluation Conference, vol. V, pp. 1743-1746. Portugal: Lisbon.

⁴ - Cf. L. Al-Sulaiti & E. Atwell (2006) Op. Cit., p. 29.

"العربي". و كذا هي الحال – أو الأخرى – كذا هو الإشكال بالنسبة للجانب المنطوق من المدونة.

ج - حجم العيّنات:

لم يحدّد الباحثان حجما معينا و واحدا في جميع العيّنات النصية، و قد عمدا إلى أخذها كاملة، بدل تشطيرها أو الحذف منها. و الملاحظ أن أغلبها عبارة عن مقالات قصيرة بحجم 4 آلاف كلمة، أما ما كان من قصار النصوص التي لا تتعدى 100 كلمة، فقد جمعها كلها ضمن ملف واحد.

د- خاصة التمثيل:

يتحقق التوازن في تصميم المدونة من وجهين: انتقاء النصوص و عدد الكلمات في كل جنس كتابي. بيد أن مسألة "حقوق التأليف" قد حالت دون تحقيق هاته الغاية؛ يضاف إليها عامل الوقت، إذ من الصعب تحقيق مشروع بهذا الحجم في ظرف سنة واحدة. رغم تلك المعوقات و الصعوبات التي قد تبدو عيوباً اعتورت المشروع برمته، إلا أنه يمكن الجزم بأن المدونة خطوة مظفار إلى الأمام في مجال الصنعة المدونانية العربية، و قد كان أن عرضت أول مرة في ملتقى TALC'04 بغرناطة، و استخدمت لإثبات ملائمة برامجية XAIRA في معالجة المدونات العربية.¹ و فوق ذلك، كانت المدونة دافعا من الدوافع إلى التفكير في مدونة أخرى، تعد أضخم مشروع عرفته اللغة العربية المعاصرة كما سنتبين ذلك في ما يلي.

10.2.II - مدونة اللغة العربية الدولية (ICA):

و هو مشروع بادرت به مكتبة الإسكندرية (Bibliotheca Alexandrina) بمصر؛ تسعى من خلاله إلى تجميع عيّنات من العربية الفصحى المعاصرة (MSA) و المكتوبة، بحجم 100 مليون كلمة، و بدافع الحاجة الماسة إلى مدونة ممثلة للغة العربية، من جهة؛ ولأن ما سبقها من محاولات لصنع مثل تلك المدونة – من جهة أخرى – لم يكن يفي بالغرض لبلوغ ذلك المرام. و الحق أن المشروع متميز عن غيره تميزا يدفع بنا إلى أن نقف عنده وقفة متأنية لاستكشافه.

➤ تصميم المدونة:

لقد حاول الباحثون تصميم المدونة بدقة متناهية، آخذين بعين الاعتبار كلا من خاصة التمثيل و التنوع و التوازن و الحجم؛ و حرصوا أيما حرص على توفرها في النصوص المزمع إدراجها، بحكم أنها العناصر الأربعة التي تقوم عليها كل مدونة. و الحال هذه عمد الباحثون إلى:

- تحديد بعض مصادر النصوص.
- جمع تشكيلة متنوعة من الأجناس الكتابية (Genres).
- تقسيم كل جنس إلى عيّنات نصية.
- تدوين كل البيانات اللازمة تدوينا دقيقا، لاسيما فيما يتعلق بتاريخ و مكان كتابة/ نشر النصوص، و فيما إذا كانت عبارة عن كتب أو مقالات مطبوعة، أو مقالات على الشبكة، ...

¹ - Cf. Ibid.

- اعتبار خاصة التناسب* (Proportions) بين مختلف أجناس المدونة وأقسامها، وذلك بالنظر إلى مدى شيوع المصدر أو الجنس الكتابي.¹
- اتخاذ القرارات بشأن عدد الفئات (Categories) التي ينبغي أن تتضمنها المدونة، و عدد العيّنات في كل فئة، و كذا عدد الكلمات في كل عيّنة. و عموما يتعلق حجم المدونة بعدد كلماتها؛ بيد أنها مسألة تتعلق أيضا بعدد النصوص التي استقيت من مختلف الفئات، و بعدد العيّنات من كل نص، و كذا عدد الكلمات في كل عيّنة. و جدير بالإشارة في هذا المقام، أن عملية تصنيع المدونة عملية دورية (Cyclical)، تتطلب إعادة تقييم على الدوام طيلة عملية التجميع. لذلك، يعتمد الباحثون إلى تغيير خطتهم الأولية في التصميم كلما اقتضى الأمر أو دعت الضرورة.

ب - المصادر و الأجناس الكتابية:

حرصا على أن تكون المدونة ممثلة للغة العربية، عمد الباحثون إلى انتقاء النصوص من مصادر مختلفة، على غرار: الصحافة، المجالات (العامة منها والمتخصصة)، الكتب، الروايات، المقالات المتاحة على الشبكة، المنشورات الأكاديمية؛ كما قاموا بتقسيم بعض تلك المصادر إلى فروعها (الجدول رقم 7-).

المصادر الأصول	المصادر الفروع
1-الصحافة.	صحف.
	مجلات: - عامة. - متخصصة.
2- مقالات الشبكة.	
3-الكتب.	
4-المصادر الأكاديمية.	

جدول رقم (7) المصادر الأصول و الفروع في (ICA).

و في مجال الصحافة – على سبيل المثال–، اعتمد الباحثون على أشهر الصحف اليومية العربية، نذكر منها: "الأهرام" و "الأخبار" و "الجمهورية" و "الأخبار الرياضية" و "السياسة الدولية" (مصر)، "الوطن" و "الجزيرة" و "الهداية" (السعودية)، "الدستور" (الأردن)، "القدس" و "الحياة الجديدة" (فلسطين)، "الراية" (قطر)، و غيرها. و يرمي الباحثون بذلك إلى تغطية الأجناس الكتابية نفسها من مختلف المصادر في جميع أنحاء العالم العربي، و ما الدول العربية التي ذكرناها للتو إلا أمثلة عما تم إدراجه، في انتظار البقية.²

كما تتضمن المدونة أيضا أجناسا كتابية شتى، مصنفة بدورها – في الأغلب الأعم– إلى أجناس فرعية، حسب ما يوضحه الجدول أدناه.

* - خاصة التناسب بالمفهوم الرياضي .

¹ - وذلك ما يقصد بمفهوم التوازن Balance في المدونات، ولا يعتقد القارئ أنه يعني التساوي الكمي بين النصوص من مختلف المصادر والأجناس.

² - Cf. S. Alansary et al. (2007) Op. Cit.

الأجناس	الأجناس الفروع
1- العلوم الإستراتيجية	السياسة - القانون
2- العلوم الاجتماعية	الاقتصاد - علم الاجتماع
3- الرياضة	
4- الدين	الإسلام- المسيحية- ديانات أخرى- الدين المقارن.
5- الأدب	النثر: الروايات - الأقاصيص - قصص الأطفال - المسرحيات
	الشعر
	دراسات في الأدب و اللسانيات
6- العلوم الإنسانية	التاريخ- علم النفس - الفلسفة- الجغرافيا.
7- العلوم الطبيعية	البيولوجيا - الفيزياء- الكيمياء- الجيولوجيا و البيئة - الفضاء.
8- العلوم التطبيقية	الطب - الهندسة- الزراعة- التكنولوجيا.
9- الفنون	
10- السير و التراجم	
11- متفرقات*	

جدول (8) الأجناس الكتابية الأصول و الفروع.

و بديهي أن اعتماد تلك المصادر أو الأجناس دون أخرى لم يكن اعتباطاً؛ بل إنه تم باعتبار بعض المقاييس، كأن تكون تلك المصادر/الأجناس شائعة و مشتركة بين مختلف الأقطار العربية، و أن يتضمن كل مصدر جملة من تلك الأجناس الكتابية، و ذلك ما يوضحه الجدول أدناه كتصنيف أولي للمدونة:

مصادر النصوص	أنواع النصوص
الصحف	السياسة - الفنون - الأدب - الثقافة - العلوم - التاريخ - الاقتصاد - الدين - الرياضة - علم الاجتماع - متفرقات.
المجلات	السياسة - الفنون - الأدب - الثقافة - العلوم - التاريخ - الاقتصاد - الدين - الرياضة - علم الاجتماع - متفرقات.
الروايات	الخيال (Fictions) - اللاخيال (Unfictions) - أقاصيص الأطفال - الدين - المسرح.

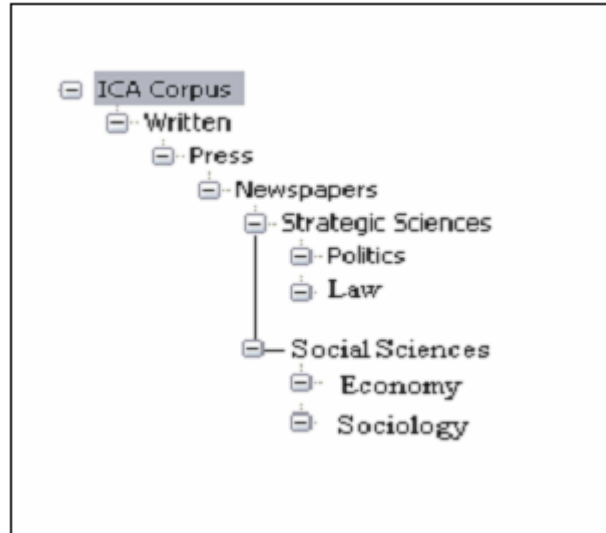
* - يمكن أن نجد في الصحف والمجلات قلة من النصوص التاريخية والثقافية والعلمية، لذلك فإن "التاريخ" مثلاً ضمن الصحف لا يمكن اعتباره جنساً كتابياً قائماً بذاته ضمن المصدر "الصحف"؛ وعليه، فإن جميع النصوص على غرار هذا المثال تجمع ضمن خانة "المتفرقات".

الكتب	الفنون – التاريخ – الرياضة – الدين – العلوم – الثقافة – الأدب.
مقالات الشبكة	السياسة – الفنون – الأدب – العلوم – التاريخ – الاقتصاد – الدين – الرياضة – متفرقات.
المنشورات الأكاديمية	العلوم.

جدول (9) التصنيف الأولي للمدونة.

- و يمكن أن نسجل بعض الملاحظات بخصوص الجداول:
- تمثل الروايات جنسا كتابيا مستقلا، و إن كانت في الأصل كتباً.
- تصنف المجالات إلى فرعين: مجالات عامة، تمثل مقالاتها أكثر من جنس كتابي؛ ومجالات متخصصة ضمن جنس واحد.
- تدرج الصحف و المجالات ضمن مصدر واحد هو "الصحافة"، يضاف إليها "الصحافة الإلكترونية".
- تتضمن المدونة أربعة مصادر رئيسية، هي: الصحافة و المقالات المتاحة على الشبكة و الكتب و المنشورات الأكاديمية. كما تتضمن إحدى عشر جنسا كتابيا رئيسيا (المبينة في الجدول - 8-).
- تقسم بعض المصادر الأصول إلى مصادر فروع، و كذا الحال بالنسبة للأجناس الكتابية الأصول.¹
- و الشكل (16) يبين أحد أهم المصادر الأصول المعتمدة (صحافة) في المدونة، و واحد من مصادرها الفرعية (الصحف). كما يبين علاوة على ذلك الأجناس الكتابية الأصول (العلوم الإستراتيجية، العلوم الاجتماعية). و الأجناس الفرعية عنها في نفس المدونة.

¹ - Cf. S. Alansary et al. (2007) Op. Cit.



الشكل (16) المصدر الأصل (صحافة) في مدونة ICA.

ج - معالجة المدونة:

فيما يخص تحشية المدونة، عمد الباحثون إلى تنقيحها تنقيحاً لغوياً على مستوى بنية كل نص (Structural Markup)، كأن تدرج البيانات العامة بخصوصه ضمن رأسية الملف الواردة في بداية كل نص، و يمكن أن تتضمن الببليوغرافيا الخاصة. و فيما يلي مثال عن التنقيح:

D/
T/
نقل ملكية عمر أفندي إلى أنوال السعودية
/T
P/
أخيراً.. تم نقل ملكية شركة عمر أفندي أمس إلى شركة أنوال السعودية . وقرر الدكتور محمود محيي الدين
وزير الاستثمار تعيين محمد وهب الله رئيس النقابة العامة ممثلاً عن حصة المال العام وقدرها ١٠ % .
/P
/D

بحيث إن: D/ بداية الوثيقة (Document).

D/ نهاية الوثيقة.

T/ بداية العنوان (Title).

T/ نهاية العنوان

P/ بداية الفقرة (Paragraph).

Q/ بداية السؤال (Question).

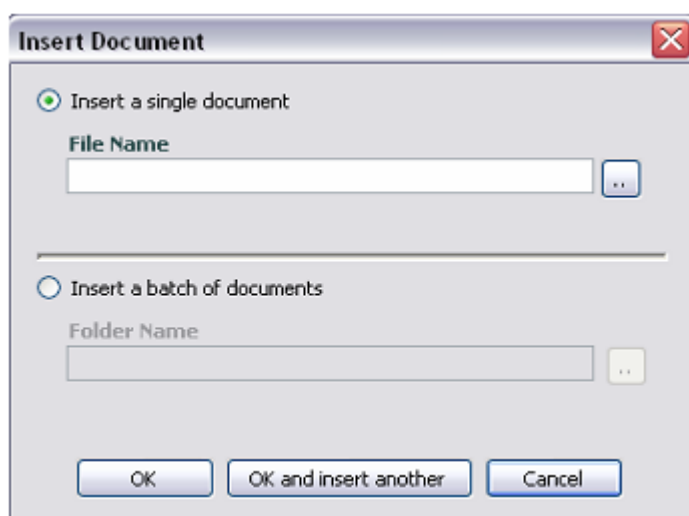
P/ نهاية الفقرة.

Q/ نهاية السؤال.¹

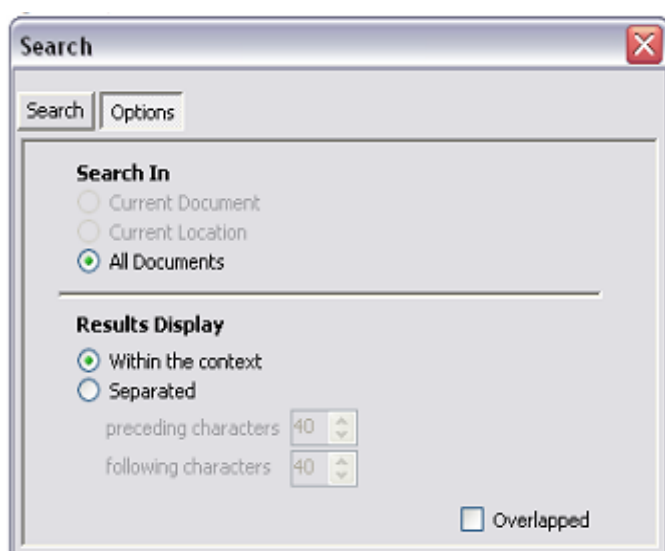
❖ أما في تحليلها، فقد شرع في صناعة برنامج خاص لذلك، و من المهام المنوطة به: إدراج وثائق جديدة واحدة فواحدة، أو جميعها دفعة واحدة، حسب ما يوضحه الشكل أدناه. و يتكفل النظام وحده باستبيان مصدر النص و الجنس الكتابي الذي

¹ - Cf. Ibid.

ينتمي إليه، تبعا لرمز اسم الملف¹ (Filename code)، ثم يدرجه ضمن العقدة المناسبة في هيكل المدونة.



الشكل (17) نافذة إدراج الوثائق.



❖ البحث في المدونة:

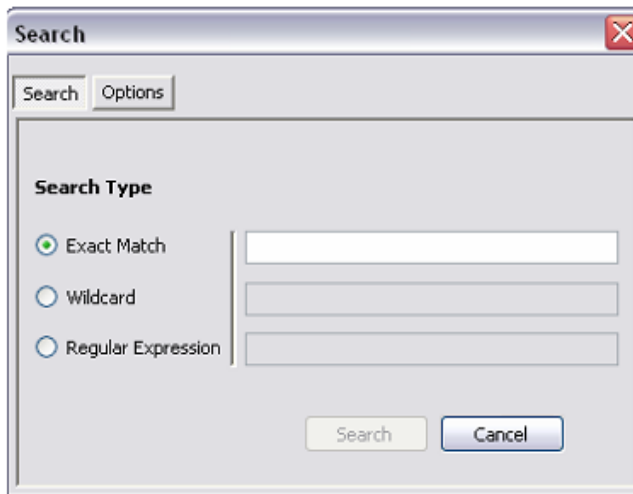
يتيح البرنامج أنواعا عدة من البحث (Search Types)، فضلا عن خيارات متعددة (Search Options). ففي خيارات البحث، يمكن للباحث أن يحدد مجال البحث في المدونة:

الشكل (18) خيارات البحث.

الوثيقة الحالية (Current Document)، الموقع الحالي (Current Location)، جميع الوثائق (All Documents). فضلا على ذلك، يمكن للباحث أن يختار الواجهة البحثية لعرض النتائج، سواء ضمن السياق أو بمعزل عنه (كما في الشكل رقم 18-). أما في ما يتعلق بأنواع البحث، فنميز: المطابق الدقيق (Exact Match)، الويلدكارد (Wildcard)، التعبير القياسي (Regular Expression)، حسب الشكل أدناه. و يمكن أن نتبين أوجه الاختلاف بين هذه الأنواع الثلاثة، من خلال المثال التالي:

¹ - يتم تجميع النصوص ضمن ملفات مرمزة (Filename code) قصد تسهيل بحثها؛ مثال ذلك، الترميز في اسم الملف AH 10- A 1.1.1-140207 ، بحيث: (AH) رمز المصدر "صحيفة الأهرام"، أما العدد (10) فيراد به أن هذا الملف هو عاشر مقال في الصحيفة ذاتها مع الجنس الكتابي الأصل والفرع نفسيهما، وكذا التاريخ نفسه. A1.1.1 : (A1) رمز الصحيفة المصدر، (A1.1) رمز الجنس الكتابي الأصل (العلوم الإستراتيجية)، (A1.1.1) رمز الجنس الكتابي الفرع (السياسة). 140 207 : (14) اليوم، (02) الشهر، (07) السنة.

إذا قمنا باختيار حرفين، " في " مثلاً، فإن استخدام هذين الحرفين يختلف حسب الأنواع الثلاثة من البحث، و كذلك النتائج تتباين من نوع لآخر:



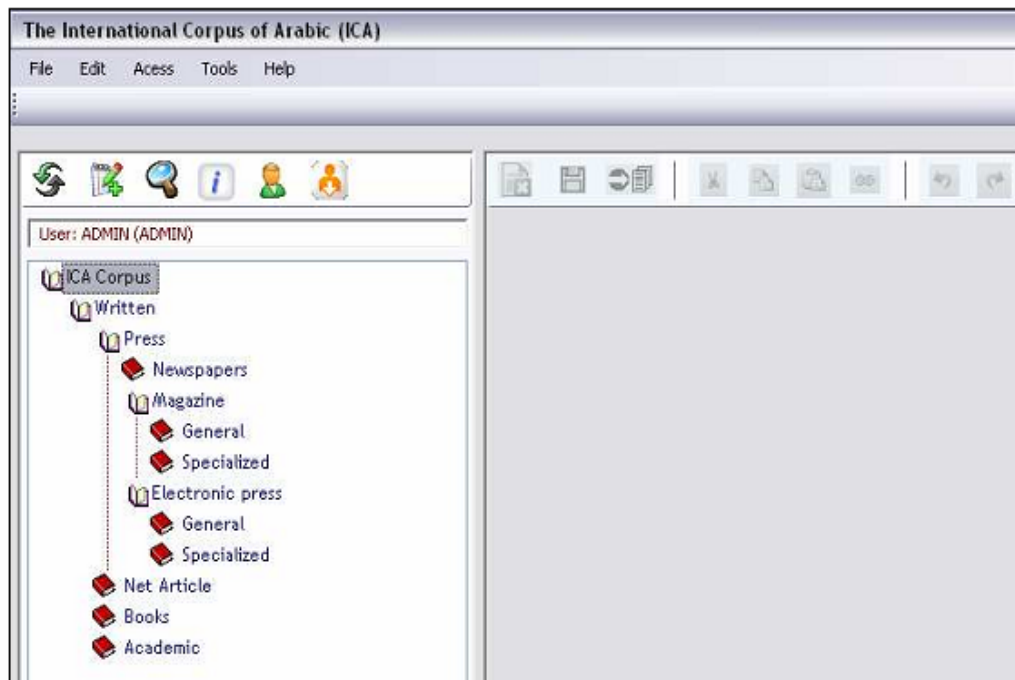
الشكل (19) أنواع البحث.

■ في حالة البحث باعتماد الويلدكارد، يمكن للباحث أن يكتب الحرفين إما متصلين، وإما أن يفصل بينهما بعلامة نجمية [*] أو علامة استفهام [؟]. تعني العلامة النجمية أي أحرف متصلة بعد هذين الحرفين أو بينهما؛ فإذا كتب الباحث (ف*ي)، فإنه يحصل على كل كلمة تتضمن هذين الحرفين، و على كل متتابعة من الأحرف بينهما. أما علامة الاستفهام [؟] فتعني (0) أو حرفاً واحداً بعد هذين الحرفين؛ فإذا كتب الباحث (في؟؟) فإنه يحصل على كل كلمة تتضمن هذين الحرفين يتصل بهما في آخرهما حرفان آخران، بمعنى كل كلمة تتضمن أربعة أحرف طولا و تبندئ بـ (في).

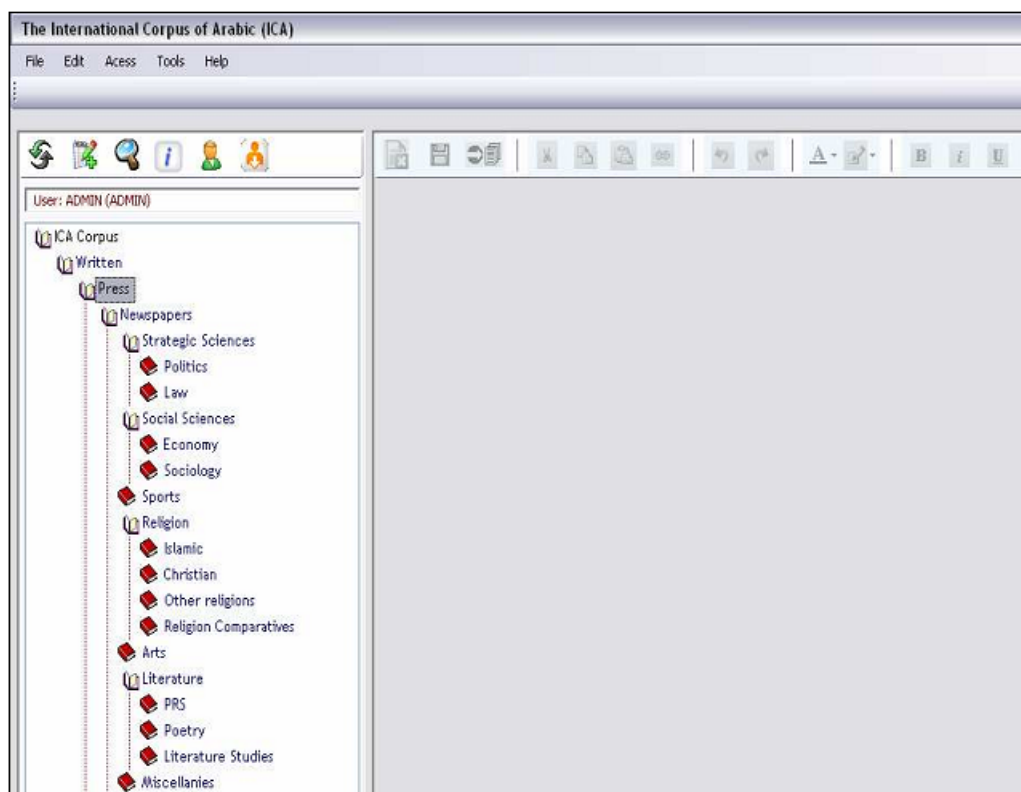
■ أما في حالة البحث "التعبير القياسي"، يمكن الحصول على أي نوع من السلاسل و لو كانت تفوق مستوى الكلمة، مثال ذلك "ف - ي[2]" تعني البحث عن كل حرفين من (ف) إلى (ي).¹

فضلا على ذلك، يمكن للبرنامج نفسه أن يعرض سلمية تصميم المدونة أي خارطتها (Corpus map)، وفق ما اختاره الباحث من عقدة (Node) في شجرة بنية المدونة. (أنظر الشكل -20-). و الشكل يبين المصادر الأساس و المصادر الفرعية، إلا أنه يمكن بسطه أو نشره أكثر لعرض الأجناس الكتابية الأساس منها و الفرعية، و ذلك من خلال فتح المزيد من العقد (أي بالنقر عليها)، و الشكل -21- يوضح لنا ذلك، من خلال عرض فروع المصادر و الأجناس الكتابية للنصوص الصحفية. و هكذا دواليك إلى أن يصل الباحث إلى آخر ملف يتضمنه المصدر الواحد (الشكل -22-)، أي النص.

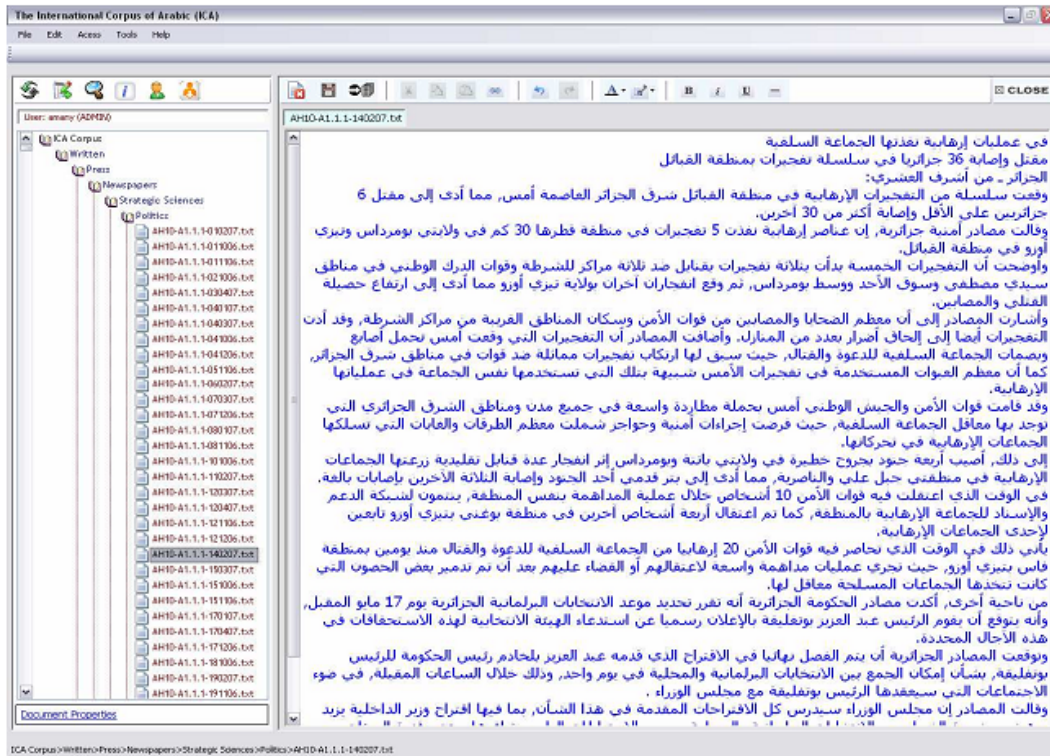
¹ - Cf. S. Alansary et al. (2007) Op. Cit.



الشكل (20) المصادر الأساس و المصادر الفروع.



الشكل (21) الأجناس الأساس و الأجناس الفروع في ICA.



الشكل (22) عرض نص من المدونة ICA.

على ضوء كل ما تقدم، يمكن القول إن مشروع هذه المدونة محاولة ناجحة لمقاربة عملية صناعة مدونة ممثلة للغة العربية الفصحى المعاصرة، مقارنة بما تقدمها من مدونات؛ و التي لا نشك في أنها قد خدمت/ تخدم – هي الأخرى – العربية. و على الرغم من عددها "المعتبر"، البالغ حوالي 29 مدونة أو يزيد حسب تقرير نملار NEMLAR لعام 2005¹، إلا أن إنتاجها كان من قبل أفراد أو فرق بحث من خارج المنطقة العربية، و بدعم من مؤسسات و هيئات أجنبية، على غرار: "إل دي سي" و "إلرا"؛ و هي إما نتاج مشروعات تجارية، و إما أنها نتاج مشروعات بحثية بحت أو بغرض نيل درجات علمية.

و تعزو سلوى حمادة سبب ذلكم القصور إلى "عدم معرفة أو استيعاب اللغويين لأهمية النصوص اللغوية المدونة في المدونات و دورها في دعم الأبحاث في مجال اللغة العربية [...]"²، اللهم إلا ما كان من المدونة الانجليزية- العربية المتوازية بقسم الانجليزية (جامعة الكويت)، حيث تستخدم في تدريس الترجمة من و إلى اللغة العربية و كذا المعجماتية. و الواقع أن كثيرا من الموارد – بغض النظر عن مسألة حقوق التأليف – هي بمثابة مرتع خصب، يحتاج إلى جهد العاملين، و رعاية الراعين القائمين على المجال.³

¹ - Cf. Mahtab Nikkhou & Choukri Khalid (2005) *Op. Cit.*

² - سلوى حمادة (2007) مر. س .

³ - Cf. Arabic Online Resources:

<http://terpconnect.umd.edu/~dlrg/clir/arabic.html>

<http://www.globalwordnet.org/AWN/Resources.html>

و كما سبق و أوضحنا، ليس الهدف تجميع النصوص كيفما تراءى للباحث، وتخزينها على شاكلة القواعد النصية العادية؛ بل إنما يرمي إلى أبعد من ذلك من حيث معالجتها آليا باستخراج البيانات و دراستها و تحليلها. غير أن المعالجة الآلية للغة العربية (لاسيما الترميز بأقسام الكلام) تعترضها بعض المعوقات، هي من صميم خصوصيات اللغة العربية.

أولى تلكم الخصوصيات، تتمثل في الخط العربي "الفريد"؛ إذ تتعدد أشكال الحرف العربي الواحد، و تختلف باختلاف موقعه ضمن المفردة (بدايتها، وسطها، أو آخرها)، كحرف العين مثلا (ع، ع، ع). كما أن الأحرف العربية ضمن المفردة الواحدة ترد متصلة بعضها ببعض. ثم إن كتابة الخط العربي تكون من اليمين إلى اليسار، لذلك عمد بعضهم إلى نسخ الخط العربي وفق نظام الأحرف الرومانية، و أشهر تلك النظم: نظام بكوالتنر¹. Buckwalter¹ و هذا يعني - بطبيعة الحال - أن الاتجاه يقلب، فيصير من الشمال إلى اليمين تماما كبقية اللغات، و تلك خطوة كانت تعد ضرورية من وجهة نظر حوسبية، لا سيما أن الحواسيب كانت تعمل وفق ASCII؛ أما اليوم، فقد تغيرت الحال بفضل نظام اليونيكود (Unicode)، و صارت أغلب أنظمة التشغيل المعاصرة تستند إلى التشفير النصي متعدد اللغات (Multi-lingual Text Encoding).

أما ثاني تلكم الخصوصيات، أن العربية تميل إلى تمديد المعنى القديم للكلمة الواحدة للدلالة على معنى/ معان جديدة، مما يعني أن الكلمات متجانسة لفظا و تؤدي معان مختلفة، أو أن المعاني المختلفة تشترك في اللفظ الواحد، كلفظ "عين" مثلا. و هذه تجعل من الترميز بأقسام الكلام صعبا بعض الشيء مقارنة باللغات الأخرى.² أما ثالث تلك المعوقات التي تعترض عملية المعالجة الآلية للغة العربية، فتتمثل في غياب الحركات؛ إذ اللغة العربية المكتوبة غير مشكولة (باستثناء الكتاب الكريم)، مما يزيد من الغموض الذي تخلقه المفردات، ما لم ينظر في سياقاتها، و يجعل تحليل العربية أعقد مقارنة ببقية اللغات. و وجه التعقيد في ذلك أن اللفظ الواحد - كما أسلفنا- ينتمي إلى أقسام مختلفة من أقسام الكلام؛³ مثال ذلك، متتالية الأحرف "عين":

فعل	عَيْنَ [يَعِين، عينا، وعينة]: عظم سواد عينه.
	عَيْنَ [الشيء، تعيينا]: خصصه من الجملة و أفرد.
	[الشيء لفلان]: جعله مخصوصا به.
اسم	عَيْن: عضو البصر، الجاسوس، ينبوع الماء،...
	عَيْن: بقر الوحش.
	عَيْن: أهل البلد، الجماعة.

كما أن هذه الخاصية تزيد أيضا من صعوبة البحث ضمن المدونة باستخدام البرنامجيات الخاصة بذلك، و تؤدي إلى نتائج محدودة أو غير ملائمة. ففي الأغلب الأعم،

¹ - Cf. Tim, Buckwalter (2002) Buckwalter Arabic Transliteration.

<http://www.qamus.org/transliteration.htm>

² - Cf. Mark van Mol (2000a) Op. Cit.

³ - Ibid.

يتم البحث عن جميع استعمالات الكلمة الواحدة انطلاقاً من الجذع (Stem)، والمفهرسات الآلية تعرض البحث وفق الخيار "ويلدكارد"، مثال ذلك:

(write*) ← writing, writers, writer, writes, write و هلم جرا.

أما بالنسبة للغة العربية، فالعملية أعقد، لأن كلماتها تشتق من جذع يتألف من ثلاثة أحرف في الأغلب الأعم، و يقلب على عدة أوزان و أشكال، بإضافة السوابق و الدواخل و اللواحق.¹ و نضرب لذلك مثال الجذع (سلب) (slb)، الذي يعني في معناه الأساسي انتزع الشيء من غيره قهراً، فإذا أردنا البحث عن جميع الكلمات التي تشتق منه و تؤدي معناه، حصلنا على: سلب، يسلب، و غيرها من الأفعال، و الاسم "السَّلب" (ما يسلب)، ومصدره "السَّلب"، و فاعله "سالب"، و الصفة "مسلوب"²...

أما إذا كان البحث بطريقة الويلدكارد، نحو *s*!b*، فسنحصل على قائمة بجميع الكلمات السابقة، بالإضافة إلى كلمات لا علاقة لها بالمعنى السابق، إلا لأنها تشتق من نفس الجذع، أو لأنها تحتوي نفس الأحرف، مثال ذلك: الأسلوب و الأساليب، و السالب (بعكس الموجب)، و السلسيل، و ما إلى ذلك؛ بمعنى أن البحث عن كلمة معينة في اللغة العربية، يعطي في كثير من الأحيان كلمات لا حاجة لنا فيها، و يجعلنا نقضي معظم الوقت في البحث، أو في النظر إلى فهرست يتضمن بالإضافة إلى اللفظ المراد ألفاظاً "خاطئة"، إن جاز التعبير.

و مثال ذلك، التجربة التي أجراها فان مول في البحث عن الفعل (ktb) "كتب" في معناه الأساسي، و ذلك ضمن مدونة خام بحجم زهاء 1500 000 كلمة، فكانت نسب النتائج كالتالي:³

الكلمة	نسبة الفعل الموافق (المبحوث عنه)	نسبة الكلمات الأخرى	نسبة الزمن الضائع في البحث
كُتِبَ	28		72%
كُتِب		18	
مكتب		42	
مكتبة		12	

الجدول (10): نسب نتائج البحث عن الفعل (ktb).

و نفس الإشكال * يطرح بالنسبة للفعل الأجوف مثلاً، إذ ستكون النتائج أضعف وإن كانت ستبدو أحسن. مثال ذلك: البحث عن الفعل "كان"، يعطي نتيجة موفقة بنسبة 84% - حسب فان مول -، بيد أن هذا الرقم لا يعكس النسبة الحقيقية، لأنها لا تتضمن جميع الصيغ الصرفية للفعل "كان"، على غرار: مضارعه في حالة الرفع، و الجزم والنصب.

¹ - Cf. Mark van Mol (2000a) Op. Cit.

A. Roberts, L. Al-Sulaiti & E. Atwell (2005) aConCorde : Towards a proper concordancer for Arabic.

² - السلب: ما يسلب، والسلب: السير الخفيف.

³ - Cf. Mark van Mol (2000a) Op. Cit., p. 3.

* - ونذكر بأن المدونة خام، ولو كانت مرمزة بأقسام الكلام، لكانت النتائج المحصل عليها أفضل دونما ضياع الوقت سدى. وهنا تكمن أهمية الترميز بالنسبة للمدونات العربية.

و الحال هذه، يقترح الباحث أن تشكل النصوص كحل للإشكال،¹ غير أن هذا الحل يبقى نسبيا أو جزئيا، بحيث أنه لا ينسب المفردة إلى القسم الحقيقي الذي تنتمي إليه مختلف حالاتها؛ مثال ذلك، كلمة "المسؤول" في جملتي:

استقبلني المسؤول. فاعل.

استقبلني الموظف المسؤول. صفة.

أما آخر خاصة، فهي خاصية الاتصال بين الكلمة العربية،² مثال ذلك: متتالية الأحرف *fhm* (فهم)، إما أنها تحيل على الفعل *فَهَمَ* (أو *فُهِمَ/ فُهِمَّ*)، أو أنها تحيل على حرف الفاء مقترنا بضمير الغائب المنفصل (*هُمَ*)، أو بالفعل "*هَمَّ*" (*فَهَمَ...*)؛ و كذلك هي الحال بالنسبة للمتتالية *whm* ("وهم")، مما يعني أن البحث – مثلا – عن استعمالات حرف العطف "الفاء" أو "الواو" باعتماد مدونة خام، يؤدي إلى ظهور جميع الكلمات التي يظهر فيها الحرفان في الوسط أو الآخر!).

لهذه الأسباب، و سعيا لاستجلاء الغموض الذي تخلقه مثل هذه المفردات، عمد فان مول إلى استحداث نظام لتصنيف الكلمات العربية، يعتمد فيه على الترميز بأقسام الكلام، و هو نظام يُمكن من تحديد الكلمات في الجملة منفصلة عن السوابق/اللواحق (سواء كانت حروفا أو ضمائر متصلة)، و لهذا النظام فائدتان:

حساب الكلمات في النصوص العربية على أنها وحدات مستقلة، و الحصول على الكلمة المراد البحث عنها دونما "تشويش"، مثلما مر معنا في مثال "فهم"، إذ يسمح النظام بالتمييز إذا ما كانت الفاء حرف عطف متصل بالفعل "*هَمَّ*" أو الضمير "*هُمَّ*"، أو أنها أصلية من حروف الفعل (*فَهَمَ/ فُهِمَّ*).

و يجدر التنبيه إلى أن حساب الكلمات آليا في النصوص العربية هو حساب لمتتابعة من الأحرف بين فراغين، بمعنى أن كلمة "*وَهُمَّ*" – مثلا – تعد كلمة واحدة، لأنه عادة ما يكتب حرف العطف متصلا بالضمير، أي دونما فراغ بينهما، و هنا تكمن أهمية النظام الذي استحدثه فان مول، من حيث إنه يتعرف إلى حروف العطف المتصلة – مثلا – الفاء و الواو، و كذلك حرف الجر (الكاف)، و حرف الاستقبال للزمن القريب (السين)، و يحسبها على أنها مستقلة عن الكلمات التي تتصل بها. و باعتماد نظام فان مول، يمكننا أيضا أن نحسب أداة التعريف "ال" كمستقلة، و كذلك الضمائر المتصلة (كتابه، كتابهم).^{*} و على هذا النحو، فطن كثير من الباحثين إلى أهمية تحشية مدوناتهم العربية، و ترميزها على مستوى أقسام الكلام، مثال ذلك: المدونة العامة للعربية العلمية (GSAC)، التي قام صاحبها بترميز ما يعادل 1 مليون كلمة ترميزا دقيقا بنسبة 92% – كما سبق وذكرنا.

و ثمة مدونة أخرى مرمزة بحجم 50 ألف كلمة، قوامها النصوص الصحافية، من صنع "شيرين خوجة" التي عمدت إلى استحداث برنامج لترميز اللغة العربية POS Tagger، يتكون من 131 شارة شارحة،³ مستعينة بالتقسيم الذي وضعه قدامى النحاة

¹ - Cf. Mark van Mol (2000a) *Ibid*.

² - *Ibid*.

^{*} - بيد أن الباحث يبدي تحفظا في هذا الأمر، باعتبار أنها تشكل وحدة واحدة.

³ - Cf. Shereen, Khoja (2001) APT: Arabic part-of-speech tagger. In: *Proceedings of Student Workshop (NAACL 2001)*, Carnegie Mellon University, Pittsburgh, Pennsylvania, U.S.

العرب (الاسم، الفعل، الحرف). و هذه الأقسام الثلاثة الرئيسية تنقسم بدورها إلى أقسام ثانوية، نحو:

الفعل: مضارع، ماض، أمر، معتل، صحيح،...

الاسم: فاعل، مفعول، اسم إشارة،...

الحرف: حروف الجر، الجزم، النصب،...

و إن هذه الخصائص هي ما يصنع اختلاف اللغة العربية عن اللغات الهندوأوروبية، كما أنها تحتم تطوير برنامج للترميز يختلف عن برامج تلك اللغات – على حد تعبير خوجة:

"[...] Arabic is a very different language from Indo-European languages, and should have its own tagset [...]."¹

أما فيما يخص الفهرسة الآلية، فقد قام أندرو روبرتس – من جامعة ليدز – بتصميم مفهرس خاص يوافق اللغة العربية إلى جانب لغات أخرى، يعرف باسم "aConCorde". و قبل أن نستعرض بعض ميزاته و خصائصه، لا ضير في أن نبدأ بمقارنة بينه و بين برنامجيات أخرى، على شاكلة: "مونوكونك" (MonoConc) و "وورد سميث" (WordSmith).²

هاتان البرنامجيتان ذواتا طبيعة تجارية، و هما من أحسن ما صمم لحد الساعة من برنامجيات تحليل النصوص؛ بيد أنهما لا تستخدمان إلا مع نظام التشغيل ميكروسوفت ويندوز Microsoft Windows. و لقد حاول كثير من الباحثين استغلالهما في تحليل المدونات العربية، و من ذلك تجربة روبرتس و السليطي و أطول من جامعة ليدز، باعتماد مدونة اللغة العربية المعاصرة (CCA)، و هي مدونة مرمزة وفقا لمقاييس TEI باستخدام التنقيح اللغوي (XML)، و جميع نصوصها مشفرة وفق UTF-8. و قد باءت التجربة بالفشل مع "مونوكونك"، و اتضح أنه لا يصلح لمعالجة اللغة العربية، نظرا للمشاكل التي اعترضت سبيل الباحثين؛ فالمفهرس المذكور لا يتعامل مع نصوص مشفرة وفق UTF-8، مما اضطرهم إلى تحويل النصوص إلى ويندوز العربية كودبايج Windows Arabic Codepage-1256.

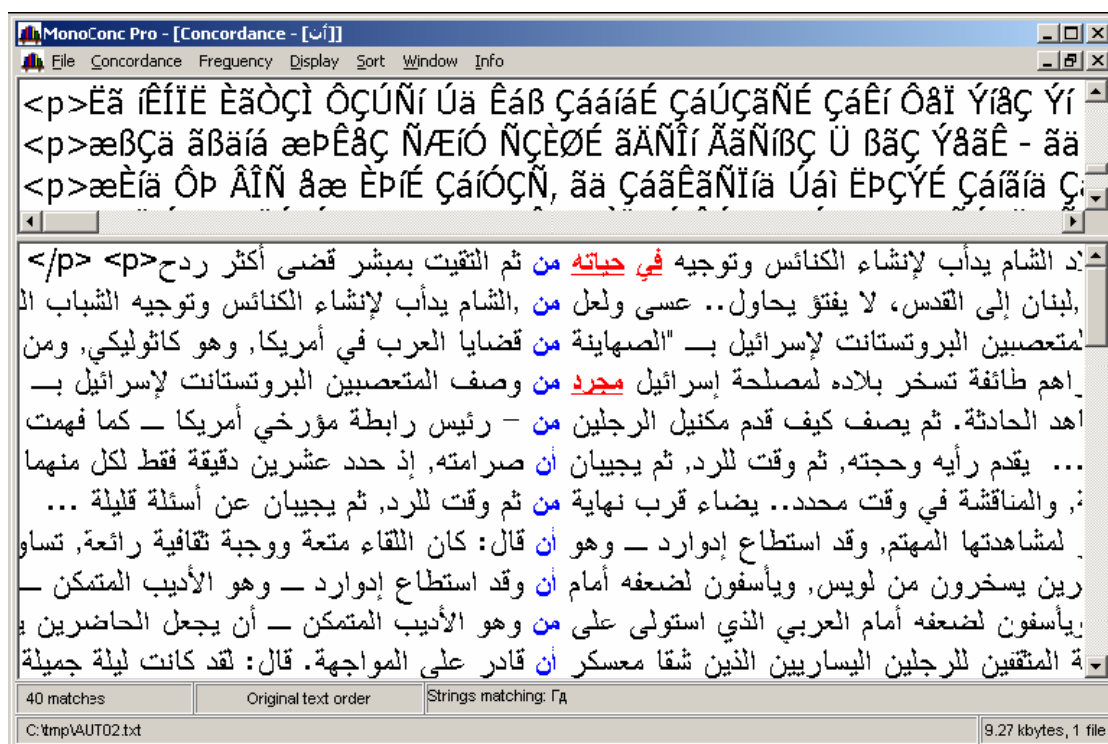
أما المشكلة الأكبر، فكانت حين ظهرت الفهارس و قد اختلط فيها الحابل بالنابل – كما في الشكل (23) أدناه –، و مرد ذلك خصوصية الخط العربي، من حيث إن اتجاهه من اليمين إلى اليسار، و مونوكونك لا يراعي هذا الترتيب في عرضه للنتائج. أما عن الفوضى التي تعم المتلازمات اللفظية فحدث و لا حرج، بحيث إن الوصيات تشغل مواقع غير مواقعها، و تتخذ تراتيبا غير تراتيبها، محدثة بذلك كثيرا من التداخل. هذا ناهيك عن ظهور بعض المفردات غير المرغوب فيها أو غير المطلوبة أصلا، و هي مشكلة ملحوظة مع بعض المفردات فقط.

¹ - Ibid.

² - Cf. MonoConc web site at:

<http://www.lexically.net/downloads/version4/html/index.html>

³ - Cf. Andrew Roberts, Latifa Al-Sulaiti & Eric Atwell (2005) Op. Cit.



الشكل (23): واجهة بحثية لفهرست عربي باستخدام مونوكونك.¹

أما في الجزء الثاني من التجربة، أي باستخدام "وورد سميث"، فقد وجد الباحثون أن هذا المفهرس يتعامل مع نصوص اليونيكود؛ بيد أنهم صادفوا المشكلة نفسها في قلب الصدر و العجز، ذلك أن وورد سميث – هو الآخر – قد صمم لمعالجة اللغات التي تكتب وفق الاتجاه (يسار - يمين). و للباحث أن يقرأ أولاً الجزء الأيسر من اليمين إلى اليسار (الصدر)، ثم يتابع قراءة الجزء الثاني و الذي يبدأ من العقدة يتلوها ما تبقى من السياق (العجز) –وفق ما يظهر في الشكل الموالي (24)–.



الشكل (24) واجهة بحثية لفهرست عربي باستخدام وورد سميث.²

¹ - Cf. Ibid.

² - Cf. Ibid.

بخلاف البرنامجيتين السابقتين، ليست *aConCorde* منتجا تجاريا، و لا هي بالمشروع البحثي، فقد طورها (و لا يزال) أندرو روبرتس، و هو طالب دكتوراه في أوقات فراغه. و قد تحرى قدر المستطاع إمكانية ملاءمتها للغات عدة، بما فيها العربية وكذا العبرية لأنها هي الأخرى تكتب من اليمين إلى اليسار، لذلك:

- لا حاجة للمستخدم في أن ينسخ المدونة العربية ضمن الأحرف الرومانية.
- صنع المفهرس باستخدام لغة البرمجة يافا (*Java*)، مما يجعله صالحا لجميع أنظمة التشغيل (ميكروسوفت ويندوز، لينكس *Linux*، ماك أوس *Mac Os*)، و لكثير من أنظمة التشفير بما في ذلك: اليونيكود (*UTF-8, UTF-16*)، ماك العربية *MacArabic*، ويندوز العربية *CP1256*، *IBM* العربية *CP420*، و غيرها.
- يسمح "إي-كونكورد" بتنزيل جميع أنواع الملفات مباشرة، على شاكلة: ملفات *txt*، *MS WORD*، *RTF*، *XML*، *HTML*.
- يمكن حفظ النتائج إما ضمن الصورة النصية (*Plain text format*)، أو ضمن صورة لغة الربط بين النصوص التشعبية *HTML*.
- يحلل تكرار الكلمات.
- يتضمن تشكيلية واسعة من خيارات البحث:

¹ *Key-word, phrase, proximity, Boolean, wildcard, Arabic stem/ root queries.*

و ثمة نقيبة أخرى ينفرد بها هذا المفهرس، من حيث إنه يتوفر على واجهتين بحثيتين، إحداها بالانجليزية و الأخرى بالعربية، و هي واجهة تتضمن ترجمة لكل القوائم (*Menus*) و الأزرار (*Buttons*)؛ بل – و فوق ذلك – تعرض جميع العناصر المكونة لها وفق الاتجاه (يمين - يسار).

ثم إن "إي-كونكورد" يتميز بسهولة الاستخدام و بساطة العرض، بمعنى أنه ما على الباحث إلا أن يختار المدونة لفتحها، ثم يقوم بكتابة الكلمة المطلوبة للبحث، أو ينتقيها من جدول تكرار الكلمات، فيحصل على نتائج دقيقة.

لقد سبق و رأينا أن استخدام طريقة الويلدكارد في الفهرسة باعتماد الجذع لا تتلاءم و اللغة العربية؛ و أن أصل الكلمات العربية متتالية ثلاثة حروف – في أغلبها الأعم* – تشتق منها كلمات أخرى. مثال ذلك: الجذع كتب (*ktb*)، الذي يمكن أن تشتق منه عدة جذور، نحو: كتب (*katab*)، و اكتتب (*iktatab*)؛ و انطلاقا من كل جذر يمكن اشتقاق مختلف الصيغ الصرفية بزيادة السوابق و الدواخل و اللواحق، نحو: تكتب (*taktub*)، يكتب (*yaktub*)، أكتب (*aktub*)، و هكذا دواليك. و لما كان استخلاص جميع مشتقات الجذع "كتب" باستخدام طريقة الويلدكارد غير ممكن، يوفر "إي-كونكورد" على المستخدم عناء البحث وفق الطريقة "التقليدية"، و يقدم له جميع الجذوع و الجذور ضمن قواعد معطيات صرفية هائلة.

¹ - Cf. *Ibid.*

* - و إن كانت ثمة كلمات رباعية و خماسية الأصل.



الشكل (25): واجهة بحثية باعتماد إي- كوناكورد.¹

و تلكم القواعد مأخوذة عن المولد أو المحلل الصرفي الذي صممه بكوالتر، و هو يتكون من ثلاثة أرصدة *Lexicons* (السوابق و اللواحق و الجذور)، و ثلاث سلاسل من القواعد (*Rules*) تحدد كيفية إحداث تقاطع بين الأرصدة الثلاثة لاستخراج جميع مشتقات الجذع الواحد. و يبدأ المولد بقراءة الملف و معالجة كل كلمة على حدى، ثم يقطع الوحدة اللغوية و يضع جميع صيغها الصرفية الممكنة مستعيناً بالأرصدة الثلاثة و القواعد.² و تشتغل قواعد المعطيات تلك، و كذا المحلل وفق نظام النسخ الذي أحدثه بكوالتر، و هو نظام قوامه رصف أو مقابلةً واحداً لواحد بين الألفبائية الرومانية و الأبجدية العربية. كما توظف فيه بعض الرموز و علامات الوقف لتمثيل الحروف العربية، نحو: & كمقابل لـ "ؤ"، و \$ كمقابل للشين.

و هكذا عمد الثلاثي إلى اقتراض تلك القواعد و تزويد "إي- كوناكورد" بها، بعد أن حولها من نظام بكوالتر في النسخ إلى نظام اليونيكود، فيعرض الجذع و الجذور في صورة اللغة العربية – كما يوضحه الشكل (26) أدناه.

¹ - Ibid.

² - Ibid.

Concordance		
Word List \ Arabic stems		
Stem	Morph	Gloss
☐ إندكالك	NduAt	crushing;leveling
☐ اندكالك	NduAt	crushing;leveling
☒ دكتور		
☐ دكتور	Nap	conferment of doc...
☒ دكتور	N/ap	doctor
☒ دكتور	Nap	doctors
☐ دكتورا	Napdu	doctorate
☐ دكتوراه	N	doctorate
☐ دكر		
☐ دكرينو	NO	decree
☐ دكرت	NAt	decrees
☐ دكن		
☐ دكان	Ndu	shop;store
Search		

الشكل (26): تزويد إي -كونكورد بقاعدة بكوالتز للجذور العربية.

و زبدة القول، إن الأبحاث العربية باعتماد المدونات اللغوية صرح قد رُفعت قواعده، و وُضِعت لبناته، و هو أحوج إلى من يكمل بناءه. و لسنا نحتاج في ذلك كله إلى سحر ساحر أو قدرة قادر، بل حري بنا أن ننظر فيما حققه الغرب في المجال، و نتأسى بهم و نتعلم من تجاربهم، و نستفيد من خبراتهم، فنجيب عن كثير من المسائل التي لا تزال عالقة في ميدان البحث العربي بعامة، و في ذلك فليتنافس المتنافسون، و ليتسابق المتنطسون.

الفصل الثالث:

المدونة في الترجمة تنظيرا و تطبيقا

I- الترجمات باعتماد المدونات: الأصول و الاتجاهات.

I-1. الشموليات الترجمة. I-2. أنواع المدونات المعتمدة في الترجمة.

II- المدونات و تدريب المترجم:

II-1. الترجمة و التدريب. II-2. تجربة الترجمة باعتماد المدونة.

II-2. 1. الفرضيات. II-2. 2. العينة المشاركة.

II-2. 3. العينة النصية. II-2. 4. الأدوات و الموارد المتاحة.

II-2. 5. التجربة. II-2. 6. نتائج الورشتين و تحليلها.

II-2. 7. تعاليق الطلبة و تساؤلاتهم.

II-2. 8. خلاصة التجربة.

"[Translation is a] communication event which is shaped by its own goals, pressures and context of production". Mona, Baker (1996:175).

"Corpus translation studies change in a qualitative as well as quantitative way both the content and the methods of the discipline of translation studies, in a way that fits with the modes of the information age." Maria, Tymoczko (1998:652).

إن المقاربة باعتماد المدونات هي مقاربة حديثة نسبياً في حقل الترجمات؛ إذ قد انتشرت منذ ما يزيد عن العقد من الزمن، و تعددت تطبيقاتها في المجال على ثلاثة مستويات: مستوى البحث و مستوى التعليمية و مستوى الممارسة المهنية. فعلى مستوى الممارسة المهنية، لا يدري أغلب المترجمين المتخصصين، إلام كانت ستؤول حالهم لولا توفر نظم الذاكرة الترجمة و التي – كما سبق و رأينا – تنبني على مدونة متوازنة من النصوص المصادر و ترجمتها. أما في ما يتعلق بمستويي البحث و التعليمية، فسنتناول كليهما بشيء من التفصيل فيما يلي.

I-الترجمات باعتماد المدونات: الأصول و الاتجاهات.

إن ما يعنى بالترجمات باعتماد المدونات CTS، هو مختلف أوجه أو صور تطبيق تقنيات تحليل المدونة، الكمية منها و النوعية على حد سوى، في دراسة الترجمة، سواءً تعلق الأمر بالعملية في حد ذاتها أو نتائجها.

"Corpus-based translation studies [are] the application of corpus analysis techniques, both quantitative and qualitative, to the study of aspects of the product and process of translation."¹

و الواقع أن الباحثين قد عكفوا على دراسة عملية الترجمة دراسة إمبريقية منذ أزيد من عشرين سنة، تأثراً بتطورات البحث الحاصلة في مجال تعليمية اللغات الأجنبية؛ فعمدوا إلى تحديد ما يجري أثناء الترجمة و وصفه و تحليله، بمعنى دراسة الخطوات الذهنية التي يتبعها المترجم بدءاً بمرحلة تلقي النص المصدر، و وصولاً إلى إنتاج النص المستهدف. و كان الاستبطان، ممثلاً في طريقة التفكير بصوت عال (Think-aloud method)، الأداة المنهجية الرئيسة في تقصي العملية.

و الملاحظ أن الدراسات الاستبطانية التي أنجزت حتئذ، كانت تعتمد على المعطيات و الوصف² (data-based and descriptive)، مركزة في كثير من الأحيان

¹ - Maeve, Olohan (2002a) Op. Cit., p.5.

² – و تلك ليست مفارقة، بل تؤكد تجاوز إشكال التعارض أو التقابل بين النزعتين العقلانية و التجريبية - كما سلف و أوضحنا.

على بعض أوجه العملية الترجمية (كاستعمال الأدوات المرجعية، ضوابط اتخاذ القرار، و نحو ذلك).

و هكذا بدأت المقاربة الوصفية التجريبية في درس العملية الترجمية بدايتها باعتماد طريقة التفكير بصوت عال؛ بيد أن كثيرا من المعوقات و الصعوبات المنهجية قد اعترضتها. و كنتيجة لذلك، لقيت هذه المحاولات تشكيكا من قبل بعض الأوساط، بخصوص المنهجية المتبعة في استنباط المعطيات و تجميعها، و كذا عجزها عن الوصول إلى العمليات الذهنية أثناء العمل الترجمي، و التي تكون ضمن الشعور الباطن أو آلية.¹

إن الإقرار بذلكم الضعف – إن جاز التعبير – الذي اعتور الدراسات الإمبيريقية زمانئذ، يعد واحدة من الحثثيات التي كانت وراء تسطير أشغال ملتقى "لوند" حول نظرية الترجمة خلال العام (1985) (Lund Symposium on Translation Theory)؛ حيث أجمع منظماه (وولن و ليندكويست) على أن الأوان قد آن لتسخير طاقة أكبر في إنجاز دراسات ترجمة حقيقية، أشمل و أدق تصميماء، و ذلك بغية التحقق من النظريات (أو الافتراضات النظرية) الجارية إما توكيدا أو تفنيذا.

"[The time had come] to devote more energy to carrying out carefully planned, large-scale investigations of authentic translation with the aim to verify and falsify existing theories and theoretical proposals."²

و جاء بيل (Bell) بالحل البديل، فاقترح أنه يمكن – بل ينبغي – التوصل إلى مقاربة تجمع بين الحُسْنَيْن: الاستقراء *induction* (أي استنباط العمليات من نتائج الترجمة)، و الاستنتاج *deduction* (أي توظيف المعطيات الاستنباطية على شاكلة اليوميات و المذكرات)؛ كما اقترح وصف "الكفاءة الترجمية في شكل تعميمات، قوامها ما تم استنباطه من ملاحظة لأداء المترجم." [الترجمة لنا]

"[describing] translation competence in terms of generalizations based on inferences drawn from the observation of translation performance".³

و ذهب إلى أن تلك الملاحظة تكون بتحليل نتائج الترجمة، و إيجاد خصوصيات تميظ اللثام عن عناصر خاصة و علائق نسقية ضمن العملية الترجمية.

تلك المقاربة التي جاء بها بيل تؤازر فكرة تجميع و استخدام مدونات، قوامها نصوص ترجمة *Corpora of translations*، مما يتيح تحليل خصوصيات نتائج الترجمة، سواء تلك التي تتم عن وعي أو ضمن الشعور الباطن، لاسيما إذا تمكنا من دراسة العلاقات بين ما هو تكراري و ما هو نموذجي، و بين المثال و المعيار – على حد تعبير Stubbs :-

¹ - Cf. M. Olohan (2002 a) *Op. Cit.*, p.5.

² - L. Wollin & H. Lindquist (eds.) (1985) *Translation Studies in Scandinavia*. Lund Studies in English 75. Lund: CWK Gleerup, p.10; qt. in: S. Granger (1996) *Op. Cit.*, p. 38.

³ - Cf. R. T. Bell (1991) *Translation and Translating : Theory and Practice*. London/New York: Longman, pp. 29-39.

"Invest[ing] relations between frequency and typicality, and instance and norm".¹

و الحق أن توظيف المدونات (باعتبارها مجموعات نصية) في درس الترجمة ليس مفهوما جديدا،² بل إن موطن الحداثة يكمن في وفرة التقنيات الحاسوبية و إتاحتها، مما أدى إلى تطور الترجمات باعتماد المدونات. ثم إن هذا الاتجاه تحصيل حاصل لتغير وجهة البحث ضمن المجال، و تحول في الاهتمام بالنص المصدر، و البحث عن التكافؤ و ما انجر عنه من ممارسة تداولية أو وظيفية أو تواصلية، إلى الاهتمام بالنص المستهدف و السعي لاستكشاف قوانين الترجمة – على حد تعبير فينوتي (Venuti):

"The literature on equivalence formulates linguistic and textual models and often prescribes a specific translation practice (pragmatic, functional, communicative). The target orientation, in contrast, focuses on actual translations and submits them to detailed description and orientation. It inspires research projects that involve substantial corpora of translated texts".³

و لقد ظفرت منى بايكر (Mona Baker) بقصب السبق و الريادة في هذا الاتجاه الجديد، و يتجلى ذلك بخاصة من خلال سلسلة من أعمالها (1993، 1995، 1996)،⁴ تقترح فيها عددا من فرص البحث المتعلقة بإمكانية تحليل كم هائل من المعطيات اللسانية؛ كما نلمس في كثير منها امتعاضها، و هي تتحو باللائمة على الدارسين اعتمادهم على طرق الاستبطان في أبحاثهم. و قد أشارت في واحدة من تنبؤاتها (1993: 243) – التي ما لبثت طويلا أن صدقت – إلى أن تاريخ الدراسات الترجمةية يشهد منعطفًا حاسما تصنعه وفرة مدونات كبيرة من الأصل و الترجمات، و ما يصحبها من تطور في المنهج الذي يعتمد على استخلاص المعطيات من المدونات (Corpus-driven methodology)، قد يتيح للدارسين في المجال استكشاف طبيعة النص الترجمة كحدث تواصلية وسيط.

¹ - Micheal, Stubbs (2001) Texts, Corpora and Problems of Interpretation: A Response to Widdowson. *Applied Linguistics* 22 (2), p. 151.

² - Cf. eg.: G. Toury (1980) *In search of a theory of translation*. Tel Aviv: The Porter Institute for Poetics and Semiotics, Tel Aviv University.

W. Wilss (1982) *The Science of translation. Problems and methods*. Tübingen: Gunter Narr Verlag.

M. Gellerstam (1986) Translationese in Swedish novels translated from English. In: L. Wollin & H. Lindquist (eds.) *Op. Cit.*, pp. 88-95.

³ - L. Venuti (2000) (ed.) *The Translation Studies Reader*. London/New York: Routledge, p.123; qt. in: S. Granger (2003) *Op. Cit.*, p. 18.

⁴ - Cf. M. Baker (1993) Corpus linguistics and translation studies: Implications and applications. In: M. Baker, G. Francis & E. Tognini-Bonelli (eds.) *Text and Technology: in honour of John Sinclair*. Amsterdam/Philadelphia: Benjamins, pp. 233-252.

- (1995a) Corpora in translation studies: An overview and some suggestions for future research. *Target* 7 (2), pp. 223-243.

- (1996) Corpus-based translation studies: The challenges that lie ahead. In: H. Somers (ed.) *Terminology, LSP and translation studies in language engineering: in honour of Juan C. Sager*. Amsterdam/Philadelphia: Benjamins, pp. 175-186.

"[to explore] the nature of translated text as a mediated communicative event".

و مذاك، تزايد الاهتمام بالمقاربة باعتماد المدونات في مجال الترجمات، على اعتبار أنها ذات شأن تتناول بالدرس الترجمة كنتاج و كعملية على حد سوى، وفق أحدث الطرق وأنجعها.

و انطلاقا من نتائج بعض الدراسات السابقة(نحو: فراولي 1984، جيلرستام 1986، بايكر 1993، توري 1995)،¹ يمكن القول إن وضع الترجمة تحت مجهر البحث يعد واحدة من أهم المرتكزات التي تنبني عليها الترجمات باعتماد المدونات، باعتبار أن لغة الترجمات تختلف لزما عن لغة الأصول، و أن بعض الخصوصيات التي تتسم بها تلك اللغة، تنطبق على الترجمات جميعها، بغض الطرف عن لغتي النص المصدر والنص المستهدف؛ و هي خصوصيات قد استرعت اهتمام كثير من الباحثين، و صنفوها ضمن الشموليات الترجمية (Translation Universals)، نذكر منها: الإضافة، الإظهار عن طريق ضرب الأمثلة، تبسيط النص المصدر، أو تطبيقه، و ما إلى ذلك من الظواهر التي يمكن اكتشافها من خلال مقارنة نص مصدر بترجماتة العديدة.²

1.1- الشموليات الترجمية:

إن مفهوم الشموليات في الترجمة – هو الآخر – ليس بالمفهوم الجديد؛ إذ يؤرخ ظهوره بظهور الفكرة القائلة إن لغة الترجمة تختلف عن اللغة المعهودة، و هي مستقلة قائمة بذاتها – وفق ما يذهب إليه فراولي (1984) –، حين يقترح فيما يقترح مقابلة النص المصدر باللغة المستهدفة أثناء عملية الترجمة، فيحصل لدينا ما يدعوه "باللغة الثالثة" (Third Code)؛ بمعنى أن اللغة التي تتطور أثناء الترجمة، و التي يتزيا بزيتها النص المستهدف، هي لغة فريدة، من حيث إنها نتاج نوع من التسوية أو التوفيق بين معايير أو بني اللغة المصدر، و تلك التي تتسم بها اللغة المستهدفة. كما سبق و وضع إيفين زهير³ (1978) ثلاثة عشر نوعا منها، تتعلق كلها بتحليل الأنظمة الأدبية و الترجمة الأدبية. بيد أن مثل تلك التأكيدات لم تكن تستند إلا إلى قلة قليلة من البيئة النصية (Textual evidence). لذلك كان غنتسلر قد أشار إلى "ضرورة إجراء تحليل مكثف للنص و العلاقات الثقافية قبلا:

¹ - Cf. William, Frawley (1984) Prolegomenon to theory of translation. In: W. Frawley (ed.) *Translation. Literary, linguistic and philosophical perspectives*. London/ Toronto: Associated University Presses, pp. 159-175.

M. Gellerstam (1986) *Op. Cit.*

M. Baker (1993) *Op. Cit.*

G. Toury (1995) *Descriptive translation studies and beyond*. Amsterdam / Philadelphia: John Benjamins.

² - لمزيد من الاستيضاح و التوضيح، نحيل القارئ على مؤلف هام:

S. Laviosa (2002) *Corpus-based translation studies. Theory, findings, applications*. Amsterdam / New York: Rodopi.

³ - Cf. I. Even-Zohar (1978) The position of translated literature within the literary system. In: James S. Holmes, J. Lambert & R. Van den Broeck (eds.) *Literature and translation: New perspectives in literary studies*. Leuven: Acco.

"A more extensive analysis of textual and cultural relations must take place before 'Universals' can be persuasively posited".⁽¹⁾

و هو المطمح نفسه الذي ترمي إلى تحقيقه الترجمات باعتماد المدونات من خلال تحليل مادة نصية هائلة.

و الشموليات في المجال المشار إليه، هي تلكم الخصوصيات اللسانية التي تتوارد على نحو نموذجي في النصوص الترجمة بخلاف الملفوظات الأصلية، فتتفرد بها وتصير تعرف من خلالها؛ و هي ليست نتاج التداخل بين الأنظمة اللسانية الخاصة، وفق ما تذهب إليه بايكر، أي أنها لا تعزى إلى تأثير النص/ اللغة المصدر.

"[Translation Universals are] features that typically occur in translated text rather than original utterances and which are not the result of interference from specific language systems."⁽²⁾

و لا يعتقد أن تلكم الشموليات هي نفسها المعايير في الترجمة (Norms)؛ إنما وجه الفرق بينهما أن الأولى نتاج إكراهات تلازم العملية الترجمة في حد ذاتها، و هو ما يضيف عليها طابع الشمولية. أما المعايير فهي جملة الخصوصيات التي يلحظ أنها تتوارد باستمرار في بعض أنواع الترجمة، ضمن سياق سوسيوثقافي و تاريخي معين.⁽³⁾ و قد توصل الباحثون إلى عدد معتبر من تلكم الشموليات التي تنطبق على جميع الترجمات، لاسيما بعد تحليلها و نصوصها المصدر تحليلًا تقابليًا؛ نذكر منها: التبسيط وتلافي التكرارات في المصدر، الإظهار، و التطبيق.

أ- التبسيط و تلافي التكرارات في النص المصدر:

يتم التبسيط (Simplification) في الترجمة على مستويات ثلاثة، هي: المستوى المعجمي و المستوى التركيبي و المستوى الأسلوبي. فالتبسيط المعجمي يقصد به ميل المترجم إلى توظيف أقل عدد ممكن من الكلمات.

"[Simplification is] the process and/or result of making do with less words."⁽⁴⁾

- و ذلك وفق ستة مبادئ أو استراتيجيات، ترتبط أساسا بكفاءته الدلالية في لغته الأم، هي:
- استخدام المصطلحات الفوقية (Superordinate terms) في حال غياب المشمولات المكافئة (Equivalent Hyponyms) في اللغة المستهدفة.
 - تقريب المفاهيم الواردة في اللغة المصدر.
 - استخدام المفردات الاعتيادية الشائعة و المألوفة.

(1) - E. Gentzler (1993) Contemporary translation theories. London / New York : Routledge, p. 122.

(2) - M. Baker (1993) Op.Cit., p. 243 ; qt. in: S. Granger (1996) Op. Cit., p. 48.

(3) -Cf. M. Baker (1993) Op. Cit., p.246; qt. in: Linn, Øverås (1998) In search of the third code: An investigation of norms in literary translation. Meta 43(4), p. 589.

(4) - Shoshana, Blum- Kulka & Eddie Levenston (1983) Universals of lexical simplification. In C. Faerch and G. Kasper (eds) Strategies in Interlanguage Communication, pp. 119-139. London and New York: Longman.

- تحويل جميع وظائف الكلمة (في اللغة المصدر) إلى ما يكافئها في اللغة المستهدفة.
 - اللجوء إلى الإسهاب/الإطناب (*Circumlocutions*) بدل التعابير أو الكلمات الغامضة (لا سيما فيما يتعلق بالمصطلحات اللاهوتية، أو التقنية أو الخاصة بثقافة معينة).
 - إعادة الصياغة/التحويل (*Paraphrase*) في حال العثور على فجوات ثقافية بين اللغتين المصدر و المستهدفة.⁽¹⁾
- أما في ما يخص التبسيط التركيبي، فثمة أمثلة عديدة عن تبسيط التراكيب المعقدة، و ذلك باستبدال المركبات اللامنتهية (*non-finite clauses*) بأخرى منتهية، بحيث تحذف عادة النقاط المتتابعة (...)*.
- أما التبسيط الأسلوبي، فيتخذ أشكالا عدة، أكثرها شيوعا كسر سلاسل الجمل الطويلة، و ذلك باستبدال التعابير الجاهزة المعقدة بمتلازمات لفظية قصيرة، التقليل من التكرارات و صور الإطناب أو حذفها، و ما إلى ذلك.⁽²⁾
- و قد عثرت شلسينجر *Shlesinger* (1991) في سياق الترجمة الفورية على مستوى المحاكم، على عدة أمثلة تبين أن التكرارات الواردة في النص المصدر كثيرا ما تحذف في النص المستهدف.⁽³⁾
- أما توري في مجال الترجمة الأدبية، فيؤكد هو الآخر على أن نزوع المترجم إلى تفادي التكرارات التي يلفاها في النص المصدر، ظاهرة حاضرة في جميع الترجمات على اختلاف اللغات المدروسة.⁽⁴⁾

ب- الإظهار (*Explicitation*):

يعنى بالإظهار في الترجمة التصريح بما كان مضمرا في النص المصدر؛ و هو ظاهرة حظيت بموفق العناية من قبل الدارسين باعتبارها خاصة من خواص النص الترجمة، فتوصلوا إلى مختلف الأساليب و التقنيات التي يتبعها المترجم في ذلك، نحو: إضافة جمل شارحة، رفع اللبس المتضمن في النص المصدر، توظيف التكرارات و ما شاكلها من أدوات الاتساق... و تلكم ظواهر حاضرة في النصوص الترجمة دون المصادر أو الأصول في اللغة ذاتها.⁽⁵⁾

⁽¹⁾-Cf. Sara, Laviosa-Braithwaite (1998) Universals of translation. In: Mona, Baker (ed.) *Routledge Encyclopaedia of Translation Studies*. London / New York : Routledge, pp. 288-291.

* أي النقاط الثلاث التي يؤتى بها لتدل على كلام محذوف من نص، لا يريد الكاتب ذكره اختصاراً.

⁽²⁾ -Cf. S. Laviosa-Braithwaite (1998) *Op. Cit.*, pp. 288-289.

⁽³⁾ - نحيل القارئ على مقال هام للباحثة:

Miriam, Shlesinger (1998) Corpus-based interpreting studies as an offshoot of corpus-based translation studies. *Meta* 43(4), pp. 486- 493.

⁽⁴⁾ - *Ibid.*

⁽⁵⁾ - نحيل القارئ على ثلاثة مقالات في الموضوع:

M. Olohan (2002a) *Op. Cit.*

M. Olohan (2002b) Leave it out! Using a comparable corpus to Investigate Aspects of Explicitation in Translation. In: S.E.O., Tagnin (Org.) *Op. Cit.*, pp. 153-169.

Linn, Øverås (1998) *Op. Cit.*, *Meta* 43 (4), pp. 571-588.

و مصطلح "الإظهار" من صنع فيني و داربلني، يحيل على تقنية أسلوبية في الترجمة، يتم من خلالها إظهار ما كان مضمرا في اللغة الأصل، و هو بَيِّن من خلال السياق أو الموقف.

"[Explicitation is] a stylistic translation technique which consists of making explicit in the target language what remains implicit in the source language because it is apparent from either the context or the situation."⁽¹⁾

و في دراسة لها عن الترجمات من اللغة الإنجليزية إلى اللغة الفرنسية و العكس، تشير بلوم كولكا (Blum-Kulka) إلى حدوث انزياحات في أنواع أدوات الاتساق (Cohesion markers) على مستوى النصوص المستهدفة، و تضرب أمثلة يقوم فيها المترجم بـ"تمديد" النص المستهدف عن طريق إدراج كلمات إضافية.

و تسهم هاتان الظاهرتان في رفع مستوى الإبانة ضمن النص المستهدف مقارنة بالنص المصدر. و تؤكد الباحثة على أن هاتين الخاصيتين الترجميتين لا تتعلقان بثنائية لغوية معينة، بل إنها نتاج عملية تأويل النص المصدر:

"The process of interpretation performed by the translator on the source text might lead to a TL text which is more redundant than the SL text. [This redundancy] can be expressed by a rise in the level of cohesive explicitness in the TL text."⁽²⁾

و بناءً على هذه الدراسة، قامت فانديروفييرا (Vanderauwera) بتحديد العديد من الأمثلة التي يطبق فيها المترجم تقنيات الإظهار، أهمها: توظيف صيغ التعجب للتعبير بوضوح عن تطور أفكار الشخصيات، أو للتشديد على تأويل معين، تمديد المقاطع النصية الموجزة، إضافة التوابع (من بدل و نعت) و كذا الوصل تقصيا لشفافية أكبر، إقحام معلومات إضافية، إدراج شروحات، تكرار ما سبق ذكره من تفاصيل توخيا للوضوح، نقل المعطيات المضمرة أو الغامضة نقلا واضحا مفصلا، الوصف بشكل أدق، إبانة أسماء المواقع الجغرافية و رفع اللبس الناجم عن استخدام الضمائر.⁽³⁾

و في السياق نفسه، تورد بايكر أمثلة عديدة عن مواقف يقوم فيها المترجم بإقحام معلومات خلفية ضمن النص المستهدف، قصد سد فجوة ثقافية ما يلفاها بين النظامين المصدر و المستهدف.⁽⁴⁾

أما الانزياحات في الاتساق، و التي تتخذ أشكالا عدة منها توظيف المرادفات أو أن يقوم التكرار مقام الاستبدال و الحذف، فهي حاضرة في الترجمة الفورية المتزامنة

⁽¹⁾-J.P. Vinay & J. Darbelnet 1995 [1958] comparative stylistics of French and English. A methodology for translation. Amsterdam / Philadelphia: John Benjamins Publishing Company, p.342.

⁽²⁾-S. Blum-Kulka (1986) Shifts of cohesion and Coherence in Translation. In: J. House & S. Blum-Kulka (eds.) *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies*. Tübingen: Gunter Narr, pp. 19-21.

⁽³⁾-Cf. Ria, Vanderauwera (1985) Dutch Novels Translated into English: The Transformation of a 'Minority' Literature. Amsterdam: Rodopi.

⁽⁴⁾-Cf. M. Baker (1992) *In Other Words- A coursebook on Translation*. London / New York: Routledge.

(Simultaneous)، من اللغة العبرية إلى اللغة الإنجليزية، و من الإنجليزية إلى العبرية – على حد سوى –، حسب دراستين لشلسينجر التي تؤكد على أن الفرضيات المتعلقة بظاهرة الإظهار تنطبق على الترجمة التحريرية منها و الشفوية؛ و عليه فإن الترجمان يميل – هو الآخر – إلى نقل المضمرات و إظهارها على نحو أدق، بغض النظر عن اللغتين المصدر و المستهدفة.⁽¹⁾

ج. التطبيع (Normalisation):

لقد كشفت لسانيات المدونة عن كثير من الأسرار اللغوية، لعلّ أهمها أن أغلب ما نكتبه أو نتفوه به هو من قبيل الروتيني (أي النمطي المتكرر) المنتظر و يمكن التنبؤ به، بحكم أنه قد سبق إنتاجه نطقاً أو كتابة، على حد تعبير هانكز:

"The creative potential of language is 'undeniable', but the concordances to a corpus remind us forcibly that in most of our utterances we are creatures of habit, immensely predictable, rehearsing the same old platitudes and the same old clichés in almost everything we say."⁽²⁾

و قولبة الترجمات على نحو يعكس جلياً روتينية اللغة المستهدفة و رتابتها، أي مواضعاتها، هي ما يعرف لدى الدارسين بالتطبيع.

و التطبيع – بتعبير أدق – يُعنى به نزوع المترجم إلى إبراز خصوصيات اللغة المستهدفة، نزوعاً قد يبلغ حد المغالاة؛ و ذلك قصد مطابقة النص الترجمة للقوالب النموذجية و الاستعمالات المعهودة التي تنفرد بها تلك اللغة، لاسيما في ما يتعلق بالبنية النحوية و علامات الترقيم و المتلازمات اللفظية أو الكلشيهات – وفق ما تذهب إليه بايكر:-

"[Normalisation is] a tendency to exaggerate features of the target language and to conform to its typical patterns. [It] is most evident in the use of typical grammatical structures, punctuation and collocational patterns or clichés."⁽³⁾

و كثيرة هي تلكم الدراسات التي تناولت الظاهرة المذكورة على المستوى المعجمي بخاصة، أهمها – على سبيل الذكر – دراسة فانديروفيرا؛ حيث عمدت الباحثة إلى مدونة قوامها 50 رواية مترجمة من الهولندية إلى اللغة الإنجليزية، و عثرت فيها على قدر هائل من أمثلة الانزياحات على مستوى علامات الترقيم، الخيار المعجمي، الأسلوب، بنية الجملة و الهيكل النصي بعامة، و هي انزياحات تعدّها الباحثة من قبيل الميل عموماً إلى التقيد بالمواضع النصية موضع اصطلاح الجمهور المستهدف؛⁽⁴⁾ من أمثلتها: بعض

⁽¹⁾-Cf. M. Shlesinger (1989b) Simultaneous interpretation as a factor in effecting shifts in the position of texts on the oral-literate continuum. Unpublished MA thesis, Tel Aviv : Tel Aviv University.

-(1995) Stranger in paradigm: What Lies Ahead for simultaneance Interpreting Research? In: Daniel Giel (ed) Target 7 (1), special issue: Interpreting Research, pp 7-29.

⁽²⁾-P. Hanks (1996) Contextual Dependency and Lexical Sets. *International Journal of Corpus Linguistics* 1 (1), pp. 75-98.

⁽³⁾-M. Baker (1996) *Op. Cit.*, p. 183; qt. in: Dorothy, Kenny (1998) Creatures of Habit? What Translators Usually Do With Words. *Meta* 43 (4), p. 516.

⁽⁴⁾-Cf. R. Vanderauwera (1985) *Op. Cit.*

التعديلات على المستوى المعجمي تشمل تكييفاً للأسماء الهولندية و الإحالات الثقافية الخاصة، و كذا التقليل إلى حد أدنى من نقل تعابير اللغة الأجنبية الواردة في النص المصدر، و تنمة الجمل غير المنتهية في النص المذكور، أما ما كان من البنى شخصياً، أو يفقر إلى السبك بُدّل بتركييب بسيطة. كما استُبدلت صيغة الحاضر أو التاريخي بصيغة الماضي، و هي الصيغة الأكثر شيوعاً في اللغة الإنجليزية المكتوبة، لاسيما الروائية منها. هذا فضلاً على ترتيب الجمل و الفقرات و الفصول ترتيباً يتوافق و منطق اللغة الإنجليزية، و ما إلى ذلك.

إن تلك الأمثلة في مجملها و غيرها كثير،⁽¹⁾ هي ما يجعل من النص المستهدف يتسم بأكبر قدر من الاتساق و الطبيعية و المقروئية، و هي ما يصنع منه نصاً مأنوساً مألوفاً لدى أهل اللغة المترجم إليها.

أما في مجال الترجمة الفورية على مستوى المحاكم، فقد قامت شلسينجر بتحليل ترجمات من اللغة العبرية إلى الإنجليزية، توصلت من خلالها إلى أشكال مختلفة من التطبيع، على شاكلة: تنمة الجملة غير المنتهية، استبدال الملفوظات المصدر اللانحوية بأخرى نحوية في اللغة المستهدفة، و هلم جرأً.⁽²⁾

و لا يخالّن القارئ أن جميع الدراسات باعتماد المدونات في مجال الترجمات مكرسة للشموليات، أو أنها تهدف إلى تقييم الترجمات؛ بل إن جلها – في الواقع – يتمحور حول استكشاف الظواهر المتعلقة باللغة الثالثة، و مقارنتها إما بالنصوص الأصل في اللغة المصدر، أو بالنصوص الأصل في اللغة المستهدفة. يضاف إلى ذلك، السعي إلى إدراك الإكراهات و الضغوط، و كذا البواعث، التي تؤثر على العملية الترجمية من جهة، و تشكل الأساس الذي تنبني عليه لغتها الفريدة.⁽³⁾

و على ضوء ما سلف، يتجلى لنا أن النصوص الترجمة هي محط الدراسة و مدار العناية في حقل الترجمات باعتماد المدونات، باعتبارها أمثلة عن الاستخدام اللغوي، قائمة بذاتها شأنها شأن النصوص المصدر. و تلك العناية هي على نقيض النظرة التي تعتبر الترجمات استعمالاً لغوياً شاذاً فلا يقاس عليه؛ و هي نظرة نجم عنها في الأغلب الأعم إقصاء لتلك النصوص من معظم المدونات المرجعية على حد قول منى بایکر.

"Le texte traduit a toujours été traité en parent pauvre de la linguistique de corpus. Il a été exclu des corpus monolingues où il est généralement considéré comme non représentatif de la langue étudiée, quelque soit le sens de la traduction." ⁽⁴⁾

و لعل أشهر المدونات التي تمّ بناؤها لحد الساعة باعتماد نصوص ترجمة هي "تيك" (TEC) مدونة اللغة الإنجليزية المترجمة Translational English Corpus

⁽¹⁾- Cf. S. Laviosa-Braithwaite (1998) Op. Cit.

⁽²⁾-Cf. M. Shlesinger (1991) Interpreter latitude vs. Due process: Simultaneous and consecutive interpretation in multilingual trials. In: Sonja Tirkkonen-Condit (ed.) *Empirical research in translation and intercultural studies : Selected papers of the TRANS-SIF seminar, Savonlinna 1988*, Tübingen : Gunter Narr, pp. 147-55.

⁽³⁾-Cf. M. Baker (1998) Réexplorer la langue de la traduction: Une approche par corpus. *Meta* 43 (4), 480-485.

⁽⁴⁾-Ibid., p. 484.

بمركز الترجمات (مانشستر)؛ و هي مدونة مكتوبة بحجم حوالي 7 ملايين كلمة، تتألف من ترجمات معاصرة إلى اللغة الإنجليزية عن عدة لغات، و قد صُممت بخاصة لغرض دراسة النصوص المترجمة، التي تنقسم إلى أربعة أنواع: الخيال، السّير و التراجم، المقالات الصحفية، و المجالات. و الترجمات من إنجاز ناطقين أصليين باللغة الإنجليزية، تمتد فترتها التاريخية من العام 1983. و قد زوّد كل نص ببيانات عن مترجمه و عملية الترجمة، تم جمعها عن طريق استبيان وُجّه للمترجمين و الناشرين؛ و هي بيانات محفوظة ضمن رأسية كل ملف.⁽¹⁾

و زبدة القول، إن الاهتمام بالمدونات في حقل الترجمة ماثل للناظرين من وجهين أو جانبين: الجانب الوصفي و الجانب التطبيقي؛ فقد عمد الدارسون إلى تصميم مدونات قوامها النصوص المترجمة، و تحليلها و مقارنتها بنصوص أصل (أي مكتوبة أصلا في لغة الترجمة نفسها)، و ذلك بغية تحديد السمات التي تتفرد بها الترجمات، و كذا المسؤوليات التي تميز النصوص المترجمة.⁽²⁾ أما ما كان من الجانب التطبيقي، فقد تزايد الاهتمام بالمدونات باعتبارها مساعدات هامة تخدم عمليات الترجمة، سواء كانت حاسوبية أم بشرية. و الحال هذه، ثمة صنفان رئيسان من تلكم المدونات، هي مدار رحي البحث في حقل الترجمة، و تعكس جليا جملة المعايير التي صممت وفقها، و الغاية التي تقبع وراء إنشائها، نستعرضها بشيء من التفصيل فيما يلي.

2.I- أنواع المدونات المعتمدة في الترجمة:

1.2.I- مدونات أحادية اللغة: (*Monolingual corpora*)

و تتضمن نصوصا ضمن لغة واحدة (أ)، و هي إما أن تكون اللغة المصدر، أي أن تكون النصوص كتبت أصلا في اللغة (أ)، و هاته تعرف بالمدونة الأصلية؛ أو أن تكون اللغة المستهدفة، بمعنى أن النصوص عبارة عن ترجمات إلى اللغة (أ) من لغة (ب) أو عدة لغات (ب، ب، ب)، و هاته تعرف بالمدونة المترجمة. و هذه المدونات إما أن تكون عامة *General monolingual corpora*، مادتها تشكيلة من الأنواع النصية المختلفة؛ أو أن تكون مدونات أحادية اللغة مختصة (*Specialized*)، تنتمي مادتها النصية إلى جنس كتابي معين و/أو موضوع محدد.

و ثمة ما يعرف بالمدونة المقارنة أحادية اللغة متعددة المصادر-*Multi-source language monolingual 'comparable' corpus*، و هي مدونة قوامها مجموعتان نصيتان: إحداهما مكتوبة أصلا في لغة (أ)، و أخراهما تتضمن نصوصا نفسها مترجمة إلى اللغة (أ) انطلاقا من عدة لغات مختلفة (ب، ب، ب)؛ هذا مع مراعاة شروط المقارنة، من حيث انتماء المجموعتين النصيتين إلى المجال نفسه، و كذا التنوع اللغوي و الزمني، و تماثل الحجمين، حسب تعريف منى بايكر:

“[...] two separate collections of texts in the same language: one corpus consists of original texts in the language

⁽¹⁾-Cf. Maeve, Olohan (2002a) *Op. Cit.*, p. 6.

⁽²⁾-Cf. M. Baker (1998) *Op. Cit.*

S. Laviosa (1998a) Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta* 43(4), pp. 557-570.

in question and the other consists of translations in that language from a given source language or languages... both corpora should cover a similar domain, variety of language and time span, and be of comparable length”¹.

و قد اعتمدت الباحثة على هذه المدونة بالذات في دراسة الشموليات الترجمية، منها أن الترجمات تنزع كثيراً إلى الإيضاح أو الإظهار على مستويات عدة، مقارنة بالنصوص الأصل، و أنها تعتمد إلى التبسيط و التطبيع و ما إلى ذلك وفق طرق شتى – كما سلف وبيّنا –².

و قيمة مثل هذه المدونة هي نظرية محض، من حيث إن ما يسعى الباحث إلى دراسته باعتمادها هو الطبيعة اللسانية للنص الترجمة، بغض النظر عن اللغة المصدر؛ مثال ذلك: المدونة التي استخدمتها سارة لافيوزا (Sara Laviosa) في دراسة الطبيعة اللسانية للنصوص الانجليزية المترجمة، و هي مدونة فرعية عن المدونة المقارنة في اللغة الانجليزية³ (ECC) *English Comparable Corpus*، تتضمن مجموعتين من النصوص النثرية القصصية: أولاهما عبارة عن ترجمات إلى اللغة الانجليزية من عدة لغات مصدر (الألمانية، الفرنسية، الإسبانية، العربية...)، و أخراهما نصوص أصل حرّرت بالانجليزية خلال الفترة الزمنية نفسها.⁴

2.2.I - مدونات ثنائية/متعددة اللغة:

تتراتب فيها اللغات مثنى مثنى (أو ثلاث، أو رباع،...) على أساس "المقارنة" و/أو "التوازي". فالمدونات (الثنائية) المقارنة تتضمن نصوصاً في اللغتين (أ) و (ب)، تشترك في نفس المعايير من حيث التركيب و الجنس الكتابي و الموضوع، دونما أن تكون إحداها ترجمة للأخرى؛⁵ بمعنى أن نصوص المدونتين الفرعيتين قد جمعت وفق معايير الانتقاء ذاتها، و هي تتماثل من حيث خاصيتي التوازن و التمثيل، كأن تتضمن كل مدونة فرعية نفس النسبة من النصوص ضمن لغتين أو أكثر، تنتمي إلى نفس الأجناس الكتابية و المجالات، و تحدد في إطار فترة زمنية معينة.

على ضوء هذا التعريف، لا يعتقَدُ – بأي حال – أن المدونات التي تتركب من مدونات فرعية تتضمن لهجات عن اللغة (أ) (مثل ICE) هي مدونة مقارنة، مثل ما يذهب إليه إيفان حين يقول:

“Two or more corpora constructed along similar parameters but each containing a different language or a

¹ - Mon, Baker (1995) *Op. Cit.*, p. 234; qt. in: M. Olohan (2002b) *Op.Cit.*, p.154.

² - Cf. M. Baker (1993) *Op.Cit.*

³ - المدونة المقارنة للغة الانجليزية تتكون من مجموعتين نصيتين باللغة الانجليزية: إحداها تتضمن ترجمات من عدة لغات، وتدعوها لافيوزا بالمدونة الانجليزية الترجمة (TEC)، وأخراهما قوامها نصوص انجليزية أصل (NON-TEC)، تنتمي إلى نفس الجنس الكتابي للترجمات، وكذا الفترة الزمنية ذاتها. والمدونة تتضمن جنسين كتابيين: النثر القصصي و الصحافة.

⁴ - Cf. S. Laviosa (1998a) *Op.Cit.*

⁵ - Cf. A. McEnery & Z. Xiao (2007) *Parallel and comparable corpora: What are they up to?* In: Gunilla M. Anderman & Margaret Rogers (eds.) *Incorporating corpora: Translation and the linguist*. Translating Europe. Multilingual Matters, Clevedon, UK, p. 20.

different variety of the same language can be regarded as comparable corpora”¹ [Emphasis added].

"إن مدونتين أو أكثر متمثلتين من حيث مقاييس التصميم، بيد أن كل واحدة منهما تتضمن لغة مختلفة أو لهجة مختلفة ضمن اللغة الواحدة، يمكن اعتبار مجموعهما مدونة مقارنة". [الترجمة لنا]

كما أن محمدي داهشاشمه تعتقد – هي الأخرى – بأن مدونة ICE مدونة مقارنة تتضمن مختلف اللهجات الانجليزية.² وفي اعتقادنا إنما ذلك الخلط ناتج عن تأويل مغلوطة لتعريف بايكر السالف الذكر:

*“[...] both corpora should cover a similar domain, variety of language and time span[...].”*³ [Emphasis added].

فما تعنيه الباحثة – هنا – أنه إذا كانت المدونة الفرعية الأولى (هي مدونة مقارنة) تتكون من نصوص أصل في اللغة الانجليزية البريطانية، وجب أن تكون نصوص المدونة الفرعية الثانية (الترجمات) في اللغة نفسها. ولا تقصد ما ذهب إليه أولئك من أن مدونة قوامها مدونة فرعية باللغة الانجليزية البريطانية، وأخرى فرعية باللغة الانجليزية الأمريكية، هي مدونة مقارنة.

و ذلك من منطلق أن جميع المدونات باعتبارها مصدرا للبحث اللساني، كانت ولا تزال الأنسب في الدراسات المقارنة سواء على مستوى اللغة الواحدة (Intra-lingual) أو بين اللغات (Inter-lingual)، فمدونات براون و لوب و فراون و فلوب – مثلا – قد صممت كلها لمقارنة اللهجات الانجليزية سانكرونيو و دياكرونيو؛ وكذا هي الحال بالنسبة للمدونة البريطانية الوطنية التي صممت بغرض تمثيل اللغة الانجليزية البريطانية المعاصرة، علاوة على أنها بمثابة الأساس لعدة دراسات مقارنة ضمن اللغة نفسها (مثل دراسة: المنطوق/المكتوب، المونولوج/الحوار، وكذا دراسة التغيرات الناجمة عن الضوابط السوسيواقتصادية)؛ بيد أنه لا يمكن اعتبارها مدونات مقارنة.⁴

أما المدونات المتوازية، فتتضمن نصوصا مصدرا مرفقة بترجمات. والواقع أن مفهوم التوازي – على النحو الذي ذكرناه – مائل للعيان عند كثير من الحضارات القديمة، ولعل أشهر الأمثلة عن ذلك: حجر الرشيد (Rosetta Stone)، الذي يعود تاريخه إلى العام 196 قبل الميلاد، وهو يتضمن نصين متوازيين باللغتين اليونانية و المصرية، ضمن ثلاث كتابات (اليونانية و الهيروغليفية و الديموطيقية بالنسبة للغة المصرية).⁵ أما في الوقت الحاضر، فكثيرة هي تلك النصوص المتوازية من عقود و معاهدات و ما شاكلها، و التي تنشأ نتيجة الاحتكاك أو التواصل بين اللغات. غير أن الاهتمام بها لم يكن بالقدر الذي تعرفه حاليا إلا في سنوات الثمانينات في إطار المعالجة الآلية للغات.

¹ - D. Evans (2006) Op.Cit.

² - Cf. M. Mohammadi Dehcheshmeh (2007) Op.Cit.

³ - M. Baker (1995) Op. Cit., p. 234.

⁴ -Cf. A. McEnery & Z. Xiao (2007) Op.Cit., p.20.

⁵ - و لقد اكتشف الحجر سنة 1799 بمدينة الرشيد على ضفة النيل، إثر حملة نابليون على مصر، و هو محفوظ حاليا بالمتحف البريطاني. وهو الحجر ذاته الذي مكن شامبليون Jean-François Champollion من فك طلاسم الخط الهيروغليفي عام 1822.

Cf. Jean, Véronis (2000) Alignement de corpus multilingues. In : Jean-Marie Pierrel (ed.) *Ingénierie des Langues*, chapitre 6, pp.151-171.1^{er} éd., Paris : Edition Hermès.

اللغة (أ)	اللغة (ب)
مدونة مقارنة.	(أ) مدونة مختصة.* (ب) مدونة مختصة ذات التصميم نفسه.
مدونة متوازنة أحادية الاتجاه.	(أ) مدونة مختصة (ب) ترجمات نصوص اللغة (أ).
مدونة متوازنة ثنائية الاتجاه.	(أ) مدونة مختصة. (ب) (1) مدونة مختصة ذات تصميم يماثل تصميم (أ). (2) ترجمات (أ).

الجدول (12) المدونات المقارنة و المدونات المتوازنة.

و جدير بالذكر أنه لما كانت المدونات التي تتضمن لغتين فأكثر ظاهرة حديثة نسبيا، إذ ظهرت أوائل التسعينات مع المدونة المتوازنة الانجليزية النرويجية (ENPC)؛¹ فلا عجب أن نجد بعض الخلط المصطلحي بخصوصها، و لا ضير في أن نفردها له بضعة أسطر نعد من خلالها إلى تبيان تلك الفوضى أو ذلك الإشكال في المصطلحات، والاحتكام إلى الحل الأمثل منها.

فإذا اعتبرنا أن المدونات التي تتضمن أكثر من لغة واحدة ثلاثة أنواع:

* النوع (أ): نصوص مصدر + ترجماتها على غرار: *Canadian Hansard* (مدونة فرنسية – إنجليزية عن البرلمان الكندي)، *CRATER*.

* النوع (ب): مدونات فرعية أحادية اللغة مصممة وفق المعايير نفسها على شاكلة: *The Aarhus corpus of contract law*.

* النوع (ج): و هو مزيج بين النوعين (أ) و (ب)، مثل: (ENPC) المدونة المتوازنة الانجليزية النرويجية، و *EMLLE*.

فإننا نجد أن بعض الباحثين على غرار (Aijmer & Altenberg (1996)، يعتبرون النوع الأول (أ) مدونة ترجمة و النوع (ب) مدونة متوازنة.

أما بالنسبة لـ (McEnery – Baker (1999, 1995, 248:1993), Wilson (57:1996), Hunston (15:2002)، النوع (أ) مدونة متوازنة و النوع (ب) مدونة مقارنة؛ أما (Johansson–Hofland (1994) و Johansson (4:1998) فمصطلح المدونة المتوازنة يطلق على النوعين (أ) و (ب).²

¹ - Cf.: S. Johansson & K. Hofland (1994) Towards an English-Norwegian parallel corpus. In: U. Fries, G. Tottie & P. Schneider (eds.) *Creating and using English language corpora*. Amsterdam: Rodopi, pp.25-37.

² - Cf. K. Aijmer, B. Altenberg & M. Johansson (1996) *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies*. Lund 4-5 March 1994. Lund: Lund University Press, pp.73-85.

A. McEnery & A. Wilson (1996) *Corpus Linguistics*. 1st ed., Edinburgh: Edinburgh University Press.

M. Baker (1993,1995) *Op.Cit.*

(1999) The Role of Corpora in Investigating the Linguistic Behavior of Professional Translators. *International Journal of Corpus Linguistics* 4 (2), pp. 281-298.

إن ذلك التباين في تعيين أنواع المدونات، يحصل بين المشتغلين في حقل اللسانيات التقابلية و نظرائهم في حقل الترجمات؛ إذ يقسم الفريق الأول المدونات إلى نوعين:

- مدونات ترجمة (Translation Corpora): تتكون من نصوص أصل في لغة (أ) مرفقة بترجمات في لغة (ب) أو لغات أخرى.
 - مدونات مقارنة: تتكون من نصوص أصل في لغتين أو أكثر، يرتبط بعضها ببعض بمعايير من قبيل: نوع النص، الجمهور المستهدف، الفترة الزمنية، الخ.
- و يوظف بعض اللسانيين مصطلح المدونة "المتوازية" للدلالة على المدونة المقارنة أو المدونة الترجمة، أو مجموعهما.¹

أما الفريق الثاني (أي المشتغلون في حقل الترجمات)، فيقسم المدونات أقساما ثلاثة هي:

- المدونات المقارنة.
 - و المدونات المتوازية (و هذان القسمان قد مرا بنا).
 - المدونات الترجمة: و يعنى بها المدونات التي تتكون من نصوص ترجمة فقط.
- بمعنى أن ما يدعى في اللسانيات التقابلية بمدونة ترجمة هو المدونة المتوازية في الترجمات. بناء على ذلك، يلحظ أن الإشكال المشار إليه يتمحور أساسا حول مصطلح "التوازي"، و حله يتوقف على المعايير المعتمدة في تعيين مختلف أنواع المدونات، نحو: عدد اللغات المتضمنة و محتوى المدونة أو شكلها.

و بديهي أنه عند وقوع الاختيار على معيار دون آخر، يجب إتباع ذلك المعيار في التعيين باستمرار، مثلا: يمكن القول إن هذه المدونة أحادية/ثنائية/متعددة اللغة، إذا ما أخذنا بعين الاعتبار عدد اللغات المتضمنة كمعيار للتعين. و يمكن القول أيضا إن أي مدونة هي ترجمة (ل2)، إذا أخذنا بعين الاعتبار معيار المحتوى. لكن إذا اخترنا تعيين أنواع المدونة انطلاقا من معيار الشكل، ينبغي أن يبقى هو الآخر ثابتا؛ و هكذا يمكن القول إن المدونة متوازية، إذا كانت تتضمن نصوصا مصدرا مرفقة بترجمات، أو إنها مقارنة إذا كانت تتضمن مدونتين فرعيتين مقارنة، من حيث إنهما تحتويان على عينات متماثلة.

و على ضوء هذه التوضيحات، من غير المعقول أن تدعى المدونة (أ) مدونة ترجمة بناء على معيار المحتوى، بينما تدعى المدونة (ب) مدونة مقارنة بناء على معيار الشكل. و الحال هذه، نرى أن ما ذهب إليه كل من ماكجري و بايكر أصوب،² حيث تُعَيَّن المدونة (أ) بمدونة متوازية و (ب) بمدونة مقارنة بناء على معيار الشكل؛ و بما أن المدونة (ج) مزيج بينهما، فإنها مدونة مقارنة خاصة.

S. Hunston (2002) Corpora in applied linguistics. Cambridge: Cambridge University Press.

S. Johansson & K. Hofland (1994) Op.Cit.

S. Johansson (1998) On the role of corpora in cross-linguistic research. In: S. Johansson & S. Oksefjell (eds.) Corpora and cross-linguistic research: Theory, Method, and Case Studies. Amsterdam: Rodopi, pp. 3-25.

¹ - Cf. S. Granger (2003) Op.Cit., pp.19-20.

² - Cf. A. McEnery & Z. Xiao (2007) Op.Cit.

و فضلا على التباين المصطلحي، ثمة تباين آخر بين الفريقين المشار إليهما من حيث نظرة كل واحد منهما إلى النصوص الترجمة؛¹ ففي الوقت الذي يراها المترجمون نصوصا قائمة بذاتها، تحلل بغية إدراك ماهية الترجمة و التعرف إلى مسارها - وفق بايكر-:

“[To] understand what translation is and how it works”²

يرى اللسانيون أنه لا يمكن اعتماد تلك النصوص في عملهم بحجة أن أوجه الالتلاف والاختلاف بينها يمكن أن تكون نتاج التداخل بين النصوص المصدر و المستهدفة. و المدونات المتوازية و المقارنة إما أن تكون مختصة، بمعنى أنها تحوي نصوصا تنتمي إلى نوع معين، نحو كتيبات استعمال جهاز ما؛ و إما أنها عامة، أي متوازنة تتضمن عدة أنواع نصية. و استثمار إحداها دون الأخرى يتوقف أساسا على طبيعة البحث و الغرض منه؛ فإذا كان الأمر يتعلق بالبحث المصطلحي،³ الأجدر بالباحث أن يستغل مدونات متوازية و مقارنة مختصة. أما إذا كان غرض البحث إحداث التقابل بين الخصائص اللسانية العامة (كالأفعال و الأزمنة مثلا)، يكتفي الباحث بمدونة متوازنة طالما أنه يفترض أنها ممثلة للغة عموما.

و المدونات المتوازية المتوفرة حاليا هي مدونات مختصة في أغلبها الأعم، نظرا لوفرة النصوص الترجمة و إتاحتها في كل المجالات و ضمن مختلف اللغات تقريبا. و كذا هي الحال بالنسبة للمدونات المقارنة، بيد أنه ليس من الهين إيجاد أنواع نصية مقارنة في لغات مختلفة بقدر وفرتها في شكل نصوص متوازية، لذلك يُنصح بتصميم مدونات مقارنة متوازنة.⁴

و ثمة نقطة جوهرية لا يجب إغفالها، لأنها هي الأخرى تصنع الفرق بين المدونتين المتوازية و المقارنة، تتعلق بعملية تصميمهما. ففي حالة المدونة المقارنة، يحرص المصمم أيما حرص على أن تكون مركباتها متماثلة من حيث نسبة النصوص المدرجة و الجنس الكتابي و المجال و الفترة الزمنية. بخلاف ذلك، يعني مصمم المدونة المتوازية أشد العناية بأن تكون مركباتها ترجمات دقيقة بعضها لبعض؛ إذ ما إن ينتقي النصوص المصدر وفق معايير محددة، ليس من الضروري أن يعتمد إلى تطبيق تلك المعايير مرة أخرى في انتقاء الترجمات.⁵

و لا يخالّن القارئ بهذا أن صناعة مدونة متوازية أسهل، لا تتعدى تجميع نصوص مصادر و ترجماتها؛ بل إنها تتطلب معالجة خاصة، تتمثل في عملية التنظير، و هي من الأهمية بمكان مما يفرض علينا أن نقف عندها فيما يلي.

¹ - Cf. S. Granger (2003) *Op.Cit.*, pp.19-20.

² - Cf. M. Baker (1993) *Op.Cit.*, p.243

³ - نشير بهذا الصدد إلى أنه إذا كان استخلاص الرصيد/المعجم ثنائي اللغة من المدونات المتوازية قد أضحى تقليدا كلاسيكيا في مجال استخلاص المصطلحات، ونشاطا قد تجاوز ميدان البحث إلى الميدان التجاري؛ فإن استخلاص الرصيد/المعجم ثنائي اللغة من المدونات المتقابلة حديث جدا، و يبقى حكرا على ميدان البحث دون سواه. ينظر بهذا الصدد:

Hervé, Déjean & Éric Gaussier (2003) Une nouvelle approche à l'extraction de lexiques bilingues à partir de corpus comparables. In: J. Véronis (ed.) *Revue électronique Lexicometrica* (2002), numéro thématique : *Alignement lexical dans les corpus multilingues*.

⁴ - Cf. A. McEnery & Z. Xiao (2007) *Op.cit.*, pp. 20-21.

⁵ - Cf. *Ibid.*, p. 20.

3.2.I- تنظير المدونات المتوازية:

تخضع الترجمات إلى مبدأ التركيبية (Compositionalité)، بحيث إن ترجمة كل مقطع مركب منوطة بترجمة أجزائه. و يحيل هذا المبدأ على بديهية جوهريّة، نصّها: إذا كان لدينا نصان تربط بينهما علاقة الترجمة، فإننا يمكن أن نلاحظ مطابقات بين وحداتهما. و تاليا، يعطينا المبدأ المذكور نظاما من المطابقات السلمية بين النصين المصدر و المستهدف، تتراتب من الوحدات النصية الأكبر (أقسام، فقرات، إلخ) إلى وحدات نصية أصغر فأصغر (الجملة، فالمركبات، فالكلمات، و حتى المورفيمات) —حسب الشكل (30) أدناه.¹

النص المصدر	النص المستهدف
القسم I	القسم I
الفقرة I	الفقرة I
الجملة I	الجملة I
المقطع I	المقطع I
الكلمة I	الكلمة I
...	...
الكلمة س	الكلمة س
...	...
المقطع و	المقطع و
...	...
الجملة ل	الجملة ل
...	...
الفقرة ي	الفقرة ي
...	...
القسم ن	القسم ن

الشكل (30) المطابقات السلمية بين نص مصدر و ترجمته.²

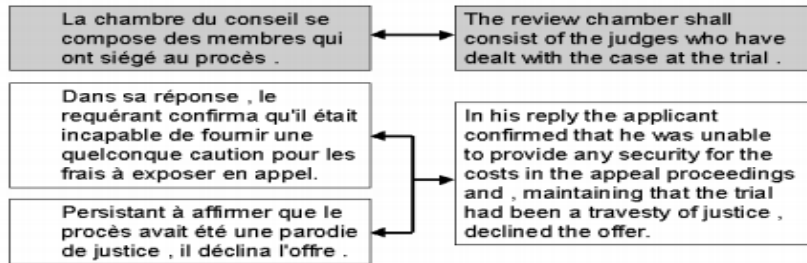
بيد أنه – و كما هو معلوم – يمكن أن نجد مطابقات بين وحدات تنتمي إلى أقسام مختلفة، مثال ذلك: الوحدة "بطاطا" تنتمي إلى القسم (كلمة)، تطابق الوحدة الفرنسية *pomme de terre* التي تنتمي إلى القسم (مركب).
كما ينبغي الحذر في بعض التراكييب الانجليزية و الفرنسية، على غرار النفي:
Max has not seen Conrad. → Max n'a pas vu Conrad.

¹ - Cf. Pierre, Isabelle (1992) *Op.Cit.*, pp. 724-726.

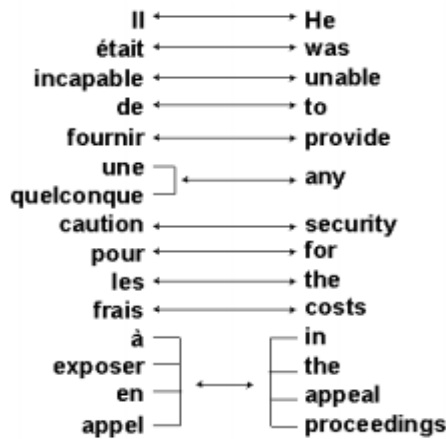
² - Cf. *Ibid.*, p.725.

و لهذا يجدر التشطير على مستوى البنى التركيبية العميقة، أي على مستوى البنى الدلالية.¹ و الربط على هذا النحو بين مختلف الوحدات النصية في النصوص المصدر وما يطابقها من وحدات في الترجمات، يعرف بالتنظير. أما تلك الوحدات، فتشكل "الوحدات الترجمية"، و هي إما أن تكون كلمات أو مقاطع أو جملا أو فقرات - حسب الشكلين (31 أ و ب) أدناه -.

إن تنظير المدونات المتوازية على مستوى وحدة ترجمة دون أخرى، يتوقف أساسا على البحث و كذا نوع المدونة؛ فإذا كانت المدونة تقنية أو قانونية - على سبيل المثال- فإنها تفرض تنظيرا دقيقا بين مصراعيها (المصادر و الترجمات) يتخذ من الجملة أو الكلمة وحدة ترجمة. و في المقابل، إذا لم تكن المدونتان الفرعيتان إحداهما ترجمة حرفية للأخرى، كما هي الحال في النصوص الأدبية مثلا، وجب تنظيرهما على مستوى أكبر، نحو الفقرات أو الفصول.²



الشكل (31-أ) التنظير على مستوى الجمل.



الشكل (31-ب) التنظير على مستوى الكلمات و المقاطع.³

¹ - Cf. *Ibid.*

² - Cf. Mathieu, Guidère (2002) *Op.Cit.*
M. Zimina-Poirot (2004) *Op.cit.*

³ - المثالان مستنبطان من مدونة " الاتفاقية " *Corpus Convention*، ينظر:

و لما كان التشطير التقليدي (أي اليدوي) إلى مطابقات عملية شاقة بالنظر إلى حجم المدونات الهائل، عمد الباحثون إلى الاستعانة ببرمجيات خاصة تخدم الغرض، واستحدثوا شتى المقاربات لتيسير ذلك،¹ و هي تندرج في إطار ما يعرف بالتنظير الآلي *L'alignement automatique* في مجال المعالجة الآلية للغات الطبيعية.

و لقد أظهرت الدراسات في مجال التنظير أن استخلاص المطابقات آليا على قدر كبير من السهولة، لاسيما فيما يتعلق بالوحدات الترجمية الكبرى، على غرار: الفصول والأقسام و الفقرات؛ حيث إن التسلسل المنطقي لتلك الوحدات، أو الخصوصيات الطباعية (العناوين و الترقيم/التعداد، العودة إلى السطر،...) تسهم كلها بحظ وافر في تيسير العملية. كما أن توظيف طرق الاحتمال قد حسن كثيرا من نتائج التنظير الآلي على مستوى الجمل بنسبة قدرها 98.5%.²

و لتنظير المدونات المتوازية أهمية كبيرة، بحيث إنه يشكل موردا هاما لإنشاء الذاكرة الترجمية و تحسين برمجياتها، و استنباط المعارف من أجل البحث التوثيقي المتعدد اللغات، هذا ناهيك عن التزويد بالأمثلة التي تخدم التعليم أو التدريس المحوسب (*Enseignement assisté par ordinateur/ EAO*) أو اللسانيات التقابلية، و كذا إنشاء مدونات تدريب تُظم رفع اللبس آليا، و تطوير محركات البحث على مستوى الشبكة. أضف إلى ذلك، استخلاص القوائم المصطلحية ثنائية اللغة، و إثراء المعاجم باستعمالات و تعابير جديدة، و قس على ذلك من التطبيقات.³ و المثال أدناه، يوضح أهمية استخدام مفهرس آلي في استخلاص المكافئ الترجمي للمصطلح القانوني الفرنسي (*Faits incriminés*) انطلاقا من مدونة متوازنة، مقارنة بما تم الحصول عليه من نتيجة باستخدام مترجم آلي (بابل فيش *BabelFish*).

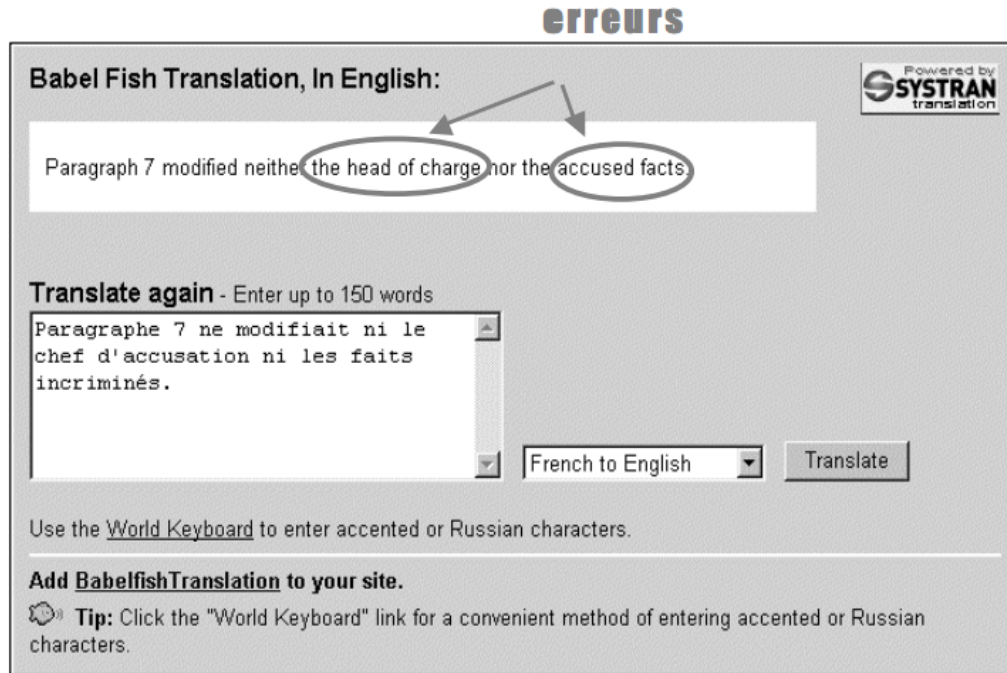
Rechercher Rechercher: faits incriminés Suivant Fermer	
L'accusé a droit à être informé non seulement des <u>faits incriminés</u> mais aussi de leur qualification juridique .	
Paragraph 7 ne modifiait ni le chef d'accusation ni les <u>faits incriminés</u> .	
Les paragraphes qui suivent se rapportent au droit tel qu'il se présentait à l'époque des <u>faits incriminés</u> .	
A person charged with a criminal offence was entitled to be informed not only of the <u>material facts alleged against him</u> but also of their legal classification .	
Paragraph 7 had altered neither the charge nor the <u>facts constituting the alleged offences</u> .	
The following relates to the law as it stood at the time of the <u>events complained of</u> .	

الشكل (27) فهرست آلي ثنائي اللغة.

¹ - La première méthode d'alignement automatique a été développée par Martin Kay (1984). Voir: M. Kay & M. Röscheisen (1988) Text-translation alignment .Technical Report. Xerox Palo AHO Research Center.

² - Cf. J. Véronis & Ph. Langlais (2000) Evaluation of parallel text alignment Systems: ARCADE. In: J. Véronis (ed.) *Parallel Text Processing*. Dordrecht: Kluwer.

³ - Cf. *Ibid*.



الشكل (28) محاولة الترجمة باعتماد (بابل فيش).¹

من خلال الشكلين، نلاحظ أن النتائج المحصل عليها باعتماد المفهرس الآلي أفضل بكثير، من حيث إن الترجمة أجود و أصوب مقارنة بالترجمة المتحصل عليها انطلاقا من (بابل فيش)، و هي ترجمة و إن كانت مقبولة تركيبيا، إلا أن مكافئي المصطلحين (*Le chef d'accusation*) و (*Les faits incriminés*) غير ملائمين. و لقد رأينا فيما تقدم* أن المفهرسات الآلية أحادية اللغة تتيح للباحث الحصول على الكلمة المفتاحية بجميع تواتراتها و سياقاتها التي وردت فيها. أما في حال المفهرسات الثنائية اللغة (أو المتعددة اللغات)، فإنه بإمكان الباحث استعراض جميع سياقات الكلمة، فضلا على تحديد ترجمتها الأنسب ضمن السياقات ذاتها، و تلك تعرف بنظم بيكويك² (Bilingual Key Word In Context) BIKWIC.

M. Zimina-Poirot (2004) *Op.Cit.*, pp. 27-29.

¹ - المثال مأخوذ عن رسالة "بوارو"، ينظر:

* - ينظر العنصر IV من الفصل الأول.

² - ثمة عدد معتبر من برمجيات الفهرسة الآلية المتعددة اللغات، نذكر منها:

TransSearch ; <http://www.tsrali.com>

Multiconcord; <http://info.ox.ac.uk/ctitext/resguide/resources/m145.html>

ParaConc; <http://www.ruf.rice.edu/~barlow/parac.html>

4.2.I- استخدامات المدونات المتوازية و المقارنة:

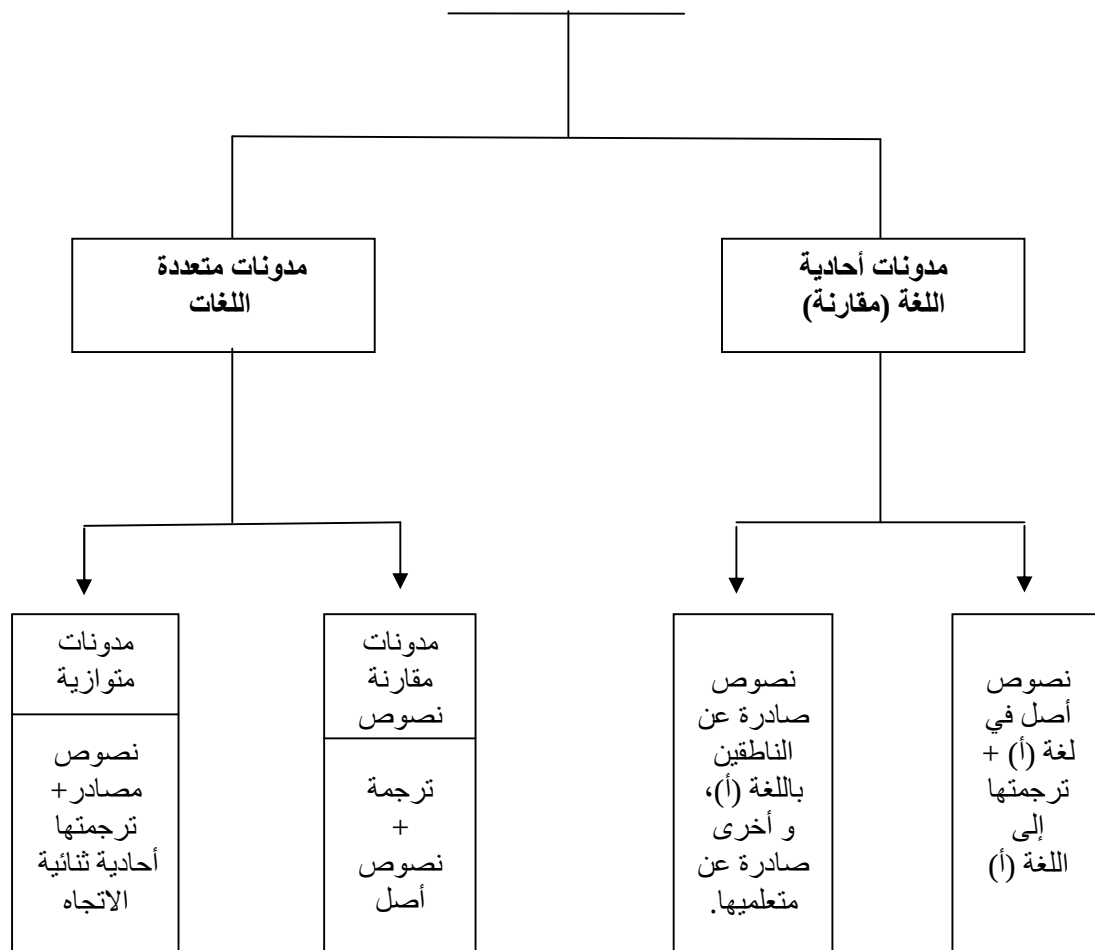
تستخدم المدونات المتوازية و المقارنة في مجالي الدراسات الترجمية و التقابلية، بحيث إنها مدونات:

- تعطي رؤى جديدة بخصوص اللغتين/اللغات المقارنة، و هي رؤى قلما تُلاحظ في الدراسات باعتماد مدونات أحادية اللغة.
 - تستخدم في كثير من الأغراض المقارناتية، و تزيد معرفتنا باللغة، لاسيما ما تتسم به من اختلافات نوعية و أنموذجية و ثقافية، هذا فضلا على إدراكنا للخصوصيات الكلية.
 - يمكن توظيفها في عدد من التطبيقات العملية، نحو: المعجماتية، تعليمية اللغات و الترجمة.
 - توضيح أوجه الاختلاف بين النصوص المصدر و ترجماتها، و بين ما كان منها صادرا عن الناطق باللغة و ما كان بخلافه¹ [(non-)native texts].
- و الحال هذه، فإن الترجمات باعتماد المدونات تهتم أيضا "بالاختلاف" الذي كان ينظر إليه دائما نظرة سلبية في تاريخ الترجمة، سواء على المستوى النظري أو التطبيقي، لأنه حَيِّدٌ عن "الأمانة"، و أمانة "الفقد" (Loss) الذي تتسم به العملية الترجمية.²
- و المخطط أدناه يلخص أنواع المدونات المقارنة منها و المتوازية الموظفة في البحث البيلساني. أما الجدول رقم (11)، فيوضح أهم أوجه المقارنة أو أنواعها باعتماد المدونات المذكورة آنفا، مع تحديد الحقل الذي ينتمي إليه كل نوع (أي الترجمة أو اللسانيات التقابلية).

¹ - Cf. K. Aijmer, B. Altenberg & M. Johansson (eds.) *Op.cit.*, p.12; qt. in: A. McEnergy & Z. Xiao (2007) *Op.Cit.*, p.18.

² - Cf. M. Tymoczko (1998) *Op. Cit.*, pp. 6-7.

المدونات ضمن البحث البيلساني



الشكل (29) أنواع المدونات في البحث البيلساني.

الحقل	نوع المدونة	نوع المقارنة
اللسانيات التقابلية	مدونة مقارنة متعددة اللغات من النصوص الأصل	اللغة الأصل (أ) ⇔ اللغة الأصل (ب)
اللسانيات التقابلية والترجمات	مدونة ترجمة/متوازية متعددة اللغات	اللغة المصدر (أ) ⇔ اللغة المستهدفة (ب)
الترجمات و اللسانيات التقابلية	مدونة مقارنة أحادية اللغة من النصوص الأصل والنصوص الترجمة	اللغة المصدر (أ) ⇔ اللغة المستهدفة (أ)
الترجمات	مدونة مقارنة متعددة اللغات من النصوص الترجمة	اللغة المستهدفة (أ) ⇔ اللغة المستهدفة (ب)

الجدول (11) أوجه المقارنة باعتماد المدونات في البحث البيلساني.¹

نلاحظ أن النوع الأول من المقارنة قائم بين مدونات نصية أصلية (أي ليست ترجمات) في لغتين مختلفتين (أ و ب)، و تختص بها اللسانيات التقابلية.* أما النوع الثاني من المقارنة، أي على مستوى المدونات المتوازية، فهو القاسم المشترك بين الحقلين، مع اختلاف في غرض الاستخدام: الكشف عن أوجه الائتلاف و الاختلاف بين لغتين/عدة لغات بالنسبة لللسانيات التقابلية، و اقتناص الخصوصيات المانزة التي تتسم بها الترجمة (سواء العملية في حد ذاتها أو نتائجها) بالنسبة للترجمات. أما النوع الثالث، المقارنة بين نصوص أصل في لغة (أ) و أخرى ترجمات إلى اللغة نفسها، فيستأثر به الباحثون في مجال الترجمات، و ذلك لاستخلاص خصوصيات النصوص الترجمة (الشموليات مثلا)؛ بيد أن هذا النوع أيضا قد صار محط اهتمام اللسانيين الذين يرون الفروق بين اللغتين الأصل و الترجمة دليلا على الفروق بين اللغتين موضوع البحث.

أما المقارنة بين نصوص ترجمة في لغتين/لغات مختلفة، فيعنى بها المشتغلون في حقل الترجمات؛ غير أنه ينبغي على اللسانيين أن يلتفتوا – هم أيضا – إلى هذا النوع من المقارنة، و أن يغيروا من نظرهم إلى النصوص الترجمة، و إلا فإنه سينتهي بهم المقام – لا محالة – إلى اعتبار بعض الفروقات بين اللغتين الأصل و الترجمة نتاج التداخل أو تأثير اللغة الأصل، بينما هي في الواقع ظاهرة من ظواهر الشموليات الترجمة.²

¹ - Cf. S. Granger (2003) Op.Cit., p.21.

* - بيد أننا نشهد اهتمام الباحثين في الترجمات بهذا النوع في الأونة الأخيرة.

² - Cf. S. Granger (2003)Op.Cit., pp. 21-22.

ما ينبغي التأكيد عليه أيضاً في هذا المقام، هو أن لكل نوع من أنواع المدونات المقارنة و المتوازية محاسن و عيوب. و لعل أكبر مزية في المدونات المتوازية أنها الأمثل للنظر في كيفية نقل المعنى أو التعبير عن المحتوى/المضمون نفسه في لغتين أو أكثر، فهي بذلك أفيد في الترجمة بحثاً و تعليمية، و إن كان ثمة من الباحثين من يعتبر النصوص الترجمة نسخاً مشوهة عن نصوص "حقيقية" – على حد تعبير منى بايكر –:

"[Translated texts are regarded as] distorted versions of 'real' texts."¹

لكن هذا الإشكال لا ينطرح إذا راعينا مسألة نوعية الترجمة أثناء بناء المدونة. أما في مجال الدراسات التقابلية، فهي لا تفيد إذا كانت أحادية الاتجاه، ذلك أن الترجمات ليست بمنأى عن آثار الضيم (Translationese)، أي تأثير اللغة المصدر على الترجمة، مما ينعكس – بلا شك – سلباً على تعليمية اللغة الثانية سواء بالنسبة لأساتذتها أو متعلميها. لذلك يرى الدارسون أن استخدام مدونة متوازية ثنائية الاتجاه كفيلاً بحل الإشكال،² هذا مع وجوب مراعاة نفس معايير انتقاء المادة النصية المصدر في اللغتين كإتاهما، ذلك أن أي تباين في النسبة أو الجنس الكتابي أو المجال – على سبيل المثال – يحول دون الحصول على نتائج مقبولة باعتماد تلك المدونة.

و تثير Malmkjaer إشكالا آخر في حال استخدام مدونة متوازية، قوامها نسخة واحدة فقط من الترجمة في اللغة المستهدفة، باعتبار أنها لا تمثل إلا استعمالاً لغوياً فردياً. و لتجاوز هذا الإشكال، ينبغي إدراج ما أمكن من نسخ ترجمة لكل نص مصدر.³ و كما سلف و ذكرنا، قد أضحت لغة الترجمة محط اهتمام التقابليين، باعتبار أنها شكل من أشكال استخدام اللغات الطبيعية قيد الاستعمال، بغض الطرف عما تنسم به من خصوصيات، على حد تعبير مورانين (Mauranen):

"[Translated language] is part of natural language in use, and should be treated accordingly. [So, translations] deserve to be investigated in their own right."⁴

و دعوى ذلك الاهتمام، أن نتائج البحث المتحصل عليها – في بادئ الأمر – انطلاقاً من الترجمات في مدونات متوازية (أحادية الاتجاه)، تخدم الدراسات التقابلية، من حيث إنها تدفع إلى البحث قُدماً في مدونات أحادية اللغة ضمن اللغتين؛⁵ و الحال هذه، يحبذ استخدام مدونة متوازية ثنائية الاتجاه منذ البدء، ربها للوقت و توفيراً للجهد.

¹ - M. Baker (1993) *Op.Cit.*, p. 233; qt. in: S. Granger (1996) *Op.Cit.*, p.38.

² - Cf. E. Teich (2002) System-oriented and text-oriented comparative linguistic research: Cross-linguistic variation in translation. *Languages in Contrast* 2(2), pp. 187-210.

- Cf. e.g. B. Maia (1998) Word order and the first person singular in Portuguese and English. *Meta* 43(4), pp. 589-601.

Cf. also: J. Ebeling (1998) Contrastive linguistics, translation, and parallel corpora. *Meta* 43(4), pp. 602-615.

³ - Cf. K. Malmkjaer (1998) Love thy neighbour: will parallel corpora endear linguists to translators? *Meta* 43(4), pp. 534-541.

⁴ - A. Mauranen (2002) Will 'translationese' ruin a contrastive study? *Languages in Contrast* 2(2), p.166; qt. in: A. McEnery & Z. Xiao (2007) *Op.Cit.*, p. 23.

⁵ Cf. *Ibid.*, p. 182.

أما البحث في كلٍّ من الترجمة المحوسبة (Machine aided Translation) MAT و الترجمة الحاسوبية (MT)، فقد استفاد كثيرا من المدونات المقارنة؛ بيد أن الفائدة كانت أعظم حين وُظفت المدونات المتوازية خاصة فيما يتعلق بنُظم الترجمة الآلية باعتماد الأمثلة (EBMT Systems).¹

تلك هي إذن المقاربة باعتماد المدونات في مجال الترجمات، و التي نرى أنها قد حققت إلى حد بعيد أمل وولن و لندكويست في رؤية الدراسات الترجمةية تنحو منحى غير الذي عرفته أول عهدها: من دراسات تتسم بالسطحية مختزلة في العمل التنظيري البحث، إلى حقل بحث يعتمد برمته على التجريب:

"[The] move away from their initial stage of impressionistic, non-professional or purely theoretical work to become a more empirically based research field."²

و لا عجب في أنها مقاربة لا تنفك تتطور من خلال التأسيس النظري و التحقيق التجريبي، إلى نسق متآلف ثري و متسق، طالما أنه يتعاطى تشكيلة واسعة من المسائل على اختلاف انتماءاتها الترجمةية، من نظرية و وصف و ممارسة. و علاوة على ذلك، يعد استثمار المدونات في الترجمات بكل أشكاله استثمارا على المدى البعيد – على حد تعبير ماريا تيموكزكو (Maria Tymoczko) :-

"The development of corpora and CTS methods represents a long-term investment for the field of Translation Studies."³

و عليه، حريّ بالدارسين التبصر في مختلف أنواع المدونات التي يمكن أن تخدم أبحاثهم خدمة جليلة مستقبلا؛ بل الأحرى بهم النظر في إمكانات استخدامها من حيث التساؤلات التي تجيبهم عنها، و الطرق و المناهج في ذلك، سواء على المستوى التنظيري المحض أو التطبيقي العملي.

إن تلك الاستشرافات و الآمال المعقودة على المدونات، هي ما يجعل – اليوم – من المقاربة باعتمادها مدار الرحي في الترجمات، و إن كانت ثمة أصوات (على رأسها تيموكزكو) قد رفعت عقيرتها بالتحذير من خطر الانسياق وراء البحث الكمي، الذي يعد الأداة لا المادة، و الوسيلة لا المطلب؛ هذا ناهيك عن خطر الوقوع في المطبات التي تكتنف السعي إلى بلوغ مدارك "العلوم"، كأن يعمد الباحث إلى استثمار المدونات في إثبات ما لا يحتاج لإثبات، لأنه واضح للعيان وضوح الشمس لمستجليها، بل الأجدر به معالجة قضايا حقيقة بالبحث و النظر باعتماد المدونات – وفق ما تذهب إليه الباحثة ذاتها:-

"Researchers must take care to ask 'the right questions': to poses questions and construct research programs that have as their goal substantive investigations that are worthy of the powerful means deployed by CTS."⁴

¹ - Cf. A. McEnergy & Z. Xiao (2007) Op. Cit., p. 21.

² - L. Wollin & H. Lindquist (eds.) (1986) Op.Cit., p. 11.

³ - Maria, Tymoczko (1998) Op. Cit.

⁴ - Ibid.

إن هذه الدعوة إلى توخي الحذر من اتخاذ الصرامة العلمية غاية في حد ذاتها عن طريق دراسات كمية لا طائل من وراءها و لا ضرورة، هي دعوة تؤيد كثيرا ما تذهب إليه منى بايكر من وجوب الفصل بين أهداف الدراسات الترجمية، و تلك التي قد سطرته لسانيات المدونة؛ بحيث إن وصف الخصوصيات اللسانية وصفا مفصلا ليس الغاية، بل الوسيلة لإدراك الضغوطات و الإكراهات التي تشكل جزء لا يتجزأ من حياة المترجم العملية، و آثارها - بلا شك - بادية في ترجمته.¹

و إن كانت تلك حقيقة يجب أن توضع في الحسبان، ثمة حقيقة أخرى لا ينتطح فيها عنزان، مفادها أن الترجمات باعتماد المدونات، تعد واسطة العقد بين النظري والبراغماتي في الدراسات الترجمية، في وقت بات انفصالهما وشيكا و الفرجة بينهما واسعة.² و لعل ما يحدد في بواكير الدراسات الترجمية باعتماد المدونات، أن ما جاءت به و غدا من قبيل النظري، تكاد آثاره العملية و التطبيقية لا تخفى سواء في الممارسة الترجمية الاحترافية أو التعليمية.

II- المدونات و تدريب المترجم:

لقد رأينا في المحطات أعلاه أنواع المدونات و استخداماتها في البحث الترجمي، أي في الدراسات الترجمية النظرية. أما في هذه المحطة، فنستعرض أهم صور استخدام تلك المدونات في واحدة من ضروب الدراسات الترجمية التطبيقية: تدريب المترجم، محاولين قدر المستطاع الإجابة عن سؤالين نراهما من الأهمية بمكان:

- لماذا تدريب المترجم باعتماد المدونات؟

- كيف ندرجه؟

لكن قبل هذا و ذاك، نرى أن لا بد من استهلال الحديث عن مفهومي "الترجمة" و "التدريب" على اعتبار أنهما نقطتان لا تقلان أهمية عن سابقتيهما، هذا لعله أولى؛ أما أخراهما، فإن الحديث عن التدريب في الترجمة يستدعي لزما الحديث عن ماهية الترجمة بالدرجة الأولى.

1-II - الترجمة و التدريب:

ما من خلاف في أن عملية الترجمة تأويل لنص مصدر قصد توليد نص مستهدف في لغة أخرى، من بين استراتيجياتها (أي اللغة) توجيه جمهورها إلى تأويل معين لذلك النص، و هو تأويل غالبا ما يطلع التأويل الذي استمد من النص المصدر و يوافقه تمام التوافق.

من هذا المنطلق، يحصل لدينا أن أشد ما يعنى به المترجم هو تطوير ملكتيه التأويلية و الاستراتيجية على حد سوى في العملية الترجمية، بالنظر إلى احتمال تضععهما أو اختلالهما في إحدى اللغتين، ذلك أنه قلما نلفى من المترجمين من يمتلك ناصية اللغتين امتلاكا يضاهي امتلاك الناطقين بها، أو أنه يبرع في المجال براعة تماثل براعة المتخصصين المتضلعين. فضلا على ذلك، مثل الترجمة كمثل التأليف، تستلزم مهارة الصانع البارعين و حذق الحاذقين في وضع مختلف الاستراتيجيات و التأويلات

¹ - M. Baker (1998) Op. Cit., p. 180.

² - Cf. M. Tymoczko (1998) Op.Cit.

و مقارنتها و تقييمها و إثارة الواحدة منها دون الأخرى، تبعا لأطر سياقية محددة لذلك الحيز أو الفضاء النصي.

و لما كان ذلكم الواقع واقع الترجمة الذي لا مفر منه، يعمل المترجم تحت جملة من الإكراهات و الضواغط يصنعها جمهور معين من المتلقين و تأويلات معينة للنص المصدر و/أو النص المستهدف.¹ و الحال هذه، يحتاج المترجم إلى موارد تمده إمدادا بكل التأويلات الممكنة و المحتملة للنص المصدر، و تدعمه دعما بالاستراتيجيات الفاعلة و الفعالية قصد التوصل إلى تأويلات معينة للنص المستهدف، و هي الموارد ذاتها التي يعول عليها في تقييم تلكم التأويلات و الاستراتيجيات، و تيسر انتقاء ما كان منها مناسباً و يخدم المقام. و لما كانت على هذا القدر من الأهمية، يقدر بعض الباحثين أن نسبة 50% من الوقت المستغرق في الترجمة، يكرسها المترجم لمعاينة ما تيسر له و ما توفر لديه من العدة و الموارد المرجعية.²

هذا عن مفهوم "الترجمة"، أما ما كان من مفهوم "التدريب" فيعني به ضمناً أن الكفاءات و المهارات المراد تلقينها للمتلم يُفترض أن يكتسبها رويداً رويداً من خلال الدربة و المران، و التعرف إلى مختلف ضروب الأدوات التي سيستخدمها و المهام التي ستناط به أثناء حياته المهنية مستقبلاً، و ذلك بدمجه ضمن بيئة تماثل تلك الحياة قدر المستطاع.

و في تدريس الترجمة، يُفترض بالأستاذ أن يدرب الطالب على أداء وظائف محددة، ضمن أوساط و باستخدام أدوات محددة هي الأخرى – وفق ما يذهب إليه غوادك (Gouadec) :-

"[W]e are supposed to train people to perform clearly identified functions in clearly identified environments where they will be using clearly identified tools and 'systems' [...]"³

بمعنى أن المتلم يُهيأ لحل إشكالات يمكن تحديد مواطنها قبلاً من خلال تطبيق إجراءات موضوعة سلفاً أو "مكتسبة".

و المدونات واحدة من تلك الموارد أو الأدوات التي يلح كثير من الدارسين المنظرين – أمثال مايا – على ضرورة إدراجها ضمن البرنامج التدريبي للمترجم، و اعتبارها جزءاً لا يتجزأ من سيرته الذاتية مستقبلاً، و بالتالي فهي تعد من بين كفاءاته المهنية.

"The methodology involved in making corpora and extracting terminology should be part of the teaching curriculum – not an

¹ - Cf. K. Reiss (1981) Type, kind and individuality of text : decision making in translation. *Poetics today* 2(4), pp. 121-131, qt. in : G. Aston (1999) *Op. Cit.*

² - Cf. K. Varantola (2000) Translators, dictionaries and text corpora. In: S. Bernardini & F. Zanettin (eds.) *I corpora nella didattica della traduzione. Corpus Use and Learning to Translate*. Bologna: CLUEB, pp. 117-136.

³ - D. Gouadec (2002) Training Translators : Certainities, Uncertainities, Dilemmas. In : B. Maia, J. Haller & M. Ulrych (eds.) *Training the Language Services Provider for the new Millenium*. Oporto: Universidade de Oporto, p. 32.

*optional extra – particularly given the pressure on translators to use computer-assisted tools."*¹

و الحال هذه، عمد كثير من الأساتذة إلى صناعة مدونات صغيرة و تبنيها ضمن خطتهم التعليمية، باعتبار أن من قوام الكفاءة الترجمية المعاصرة معرفة كيفية صناعة المدونة، و أنواعها و استخداماتها.

و بناء عليه، فإننا نتصور أن رسم خطة أولية لدرس الترجمة باعتماد المدونات يقوم على ثلاثة عناصر رئيسة، يعد مجموعها بمثابة مدخل في اللسانيات المدوناتية، هي: أ- استخدام المدونات: و نقصد بذلك مختلف التطبيقات التي يمكن إنجازها باعتماد المدونات.

ب- تجميع المدونات: أي بناؤها، حيث يعنى فيه بتلقين الطالب مبادئ صناعة المدونة و معايير تصميمها على اختلاف صنوفها، و كذا التعرف إلى مختلف مصادرها ومعايرتها و انتقاء الأنسب منها، و صقل قدرته على اختبار ملاءمتها للتطبيق المراد إنجازها.

ج- برمجيات معالجة المدونات: و هو درس يتعرف من خلاله الطالب إلى مختلف أنواع البرمجيات المستخدمة في معالجة المدونات، و كيفية توظيفها بما يتماشى و نوع التطبيق المنجز.²

و ما أكثر تلكم الدراسات التي يعتمد فيها أصحابها (سواء في صفوف اللغات أو الترجمة) على صنع مدونات قوامها صفحات الويب (على غرار دراسة مايا 2000، و فارانتولا 2000)، أو أنها مدونات قوامها نصوص مستنبطة من مصادر إلكترونية أخرى كالصحف أو المجلات أو الأقراص المضغوطة (زانتان 2001، باوكر 1998)،³ و الملاحظ أن صناعة هذه المدونات تستغرق من الوقت و الجهد أكثر مما تستغرقه صناعتها انطلاقا من صفحات الويب. بيد أنه جدير بالذكر أن ذلكم الاستغراق سواء في صناعة المدونة أو تصفحها – على اختلاف صنوف مصادرها – هين إذا ما قورن بنظيره باستخدام الأدوات أو المساعدات الترجمية المعهودة كالقواميس الورقية و الالكترونية، أو الإستعانة بأهل الاختصاص في المجال موضوع الترجمة.

II – 2 – تجربة الترجمة باعتماد المدونة:

يُعمل المترجم فُكره على مستوى نصين: نص أول سابق الوجود في لغة مصدر، و آخر يُنتظر منه أن يحدثه و يبتدعه في لغة ثانية هي اللغة المستهدفة؛ و الموقف أشبه

¹ - Belinda, Maia (2003) Some languages are more equal than others. Training translators in terminology and information retrieval using comparable and parallel corpora. In: Federico, Zanettin, Silvia Bernardini & Dominic Stewart (eds.) *Corpora in Translator Education*. Manchester: St. Jerome, p. 52.

² - For more explanation, see :

Krista, Varantola (2002) Disposable corpora as intelligent tools in translation. In : S.E.O, Tagnin (Org.) *Op. Cit.*, pp. 171-189.

³ - Cf. B. Maia (2000) Making corpora: A learning process. In: S. Bernardini & F. Zanettin (eds.) *Op. Cit.*, pp. 47-60.

K. Varantola (2000) *Op. cit.*

F. Zanettin (2001) Swimming in words: Corpora, translation and language learning. In: G. Aston (ed.) *Learning with corpora*. Bologna/ Houston, TX: CLUEB / Athelstan, pp. 177-197.

L. Bowker (1998) Using specialized monolingual native-language corpora as a translation resource: a pilot study. *Meta* 43(4), pp. 631-651.

بحل معادلة يحرص فيها أيما حرص على علاقة التكافؤ بين طرفيها، و هي العلاقة – المحور في كل عملية ترجمية، إذا وُفق المترجم في إقامتها استقام إنتاجه و كان ذا جودة و ريع.

و بين سندان اللغة المنقولة و مطرقة اللغة المنقول إليها، يقف صاحبنا حائرا: كيف يحقق تلكم العلاقة؟ إنه مطالب بأن يكون أعلم الناس باللغتين من جهة، و أن يكون ملما بموضوع الترجمة، من جهة أخرى. و هنا يقبع الإشكال، لاسيما إذا سلمنا أن أقسام الترجمة و معاهد التدريب لا يمكنها – بأي حال – تحقيق معجزة تخريج مترجم يجمع بين تلكم المهارات الثلاث معا، فإن هو امتلك ناصية اللغتين و صار له بعض العلم بموضوع أو اثنين، جهل البقية.

و بالحديث عن قسم الترجمة بجامعتنا، نلاحظ لدى الطلبة ضعفا لاسيما فيما يتعلق بالنقل إلى اللغة الأجنبية، إذ كثيرا ما تأتي ترجمة الطالب تَعْتَوِرُها العيوب من حيث إنها ركيكة الحبكة، متضعضعة السبك، مستعصية الفهم، لأن رصيده الفردي من تلكم اللغة لم يسعفه من ناحية، و لأنه "متكلم غر"¹ من ناحية أخرى، أي أنه يعتمد "الحرفية" في التعبير عن مقاصد اللغة الأجنبية، لأنه لا يلم بمواضعاتها كالفرق بين ما يخاله من المترادفات، أو جهله ببعض الصيغ و التراكيب و المسكوكات،* مما يعني أن الطالب – المترجم يمكن أن يكون هو الآخر "مترجما غرا"² إذا لم يدرك (نظريا و عمليا) بأن القسط الأكبر من اللغة عبارة عن قوالب موضوعة و تراكيب مصوغة، تكتسب بالاستعمال و التكرار إلى أن تترسخ لديه.

كما أنه ليس من المستبعد أن يكون "مترجما غرا" ما لم يدرك أنه أمام كل ذلك الزخم من "الاحتمالات" الترجمية المبنوثة في ثنايا قواميسه و معاجمه، ثمة منها ما يدعى بـ "الخيارات المحبذة" أو "المفضليات" (*Preferred options*)، يتوارد استعمالها عند أهل اللغة المستهدفة على نحو يجعل منها خيارات نمطية ثابتة، بحيث إن المترجم لو اصطفى خيارا آخر على حسابها، لجاءت ترجمته أبعد ما تكون عن الأصول في تلكم اللغة، لأنها تفتقر إلى الانسيابية و الطبيعية.

و الحال هذه، يجب البحث عن سبل دعم ذلكم الرصيد الفردي و تحسينه لدى طلبتنا، و ليس يجدي نفعا التفكير في إضافة ساعات الدعم أو التحسين اللغوي كحل للمشكلة، بالنظر إلى كثافة البرنامج الدراسي و ضيق الحجم الساعاتي؛ بل ينبغي النظر في تزويد الطلبة و إمدادهم بأدوات و موارد تساعد على تلقن اللغة الأجنبية و اكتسابها قواعد و أسلوبا، بما يضمن تمكنهم من الترجمة إليها وفقا لمواضعات الإنتاج النصي التي تنسم بها، أيا كان موضوع الترجمة أو مجاله. و في اعتقادنا أن للمدونات الشيء الكثير

¹ - المصطلح من استحداث (FILLMORE) فيلمور لتعيين متعلم اللغة الأجنبية (Innocent speaker). ينظر:

Charles J., Fillmore (1979) Innocence: A Second Idealization for Linguistics, Berkley Linguistic Society 5, pp. 63-76.

* - الواقع أن كل واحد منا "غر" حتى في لغته الأم، لاسيما أمامك مجال تخصص معين، نظرا لجهلنا مثلا بالتعابير الاصطلاحية في ذلكم المجال.

² - مصطلح *Innocent translator* من استحداث Stella E.O. Tagnin، ينظر مقالها :

Stella E. O., Tagnin (2002b) Corpora and the Innocent Translator: How can they help him? In: M. Thelen (ed.) Translation ad Meaning, Part 6, Proceedings of the Lodz Session of the 3rd Maastricht – Lodz Duo Colloquium on "Translation on Meaning" (489-496), Lodz, Poland, September 22-24, 2000, Maastricht: Universitaire Pers Maastricht.

لتقدمه للطالب بهذا الصدد، و هو ما سنوضحه من خلال تجربة ترجمة نص من اللغة العربية إلى اللغة الانجليزية، باعتماد مدونة أحادية اللغة.

II-2. 1. الفرضيات:

نفترض أن استخدام مدونة أحادية اللغة قوامها نصوص مكتوبة أصلا في اللغة المستهدفة (الانجليزية) إلى جانب برنامجية خاصة للمعالجة، يساعد المتمرّن في الترجمة إلى تلك اللغة انطلاقا من لغته الأم (العربية)، و هي أجدى من الأدوات التقليدية (القواميس) من حيث إنها تمكنه من تحسين أدائه الترجمي من عدة أوجه:

- ✓ تصيد المتلازمات اللفظية و الاهتداء إليها.
- ✓ اعتماد/ رفض خيار ترجمي مستمد من أدوات و موارد أخرى (كالقواميس مثلا).
- ✓ التحقق من الخيارات المبنية على الحدس.
- ✓ اقتناص القوالب المألوفة و التعابير المعهودة و الاستعمالات الشائعة في اللغة المستهدفة.
- ✓ ضبط استعمالات الكلمات/ التعابير الجديدة ضمن سياقاتها.
- ✓ استخلاص المصطلحات التي تمثل نظاما مفاهيميا يختص به موضوع الترجمة.

II-2. 2. العينة المشاركة:

لما كان موضوع الدراسة يتعلق أساسا بتدريب المترجم، و أحد أهدافها الرئيسية الأخذ بيد طلبتنا بما يساعدهم على تحسين أدائهم الترجمي و تجويد ترجماتهم، و جب أن تكون العينة المشاركة في التجربة من فئة الطلبة، على أن تتوفر فيهم الشروط التالية:

- ينتمي الطلبة جميعهم إلى المستوى الدراسي نفسه (السنة الرابعة ترجمة)، بحيث إنهم قد اطلعوا على كثير من المفاهيم و المبادئ الترجمية طيلة مشوارهم الدراسي، و كان لهم حظ من الحصص التدريبية على الترجمة من اللغة العربية إلى اللغة الانجليزية.
- جميعهم ناطقون أصليون باللغة العربية، لأن من أهداف التجربة اختبار مدى فعالية المدونة في إعانة المتمرّن على تحسين بعض قدراته اللغوية فيما يتعلق باللغة الأجنبية (الانجليزية).
- أن تكون لهم سابق تجربة بالنسبة لترجمة بعض النصوص الصحفية، دون أن يعني ذلك تطرقهم قبلا إلى الموضوع الذي يعالج نص الترجمة.
- و بما أن عماد التجربة هو استثمار المدونة "الإلكترونية" في التدريب، و جب على كل طالب أن يكون متمكنا من المبادئ الأساسية في استعمال الكمبيوتر، و ذلك لتسهيل التعامل مع المدونة، و التحكم في البرماجية.

II-2. 3. العينة النصية:

تحمل العينة النصية موضوع الترجمة عنوان (روسيا تختبر "أبو القنابل جمعاء")، و هي عبارة عن نص صحفي مقتبس عن وكالات مختلفة مع بعض التصرف. و قد راعينا في انتخابه تحقيق شرطين:

➤ أن يكون على قدر من التخصص، ذلك أن العينة المشاركة في التجربة ممثلة في طلبة. و تعرفهم إلى هذا النوع من النصوص قبلاً، لا يعني بالضرورة تطرقهم إلى الموضوع أصلاً؛ لذلك فإن كثيراً من المفاهيم و المصطلحات والتعابير التي تضمنها النص هي من قبيل المستجدات بالنسبة إليهم (يمكن النظر في النص ضمن الملحق 5).

➤ أن يكون موجزاً، إذ لا يتعدى طوله 215 كلمة، مراعاة لحال الطلبة الذين أبوا إلا أن يتطوعوا للتجربة، رغم انشغالهم على مستوى مذكراتهم و كثافة برنامجهم.

II-2.4. الأدوات و الموارد المتاحة:

أ- الأدوات "التقليدية": رُحِّص للطلبة استعمال كل ما توفر من قواميس و معاجم أحادية/ ثنائية اللغة، على غرار: المغني الوسيط، قاموس إلياس العصري، القاموس المزدوج (توينز)، المورد، و قاموس اللغة الانجليزية أكسفورد، و كذا مقالات طويلة نسبياً في الموضوع ذاته مكتوبة أصلاً في اللغة الانجليزية، بالإضافة إلى النت.

ب- المدونة و برماجية المعالجة: بالنظر إلى عمر هذه الدراسة المحدد وإمكانياتها المحدودة مقارنة بما تقتضيه صناعة مدونة تخدم بحثنا، اهتدينا إلى فكرة بديلة توفر علينا مضیعة الوقت و تخدم غرض الدراسة، تتلخص في استخدام "ويكوروب" (WebCorp). "ويكوروب" طقم من الأدوات التي تمكن الباحث من الولوج إلى الشبكة العنكبوتية و معالجتها باعتبارها أضخم مدونة نصية على الإطلاق، و هو من إنتاج وحدة البحث و تنمية الدراسات الانجليزية *Research and Development Unit for English Studies* (RDUES) بجامعة برمنغهام.¹

و إمكانية الإبحار على مستوى الشبكة العنكبوتية لا تعني أن ويكوروب مجرد محرك بحث، شأنه كشأن محركات البحث "التقليدية"، على غرار: جوجل، ألتافيسا و غيرهما، أو أنه مجرد أرشيف من الصحف في شكلها الالكتروني. بل إن طقم الأدوات التي يتميز بها تجعله متفرداً من حيث مبدأ العمل.

ويكوروب بخلاف محركات البحث البسيطة و المعهودة، مُصمَّم بما يتوافق ومتطلبات البحث في مجال لسانيات المدونة، و جميع الحقول المعرفية التي تستثمر تلكم اللسانيات؛ إذ يتولى عملية استرجاع المعطيات اللسانية من صفحات الويب في شكل أسطر فهرست مرتبة ألفبائياً، حيث تتمركز وحدة البحث (الكلمة المفتاحية أو العقدة) مبرزة بلون أحمر ضمن جميع سياقاتها المحتملة، كما يمكن لويكوروب أن يصلك بجميع المواقع و الصفحات التي استقيت منها الكلمات المفتاحية.

أما محركات البحث البسيطة فتعطي قائمة بما عثرت عليه من صفحات تحتوي على وحدة البحث. يتصفح الباحث كل صفحة على حدى بالنقر على الرابط، و يُمَوِّع

¹ - CF. WebCorp site at:
<http://www.webcorp.org.uk/>

الوحدة المطلوبة ثم يلجأ إلى طريقة القص/ اللصق ضمن صفحة من معالج النصوص، وفي ذلك مشقة و مَـثَـبَةٌ كبرى مقارنةً بويكوروب الذي يعتمد إلى تنفيذ العملية بشكل آلي تماماً.

يتضمن ويكوروب نوعين من خيارات البحث: خيارات أساسية و أخرى متقدمة.

أ- الخيارات الأساسية (Basic Options):

أول ما يظهر من ويكوروب هو واجهة البحث الأساسية (الشكل -30-)، و تتضمن الخيارات التالية:

■ وحدة البحث: و تكون إما كلمة أو جملة أو قالباً، دونما الحاجة إلى إضافة علامتي التنصيص.

WebCorp

An improved version of the original WebCorp,
designed to search the web for concordances in real time

Search term:		
Enter a word, phrase (no quotes necessary) or pattern		
Search engine:	Concordance Span:	Case options:
Google	5 word(s) to left & right	Case Sensitive
Advanced Search Options		
Submit		

الشكل (30) الواجهة البحثية الأساسية لويكوروب.

و البحث عن القوالب يكون وفق الويلدكارد (*)، بحيث تُدَوَّل سلسلة الأحرف بالعلامة (*)، مثل: *run** تعطينا *runners, running* و ما إلى ذلك. كما يمكن استخدام الويلدكارد لتقوم مقام كلمة بأكملها في جملة ما، نحو: *the * sank*، تعطينا الجمل:

The boat sank, the ship sank, the ferry sank,...

أو في المثال: *the * * sank*، نحصل على جمل من قبيل:

The ship had sank, the 'unsinkable' ship sank, etc.

أو في بداية سلسلة أحرف ضمن جملة، على شاكلة: *the * ing man* و تكون النتائج من قبيل:

the running man, the laughing man, etc.

■ **محرك البحث:** ويكوروب بمثابة "الأمر" بالنسبة لمحركات البحث المعهودة:

Ask, live Search, Yahoo, AltaVista, Google, Open Directory, Metacrawler.

هذا الخيار يسمح للباحث بتحديد أي محرك من تلك المحركات يرغب في أن يوظفها ويكوروب.

■ **مسافة الفهرست:** عدد الكلمات المطلوب عرضها على يمين الكلمة المفتاحية و يسارها. و القيم المتوفرة ضمن واجهة البحث الأساسية هي: 1، 5، 10، 15، 20 كلمة يمينا و يسارا.

■ **خيارات الحالة:** يحدد المستخدم حالة البحث التي تكون إما دقيقة (*Sensitive*) أو غير دقيقة (*Insensitive*). و حالة البحث الدقيق تعني أنه سيتم استخراج وحدة البحث المطوبة فقط.

ب- الخيارات المتقدمة (*Advanced Options*): تتضمن الواجهة البحثية المتقدمة

جملة من الخيارات زيادة على خيارات البحث الأساسية السابقة، هي:

1- **الشكل (*Format*):** و يتضمن سلسلة من الخيارات تسمح للباحث بتحديد الشكل الذي يرغب في أن تظهر عليه نتائج البحث (ينظر الشكل -31-).

Output Format:	Web Addresses (URLs):
HTML	Show for all pages accessed
HTML Tables (KWIC)	Never show
HTML	Show for concordance lines
Plain Text (KWIC)	Show for all pages accessed

الشكل (31) خيارات الشكل.

● **شكل المخرج:** تظهر العقدة في مركز كل سطر من الفهرست بخط أحمر ذي نمط عريض، ضمن لغة HTML.

● **عناوين المواقع (*URLS*):** للباحث خيار إظهار / إخفاء عنوان موقع كل صفحة ويب ضمن صفحة نتائج البحث، كما يمكنه اختيار إظهار عنوان مواقع الصفحات التي استنبطت منها أسطر الفهرست دون غيرها. و في الخيار (عدم إظهار) (*Never show*)، تعرض أسطر الفهرست ضمن قائمة مرقمة، متضمنة الكلمة المفتاحية بخط أحمر ذي نمط عريض، و بالنقر عليها تظهر صفحة الويب التي استخرجت منها.

2- **الفهارس (*concordances*)**، و يتضمن هذا الخيار: (الشكل-32-)

● **مسافة الفهرست:** يمكن للباحث أن يختار من كلمة واحدة إلى عشرين (20) كلمة كمسافة محيطة بوحدة البحث، سواء على يمينها أو يسارها. كما يمكنه أن يختار (*Max*) حيث يظهر 200 حرف على جانبي العقدة.

● **الجملة كاملة:** بدل تحديد مسافة الفهرست بعدد معين من الكلمات يمينا و يسارا، يمكن اختيار عرض الأمثلة عن وحدة البحث ضمن الجمل التي وردت فيها.

- عدد الصفحات المراد استرجاعها: يمكن استخدام هذا الخيار لتسريع نتائج البحث عن طريق تحديد عدد صفحات الويب التي ينبغي لويكوب تصفحها، و القيم المتاحة: 10، 25، 50، 100، 200، 500 صفحة.

Concordance Span: 5 word(s) to left and right OR Full sentences? ☐	Number of Pages to Retrieve: 100 ☐ One concordance line per web site
--	--

الشكل (32) خيارات الفهرس.

- سطر فهرست واحد من كل صفحة ويب: يعطي هذا الخيار أول سطر فهرست من كل موقع، و هذا تفاديا للتكرار.
- 3- المجالات (Domains): تقدم الواجهة البحثية المتقدمة ثلاثة أنواع من محددات مجال البحث (الشكل-33-):

Site Domain: Leave blank to search the whole web.
Newspaper Domains: None selected This will override any domain set above
Textual Domain: All Select Open Directory category

الشكل (33) المجالات.

- مجال الموقع: عند استعمال هذا الخيار إلى جانب محركات البحث (جوجل، ألتافيسا، لايف سورتش، أسك)، فإن ذلك يؤدي إلى:
- ✓ تحديد البحث في جميع الصفحات ضمن الموقع الواحد بإدخال عنوانه دون كتابة البروتوكول (http: //). مثال ذلك: إدخال العنوان (www.bbc.co.uk) لبحث جميع الصفحات ضمن موقع الـ (بي بي سي).
- ✓ تحديد البحث في جميع المواقع ذات مجال معين، و ذلك بإدخال جزء فقط من عنوان الموقع URL. مثال ذلك: إدخال (ac.uk) يعني تحديد البحث ضمن مواقع مؤسسات المملكة المتحدة، و إدخال (.fr) يعني تحديد البحث ضمن مواقع الويب في فرنسا، و هكذا دواليك. و لقد زود ويكوب بقائمة من المجالات الشائعة الاستعمال، تظهر أسفل الخيار "مجال الموقع".
- مجالات الصحف: يسمح هذا الخيار بتحديد البحث ضمن مجموعة من مواقع الصحف محددة سلفا (على غرار: الصحف البريطانية، الفرنسية، الأمريكية...).

• المجال النصي: يسمح هذا الخيار بتحديد البحث في صفحات الويب ضمن فئة خاصة في المحرك (Open Directory).

4- ترشيح الكلمات (Word Filter):

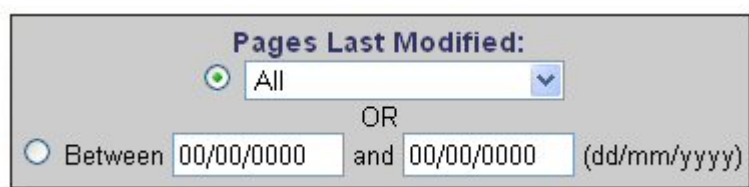
يسمح هذا الخيار بتضمين كلمات أكثر يرغب الباحث في ظهورها/ إخفائها في نفس صفحات الويب إلى جانب وحدة البحث، وذلك بوضع علامة (-) أمام الكلمات غير المرغوب فيها. مثال ذلك: لدينا الوحدة (plant)، يمكن أن نضيف ضمن خيار ترشيح الكلمات (nuclear-flower)، وذلك لتحديد معنى وحدة البحث (plant)، أي أن ويكوروب سيعرض جميع أمثلة الكلمة (plant) الواردة في صفحات الويب التي تتضمن أيضا كلمة (nuclear) مع إسقاط الأمثلة عن كلمة (flower) وإغائها من النتائج. (ينظر الشكل -34-).



الشكل (34) ترشيح الكلمات.

و ثمة من الخيارات ما يتم بعد المعالجة (Post-Processing)، أي بعد انتهاء البحث و الحصول على صفحات النتائج، نذكر:

❖ ترشيح التاريخ (Date Filter): حيث نجد في صفحة نتائج البحث طريقتين لترشيح النتائج بواسطة التاريخ، و هذا ضمن خيار "صفحات معدلة مؤخرا" – مثلما يوضحه الشكل 35-: إما باختيار فترة زمنية محددة (الشهر الماضي، الأشهر الثلاثة الماضية، الأشهر الستة الماضية، السنة الماضية، أكثر من سنة مضت، أكثر من سنتين مضتا، أكثر من خمس سنوات ماضية)، و إما بإدخال الفترة الممتدة بين تاريخين.



الشكل (35) خيارات التاريخ.

❖ خيارات التلازم اللفظي (Collocation Options): نجدها هي الأخرى في صفحة نتائج البحث، و هي جملة من الخيارات التي تسمح لمستخدم "ويكوروب" بتحليل وصيالات وحدة البحث، أي جميع الكلمات التي يتواتر ظهورها كثيرا حول العقدة ضمن النصوص المعالجة. (الشكل -36-)



الشكل (36) خيارات المتلازمات اللفظية.

- الوصلات الخارجية: بالنقر على هذا الخيار، يكون المخرج عبارة عن جدول لتكرارات الكلمات التي تنتظم بجوار العقدة في أربع مراتب متتالية سواء يمينا أو يسارا.
 - الوصلات الداخلية: إذا كانت وحدة البحث عبارة عن قالب يتضمن ويلدكارد، فإنه بالنقر على خيار(الوصلات الداخلية)، يكون المخرج جدولاً بجميع الكلمات الأكثر تواردا و التي تحتل مراتب الويلدكارد تلك.
 - إلغاء كلمات الوقف: من خلال هذا الخيار، يتسنى للباحث ترشيح الكلمات ذات التكرارات الكبيرة (مثل: the, of)، و إسقاطها من جدول الوصلات.
- ❖ النظم (Sorting):
- بعد انتهاء البحث، تعرض صفحة النتائج على الباحث طريقتين لنظم المخرج (Output) – الشكل (37)–:
- نظم الفهارس وفق التاريخ: يُمكن هذا الخيار الباحث من نظم الفهارس وفق ترتيب تاريخي تنازلي أو تصاعدي.
 - نظم الفهارس أبجديا: يعطي هذا الخيار إمكانية نظم الفهارس أبجديا تبعا لكلمات التي ترد على يمين/ يسار العقدة الموافقة، و هو خيار بالنسبة للبحث وفق الويلدكارد.



الشكل (37) خيارات النظم.

أدوات أخرى:

بالإضافة إلى الفهرست الآلي، يتوفر "ويكوروب" على أدوات أخرى، هي:

❖ كشاف وحدة البحث (Search term highlighter):

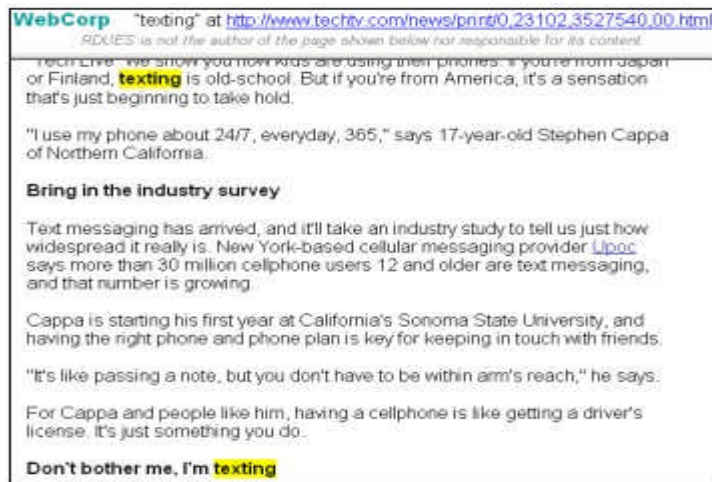
بالنقر على العقدة ضمن صفحة نتائج البحث، فإن ويكوروب يصل المستخدم بالموقع الذي استنبط منه سطر الفهرست (الشكل-38-)، حيث يجد جميع تواترات الوحدة مظهرة باللون الأصفر.

❖ ناظر النص العادي (Plain Text viewer):

إذا كان المخرج في صورة HTML مشتملا على عناوين المواقع (URLs)، يمكن الوصول إلى نسخة ASCII من كل صفحة ويب ثم تصفحها و استنباط الفهرست منها، و ذلك بالنقر على الرابط (Plain Text) – حسب الشكل-39-.

❖ ناظر قائمة الكلمات (Word List viewer):

و هو أداة تعطي قائمة بجميع الكلمات (الأنواع Types) الموجودة ضمن صفحة الويب، وفق ترتيب ألفبائي أو ترتيب التكرارات تنازليا (ينظر الشكل-40-).



الشكل (38) كشف وحدة البحث.



الشكل (39) الرابطان Plain Text / Word List

WebCorp
Live

Enter the address (URL) of a web page and the WebCorp Word List Generator will produce a frequency or alphabetical listing of the words or phrases found on that page.

Enter the URL:

Order Output:

By frequency

Stopword Filter: ☒

N-gram Size:

Submit

المعايير	النصوص المنتخبة (باعتماد ويكويرب)
حجم المدونة: يجب أن تحتوي المدونة على كمّ معتبر من النصوص لتحديد ما إذا كان الاستعمال شائعاً أم أنه شخصي .	يفوق عددها المئة مقالا، و يتجلى ذلك من خلال عدد صفحات الويب التي يقوم ويكويرب بزيارة مواقعها (حوالي 500).
حجم النص: يجب أن تكون النصوص المتضمنة كاملة غير مبتورة، حرصاً على عدم إسقاط أمثلة الاستعمال أو شروحات المفاهيم .	جميع المقالات تامة.
نوع النص: نصوص هجينة تتراوح بين المختصة و التبسيطية، كي تساعد المترجم على فهم المجال و تمكنه من التعرف إلى مختلف السجلات .	مقالات مختصة و أخرى تبسيطية.
تاريخ النصوص: ينبغي أن تتراوح النصوص بين القديم منها و المعاصر، فالأولى مفيدة لأنها قد تتضمن شروحات المفاهيم حيث ظهرت أول مرة، أما النصوص المعاصرة فلا بد منها لأنها المرآة التي تعكس واقع المجال و حاضره.	تتضمن مقالات قديمة نسبياً و أخرى معاصرة.
المؤلف: لا بد من إدراج نصوص لعدة مؤلفين، بغرض معرفة ما إذا كان الاستعمال شائعاً أم أنه مجرد أسلوب شخصي.	المقالات من تسطير مختلف أقلام الكتاب.
اللغة: يحبذ أن تكون النصوص من إنتاج كتاب في لغتهم الأصل، و هذا قصد تبيان تنميّطات الاستعمال .	المقالات باللغة الإنجليزية مستنبطة من صحف بريطانية و أمريكية.
عنصر الثقافة (موقعة النصوص): ينبغي أن يكون كتاب النصوص ذوي خلفيات ثقافية متباينة، و هذا لإبراز الاستعمال "الجهوي"، أي الموافق لكل منطقة.	تحقق المقالات هذا المعيار، لأنها مستقاة من الصحف البريطانية و الأمريكية.

143

II-2-5- التجربة

نُظِمَ العمل ضمن ورشتي اختبار:

الورشة الأولى:

- انتظم الطلبة مثني مثني ضمن خمسة أفواج، ثم طلب منهم ترجمة النص باعتماد الأدوات و الموارد "التقليدية" المتاحة لمدة ساعتين، على اعتبار أنه مجرد تمرين ترجمة، بمعنى أننا لم نطلعهم على الهدف من التجربة و لا حدثناهم بخصوصها. بعد انقضاء المدة الزمنية، جمعت الترجمات الخمس، دون إبداء تعاليق أو تصحيحها.

الورشة الثانية:

- عمدنا إلى شرح مبدأ العمل بواسطة المدونات، و قد ركزنا أكثر على تقديم "ويكوروب" و كيفية استخدامه بصفة عامة، دون إطلاع الطلبة على الغرض من هذا العرض، هذا في مرحلة أولى.
 - أما المرحلة الثانية فهي مرحلة الترجمة. و قد تعمدنا تأخيرها قدر ما استطعنا لحاجة في نفس يعقوب؛ فقد نسي الطلبة ما كان من ترجماتهم للنص الأول باعتماد قواميسهم و معاجمهم. و لم يكن أي منهم يتوقع أن يُسألَ العمل عليه مجددا؛ لذلك، و نحن نمرر النص بينهم، سمعنا بعض اللغط و التعاليق من قبيل: "لقد سبق و أن ترجمنا هذا النص!"، "لماذا يتوجب علينا العودة إليه؟". و سرعان ما تبددت علامات الاستغراب و التساؤل تلك، لما طلبنا منهم ترجمة النص باعتماد ويكوروب.
 - و في المرحلة الأخيرة، طلبنا من كل متمرّن تسجيل ملاحظاته بخصوص التجربة و أدواتها.
- جدير بالإشارة أن العمل في هذه الورشة، قد استغرق من الوقت خمس ساعات، خصصنا الساعات الثلاث الأولى لمرحلة الشرح و التوجيه، و البقية للترجمة.

II-2-6- نتائج الورشتين و تحليلها:

1- الورشة الأولى:

- بعد فحص الترجمات الخمس نتاج العمل باستخدام الأدوات "التقليدية" (ينظر الجدول رقم - 12 -)، و بالنظر إلى بعض تعليقات المتمرّنين، سجلنا الملاحظات التالية:
- أ- يعتمد طلبتنا على قواميسهم الثنائية اللغة اعتمادا هو أظهر من أن يُنَبَّه عليه، و أغلبهم ينكب عليها باعتبارها أهم الأدوات المرجعية بالنسبة إليه، و هي أدوات وإن كانت غالبا ما تورد مقابلات في لغتي الترجمة، إلا أن من عيوبها:
- * أن واضعيها لا يولون فكرة "اللغة بنت الاستعمال" أهمية تذكر، لذلك تأتي مداخلها مجردة من السياقات، و هو ما يفسر عدم تمكن الطالب من القبض على الاستعمال الصحيح و الدقيق للمفردة.
- * انتقاء تلكم المداخل دون غيرها، و إن كان يسوَّغ بمعيّار الاستعمال، إلا أنه لا يستند إلى أي دراسات ميدانية ترصد المستعمل حقا و تثبته، و تحدد المهمل و تحذفه.

المصدر	الترجمة (1)	الترجمة (2)	الترجمة (3)	الترجمة (4)	الترجمة (5)
روسيا تختبر "أبو القنابل جمعاء"	Rusia tried out "father of all bombs"	Russia Test "*"*****"	Russia tries the "world's strongest bomb".	Russia tests the "father of bombs".	Russia examines "the father of all the bombs".
أعلنت روسيا أنها اختبرت بنجاح قنبلة جديدة محمولة جوا، والتي وصفتها بأنها أقوى سلاح غير نووي في العالم.	Russia declared its succesful trial of a new airborne bomb which they described it as the strongest unnuclear weapon in the world.	Russia announced it had successfully tested a new airborne ×, which it described as the most powerful non-nuclear weapon in the world.	Russia declared that it had tried successfully a new bomb lifted in the air, which it described as the world's strongest weaponry that is not nuclear.	Russia declared that it had tested successefully a new airborne bomb, considering it as the non-nuclear strongest weapon.	Russia proclaimed that it examined with success a new bomb that is airy carried. It described it as the most robust weapon in the world that isn't nuclear.
و القنبلة الجديدة "لا مثيل لها في العالم" على حد قول الجنرال ألكسندر روكشين، نائب هيئة الأركان العامة للقوات المسلحة الروسية.	The new bomb is "unprecedented in this world" declaring the Army commander, Alexander Roxin, vice president of plenary session of the Rusian Army [...]	And the new bomb "isn't unprecedented in the world", said General Alexander Rockchin, Deputy chief of General staff weapon of the Russian armed forces.	The new bomb is "unexampled in the world", as General Alexander Rokchin said, who's vice-head of General Pillars of Russian Armed Forces.	According to the general Alexander Rocchin, the vice- president of general staff organization of Russian armed forces, the new bomb is incomparable.	General Alexander Rokchine, the vice-president of general staff organization Russian armed forces, said that the new bomb is "peerless in this world".

الجدول (12- أ) النص المصدر و ترجماته.

عدم إيلاء أهمية كبرى للتعابير المسكوكة أو الاصطلاحية و الأمثال، هذا ناهيك عن عدم الاهتمام بإقامة علاقات دلالية بين المفردات تبين ما بينها من فروق دقيقة، و تعين المترجم على اختيار المقابل الدقيق و المناسب.

* أضف إلى ذلك، أنها معاجم تفتقر إلى خدمة التحيين بمواكبة الحركة الدائبة التي تشهدها كل لغة، من حيث تغذيتها بكل ما استجد من المفاهيم و المصطلحات، و استئصال ما أهمل استعماله.

و الحق إن من تلكم العيوب ما يبرره الحجم، إذ لا يعقل أن يشمل أي معجم جميع مفردات اللغة و ألفاظها مهما بلغ حجمه عظمة، هذا من جهة؛ و بالنظر إلى عامل الوقت المستغرق في تصنيف المعاجم من جهة أخرى، إذ ما إن تُخرَج حتى يصدر زخم هائل من المستجدات التي لا سبيل إلى تجميعها و زيادتها إلا ضمن طبقات لاحقة.

He told also to 'O.R.T' the first channel of Russia that: "the results of the experiments ... showed that this bomb is the same as a nuclear one in its effectiveness and fitness". At the same time, he emphasized that the new weapon doesn't menace environment on the contrary to nuclear weapons.	"The results of experiences... shown that the bomb neutralizes a nuclear bomb concerning its efficiency", He declared to the first canal of Russia 'O.R.T.N', and He confirmed at the same time that unlike nuclear weapons, the new weapon has no negative points on the environment.	He said also to the Russian First channel (ORT) that "the results of the probations... had manifested that the bomb is equal to a nuclear one considering its efficiency and competence». He also confirmed at the same time, that the new weapon doesn't threaten the environment on the contrary to nuclear ones.	Has been added to the first Russian channel 'O.R.T' "the result of tests showed that the bomb ...equivalent to a nuclear bomb in the effectiveness and efficiency". He also stressed at the same time that the new weapon wasn't a threat to the environment other than nuclear weapons.	Who said further to the First Russian channel (O.R.T.N) "The experience results show that the bomb is effective such as any other nuclear bomb, but at the same time the new weapon doesn't have any impact to the environment comparing it with nuclear weapons".	و قد أضاف قائلا للقناة الأولى الروسية (أو آر تي إن) "نتائج التجارب".... بينت أن القنبلة تعادل قنبلة نووية في فاعليتها و كفاءتها". كما أكد في الوقت نفسه أن السلاح الجديد لا يشكل تهديدا للبيئة بخلاف الأسلحة النووية.
ORT said that the new bomb, surnamed "the father of all the bombs", is four times stronger than the "Mother of all the bombs" which is an American bomb of air blast of great mass (MOAB).	The (O.N.R.T) said that the new bomb which is surnamed the father of all bombs is stronger four times from the American lump superior aerial exploding (M.O.A.B) bomb which is surnamed 'the Mother of all bombs'.	Russia's First channel said that the new bomb, named 'the world's strongest bomb', is more powerful four times than the bomb of air burst which is of big mass (MAOB) made in America and known as the 'Mother of all the bombs'.		The First Russian channel said as well that the new bomb, which called 'Father of all bombs', is four times stronger than the (american aerial exploding of extensive lump bomb) known as 'Bomb's Mother'.	و قالت القناة الأولى الروسية إن القنبلة الجديدة، الملقبة بـ "أبو القنابل جمعاء"، أقوى أربع مرات من قنبلة التفجير الهوائي الفائقة الكتلة (أم أو أي بي) الأمريكية المعروفة بـ "أم القنابل جمعاء".

الجدول (12- ب) النص المصدر و ترجماته.

After that, it showed a strategic bomb thrower of Toboliv TU 160 shooting the bomb in the examination spot. There was a huge explosion.	during its throwing the bomb, the bomber Topoliva TU 160 has show a strategy, followed by a big explosion.	Thereafter, it showed a thrower of strategic bombs called Tobolive TU 160 when it was throwing the bomb at probation site. It was followed by a big explosion.	Then came bombers of the type of strategic Topolive TU 160 to throw the bomb at the test site, and lead out on to a large explosion.	A strategic bomb thrower type 'ToBoLiF TU 160' showed during its activity in test location. Then it came after a huge explosion.	ثم أظهرت قاذفة قنابل استراتيجية من نوع توبوليف تي يو 160 أثناء رميها القنبلة على موقع اختبار، وتلا ذلك انفجار كبير.
It is known that weapons of compression and temperature, known also as "inanity/ vacuity bombs" or "airy fuel stores" are distinguished from traditional explosives by using oxygen existing in the atmosphere in place of on oxidized element filled in the explosive material.	As it is known, pressure and heat weapons which is surnamed 'Vacuum bombs' or "Aerial fuel munitions" is different than traditional explosifs since it use the O2 of aerial cover instead loading an oxidized factor into the explosif matter.	It's known that weapons pressure and temperature, called also 'space bombs' or 'fuel air ammunitions' differ from old explosive matters that they use oxygen from the atmosphere instead of stuffing an oxidizing factor in the exploding materials.	And known that heat and pressure weapons, also known as the "Vacuum bombs" or "Dechain fuel-air". as distinct from conventional explosives to be used in atmospheric oxygen instead of air-filling factor or the explosive material within the oxidizer.	We all know that pressure and thermic weapons, called too as 'empty bomb' and 'aerial fuel stores', which are different from old exploding in using oxygen of the atmosphere instead of loading oxidized element inside of the explosive matter.	و المعلوم أن أسلحة الضغط والحرارة، المعروفة أيضا باسم "القنابل الفراغية" أو "ذخائر الوقود الهوائي"، تتميز عن المتفجرات التقليدية بأنها تستخدم الأكسجين الموجود في الغلاف الهوائي بدل حشو عامل مؤكسد داخل المادة المتفجرة.

الجدول (12- ج) النص المصدر و ترجماته.

ب- رغم محدودية تلك المعاجم و تنبيه الأساتذة لذلك، إلا أن الطالب يثق بها ثقة الأعمى بمن يقوده – و إن كان أعمى هو الآخر –، إذ أننا لم نلاحظ و لو طالبا واحدا قد عمد إلى استخدام معجم أكسفورد في اللغة الإنجليزية، بل إنهم لم يكلفوا أنفسهم حتى عناء قراءة بضعة أسطر منها. و حجتهم في ذلك أنه لم يسبق لهم و أن اعتمدوا على مثل تلك الأدوات عدا المعاجم الثنائية اللغة، و لما تبين لبعضهم أنها حجة واهية، قالوا إنه و بحكم عملهم المحدد زمنيا بساعتين، فإنهم يرون أن مجرد تصفح تلك المقالات "الطويلة" يستهلك من الوقت ما يحذون استغلاله في تصفح معاجمهم و قواميسهم.

و كذا رأي فريق عمد إلى استخدام محرك البحث Google، حيث قام الطلبة بإدخال عدد من الكلمات المفتاحية، وحصلوا على عدة نتائج، و لما بدأوا بتصفح كل صفحة ويب

على حدى – كما هي العادة – راودهم قلق شديد بالنظر إلى أن كثيرا من الصفحات التي قاموا بزيارتها لم تسعفهم في شيء. و قد تعاضم قلقهم لما بدت لهم بعض الصفحات ملائمة، إلا أن تحميلها استغرق منهم وقتا طويلا، و لما يتبق لهم لترجمة النص إلا قليلا، لذلك لم يجدوا متسعا لقراءة ما حصلوا عليه، فألقوا بها جانبا و عادوا أدراجهم إلى معاجمهم و قواميسهم.

Bombs such as these explode according to 2 stages: a little explosion sprinkles a principal freight of explosive materials like a cloud, then it explodes in the atmosphere.		These bombs burst forth into 2 phases: a small explosion scatters a principal charge of exploding material as a cloud, which explodes later in the atmosphere spontaneously.	And of these, such as bombs explode in two phases: a small explosion scattered president shipment of explosive materials cloud and explode in the next phase of those components automatically cloud in the air.	Such as bombs detonate in over two steps: a small exploding is arising a high volume of explosive matters seemed to a cloud. While in the next step that cloud together with its components will explode automatically in the air.	و مثل هذه القنابل تنفجر على مرحلتين: ينثر انفجار صغير شحنة رئيسية من المواد المتفجرة كغيمة؛ وفي مرحلة تالية، تنفجر تلك الغيمة بمكوناتها تلقائيا في الجو.
The explosion generates a wave of pressure stronger than of old explosive materials and the consumption of the gaz in the explosion makes a small vacuity and elevated temperatures. Like that “Each living being evaporates without fail”, as a Russian responsible declared.		The explosion engenders a wave of pressure that is stronger than that of traditional exploding matters. The consumption of gases in the explosion also generates a small space and a high heat. So, “every living thing inevitably evaporates”, according to one of the Russian responsible.	And generate the blast wave pressure exceeds the intensity of conventional explosives, and that the consumption of gas within the blast also generates a vacuum and high temperature part. “and the every living organism inevitably evaporates” as one of the Russian officials.	The explosion causes a high pressure effect superior than the power of old explodings beside of the consumption of gases including the explosion, this later make partial vacuum and high temperatures, this way ‘Each live being will certainly die’ in declaration of Rusian Responsable.	و يولد الانفجار موجة ضغط تفوق شدتها قوة المتفجرات التقليدية، كما أن استهلاك الغازات ضمن الانفجار يولد أيضا فراغا جزئيا وحرارة عالية. وهكذا فإن “كل كائن حي يتبخر حتما”، حسب أحد المسؤولين الروس.

الجدول (12- د) النص المصدر و ترجماته.

ج- ثقّتهم بتلك المعاجم و ما حوت، جعلتهم يوظفون في ترجماتهم كل ما وقعت عليه أبصارهم دونما تبصر في توظيفاتها الحقيقية، من منطلق أن هذه المفردة أو تلك التي وقع اختيارهم عليها، لم يكن إيرادها في المعجم عبثا، بل لأنها "مقبولة" ترجميا. و هي "المقبولية" ذاتها التي جعلت من ترجماتهم في الكثير من المواطن أشبه بمرصوفة، رُصّت فيها الكلمات جنبا إلى جنب دون مراعاة المعنى أو التركيب، كما نلاحظه في الترجمة الموالية:

النص:	كمقابل لـ:
« During its throwing the bomb, the bomber Topoliva TU 160 has show a strategy, followed a big explosion ».	"ثم أظهرت قاذفة قنابل استراتيجية من نوع توبوليف تي يو 160 أثناء رميها القنبلة على موقع اختبار، و تلا ذلك انفجار كبير".

هذا ناهيك عن الحذف و القص اللذين لا مبرر لهما، و كثير من الأخطاء اللغوية مثلما هي موضحة في الجدول.

د- لم يستطع بعض الطلبة إخفاء سخطهم و استيائهم من معاجمهم التي لم تسعفهم في إيجاد الحل الترجمي المناسب لكثير من الإشكالات، مما ولد لديهم ما تجوز دعوته بـ "الرهاب الترجمي". أدى إلى فقدان ثقّتهم بأنفسهم و حال دون إتمامهم العمل – كما هو ملاحظ في الترجمة الثانية و الرابعة –، أو أنه جعلهم يفصحون عن عدم رضاهم عما قدموه من ترجمات في الأغلب الأعم.

هـ- و فريق آخر، لما خانته تلك المعاجم و لم يجدوا ضالّتهم فيها، لجأوا إلى حلول ترجمية استحدثوها وفق ما يليه عليهم حدسهم، و رأوا أنها "مقبولة" إنجليزيا. مثال ذلك:

في ترجمتهم لمصطلح "أسلحة الضغط و الحرارة"، نجد:

pressure and thermic weapons, heat and pressure weapons, weapons of pressure and temperature, weapons of compression and temperature.

بدلا من المقابل الصحيح (Thermobaric Weapons)، و حجتهم في ذلك أن المعاجم التي بين أيديهم لا تتوفر على مدخل مركب كهذا، لذلك عمدوا إلى تركيبه بأنفسهم كيفما اتفق أو بحسب ما تراءى لهم، و وظفوه في ترجماتهم على أنه مصطلح إنجليزي مركب "مقبول" و "مستساغ". و قد قال بعضهم مبررا استعماله كلمة *Heat* مثلا: "أنا لم أوظف هذه الكلمة لأنها موجودة في القاموس، بل لأنني على يقين من أنها المقابل الإنجليزي لكلمة (حرارة)". و كذا كان تبرير متمرّن آخر لاستعمال كلمة « *Compression* » بدل « *Pressure* ».

و لما كان السؤال: "لكن لماذا وقع اختياركما على كلمة *Heat* بدل *Temperature*، أو *Compression* بدل *Pressure*؟"، أجابا أنهما فعلا ذلك بحكم ما أملاه عليهما حدسهما. "و هل حدث و أن صادقتما مثل هذا المصطلح أو ذاك من قبل؟"

أجاب أحدهما: "لا أذكر ذلك، و لكن يبدو لي مقبولا طالما أنه تركيب إنجليزي، وثمة الكثير من التراكييب التي تنسج على هذا المنوال".
صحيح أن بعضا من "الحلول" التي يبتدعها المترجم في كثير من الأحيان، قد تبدو مناسبة و ملائمة لا لشيء إلا لأن لها مثيلات في اللغة المترجم إليها، بيد أنه يجب أن يكون حذرا فيما يخص كثيرا من الاستثناءات على غرار المثال المذكور؛ لذلك، لا مناص من أن يعتمد إلى البحث و يجد في الاستقصاء حتى يتأكد و يستوثق، بدل التصرف بتلقائية و آلية.

2- الورشة الثانية:

لقد لاحظنا من خلال الورشة الأولى أن أعوص الإشكالات التي تنطرح في كل ترجمة، تتعلق أساسا بقضايا: المصطلحات و المتلازمات اللفظية و استعمال المفردات ضمن سياقاتها، و كذا التوظيف الشائع و المؤلف في اللغة المستهدفة.
و تلکم الإشكالات لم يوفق الطلبة في إيجاد حلولها و لا اهتموا إليها بمعونة أدواتهم التقليدية، بالنظر إلى أنه يستحيل أن نجد في تلك المعاجم – على بكرة أبيها – جردا بجميع المفاهيم المصطلحية لاسيما المستجدة منها، أو المتلازمات اللفظية التي – و إن عثرنا على بعض منها – فإنها قلما ترد ضمن مدخل مستقل.
هذا فضلا على أن مصنفي المعاجم يعمدون إلى جرد المفردات مجردة من سياقاتها الطبيعية، و لا يضربون أمثلة عن استعمالاتها إلا لماما؛ و إذا حدث و أن عثر مستخدمها على مثال، فإنه سرعان ما يعود خائبا صفر اليدين لأنه لم يجد ضالته، من حيث إن ذلكم المثال لا يتوافق و الحل الأمثل للإشكال الذي بين يديه.
بيد أننا في الورشة الثانية، لمسنا فرجا لهم البحث و انفراجا لغم الترجمة، بفضل استخدام ويكويرب، و يتجلى ذلك في الأمثلة الموالية.

① "قنبلة محمولة جوا": يستحيل أن نجد مقابلا لهذه التركيبية ضمن مدخل مستقل في أي معجم، لذلك عمد الطلبة إلى الترجمة كل عنصر منها على حدى، فكان أن حصل بعضهم على المقابلين: *a bomb lifted in the air, airborne bomb* و قد وظفوهما في ترجماتهم، و إن كانوا في ريب منهما و على غير قناعة بهما.
في المرحلة الثانية، أدخلنا وحدة البحث باستعمال الويلدكارد (*air * bomb*)، كما هو موضح في الشكل (41).

قام ويكويرب بتفحص 500 صفحة ويب باستخدام محرك البحث (Google)، فكان عدد الفهارس الناتجة 30 سطرا، من أمثلتها:

- of a 21,000lb bomb known as the massive ordnance **air blast bomb** or Moab. Far more powerful than the 15,000lb daisy
- British can be anything from rifles to a 1,000-pound **air-dropped bomb**. Afghan children often ask the soldiers for sweets The only
- shot into the sky in front of us, an Israeli **air-dropped bomb** that exploded on the road scarcely 80 metres away with

- also the ones that create widows in Afghanistan when their **air forces bomb** the south and the east of the country in the
- nicknamed the Mother Of All Bombs, is a large-yield satellite-guided, **air-delivered bomb** described as the most powerful non-nuclear weapon in history. Channel

WebCorp
Live

An improved version of the original WebCorp, designed to search the web for concordances in real time

Search term:

air * bomb

Enter a word, phrase (no quotes necessary) or pattern

الشكل (41) البحث بالويلدكارد.

من خلال الأسطر أعلاه، نلاحظ أن المثالين (03) و (05) يطابق سياقهما سياق التركيبية التي بين أيدينا، و هما يتضمنان المقابلين:

air-delivered bomb / air-dropped bomb

بإدخال هاتين الوجدتين كلاً على حدى، نحصل على نتيجة 15 سطر فهرست بالنسبة لوحدة البحث (*air-delivered bomb*) ضمن 10 صفحات ويب، و 4 أسطر بالنسبة للوحدة الثانية ضمن 4 صفحات.

تالياً، أراد أحد المتمرنين اختبار الوحدة (*airborne bomb*) و التأكد منها، فقام ويكروب بتصفح ست (06) صفحات باستخدام محرك البحث نفسه دائماً، و حصل المتمرن على أربعة أسطر، من أمثلتها:

- who, in the face of certain death, took quick action and successfully avoided having their aircraft possibly become the fourth **airborne bomb** on Tuesday. The memory of their sacrifice will remain with me for the rest of my life. I hope that
- it is a fine distinction to draw between death resulting from rocket attack, and death resulting from tank shell and **airborne bomb**. And for every dead Israeli citizen there have been at least ten dead Lebanese citizens, most of them children. Your

من خلال تلك الفهارس حصلنا على ثلاثة مقابلات محتملة للتركيبية (قنبلة محمولة جوا). وبالنظر إلى توارد كل مقابل على حدى، يتأكد لدينا أن المقابل الأنسب و الأكثر شيوعاً هو: (*air-delivered bomb*).

② في ترجمة العبارة "أعلنت روسيا أنها اختبرت بنجاح قنبلة جديدة"، كانت مقترحات الطلبة بالنسبة للفعل "اختبر" كالآتي: *tested, tried (out), examined.*

لملاحظة هذه الأفعال الثلاثة و التعرف إلى استعمالاتها، و كذا التأكد من تلكم الاختيارات الترجمية أعلاه، قمنا بإدخال وحدات البحث (*try a bomb*), (*test a bomb*), (*examine a bomb*) فحصلنا على النتائج التالية:
بالنسبة للوحدة (*test a bomb*)، أدى البحث ضمن 43 صفحة ويب إلى توليد 24 سطرا، و نذكر منها:

- North. Meanwhile Reuters quoted a Chinese official, briefed by Pyongyang, saying that North Korea was "more or less ready" to **test a bomb**. The Chinese official said the plan was to set off the device in an abandoned coal mine, deep underground, near
- Washington has pursued a policy of not demanding transparency from its close ally, and in return Israel agreed not to **test a bomb** or declare its nuclear capability - a policy of "strategic ambiguity". "As far as we are concerned, there is no change
- governments in the region to increase defense spending.' Defense Secretary Donald H. Rumsfeld warned Thursday that allowing North Korea to **test a bomb** would provoke far-reaching fallout. 'The lack of cohesion and the inability to marshal sufficient leverage to prevent North Korea from

- بإدخال وحدة البحث (*examine a bomb*)، حصلنا على ثلاثة أسطر ضمن 13 صفحة منها:

- the wreckage of a house hit by an Israeli air strike in Jabaliya, northern Gaza. Sri Lankan army explosives experts **examine a bomb** that was found on a bicycle in the capital, Colombo, and defused. A wrecked Hezbollah bunker on the Lebanese side
- police need to acquire the technology. 6: The forensic burden has become greater. It took two weeks before officers could **examine a "bomb" factory** in Yorkshire after 7 July Butler: "An argument which, on the face of it, appears to have some merit."

- أما بالنسبة للوحدة الأخيرة (*try a bomb*)، فقد حصلنا على سطر واحد فقط بعد البحث ضمن صفحتين ، و المثال هو:

- in the first half, Edwards sent Hall out to attempt a 56-yard field goal rather than let quarterback Vinny Testaverde **try a bomb** into the end zone. Hall's kick had enough distance but tailed to the left at the last moment. In the

بالمقارنة بين السياقات المقترحة ضمن كل فهرست، و السياق الذي بين أيدينا، نجد أن الخيار الأنسب هو (*test a bomb*) بمعنى التجريب و الاختبار، أما الوحدة الثانية فهي أنسب لسياق يتضمن معنى الفحص و المعاينة، في حين أن الوحدة (*try a bomb*) قد وردت بالمعنى المجازي في مجال الرياضة، كما يتوضح لدينا بالعودة إلى النص الذي استمد منه سطر الفهرست عن النيويورك تايمز، و الموسوم بعنوان:

"PRO FOOTBALL; Renewing Their Faith in Hall, Jets Revive Their Kicking Game."

③ و كذا هي الحالة بالنسبة لترجمة " نتائج التجارب"، إذ كانت مقترحات الطلبة كالآتي:

the experience results, the results of tests, the results of the probations, the results of the experiments.

و قد تساءل بعضهم أي التعبير التالية أصح:

Bomb experiment, bomb experience, bomb probation, bomb test ?

و عمدنا إلى البحث عن كل وحدة بواسطة ويكويرب.

- عند البحث عن وحدة (*bomb experiment*) ضمن 31 صفحة ويب كان عدد الأسطر الناتجة 20 سطرا، توافق سياقاتها السياق نفسه ضمن نص الترجمة، و هو ما نتبينه خلال الأمثلة الثلاثة:

- remembered Japanese survivor Matashichi Oishi wants government compensation Peace activists in Japan have marked the 50th anniversary of an atomic **bomb experiment** in the Pacific which killed one person and injured dozens more. The US test on tiny Bikini Atoll in the
- case, there was a huge, and as yet unexplained, explosion at the compound. It is believed that this was a **bomb experiment** that went wrong. In January came the first official reports of child abuse by cult members - mainly by Koresh, who
- that the Government was negligent in "fulfilling its active and ongoing duty to examine, monitor, control and conduct each atomic **bomb experiment**." The claim is essentially the same as that argued in the earlier case. Project Sedan, detonated on July 6, 1962,

- بالنسبة للوحدة (*bomb test*)، قام ويب كورب بتصفح 500 صفحة ليعطي 367 سطرا، توافق - هي الأخرى- سياقاتها السياق المطلوب، نحو:

- devotes so many of its scarce resources to missiles and nuclear weapons progress." The European Union presidency said the nuclear **bomb test** was a threat to regional stability and that it would work with the world's major powers for a "decisive international
- of mass destruction. No choice The US had taken the lead during World War Two with its first successful atom **bomb test**, but in 1949 the Soviet Union caught up, detonating its first A-bomb. Edward Teller, the scientist often labelled "the father
- dropped on Hiroshima. In 1954, again in the South Pacific, Dr. Rosenbluth was aboard a Navy destroyer when a hydrogen **bomb test** turned out to be unexpectedly strong and showered his ship with radioactive fallout. "It was pretty frightening," he recalled in

و بذلك صار لدينا احتمالان اثنان.

و قد اكتشف متمرر مصادفة أمثلة أخرى تدعم أكثر الاحتمال الثاني، حيث قلب بين مركبتي الوحدة فصارت (*test bomb*) أي "قنبلة اختبار"، و قام بالبحث عنها ضمن 121 صفحة فحصل على 73 سطرا، منها:

- commercial jetliners and tried to kill Americans traveling outside the United States. Mr. Yousef was convicted separately of planting a **test bomb** on a Philippine Airlines jetliner bound for Tokyo that killed a Japanese passenger in late 1994. After the trial, James

- the process, the Oak Ridge fuel was used. Later, fuel from Hanford was used to produce plutonium for the first **test bomb**, detonated in New Mexico in July 1945, and for the bomb that was dropped on Nagasaki, Japan, a month later.

- أما عند إدخال الوحدة (*bomb probation*)، فلم نعثر على أي نتيجة. و في محاولة ثانية باستخدام الوحدة (*probation*) عثرنا على 2002 سطر فهرست بعد تصفح 500 صفحة ويب. فتبين لنا أن هذا المصطلح قانوني، يظهر ضمن عدة متلازمات لفظية من قبيل:

On probation, probation department, probation officers, probation (and correction) commissioner, probation service, probation supervision...

و من أمثلة توظيفاته، نجد:

- addition to running decent prisons and jails." Dominic Coluccio, president of the United **Probation** Officers Association, which represents about 950 probation officers in the city, said that the merger might do his union some good. "Corrections gets a lot of money,"
- another, he said, and the Probation Department is chronically underfinanced. "And by pairing it with this 800-pound gorilla, it lays **probation** open to an even more tenuous budget situation," he said. Robert Gangi, executive director of the Correctional Association of New
- at John Jay College of Criminal Justice, who himself has held both positions at the same time between appointments as **probation** and correction commissioner. The Correction Department is always facing a crisis of one sort or another, he said, and the
- participating in a conspiracy to smuggle aliens into the country illegally and instead suspended their sentences and placed them on **probation**. But the atmosphere in Federal District Court surrounding Judge Earl ... How to Read the Full Article Get Free Archive Access
- Manchester and the West Midlands had no places left, says the study by the National Association of **Probation** Officers. The probation service in the West Midlands has suspended the bail information scheme, which helps lawyers to find places for their clients
- July 2, 2004 A 20-year-old college student who faked her own abduction in March was sentenced to three years of **probation** and must pay the Madison Police Department at least \$250 per month in restitution during her probation. The woman, Audrey

و هكذا، عمدنا إلى إلغاء هذا الاحتمال و إقصائه لأنه لا يتوافق و السياق المرغوب.

- في المقابل كانت نتائج البحث عن وحدة (*bomb experience*) سطرًا واحدًا هو:

- figures during my dot.com adventure, and then learning the true value of stock options as it became a dot.**bomb experience**. But I still believe my proudest moment is yet to come. The people who invest their careers or money in

و كما نلاحظ في المثال إن تركيبة (*dot.bomb experience*) تشكل وحدة واحدة، و قد دفع الفضول المتمرنين إلى إيجاد أمثلة أخرى عنها، ضمن مختلف مواقع الويب باعتماد ويكويرب عدا المدونة التي اعتمدناها في التجربة؛ فكان أن حصلنا على الأمثلة التالية:

- sustain their business model. Most of these companies were among the hundreds of companies that bellied up in the 2000 **dot bomb experience**. Don't blindly believe in your plan. Research others on the web who have undertaken similar business models and see if
- of 1929 are forgotten, the speculative excesses that led to that debacle will recur. I am sure that when the **dot-bomb experience** is forgotten, it will be repeated with some new class of speculation in some future generation. With the recent experience
- that this session will awaken the entrepreneurial spirit in you - the spirit that might have gone dormant through a challenging **dot-bomb experience**, or through an extended experience working in corporations which stifled innovation and initiative. Below are testimonials from workshop attendees: Linda
- of protection never hurts. , Ph.D. is in the process of looking for another infosec job after having his fun **dot-bomb experience**. He spends his plentiful pastime running Linux, building a small and preparing for CISSP. Copyright 2006, SecurityFocus

و المصطلح يستخدم بخاصة في عالم الاستثمار، و بالمعنى المجازي للتعبير عن تجربة فاشلة.

بالعودة إلى معدل تواتر الوحدات السابقة، نجد أن المفردة (bomb) تنزع كثيرا إلى الورود مع كلمة (test)؛ لذلك، فإن المقابل الأنسب لعبارة (نتائج التجارب) (test results) و قياسا على ذلك، فإن ما يكافئ عبارة (موقع اختبار) الواردة في النص، هو عبارة testing ground أو test (ing) site بدل (test location), (probation site) أو (examination spot)؛ و يمكن التحقق من ذلك باعتماد ويكويرب دائما:

- عند إدخال الوحدة (test site) و البحث ضمن 500 صفحة ويب حصلنا على 818 سطرا من أمثلتها:

- simulated nuclear explosion near the Nevada test site. American scientists conducted similar experiments in September 1987 near the Soviet nuclear-weapons **test site** in Kazakhstan. "This is a very important part of our work with American scientists to verify nuclear explosions of low
- also called for a second 10-ton explosion to be set off later tonight about 50 miles south of the Nevada **test site**. A third explosion is scheduled for noon tomorrow about 120 miles northwest of the site. The tests have several purposes,

غير أن الباحث يكتشف أيضا التركيبة (test-site)، بحيث تعمل عمل الصفة، كما في المثالين:

- have said they would drop their proposal if the \$150 million was added to the new trust fund and covered **test-site** workers. Members of the military who were affected by test-site radiation are not covered by the new trust fund or
- million was added to the new trust fund and covered test-site workers. Members of the military who were affected by **test-site** radiation are not covered by the new trust fund or the proposed addition to it because a 1988 law provided

④ (...) الجنرال ألكسندر روكشين نائب رئيس هيئة الأركان العامة للقوات المسلحة الروسية:"

إن هذه الجملة لها خير مثال على الخلط الذي وقع فيه الطلبة، لاسيما فيما يتعلق بترجمة الرتبة العسكرية، إذ لم يتمكن أحد من إيجاد الحل المكافئ، لأن قواميسهم ومعاجمهم أعجزتهم عن ذلك، فراحوا يرصون الكلمات بعضها إلى بعض، و كانت النتيجة:

- Army commander Alexander Roxin, vice president of plenary session of the Russian Army.
- General Alexander Rockchin, Deputy chief of General staff weapon of the Russian armed forces.
- Alexander Rokchin, vice head of General Pillars of Russian Armed Forces.
- general Alexander Rocchin, the vice-president of general staff organization of Russian armed forces.

و الحال هذه عمدنا إلى إدخال وحدة البحث (Alexander Rukshin) فحصلنا على 13 سطر فهرست ضمن 11 صفحة ويب، توفرت لدينا من خلالها ثلاثة خيارات هي:

- a- Alexander Rukshin, Russian deputy armed forces chief of staff.
- b- Alexander Rukshin, a deputy chief of the Russian armed forces staff.
- c- Alexander Rukshin, a deputy chief of the Russian Military's General Staff.

⑤ كمقابل للتعبير (غير نووي)، استعمل بعض المتمرنين الصفة (non-nuclear) و هو الاستعمال الجائز و الصحيح، بينما وظف بعضهم الآخر الصفة (unnuclear) و هذا خطأ لغوياً، و لعل مرد ذلك الجهل بكيفية اشتقاق هذه الصفة؛ بيد أننا لاحظنا إصرار الطلبة على فكرة أن هذا الاستعمال صائب. و الحال هذه، رأينا أنه لا بد من قطع الشك باليقين، و قمنا بإدخال وحدة البحث (unnuclear bomb)، فلم نعثر على أي نتيجة، مما يعني أنه خيار مرفوض.

في حين إن نتائج البحث باستخدام الوحدة (non-nuclear bomb)، كانت تعادل 35 سطرا ضمن 30 صفحة ويب، من أمثلتها:

- propaganda move, the Pentagon issued a thinly disguised threat to deploy - for the first time in the war - the biggest **non-nuclear bomb** in its arsenal, the 21,000lb massive ordnance air burst. The warning came as Allied forces continued the aerial bomb-
- explosions, many in the capital speculated that the US military had used the Massive Ordnance Air Blast (MOAB), the biggest **non-nuclear bomb** in the world. If the Allied troops were able to get any rest, the delay would have been welcomed. Correspondents

- for "mass ordnance air blast" or, unofficially, "mother of all bombs"; specifically a 21,000lb bomb which is the largest **non-nuclear bomb**; also has Biblical associations N new Iraq - Tony Blair's goal for the shattered Middle East state O OBE-sity - the giving

و قس على ذلك بالنسبة لـ (سلاح غير نووي)، إذ نجد أن المقابل الأنسب هو (non-nuclear weapon)، على غرار ما ورد ضمن سطر الفهرست أدناه:

- indicate that Article VI is a provision 'essential to the accomplishment of the object or purpose of the treaty.' The **non-nuclear weapon** states required commitments from the nuclear weapon states as part of their willingness to accept non-nuclear status under the NPT

⑥ في الأسطر الثلاثة التي حصلنا عليها بإدخال وحدة البحث (non-nuclear bomb)، عثر الطلبة بالصدفة على بعض المكافئات الترجمية التي يحتاجونها في ترجمة النص، مما دفع بهم إلى توظيفها و الاستغناء عن البحث، مثال ذلك: Massive Ordnance Air Blast (MOAB) كمقابل لـ: (قنبلة التفجير الهوائي الفائقة الكتلة). و هو الاستعمال نفسه الذي لاحظناه أثناء البحث عن الوحدة (air-delivered bomb)، حيث عثرنا على الفهرس الموالي:

- over global crises. The U.S. Massive Ordnance Air Blast, nicknamed the Mother Of All Bombs, is a large-yield satellite-guided, **air-delivered bomb** described as the most powerful non-nuclear weapon in history. Channel One said that while the Russian bomb contains 7.8 tons

و من المكافئات أيضا: "أم القنابل جمعاء" (the Mother of All Bombs).
و قد قام بعضهم باقتطاع الجملتين المفيدتين:

• 'The world most powerful non-nuclear bomb.'

• 'The most powerful non-nuclear weapon in history.'

و وظفوهما مع بعض التصرف كمكافئ للجملة: "أقوى سلاح غير نووي في العالم".

⑦ و في إحدى صفحات نتائج البحث، اكتشف الطلبة ضمن بعض الفهارس ما نصه:

-[...] Channel One television said the new ordnance, nicknamed the "dad of all bombs" is [...].

-State TV described it as the "dad of all bombs". [...]

حيث إن عبارة (dad of all bombs) تضيف احتمالا ترجميا آخر إلى الاحتمال (Father of All Bombs).

و من الصيغ و الجمل التي اقتطعها المتمرنون من الفهارس التي عثروا عليها، نجد الأمثلة التالية المأخوذة من صفحات نتائج البحث عن الوحدة (Alexander Rukshin):

- [...]its efficiency and power is commensurate with a nuclear weapon.

- [...]it is comparable with nuclear weapons in its efficiency and potentiel[...]
- [...]is comparable to a nuclear weapon in is efficiency and capability[...]

كمكافئات لـ: " تعادل قنبلة نووية في فعاليتها و كفاءتها". و الجملة:

It has, he says, « no match in the world ».

كمكافئ لجملة: "لا مثيل لها في العالم".

و كذا الجملة: [...] Unlike a nuclear weapon, the bomb doesn't hurt the environment.

كمكافئ للجملة: "السلاح الجديد لا يشكل تهديدا للبيئة بخلاف الأسلحة النووية".

⑧ و ثمة من المفردات و التراكيب ما أرق كثيرا المتمرنين لعدم يقينهم من صحة استخدامها، على غرار (Pressure wave) كمقابل لـ (موجة ضغط)، فلم يجدوا بدا من اللجوء إلى ويكوب كى يستوثقوا و يتأكدوا من هذا الاستعمال، و كان أن حصلوا على نتيجة 119 سطر فهرست ضمن 144 صفحة بحث، تؤكد خيارهم المبني على الحدس. و كذلك هي الحال بالنسبة لمصطلح (المتفجرات التقليدية)، إذ احتاروا بين (Traditionnel explosives)، (Conventionnel explosives) و (Old explosives).

و بالعودة إلى نتائج البحث باعتماد ويكوب، لاحظ الطلبة إمكانية توظيف الاحتمالات الثلاثة، و إن تفاوتت نسب شيوعها، بحيث حصلنا على 487 سطر فهرست ضمن 500 صفحة فيما يخص (conventionnel explosives)، و 22 سطرا ضمن 20 صفحة فيما يخص (traditionnel explosives)؛ أما فيما يتعلق بالاحتمال الأخير، فقد حصلنا على 14 سطرا من 38 صفحة.

⑨ و من الاحتمالات التي اعتمدنا على ويكوب في التحقق منها، نذكر: ما اعتقدناه من مكافئات المصطلح (القنبلة الفراغية)، أي (vacuum bomb) و مصطلح (empty bomb). فمن أمثلة الفهارس التي وجدناها بالنسبة للمصطلح الأول، نذكر:

- where the children had been lying before the blast, and the remains of human intestines lying on the ground. The **vacuum bomb** is dropped by a parachute. As it falls to the ground, it releases a cloud of petrol vapour, which ignites,
- It showed a Tupolev 160 strategic bomber dropping the bomb over a testing ground. A large explosion followed. The aviation **vacuum bomb**, which is also known as a fuel-air bomb, was the mightiest ever created, it added. Last night's announcement comes at
- flattened out to make a mushroom cloud. The partial vacuum at ground level sucked in more air - hence the name '**vacuum bomb**' sometimes given to these weapons. An SAS patrol is reported to have mistaken the cloud for a nuclear attack, and
- euphemism beloved of the military men - must be kept to a minimum. (Britain, by developing a version of the thermobaric "**vacuum bomb**" beloved of the Russians in

Chechnya, appears to disregard such concerns.) Nato boasted during the Kosovo war of its glasnost,

أما بالنسبة لمصطلح (empty bomb)، فقد عثرنا على أمثلة من قبيل:

- Afghanistan. On Monday a Nimrod surveillance aircraft sent out a mayday call after crew members spotted fuel spraying into an **empty bomb** bay. The crew's log, leaked to the BBC, reported the bomb bay doors "to be wet with fuel". The aircraft
- forest have an erotic undercurrent, some have a darker feel. Take for instance Txernóbil (Chernobyl), a monument built from gigantic **empty bomb** shells, with a sea of large pebbles scattered at its feet, some carved into distinguishable skulls, some left smooth and
- a base in Cyprus. "We threw out everything we could tear loose from the ship," he said. "We ditched one **empty bomb** bay tank, our heavy flying clothes, all our guns that were shot up, frequency-meter tools, a ladder, everything except food,

و قد تبين لنا أنه ثمة فرق بين المصطلحين في المعنى، بحيث إن المصطلح الأول يقابل مصطلح "القنابل الفراغية" (و هي التي يولد انفجارها فراغا جزئيا)، و مصطلح (empty bomb) يقابل مصطلح "القنابل الفارغة" (و تستخدم لحمل القنابل الكيماوية).
 و في بحثنا عن الوحدة (vacuum bomb)، عثرنا على فهارس أفادتنا في إيجاد كثير من الحلول الترجمية التي يستدعيها النص، مثال ذلك:

[...] the vacuum bomb is dropped by a parachute.

فالفعل (drop) مرادف للفعل (throw)، و قد احتفظ به الطلبة لتوظيفه في النص.

و في الفهرست نفسه، نجد:

As it falls to the ground, it releases a cloud of petrol vapor, which ignites.

و الفعل (ignites) يكافئ " تنفجر (الغيمة بمكوناتها في الجو)".

و في فهرست آخر، وجدنا:

"by developing a version of the thermobaric (...)"

و المصطلح (thermobaric (weapons) يكافئ مصطلح (أسلحة) الضغط و الحرارة، بدل ما وظفه الطلبة على غرار:

Pressure and thermic weapons, heat and pressure weapons, weapons of compression and temperatures.

⑩ من ناحية أخرى، أفادت المدونة الطلبة كثيرا فيما يتعلق بالبحث عن المختصر الوارد في النص (تي يو 160)، من أمثلة النتائج المحصل عليها:

- the Soviet Union's proxy battles with the United States in the region during the Cold War. Last week, two Russian **Tu-160** strategic bombers flew to Venezuela for exercises over the Caribbean Sea, and a Russian delegation led by Igor I. Sechin,
- its 1989 report on Soviet military power, the Pentagon warned about the deployment of the Blackjack bomber, the Russian supersonic **Tu-160**. With low-mounted, swept-back wings and a long, pointed nose, the plane was the most powerful combat aircraft in the Soviet
- navy's strike force in the Atlantic," the Russian air force said in a statement reported by Reuters. "Today, two strategic **Tu-160** bombers departed for exercises in the Bay of Biscay, which ... will carry out a number of missions and will conduct

و ما يؤكد هذه الأمثلة هو سطر الفهرست الذي عثر عليه أحد المتمرنين، و هو يتفحص صفحة نتائج البحث عن الوحدة (Vacuum bomb)، و قد استفاد منه مع بعض التصرف في ترجمة العبارة: "قاذفة قنابل إستراتيجية من نوع توبوليف تي يو 160..."
[...] It showed a Tupolev 160 strategic bomber dropping the bomb [...].

و في الفهرست نفسه، عثر على مصطلح (fuel-air bomb) أي (قنبلة الوقود الهوائي)، ثم قام بإدخال وحدة البحث (fuel-air)، فتحصل على 451 سطرًا ضمن 500 صفحة، من أمثلتها:

- Grozny Patrick Cockburn in Moscow Friday, 26 November 1999 THE RUSSIAN air force may plan to use powerful "vacuum" or **fuel-air** bombs in Chechnya for the first time in the present war. The bombs explode above ground but exert enough pressure
- Russia's main TV channels carried reports on the test Launch and aftermath The Russian air force has tested a giant **fuel-air** bomb which the military says is the biggest non-nuclear explosive device in the world. Russian TV showed a Tupolev bomber
- stages: a small blast creates a cloud of explosive material, which is then ignited with devastating effect. Graphic: How a **fuel-air** bomb works The name Moab officially stands for Massive Ordnance Air Burst but has unofficially been interpreted as Mother Of
- agent in their explosives. They produce more energy than normal weapons but are hard to control. The US used similar **fuel-air** munitions to clear jungle for helicopter landings during the Vietnam War. The Soviet Union also developed its own fuel-air weapons,
- 1984 Jane's Defense Weekly reported today that Soviet planes are dropping "liquid fire" bombs on Moslem guerrillas in Afghanistan and **fuel-air** explosives that set off shock waves that kill anyone within a quarter-mile. The magazine, issued by the authoritative Jane's Publications,
- was written by Yossef Bodansky, identified by Jane's as a consultant to the United States Defense and State Departments. The **fuel-air** explosive bombs detonate in the air, releasing a volatile chemical cloud about 15 yards in diameter. That, in turn, is
- physics teachers shows that even custard powder is explosive under the right conditions. The effect is referred to as a **fuel-air** or thermobaric ("heat pressure") explosion, and the shockwave from it is unusual in two ways. The extremely high-speed burning means
- pressure on Taliban and Al Qaeda fighters. On Sunday, for the first time, the Air Force dropped a 15,000-pound **fuel-air** explosion bomb on Taliban positions behind the front lines. The weapon, a "Daisy Cutter," was used in the last year

و كانت "سعادته غامرة" حين وجد الكثير من الاحتمالات الترجمية التي وظفها في صياغة النص الانجليزي، نذكر منها:

(fuel-air munitions) كمقابل لـ (ذخائر الوقود الهوائي).

Thermobaric (weapons) ← (أسلحة) الضغط والحرارة.

و الفعل (detonate) كمرادف للفعل (explode).

كل هذا فضلا على ذلكم الكم الهائل من المصطلحات التي استخلصها الطلبة من أسطر الفهارس، على غرار:

Timebomb, car bombs, suicide bombers, long-range bomber, strategic nuclear bomber, "Blackjacks" supersonic bomber, the White Swan, TU-160

Blackjacks, TU-95 Bears, the TU-160 jet nuclear bomber, Dirty bombs, Ordnance, lethal/nonlethal weapons, homemade fertilizer bombs, anti-tank weapons, Daizy Cutter, Bunker-buster, empty Bombay, WC-135 reconnaissance planes, microwave bomb, Tomahawh cruise missile.

و من التعابير المفيدة التي أضافها الطلبة إلى رصيدهم، قصد توظيفها في ترجمة النص الذي بين أيديهم أو نصوص أخرى:

"The US Massive Ordnance Air Blast nicknamed (dubbed/christened) the Mother of All Bombs".

و منها أيضا:

"It was the latest show of Russia's military muscle amid chilly relations with the United States".

ما يكافئها عربيا:

"هي [أي القنبلة] الأحدث في استعراض القوة العسكرية الذي تقدمه روسيا في خضم علاقتها الباردة مع الو. م. أ." و منها عبارتا:

- *It is a fine distinction to draw between ...*
- *The memory of their sacrifice will remain with me for the rest of my life.*

II-2-7- تعليقات الطلبة و تساؤلاتهم:

- "عندما أنظر في القواميس و المعاجم، أحتار في جميع تلك المقابلات الموضوعة وأتساءل أيها أختار، و قلما أوفق في اختيار الأنسب منها".
- "المدونات – بخلاف القواميس – تعطينا المفردات ضمن سياقاتها الطبيعية، مما يجعلنا واثقين من توظيف أو استعمال دون آخر بحسب ما يتوافق و المقام في ترجمتنا".
- "لقد تعرفت إلى كثير من القوالب و الصيغ التي تميز اللغة الإنجليزية، فضلا عن بعض التعابير المفيدة التي يمكن أن أوظفها مستقبلا".
- "مكنتني المدونة من معرفة مدى شيوع مفردة على حساب أخرى، و لما تأكد لدي ذلك بالنظر إلى الكم الهائل من النتائج، رأيت أن لابد من توظيفها".
- "إنني أشعر بالرضا عن الترجمة التي حصلنا عليها باعتماد وبيكورب، لأنني أراها أقرب ما تكون إلى الأصول في اللغة الإنجليزية".
- "تعجبني فكرة استعمال المدونة لأنها – في رأيي – أداة هامة لإثراء الجانب اللغوي لدي، و أنا على يقين من إمكانية تحسين رصيدي من اللغة الإنجليزية أو غيرها".
- "المدونة مفيدة لأنها تساعد في اتخاذ القرار الترجمي".
- "أنا أثق في اختياراتي باعتماد المدونة، لأنني على يقين من صحتها و ملاءمتها، طالما أنها من إنتاج متكلم / كاتب أصلي في اللغة الإنجليزية".
- "لا شك أن المدونة ستفيدني مستقبلا في ترجمات أكثر تخصصا، بالنظر إلى أن مادتها النصية من وضع متخصصين في المجال موضوع الترجمة".

- "تكمّن المتعة الكبرى في اكتشاف ما لم تكن تتوقعه، و هو اكتشاف إما جديد لا عهد لك به من قبل، أو أنه يأتي لتصحيح تخمينات أو معطيات خاطئة، و تقديم نظرة جديدة صائبة."

- "ذلك الكم الهائل من المعلومات الذي توفره المدونة بمجرد لمسة أو اثنتين، يجعلني أشعر بارتياح كبير لم أعده من قبل و أنا أتصفح القواميس بلا فائدة ترجى في كثير من الأحيان."

- "الواقع أنني - في البداية - كنت في شك من جدوى العمل باعتماد ويكوروب، لكن شكوكي سرعان ما تبددت و تبدلت يقينا بأهميته."



من خلال تعليقات الطلبة أعلاه، لمسنا إقرارا بالنقائص التي تعتور المعاجم المعتمدة، و سجلنا تفاؤلا كبيرا و ارتياحا عظيما لما حصلوا عليه من نتائج باعتماد ويكوروب، الذي ابتذر في أنفسهم الثقة و خلصهم مما كانوا فيه من "رهاب ترجمي". و يتجلى ذلك خاصة من خلال نبراتهم و العبارات الواردة في تعاليقهم:

"يجعلنا واثقين..."، "تأكد لدي..."، "أنا أثق في اختياراتي..."، "أشعر بارتياح كبير..."، "أشعر بالرضا..." و قد بدا جلهم متحمسين لفكرة توظيف المدونات مستقبلا، سواء بغرض التحسين اللغوي أو الترجمة.

بيد أننا سجلنا امتعاض طالب من استخدام ويكوروب أو المدونات بصفة عامة، و تساءل عن جدوى ذلك إذا كان ما توفره من معطيات يمكن الحصول عليه في مجموعة من النصوص / المقالات التي يمكن تنزيلها من النت.

و كان جواب زملائه (من المؤيدين) أن استخدام ويكوروب يُسرّع من نتائج الحصول على تلك المعطيات، و هي نتائج أدق مقارنة بغيرها، فضلا على أن مثل هذه المدونة أضخم من مجموع تلك المقالات.

إن ما يذهب إليه هؤلاء صحيح من وجهين:

- يقتضي جمع المقالات / النصوص في المجال ذاته وفق الطرق "التقليدية" جهدا جهيدا، و يستغرق وقتا كثيرا، و النتيجة: لم يتوصل الباحث إلا إلى نزر يسير من تلك النصوص؛ هذا إن لم يعد منها بخفي حنين، إذا كانت لا تتوفر على المعطيات المطلوبة وفق النسبة المرغوبة.

- بالمقابل، يفوق عدد النصوص التي تتضمنها المدونة "الإلكترونية" المئات، قد جمعت في مدة زمنية قياسية، و هي بذلك تعطي المترجم فرصة الحصول على معطيات أكبر مما كان يتصور. أضف إلى ذلك، فإن تلك المدونة ذاتها تخدمه في ترجمات أخرى ضمن المجال ذاته.

و هنا انقدح تساؤل آخر: لماذا لا نستخدم ذاكرة ترجمية بدل مدونة؟ و السؤال وجيه و في محله، أما الإجابة عنه فتكمّن في الفرق بين الأدوات؛ بحيث تصنع الذاكرة الترجمة ربحا للوقت الذي يقضيه المترجم - إن لم نقل يضيعه - في العمل على نصوص و ترجمات تتكرر معه في كل مرة، و في ذلك فائدة براغماتية أيضا من حيث ربح المال.

و هكذا، فإن تلكم الذاكرة أداة من أدوات المترجم "الاحترافي"، و هي محدودة لأنها "ظرفية" – أي تخدم ظروفًا معينة –، أكثر من أنها تلائم أبحاث الفئة الأخرى من المترجمين.

و في سياق آخر، قال طالب إنه و إن كان يحترم ما يذهب إليه رفاقه من رأي، إلا أنه يحبذ استخدام القواميس و المعاجم، بحكم "تعوده على ذلك". و هو يرى أن التغيير ينبغي أن يكون على مستوى الصنعة المعجمية، و ذلك بتطوير القواميس المتوفرة، بدل استحداث طرق جديدة في استنباط المعلومات و المعطيات تقتض كفاءة تكنولوجية هامة. الحق أن فكرة "التعود" تلك ليست مبررا كافيا بسببه يعزف (الطالب-) المترجم عن اكتساب "عادات" أخرى تخدمه من حيث لا يدري. ثم إنه لا شك في أنه سيأتي زمن تغدو فيه المدونات ألمع أدواته و أهم معيناته، تماما كما كانت القواميس و المعاجم، و ستصير لديه "عادة" يصعب إقلاعه عنها، و تلك سنة لن تجد لها تبديلا.

II-2-8- خلاصة التجربة:

توسمنا من خلال هذه التجربة تبيان أهمية توظيف مدونة اللغة المستهدفة في تدريب طلبة الترجمة، و هذا من حيث دعم كفايتهم في تلكم اللغة و تحسين أدائهم الترجمي، مقارنة بالأدوات المتعارف عليها. و فيما يلي نتائج ورشتي الاختبار: من خلال الورشة الأولى، تأكد لدينا أن طالب الترجمة يفتقر إلى اللغة الإنجليزية و مواضعاتها فقرا مدقعا، يجعله ضيق الأفق الترجمي قليل الحيلة؛ مما يدفعه إلى الاعتماد كلية على قواميسه و معاجمه، و توظيفه في ترجماته مقابلات مباشرة يراها مكافئات ترجمة مقبولة و مستساغة، أو تبعا لما يمليه عليه حدسه. و هكذا تزل به قدمه في كثير من المواطن، مما يشوه نتاجه، فيأتي بخسا رديئا إن لم يكن بالأصل لصيقا، موسوما بسماته و مواضعاته.

و في حال ما لم تسعف الطالب تلك العدة، رأيته ساخطا ناخطا عليها، مهزوز الثقة بقدراته الترجمية، فيقدم ترجمة فيها من الفجوات و "الفراغات" ما يشابه تلك الثغرات الملحوظة في كفاءته، و المنعكسة سلبا على أدائه.

أما من خلال الورشة الثانية، فقد صح لدينا ما افترضناه في بداية التجربة من أن استخدام المدونات إلى جانب برمجيات المعالجة في التدريب، يساعد على سد تلك الثغرات ما أمكن، إن على مستوى الكفاءة اللغوية لدى المتمرن، أو على مستوى نوعية نصوصه الناتجة؛ من حيث إنها توسع أفقه الترجمي بهديه إلى العديد من الحلول لكثير من الإشكالات، و تزويده باستراتيجيات تنمي لديه القدرة على الانتقاء و الفرز و الترشيح و الغرلة.

و الطالب أحوج ما يكون إلى تلك القدرة، لاسيما بالنظر إلى تلكم الكثرة الكاثرة مما يخاله من المكافئات الترجمية المباشرة، المبثوثة في أعطاف قواميسه و ثنايا معاجمه، و التي ينزع إلى توظيفها كيفما اتفق أو بدا له، و بحسب ما يوحي إليه حدسه. و لقد ثبت لدينا – من خلال الورشة – أن المدونة تساعد المتمرن (-المترجم) على اتخاذ القرار الصائب فيما يتعلق باعتماد خيار من تلكم الخيارات و التحقق من ملاءمته – لاسيما بالنظر إلى موافقته للسياق تمام التوافق –، أو رفضه مع اقتراح الحلول البديلة،

و للمتمرن أن يتخير منها ما يشاء آمنا مطمئنا، لأنها موثوقة صادرة عن متكلمين أصليين باللغة المنقول إليها.

و على صعيد آخر، تأكد لدينا أن الكلمات تميل إلى التوارد مع كلمات أخرى و الاقتران بها، بحيث تستمد من بعضها بعض معانيها، مما يجعلها ثابتة محددة، يفترض بالطالب (-الترجم) أن يحصيها، فإن هو عجز عن ذلك أو أعجزته موارده، كانت له المدونات خير معين في تصيدها و الاهتداء إليها.

من ناحية أخرى، تزخر المدونات بـ "الوحدات المعنوية" المثبت استعمالها لدى أهلها، و نقصد بها ما تعارفوا عليه من طرق في الإفصاح عن المعاني، فصاغوها في قوالب صارت لديهم مألوفة، و تعابير أضحت عندهم معروفة. و لقد رأينا كيف اهتدى الطلبة إلى اقتناص تلكم الوحدات و اجترائها من فهارسها، و المؤلفات بينها لتوظيفها في ترجمتهم، توظيفا وقاهم شر الوقوع في شرك الترجمة الحرفية، و كفاهم نائبة هيمنة الأصل و التعلق به.

فضلا على ذلك، عثر الطلبة فيما عثروا على دُرر مصطلحية، ما كان لهم أن يظفروا بها في مواردهم التقليدية، فعمدوا إلى استخلاصها و تجميعها و تدقيقها، وادخروها بنى العودة إليها كلما دعت الضرورة الترجمية، سواء اليوم أو غدا. و لم يكن حظ المتمرنين من المدونة استخلاص المصطلحات فحسب، بل صارت لهم بعض المعرفة بخصوص الموضوع، و تولد لديهم أنس بالأسلوب و ألفة بالسجل وفقا لمحددات اللغة المستهدفة. كما تبين لهم أن تلكم المدونة أشبه بـ "أجمة" لغوية، كلما درجوا في مدارجها و سلكوا مسالكها، اكتشفوا بالصدفة صدقا جديدا و اقتنصوا صيدا فريدا.

إن ما وسمناه في الورشة الأولى بـ "الرهاب الترجمي" لدى طلبتنا، قد تبدل في الثانية "ثقة بأنفسهم"، دفعتهم إلى ترجمة النص بنقاطه و فواصله، ترجمة لا يلفى قارئها أثرا لتلكم الفجوات و "الفراغات"، و لا يَشْتُمُ فيها رائحة الأصل و ما ينجر عنه من تشويهات.

و تلكم الثقة و إن كانت عمياء، إلا أنها تختلف كمّا و كيفا عن ثقتهم بقواميسهم و معاجمهم، لأن لها ما يدعمها و يؤكدّها بما لا ريب فيه؛ بحيث إنها وليدة النظر فيما سلكه الغير من مسالك، و هو ابن لغته عارف بقواعد كتابتها، واع لدقائقها و تفاصيلها، فيعمد أولئك تارة إلى تقليده بحيث لو سلك جحر ضب لسلكوه، و يعمدون تارة أخرى— إلى أعمال الفكر و تقليب النظر فيما توفر لهم عنه من استراتيجيات و حلول، ليتخيروا أوجه الاستفادة منها، و يكيّفوها بحسب مقتضى الحال.

خاتمة

خاتمة

تهتم لسانيات المدونة بدراسة اللغة انطلاقاً من المدونات، و هي بناء كبير من النصوص المحفوظة الكترونياً، و المنتخبة وفق معايير خارجية و أخرى داخلية يحددها الهدف من إنشاء المدونة، مع اعتبار كل نص بمثابة عينة تمثل السلوك اللغوي، سواء في شكله المكتوب أو المنطوق. بمعنى أن هذه المقاربة اللسانية تجريبية وصفية، مادتها أمثلة الاستعمال اللغوي الحقيقية و المثبتة؛ و أدواتها برمجيات خاصة تتيح للباحث المعالجة الآلية لتلك المادة، من حيث إنها تنظمها و تهيئها بشكل ييسر عليه معابنتها و دراستها (الفهارس الآلية)، كما أنها تضبطها في شكل معطيات إحصائية (قائمة الوصلات، جدول التكرار).

و لقد عرفت هذه المقاربة كيف تشق طريقها نحو الترجمة، لتؤدي إلى ظهور الترجمات باعتماد المدونات التي تعنى بالترجمة نتاجاً و عملية على حد سواء، قوامها الدراسات التي أنجزها أنصار المقاربة الوصفية في الترجمة (لاسيما أشياح النظرية المتعددة الأنساق، أمثال ابن زهير، جدعون توري)، و كذا الدراسات باعتماد المدونات التقليدية (سنيل هورنبي 1988).

و المقاربة المدوناتية تهتم بالنص-الترجمة، باعتباره "حدثاً تواصلياً وسيطاً"، و تسلط الضوء على الشموليات الترجمية، أي تلك الخصوصيات التي تتفرد بها لغة الترجمات جميعها و تميزها عن لغة الأصول؛ هذا عن جانب البحث و التنظير في الترجمة.

أما فيما يتعلق بالجانب التعليمي، فإن لتلك المقاربة من الأهمية ما يستدل عليها من واقع أن الكفاءة الترجمية المعاصرة تستوجب معرفة بكيفية صناعة المدونة و أنواعها و استخداماتها؛ هذا فضلاً على ضرورة التحكم في مختلف أنواع البرمجيات و آليات البحث باعتمادها.

و الحق إنه لمن السهل التحكم في المفهرسات الآلية، لاسيما فيما يتعلق بأبسط وظائفها، إذ بعد حصص تدريبية معدودة، بإمكان كل متمرّن استخدام مفهرس لاسترجاع الكلمات أو المصطلحات أو الجمل، و فحص أنماطها التكرارية محل الاستعمال.

إن استخدام المدونات في تدريب المترجم نابع من قناعات عدة لا مباحكة فيها و لا مشاحة، لعل أهمها اثنتان:

- ❖ إن المرء يسلك سلوكاً معيناً، أو يتعلم كيف يتصرف حيال وضعية ما، عن طريق النظر في سلوك غيره و تصرفاتهم حيال الوضعية نفسها. و قس على ذلك مثل (الطالب-) المترجم و هو أمام النص، إذ لا يدري كيف يتصرف معه على النحو الذي ينبغي، أو أي سلوك أمثل يسلك، إلا إذا نظر فيما كان من سلوك غيره الترجمي. و هذا يقودنا إلى القناة الثانية التي مفادها:
- ❖ إنه ما من حلول ترجمية سابقة تقدمها الطرائق النظرية لصاحبنا على طبق من ذهب؛ بل إن كثيراً من الحلول الممكنة و الاستراتيجيات -في أغلبها الأعم - متناثرة في ثنايا نصوص سابقة الإنتاج، أصولاً كانت أم ترجمات.

و الحال هذه، ثمة صنفان رئيسان من تلكم المدونات، يعتمدان في تدريب المترجم: مدونات أحادية اللغة (و تكون إما في لغته الأم، أو اللغة الأجنبية)، و مدونات ثنائية اللغة متوازية و أخرى مقارنة.

و بعضهم يعتمد إلى مقارنة هذين الصنفين بالقواميس و المعاجم الأحادية أو الثنائية اللغة كل حسب عدد اللغات المتضمنة، إلا أنه ثمة فرق بين هذين الضربين من الموارد؛ إذ بينما ينتهج مستخدم القواميس و المعاجم مقارنة تركيبية في الحصول على المعاني عن طريق التعريف، فإن مستخدم المدونات يتبع مقارنة تحليلية و ذلك بالنظر في مختلف السياقات المقترحة.

و لعل أشهر التجارب بخصوص استثمار المدونات في تدريب المترجم، كانت تتم بواسطة المدونات المتوازية و المقارنة، سواء العامة أو المختصة.

أ- المدونات المتوازية: إذا كانت الترجمة المثلى تقتضي اقتناص المعنى في النص المصدر، و التعبير عنه في اللغة المستهدفة بما يتوافق و مواضعها، فإن للمدونات المتوازية الشيء الكثير لتقدمه للمتمرن لتحقيق هذه المنية، و هو ما نتبينه من خلال الفرق بينها و بين المعاجم الثنائية اللغة؛ بحيث إذا كانت تلكم عبارة عن مسارد تتضمن مكافئات معجمية (المعاجم العامة) أو المصطلحية (المعاجم المختصة) باعتبارها احتمالات ترجمية؛ فإن المدونات المتوازية بمثابة مسارد تتضمن تلك المكافئات، يضاف إليها الاستراتيجيات التي اتبعها المترجم.

و لانتقاء المكافئ المستمد من المعجم، ينبغي معاييرة مدى ملاءمته للسياق، فليجأ المتمرن إلى تعريفه الذي غالبا ما يكون موجزا مقتضبا، مرفقا ببضع أمثلة عن توظيفاته، قد لا تفي بالغرض. بخلاف ذلك، تضع المدونة بين يدي المتمرن كماً هائلاً من الاستراتيجيات التي اعتمدها المترجم و الحلول التي أوجدها، عندما واجه الإشكالات الترجمية نفسها. كما أن هذا النوع من المدونات يعطي المتمرن فكرة عن كيفية تصرف المترجم في حالة عدم العثور على مكافئ مباشر.

ثم إن استخدام برمائية خاصة بمعالجة المدونات المتوازية - على غرار: براكونك (ParaConc)-، يتيح له فرصة موقعة جميع تواردات أي وحدة بحث (كلمة، عبارة، جملة) ضمن اللغة المصدر، و ما يكافئها في اللغة المستهدفة.

ب- المدونات المقارنة: هي مدونات قد حررت مادتها النصية مستقلاً بعضها عن بعض، أي أنها مادة أصلية "طبيعية" في كل لغة، غير أنها متماثلة من حيث المحتوى و المجال و الوظيفة التواصلية؛ بمعنى أن الظروف و الملابسات الأصلية التي تولد في خضمها كل نص، قد نتج عنها وحدات و أفكار و مفاهيم متماثلة عرضاً، يمكن للمتمرن أن يقارن بين سلوكاتها في اللغتين.

ثم إن تحليل النصوص نتاج وضعيتي تواصل مقارنتين، يتيح له فرصة اكتشاف توقعات الجمهورين بشأن الخصوصيات اللسانية ضمن النصين المصدر و الترجمة في جنس كتابي معين، فضلاً على خبرة كل جمهور و معرفته.

و لا يعتقد أن هذه المدونات تعطي مقابلات ترجمية مباشرة، بل إنها تقدم مسرداً بالخصوصيات التكرارية، التي تعد أساس إقامة علاقة التكافؤ بين المقاطع النصية المقارنة. في هذه الأثناء، يمكن للمتمرن أن يتعرف إلى كيفية تشكل الخطاب

في كل لغة، بما يضمن له تحسُّن فهمه للنص المصدر، و تطوُّر قدرته على إنتاج نص مكافئ في اللغة المستهدفة، بفضل ما استخلصه من شواهد بصفتها الاحتمالات الأنسب لإدراجها في الترجمة، وفق ما تقتضيه مواضع ثقافة المتلقي، إن على المستوى اللغوي أو على مستوى الجنس الكتابي.

فضلا على تلكم الأهمية و نواحي الاستفادة من المدونات المتوازية و المقارنة، يمكن للمتمرن أن يستخدمها في مراجعة ترجمته، باعتبارها مرحلة هامة من مراحل العمل الترجمي.

أما دراستنا، فقد تمثلت في محاولة اختبار فعالية المقاربة المشار إليها في التدريب على الترجمة من اللغة العربية إلى اللغة الانجليزية. و قد وقع اختيارنا على هذه الثنائية اللغوية، لإيماننا بأنه إذا كانت جل الدراسات السابقة تؤكد على أن توظيف المدونات في الترجمة يساعد على تحسين الإنتاجية، و تفادي كثير من الأخطاء المرتكبة عند الترجمة إلى اللغة الأم، فإنه عند الترجمة انطلاقا منها إلى لغة أجنبية، تزداد الحاجة إلى تلكم المدونات أكثر فأكثر.

و لقد اعتمدنا على مدونة أحادية اللغة، نصوصها أصلية في اللغة الانجليزية، إلى جانب برمائية خاصة للمعالجة (ويكورب). فخلصنا إلى أن المدونات تشكل موردا رئيسا من موارد (الطالب-) المترجم، و أداة هامة من أدواته، تعطيه من الحلول ما يجعله يتوصل إلى نصوص "ترجمة" عديدة، تختلف مبنى و تتفق معنى. و هي نصوص تتسم بالطبيعية، لأنه قد تحرر من الأصل، و قد تحرر فيها النحو الأمثل للتعبير عن المعاني في اللغة المستهدفة، حسب ما هو متعارف عليه عند أهلها، و جرى على أساليبهم في ذلك.

و السر أنه بخلاف عدة المتمرن التقليدية، تتيح له البرمجيات فرصة اكتشاف كيفية/كيفية توظيف وحدات البحث في مختلف سياقاتها دفعة واحدة من خلال الفهارس، سواء كانت تلك السياقات مجتزأة أم أنها كاملة (بالنقر على الوحدة)، هذا فضلا على إمكانية التعرف إلى الكلمات التي تنزع إلى التوارد معها و الاقتران بها، من خلال قائمة الوصلات. أما ما كان من النتائج الإحصائية التي تتعلق بتكرار تلكم الوحدات، فتفيدة في الكشف عما إذا كان استعمالها لا يعدو استعمالا شخصيا، أم أنه المستساغ لدى أهله الشائع بينهم.

و على الرغم مما تحظى به المقاربة -اليوم - من اهتمام الباحثين و المدرسين على حد سوى، إلا أنه ثمة بعض الأصوات المعارضة ترى أن المدونات تدفع بمستخدمها (باحثا كان أم متمرنا) إلى البحث فقط عما يسهل العثور عليه، أي أنه لا يلحظ إلا ما يرغب في لحظه، مهملا ما تبقى. بيد أن تجربتنا في صف الترجمة؛ قد أكدت لنا بما لا ريب فيه أن البحث في المدونات يؤدي إلى اكتشافات بالصدفة، و حلول لإشكالات لم يكن يفكر في طرحها المتمرن قبل أن تقع عينه على حلولها.

و من تلكم الأصوات أيضا، ثلة تعيب على المدونات أنها "رسكلة" للملفوظات السابقة، مما قد يؤدي باللغة إلى الهلاك في شرك "السطحية"، و الاستهلاك في قوقعة "التنميط". بيد أننا نرى خلاف ذلك الرأي صوابا، إذ إن المفهرسات الآلية تعرض أمام المتمرن ما لا حد له من الاحتمالات و الخيارات التي تحظى بخدمة التحيين و التغذية

بكل ما استجد، نظرا لكون المدونة مفتوحة، مما يحفز لديه ملكة الانتقاء و التخيير و يشحذها.

إنما خطر "الرسكلة"، و ما ينجر عنها من هلاك و استهلاك، يُخشى على المترجم الذي يعتمد على الذاكرة الترجمية، بحيث يلجأ دوماً إلى الحلول نفسها، و يجد في نفسه "كسلا" يحول دون تقصيه للجديد من المفاهيم و المصطلحات و بحثه عنها. على صعيد آخر، قد يجد طلبتنا صعوبة في التأقلم مع "الوافد" الجديد، و قد يبدو لهم أن في تمارين الترجمة باعتماد المدونات غُبْنًا و مشقة، لأنها تمارين تقوم على المقاربة بالاستقراء التي "ينفر" منها بعض المتعلمين. و هنا يتدخل الأستاذ ليقوم وجهة النظر هذه، بحيث ينبغي عليه أن يجد طرقا مستطرفة تحمس طلبته للتعلم وفق هذه المقاربة.

و عموما، تلك هي حال كل طريقة جديدة في العمل، تلقى بعض الامتناع أو "التمنع" – إن جاز القول –، بيد أننا لا نشك في أن شيئا من الدربة و المران، و بعض الحيل المفيدة قد يساعد أمثال هؤلاء من طلبتنا على تجاوز فكرة المعارضة تلك.

و لضمان نجاح عملية دمج المدونات ضمن الخطة التعليمية، و جب على الأستاذ-المدرّب نفسه أن يتمكن منها؛ هذا ناهيك عن ضرورة إدراج مقياس في هذا الصدد ضمن البرنامج التدريبي. و يؤكد عدد من الباحثين أنه – وفق تجاربهم – كثيرا ما يبدي أساتذة اللغات حماسهم لاستخدام هذه الأداة، بعد تعرفهم إليها و احتكاكهم بها، و يتحسرون على ما مضى من حياتهم التدريسية دونما مدونات.

و الحال هذه، فإننا ندعو أساتذتنا إلى دخول معترك هذه المقاربة، و نوصيهم بـ:

- أن يبينوا لطلبتهم أن الترجمة قوامها نظرة ثاقبة و إمعان، و أن النظرة دربة و مران؛ و قصدنا أن أغلب الحلول الترجمية مصدرها النظر فيما سبق من نواتج نصية، قد جمعت المدونات شتاتها، و أن النظرة الفاحصة الممحصة لا تحصل ملكتها إلا بالممارسة.

- أن يشجعوهم على تصميم مدوناتهم الخاصة، و استخدامها في البحث التوثيقي و الترجمة و المراجعة؛ ذلك أن بناء المدونة مهارة من المهارات المطلوبة في المقاربة المدوناتية. ثم إنه – على هذا النحو – تسنح لهم فرصة أن يصبحوا مشاركين فاعلين في عملية التعليم/التعلم، بخلاف المقاربة التقليدية.

- أن يدفعوا بهم إلى ملاحظة السلوكات اللغوية على مختلف المستويات: المفردة، فالجملة، فالنص باعتباره أصغر وحدة في المدونة، يستمد معناه و هويته من بقية النصوص.

- أن يشرحوا لهم أهمية الاستخدام الأمثل للمدونات و برمجيات المعالجة، من حيث إنها تساعدكم – كطلبة و كمتترجمين مستقبلا – على تحسين نوعية إنتاجهم و إنتاجيتهم و تجويدها، لاسيما فيما يتعلق بالترجمة المختصة إلى اللغات الأجنبية أو اللغة الأم.

- أن يؤكدوا لهم على وجود تكامل بين مهام كل من العنصر البشري و الحاسوب؛ من حيث إن مهمة الحاسوب (البرمجيات) تتمثل أساسا في إيجاد القوالب/

التعبير المحتملة، و تنظيمها في أسطر الفهرست و ترتيبها، أما مهمة المترجم فتتمثل في غرلة نتائج البحث و انتقاء الاحتمال الملائم.

- أن يلقنهم مهارات القراءة العمودية لأسطر الفهارس، طالما أنها تختلف لزاما عن المهارات التي تتطلبها القراءة الأفقية أو الخطية لنص تام.
- أن يبينوا لهم أن الوقت الذي يستغرقونه في صناعة مدوناتهم، ليس من المضیعة في شيء. إنما هو وقت إضافي كان من الممكن أن يستغرقه في معاينة قواميسهم أو البحث في صفحات الويب، دونما أن يعود عليهم بفائدة أكيدة. علاوة على ذلك، يمكنهم الاستفادة من المدونة لاحقا و الاعتماد عليها في ترجمة نصوص أخرى تنتمي إلى المجال نفسه.

و لا بأس في أن نشير إلى أنه يمكن للمدرب أن يدمج المدونات في درسه، وفق طريقتين:

✓ إما أن يقوم باستخلاص الفهارس بنفسه و طبعها، و من ثم يطلب من المترنين تفحصها داخل الصف، باعتبارها أداة مكملة لبقية الموارد الترجمة، تساعد في تحليل خصوصيات النصين الأصل و المستهدف.

✓ و إما أن يطلب منهم استخدام البرمجية بأنفسهم، و استخلاص الفهارس منها، بصفتها أهم مورد ترجمي.

زيادة على ذلك، يمكن للمدرب أن يعتمد على المدونات في تقييم ترجمات طلبته تقييما يتسم بكثير من الموضوعية و الدقة.

و قبل أن نضع نقطة آخرَ آخر سطر من هذه الدراسة، لا بأس أن نذكر بأنه لمن السذاجة و الاغترار الاعتقاد بأن التكنولوجيا الحديثة تصنع من المترجم السيء مترجما بارعا، بل إن توظيف الأدوات المناسبة على النحو الصحيح، يَمَكِّنُ – بلا ريب – المترجم الكفاء من تحسين أدائه.



إطلالة خاطفة تلك التي قمنا بها عبر ما سلف من وريقات، حاولنا من خلالها التعرف إلى طبيعة المقاربة التي ارتأينا تبنيها في تدريس الترجمة. و لا نزعم بتاتا أننا قد وفينا الموضوع حقه، لاسيما أن كثيرا من جوانبه تحتاج إلى المزيد من إمعان النظر، و أعمال الفكر.

و آخر دعوانا أن الحمد لله رب العالمين.

الملاحق

Annex n°1 : POS Taggers.

الملحق رقم 1: الرموزات الآلية.

AMALGAM tagger at:

<http://www.scs.leeds.ac.uk/ccalas/amalgam/amalgtag3.html>

Brill's tagger at :

<http://www.cs.jhu.edu/~brill/code.html>

CLAWS tagger at :

<http://www.comp.lancs.ac.uk/computing/research/ucrel/claws/tagservice.html>

GoTagger at :

<http://uluru.lang.osaka-u.ac.jp/~k-goto/-download.cgi?name=GoTagger07.zip>

TreeTagger at :

<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/DecisionTreeTagger.html>

Annex n°2 : Penn Treebank Tagset

POS Tag	Description	Example
CC	coordinating conjunction	and
CD	cardinal number	1, third
DT	determiner	the
EX	existential there	<i>there is</i>
FW	foreign word	d'hoevre
IN	preposition/subordinating conjunction	in, of, like
JJ	adjective	green
JJR	adjective, comparative	greener
JJS	adjective, superlative	greenest
LS	list marker	1)
MD	modal	could, will
NN	noun, singular or mass	table
NNS	noun plural	tables
NNP	proper noun, singular	John
NNPS	proper noun, plural	Vikings
PDT	predeterminer	<i>both</i> the boys
POS	possessive ending	friend's
PRP	personal pronoun	I, he, it
PRP\$	possessive pronoun	my, his
RB	adverb	however, usually, naturally, here, good
RBR	adverb, comparative	better
RBS	adverb, superlative	best
RP	particle	give <i>up</i>
TO	to	<i>to go, to him</i>
UH	interjection	uhhuhhuhh
VB	verb, base form	take
VBD	verb, past tense	took
VBG	verb, gerund/present participle	taking
VBN	verb, past participle	taken
VBP	verb, sing. present, non-3d	take
VBZ	verb, 3rd person sing. present	takes
WDT	wh-determiner	which
WP	wh-pronoun	who, what
WP\$	possessive wh-pronoun	whose
WRB	wh-abverb	where, when

الملحق رقم 3: برنامجيات الفهرسة. **Annex n° 3 : Concordancing software**

Aconcorde (ACC) by Andrew Roberts, available at :

<http://www.andy-roberts.net/software/>

AntConc 3.1 by Laurence Anthony, available at :

<http://www.antlab.sci.waseda.ac.jp/software.html>

ConcApp4 (CA4) by Chris Greaves ; available at :

<http://www.edict.com.hk/pub/concapp>

Concordance by R. J. C. Watt, available at :

<http://www.concordancesoftware.co.uk/>

MonoConc Pro 2.2 (MCP 2.2) by Michel Barlow, available at :

<http://www.athel.com/mono.html>

Multi Language Corpus Tool (MLCT) by Scott Piao, available at :

<http://www.lancs.ac.uk/staff/piaosl/research/download.htm>

Simple Concordance Program (SCP) by Alan Reed :

<http://web.beham.ac.uk/a.reed/textworld/scp/>

TextSTAT 2.6 (TS 2.6) by Matthias Hüning, available at :

<http://www.niederlandistik.fu-berlin.de/textstat>

WordSmith Tools 4 by Mike Scott, available at :

<http://www.lexically.net/wordsmith>

استثمار المدونة في تدريب المترجم

الملحق رقم 4: مدونات عربية.

اسم المدونة	من صنع	الحامل اللساني	حجمها	الغرض منها	مصدر مادتها
مدونة بكوالتير العربية <i>Tim Buckwalter</i>	تيم بكوالتير	مكتوب	2.5 - 3 بلايين كلمة	المعجماتية	الموارد المتاحة على مستوى الشبكة.
مدونة لوفن Leuven 2004-1990	جامعة لوفن الكاثوليكية (بلجيكا)	مكتوب و منطوق	3 ملايين كلمة (و المنطوق منها 700000)	في صناعة معجم عربي- هولندي/ هولندي- عربي لمتعلمي اللغة العربية	الكتب المدرسية، الشبكة، الإذاعة و التلفزيون.
مدونة كلفرند (1995) <i>CALLFRIEND</i>	جامعة بنسلفانيا (LDC)	محادثات	60 مكالمات هاتفية	تطوير تقنيات المماثلة اللغوية	متحدثو اللهجة المصرية
مدونة نيميغن (1996) <i>Nijmegen</i>	جامعة نيميغن (هولندا)	مكتوب	أكثر من 2 مليون كلمة	صناعة معجم عربي- هولندي/ هولندي- عربي	مجلات و روايات
مدونة كولهوم (1997) <i>CALLHOME</i>	جامعة بنسلفانيا (LDC)	محادثات	120 مكالمات هاتفية	التعرف إلى الكلام، مصدره الخطوط الهاتفية	متحدثو المصرية
مدونة مصر (1999)	جامعة جون هوكينز John Hopkins	مكتوب		الترجمة الآلية	مدونة متوازية إنجليزية - عربية (القرآن الكريم)
الأخبار الإذاعية (2000)	جامعة بنسلفانيا (LDC)	منطوق	أكثر من 110 برامج إذاعية	التعرف إلى الكلام	الأخبار المذاعة على أمواج "صوت أمريكا"
مدونة دينار (2000)	جامعة نيميغن و سوتيتل بالتعاون مع جامعة ليون2	مكتوب	10 ملايين كلمة	المعجماتية، المعالجة الآلية للغة، و أبحاث أخرى عامة	
المدونة المتوازية إنجليزية - عربية (2003)	جامعة الكويت	مكتوب	3 ملايين كلمة	تدريس الترجمة و المعجماتية	منشورات المجلس الكويتي الوطني
المدونة متعددة اللغات (2004)	جامعة مانشستر - معهد العلوم والتقنيات UMIST	مكتوب	11.5 مليون كلمة (العربية منها 2.5 مليون)	الترجمة	مواقع تقنية المعلومات المتخصصة...
مدونة سوتيتل SOTETEL	سوتيتل لتقنيات المعلومات (تونس) SOTETEL-IT	مكتوب	8 ملايين كلمة	المعجماتية	مواد أدبية أكاديمية و صحافية

الملحق رقم 5:

النص باللغة العربية:

روسيا تختبر "أبو القنابل جمعاء"

أعلنت روسيا أنها اختبرت بنجاح قنبلة جديدة محمولة جواً، و التي وصفتها بأنها أقوى سلاح غير نووي في العالم.

و القنبلة الجديدة "لا مثيل لها في العالم" على حد قول الجنرال ألكسندر روكشين، نائب هيئة الأركان العامة للقوات المسلحة الروسية.

و قد أضاف قائلاً للقناة الأولى الروسية (أورتي) إن "نتائج التجارب".... بينت أن القنبلة تعادل قنبلة نووية في فاعليتها و كفاءتها". كما أكد في الوقت نفسه أن السلاح الجديد لا يشكل تهديداً للبيئة بخلاف الأسلحة النووية.

و قالت القناة الأولى الروسية إن القنبلة الجديدة، الملقبة بـ "أبو القنابل جمعاء"، أقوى أربع مرات من قنبلة التفجير الهوائي الفائقة الكتلة (أم أو أي بي) الأمريكية المعروفة بـ "أم القنابل جمعاء".

ثم أظهرت قاذفة قنابل استراتيجية من نوع توبوليف تي يو 160 أثناء رميها القنبلة على موقع اختبار، و تلا ذلك انفجار كبير.

و المعلوم أن أسلحة الضغط و الحرارة، المعروفة أيضاً باسم "القنابل الفراغية" أو "ذخائر الوقود الهوائي"، تتميز عن المتفجرات التقليدية بأنها تستخدم الأكسجين الموجود في الغلاف الهوائي بدل حشو عامل مؤكسد داخل المادة المتفجرة.

و مثل هذه القنابل تنفجر على مرحلتين: ينثر انفجار صغير شحنة رئيسة من المواد المتفجرة كغيمة؛ و في مرحلة تالية، تنفجر تلك الغيمة بمكوناتها تلقائياً في الجو.

و يولد الانفجار موجة ضغط تفوق شدتها قوة المتفجرات التقليدية، كما أن استهلاك الغازات ضمن الانفجار يولد أيضاً فراغاً جزئياً وحرارة عالية. و هكذا فإن "كل كائن حي يتبخر حتماً"، حسب أحد المسؤولين الروس.

الترجمة المقترحة باعتماد "ويبكورب":

Russia tests ‘the father of all bombs’.

Russia has announced that it had successfully tested a new air-delivered bomb, which it described as the world’s most powerful non-nuclear weapon.

The new bomb ‘has no match in the world’, said Alexander Rukshin, a deputy chief of the General Staff of the Russian Armed Forces.

He told Russia’s state ORT First Channel television that ‘the results of tests ... have shown that it is comparable with nuclear weapons in its efficiency and potential’. Rukshin stressed at the same time that the new weapon is environmentally friendly, in contrast to a nuclear one.

ORT First Channel television said the new ordnance – dubbed the Father of all Bombs – is four times more powerful than the American Massive Ordnance Air Blast Bomb (MOAB).

Then, it showed a Tupolev TU-160 strategic bomber dropping the bomb over a testing ground. A large explosion followed.

Thermobaric weapons, which are also known as ‘vacuum bombs’ or ‘fuel-air munitions’, differ from conventional explosive weapons by using oxygen from the atmosphere, rather than carrying an oxidizing agent in their explosives.

Such bombs detonate in two stages. First, a small blast disperses a main load of explosive material into a cloud, which then spontaneously ignites.

This explosion generates a pressure wave that reaches much further than that from a conventional explosive. The consumption of gases in the blast also generates a partial vacuum and a high temperature. ‘All that is alive merely evaporates’, according to a Russian official.

معجم البحث

Glossary مسرد

Performance	أداء (لغوي)
Retrieval	استرجاع
Induction	استقراء
Deduction	استنتاج
Explicitation	إظهار
Paraphrase	إعادة الصياغة
Parsing	إعراب (تحليل الجمل)
(Texts) Selection	انتخاب النصوص
Shift	انزياح
Software	بر (نا)مجيات
Simplification	تبسيط
Empiricism/ Experimentation	تجريب
(Corpus) compilation	تجميع (المدونة)
Enriching (corpus)	تحشية (المدونة)
Quantitative analysis	تحليل كمي
Qualitative analysis	تحليل نوعي
Machine translation	ترجمة حاسوبية
Machine – aided translation	ترجمة محوسبة
Corpus–based translation Studies (CTS)	ترجمات باعتماد المدونات
Annotation	ترميز
POS tagging	الترميز بأقسام الكلام
(Corpus) Design	تصميم (المدونة)
Normalization	تطبيع
Generalization	تعميم
Recurrence	تكرار
Recurrent pattern	قالب متكرر
Representativeness	تمثيل (خاصية التمثيل)
Alignment	تنظير (المدونة)
Markup	تنقيح
Occurrence	توارد
Balance	توازن
Frequency list	جدول التكرار
Distribution Frequency	جدول توزيع التكرار
Stem	جذع (الكلمة)
Intuition	حدس
Computational Linguistics	حوسبية (لسانيات)

Tagset	سلسلة الشارات الشارحة
Context	سياق
Semantic tag	شارة دلالية
Tag	شارة شارحة
Translation universals	شموليات ترجمية
Right co-text	صدر (مرافق العقدة اليمين)
Grammaticality	صواب (نحوي)
Translationese	ضيم ترجمي
Left co-text	عجز (مرافق العقد اليسار)
Node	عقدة
Concordance	فهرست آلي
Collocate list	قائمة الوصيلات
Lexicometry	قياسة معجمية
Competence	كفاءة (لغوية)
Key Word in Context(KWIC)	كلمة مفتاح ضمن السياق (كويك)
Universals	كليات نحوية
Corpus Linguistics	لسانيات المدونة
Collocations	متلازمات لفظية
Corpus, pl· Corpora	مدونة
Monolingual ~	~ أحادية اللغة
Bilingual ~	~ ثنائية اللغة
Multilingual ~	~ متعددة اللغات
Synchronic ~	~ سانكرونية (آنية)
Diachronic ~	~ دياكرونية (تاريخية)
Balanced ~	~ متوازنة
Specialized ~	~ مختصة
Open –ended ~	~ ممتدة (مفتوحة)
Finite ~	~ ثابتة (مغلقة)
Raw / Unannotated ~	~ خام
Annotated ~	~ مرزمة
Spoken ~	~ منطوقة
Written ~	~ مكتوبة
Learner ~	~ متعلم اللغة
Native –Speaker ~	~ المتكلم الأصلي باللغة
Monitor ~	~ الضبط و المراقبة
Sub-corpus ~	~ فرعية
Translation ~	~ ترجمة
Comparable ~	~ مقارنة
Multi-source-language monolingual comparable corpus	~ ~ أحادية اللغة متعددة المصادر

استثمار المدونة في تدريب المترجم

Parallel ~	~ متوازية
Unidirectional ~ ~	~ ~ أحادية الاتجاه
Bidirectional/ Reciprocal ~ ~	~ ~ ثنائية الاتجاه / متبادلة
Multidirectional ~ ~	~ ~ متعددة الاتجاه
Corpus size	حجم المدونة
Corpus Adequacy	ملائمة المدونة (للمشروع)
Tagger	مرمز آلي
Treebanks	مشجرات (بنوك شجرية)
Concordancer	مفهرس آلي
Natural language Processing/ NLP	معالجة آلية للغات الطبيعية
Criteria	معايير
External ~	~ خارجية
Internal ~	~ داخلية
Parser	معرب آلي
Token	نموذج (وحدة)
Type	نوع (شكل)
Collocate	وصيلة (العقدة)

مكتبة البحث

استثمار الحدوث في نشر المخرجات

المراجع باللغة العربية:

- أرسطو، منطق أرسطو، تحقيق عبد الرحمن بدوي، ج3، مكتبة النهضة المصرية، 1952
- بدوي، عبد الرحمن، مناهج البحث العلمي، وكالة المطبوعات، ط3، الكويت، 1977
- بوقلي حسن، جمال الدين، إشكاليات فلسفية، ط1، الديوان الوطني للمطبوعات المدرسية، الجزائر، 2007.
- حمادة، سلوى، المدونات النصية و دور اللغة العربية في التعامل معها. مجلة واتا للترجمة و اللغات، السنة الأولى، العدد4، خريف 2007 . <http://www.watajournal.com/mag4/index.html>
- الديداوي، محمد، الترجمة و التواصل: دراسات تحليلية عملية لإشكالية الاصطلاح و دور المترجم، ط.1، المركز الثقافي العربي، الدار البيضاء - المغرب/بيروت - لبنان، 2000.
- منهاج المترجم: بين الكتابة و الاصطلاح و الهوية و الاحتراف، ط.1، المركز الثقافي العربي، الدار البيضاء - المغرب/بيروت- لبنان، 2005.
- الطويل، توفيق، أسس الفلسفة، ط.5، دار النهضة العربية، 1967
- عيسى حبيب، ماهر، الارتقاء بتعليم العربية التقليدي إلى المستوى الإلكتروني. بحث مقدم إلى ندوة تطوير المناهج و الاختصاصات الجامعية - جامعة حلب 30-31 أيار 2007.
- فضل الله، مهدي، مدخل إلى علم المنطق، ط3، دار الطليعة، بيروت، 1985
- كولبي، أزفد، المدخل إلى الفلسفة، تر: أبو العلا عفيفي، ط 5، مكتبة النهضة المصرية القاهرة، دون سنة.

المراجع باللغات الأجنبية:

Aarts, Jan (2000) Towards a new generation of corpus-based English grammars. In: B. Lewandowska-Tomaszczyk & P.J. Melia (eds.) *PALC'99: Practical applications in language corpora* (17-36). Frankfurt am Main: Peter Lang.

- (1991) Intuition-based and observation based grammars. In: J. Aijmer & B. Altenberg (eds.) *Directions in English Corpus Linguistics* (44-62). London/New York: Longman.

Abdelali, Ahmed, Jim Cowie & Hamdy S. Soliman (2005) Building a Modern Standard Arabic Corpus. *Workshop on Computational Modeling of Lexical Acquisition*. The Split Meeting, Croatia, 25th to 28th of July 2005.

Aijmer, Karin, Bengt Altenberg & Lats Johansson (1996) *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies* (73-85). Lund 4-5 March 1994. Lund: Lund University Press.

Al-Ansary, Sameh, Magdy Nagi & Noha Adly (2007) Building an International Corpus of Arabic (ICA): Progress of Compilation Stage. *7th International Conference on Language Engineering*, Cairo- Egypt, 5-6 December 2007.

Al-Ansary, S. (2003) NP-Structure Types in Spoken and Written Modern Standard Arabic (MSA) Corpora. In: Parkinson, Dilworth B. & Samira Farwanah (eds.) (2003a) *Perspectives on Arabic Linguistics XV: Papers from the Fifteenth Annual Symposium on Arabic Linguistics* (149-180). Salt Lake City.

Al-Muhanna, Amin (2003) *Scientific and technological terms transfer into Arabic: A corpus –based study of Arabic noun+ noun and noun+ adj. compounds*. –Ph.D. thesis, UMIST, UK

Al-Sulaiti, Latifa & Eric Atwell (2006) The design of a corpus of contemporary Arabic. *International Journal of Corpus Linguistics* 11(1), pp. 5-7.

Al-Sulaiti, Latifa (2005) Arabic Corpora. Viewable online at : http://www.comp.leeds.ac.uk/eric/latifa/arabic_corpora.htm (Retrieved: 10/02/2007)

Aston, Guy (1999) Corpus use and learning to translate. *Textus* 12, pp. 289-314.

- (1997) Small and large corpora in language learning. In: B. Lewandowska-Tomaszczyk & P.J. Melia (eds.) *Proceedings of the First International Conference on Practical applications in language corpora* (51-62). Lodz, Poland: Lodz University Press.

Baker, Mona (1999) The Role of Corpora in Investigating the Linguistic Behavior of Professional Translators. *International Journal of Corpus Linguistics* 4 (2), pp. 281-298.

- (1998) Réexplorer la langue de la traduction: Une approche par corpus. *Meta* 43 (4), 480-485.

- (1996) Corpus-based translation studies: The challenges that lie ahead. In: Harold Somers (ed.) *Terminology, LSP and translation studies in language engineering: in honour of Juan C. Sager* (175-186). Amsterdam/ Philadelphia: Benjamins.
- (1995a) Corpora in translation studies: An overview and some suggestions for future research. *Target* 7 (2), pp. 223-243.
- (1995b) Stranger in paradigm: What Lies Ahead for simultaneous Interpreting Research? In: Daniel Giel (ed.) *Target* 7 (1), special issue: *Interpreting Research*, pp 7-29.
- (1993) Corpus linguistics and translation studies: Implications and applications. In: M. Baker, G. Francis & E. Tognini-Bonelli (eds.) *Text and Technology: in honour of John Sinclair* (233-252). Amsterdam/ Philadelphia: Benjamins.
- (1992) *In Other Words- A course book on Translation*. London/New York: Routledge.

Ball, N. Catherine (1997) *Tutorial: Concordances and Corpora* (Online tutorial), viewable online at: <http://www.georgetown.edu/faculty/ballc/corpora/> (Consulted 9 June, 2005).

Bell, Roger Thomas (1991) *Translation and Translating: Theory and Practice*. London/New York: Longman.

Bernardini, Silvia (2006) Corpora in translator education and translation practice: Achievements and challenges. *Proceedings of LREC 2006*. Viewable online at: http://mellange.eila.jussieu.fr/bernardini_lrec06.pdf (Last visited: 28 December, 2006)

- (2002a) Exploring new directions for discovery learning. In: Ketteman, Bernard & Georg Marko (eds.) *Teaching and learning by doing corpus analysis. Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz 19-24 July, 2000* (165-182). Amsterdam: Rodopi.
- (2000b) Systematising serendipity: Proposals for concordancing large corpora with language learners. In: Burnard, Lou & Tony McEnery (eds.) *Rethinking language pedagogy from a corpus perspective* (225-234). Frankfurt am Main: Lang.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan (1999) *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Ltd.

Blum-Kulka, Shoshana (1986) Shifts of cohesion and Coherence in Translation. In: J. House & S. Blum-Kula (eds.) *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies* (19-21). Tübingen: Gunter Narr.

Blum-Kulka, Shoshana & Eddie Levenston (1983) Universals of lexical simplification. In C. Faerch and G. Kasper (eds) *Strategies in Interlanguage Communication* (119-139). London and New York: Longman.

Bob-Godwin, Jones (2001) Emerging Technologies: Tools and Trends in Corpora Use for Teaching and Learning. *Language Learning and Technology* 5(3), September 2001, pp.7-12.

Bouquet, S. (2005) Après un siècle, les manuscrits de Saussure reviennent bouleverser la linguistique. *Texto!* Viewable online at :

http://www.revue-texto.net/saussure/SurSaussure/Bouquet_Après.html (Visited: April 2006)

Bowker, Lynne (2002 a) Exploring new directions for discovery learning. In: B. Kettemann & G. Marko (eds.) (2002) *Teaching and learning by doing corpus analysis* (165-182). Amsterdam/ New York: Rodopi.

- (2002 b) Working Together: A Collaborative Approach to DIY Corpora. *Paper presented at the First International Workshop on Language Resources for Translation Work and Research*, Gran Canaria, 28/05/2002.
- (1999) Exploring the potential of corpora for raising language awareness in student translators. *Language Awareness* 8 (3/4), pp. 160-173.
- (1998) Using specialized monolingual native-language corpora as a translation resource: a pilot study. *Meta* 43(4), pp. 631-651.

Brown, P., V. Della Pietra, S. Della Pietra, J. Lafferty & R. Mercer (1993) Analysis, Statistical Transfer and Synthesis in Machine Translation. Submitted to *TMI-92, 4th International Conference on Theoretical and Methodological Issues in Machine Translation*, 1992.

Brown, Peter F., John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick. Jelinek, John D. Lafferty, Robert L. Mercer & Paul S. Roossin (1990) A Statistical Approach to Machine Translation. *Computational Linguistics* 16(2), pp. 79-85.

Buckwalter, Tim (2002) Buckwalter Arabic Transliteration. (Consultée le 24/10/2007)

<http://www.qamus.org/transliteration.htm>

Burnard, Lou & Tony McEnery (eds.) (2000) *Rethinking Language pedagogy from a corpus perspective: Papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt: Peter Lang.

Chambers, Angela (2005) Integrating Corpus Consultation in Language Studies. *LLT* 9(2), pp. 11-125 .

Chambers, A. & Ide O'Sullivan (2004) Corpus consultation and advanced learners' writing skills in French. *ReCALL* 16(1), pp. 158-172.

Cheng, W., M. Warren & X. Xun-feng (2003) The language learner as language researcher: Putting corpus linguistics on the time-table. *System* 31(2), pp. 173-186.

Collins Cobuild English Dictionary (1996) London: Harper Collins Publishers.

Condamines, Anne (2006) Modes de construction du sens en corpus spécialisé. *Cahiers de Grammaire* 30, 'Spécial Anniversaire', pp. 75-88.

Crystal, David (1991) *A Dictionary of Linguistics and Phonetics*. Blackwell, 3rd edition. Viewable online at: <http://www.engl.polyu.edu.hk/corpuslinguist/corpus.htm> (Consultée le 24/04/2006)

Déjean, Hervé & Éric Gaussier (2003) Une nouvelle approche à l'extraction de lexiques bilingues à partir de corpus comparables. In: J. Véronis (ed.) *Revue électronique Lexicometrica* (2002), numéro thématique: *Alignement lexical dans les corpus multilingues*.

Dodd, Bill (2000) Introduction. The Relevance of Corpora to German Studies. In: Bill, Dodd (ed.) *Working with German Corpora* (69-95). Birmingham: Birmingham University Press.

Ebeling, Jarle (1998) Contrastive linguistics, translation, and parallel corpora. *Meta* 43(4), pp. 602-615.

Evans, David (2006) Unit 1: Introduction. Corpus Building and Investigation for the Humanities (*Online introduction to Corpus Investigative Techniques*), AHRC. Viewable online at: <http://www.humcorp.bham.ac.uk/> (Visited: 9 June, 2007)

Even-Zohar, I (1978) The position of translated literature within the literary system. In: James S. Holmes, J. Lambert & R. Van den Breock (eds.) *Literature and translation: New perspectives in literary studies*. Leuven: Acco.

Fillmore, Charles J. (1979) Innocence: A Second Idealization for Linguistics, *Berkley Linguistic Society* 5, pp. 63-76.

Fletcher, W. H. (2007) Concordancing the Web: Promise and Problems, Tools and Techniques. In: Marianne Hundt, Nadja Nesselhauf & Carolin Biewer (eds.) *Corpus Linguistics and the Web* (25-45). Amsterdam: Rodopi.

Francis, W. & H. Kucera (1979) *Brown Corpus Manual: Manual of Information to Accompany A Standard Corpus of Present-Day edited American English, for use with Digital Computers* (3rd ed.) Providence: Brown University.

Frawley, William (1984) Prolegomenon to theory of translation. In: W. Frawley (ed.) *Translation. Literary, linguistic and philosophical perspectives* (159-175). London/ Toronto: Associated University Presses.

Gellerstam, M (1986) Translationese in Swedish novels translated from English. In: L. Wollin & H. Lindquist (eds.) *Translation studies in Scandinavia* (88-95). Lund Studies in English 75. Lund: CWK Gleerup.

Gentzler, Edwin (1993) *Contemporary translation theories*. London/New York: Routledge.

Ghazali, Salem & Abdelfattah, Braham (2001) Dictionary Definitions and Corpus_ Based Evidence in Modern Standard Arabic. *Arabic NLP Workshop at ACL/EACL*. Toulouse, France. Viewable online at: <http://www.elsnet.org/arabic2001/ghazali.pdf> (15/12/2006)

Godwin-Jones, Bob (2001) Emerging Technologies: Tools and Trends in Corpora Use for Teaching and Learning. *Language Learning and Technology* 5(3), September 2001, pp. 7-12.

Gouadec, Daniel (2002) Training Translators: Certainities, Uncertainities, Dilemmas. In: B. Maia, J. Haller & M. Ulrych (eds.) *Training the Language Services Provider for the new Millenium* (31-41). Oporto: Universidade de Oporto.

Granger, Sylviane & Viana Vander (2007) Corpus Linguistics, Language Learning and ELT: Interviewing Sylviane Granger, *Mindbite*, 2007, 1.

Granger, Sylviane (2003) The corpus approach: a common way forward for Contrastive Linguistics and Translation Studies. In: S. Granger, Jacques Lerot & Stephanie Petch-Tyson (eds.) *Corpus-based Approaches to Contrastive Linguistics and Translation Studies* (17-29). Amsterdam/ New York: Rodopi.

- (1996) From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In: Aijmer K., Altenberg B. & Johansson M. (eds.) *Languages in Contrasts: Papers from a Symposium on Text-based cross-linguistic Studies* (37-51), Lund, 4-5 March 1994. Lund: Lund University Press.

Greenbaum, Sidney (ed.) (1996) *Comparing English Worldwide. The International Corpus of English*. Oxford: Clarendon Press.

Guidère, Mathieu (2002) Toward Corpus-Based Machine Translation for Standard Arabic. *Translation Journal* 6(1), January 2002.

Halliday, M. A. K. (1985) *An Introduction to Functional Grammar*. London: Baltimore, Md / UAS: Edward Arnold. *Preface*: xxxi.

Halliday, M. A. K. & R. Hasan (1976) *Cohesion in English*. London: Longman.

Halverson, Sandra (1998) Translation studies and representative corpora: Establishing links between translation corpora, theoretical/descriptive categories and a conception of the object of study. *Meta* 43(4).

Hanks, P. (2002) Literal and Metaphorical Word Meaning. Toscan Word Center Document.

- (1996) Contextual Dependency and Lexical Sets. *International Journal of Corpus Linguistics* 1 (1), pp. 75-98.

Harris, B. (1988a) Are you bi-textual? *Language Technology* 7, p.41.

- (1988b) Bi-texts: A new concept in translation theory. *Language Monthly* 54, pp. 8-10.

Hoey, Michael (1997) From concordance to text structure: New uses for computer corpora. *Paper presented at the 1997 PALC Conference*. Poland: Lodz.

Hoogland, J. (1996) The use of OCR Software for Arabic in order to create a text corpus of Modern Standard Arabic for lexicographic purposes. In A. Ubaydly (ed.) *Proceedings of the*

International Conference and Exhibition on Multi-Lingual Computing(2701-2716), Cambridge University.

Hunston, Susan (2002) *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Isabelle, Pierre (1992) La bi-textualité : Vers une nouvelle génération d'aides à la traduction et terminologie. *Meta* 37 (4), pp. 721-737.

Johansson, S. (1998) On the role of corpora in cross-linguistic research. In: Johansson S. & S. Oksefjell (eds.) *Corpora and cross-linguistic research: Theory, Method, and Case Studies* (3-25). Amsterdam: Rodopi.

Johansson, S & J. Ebeling (1996) Exploring the English-Norwegian parallel corpus. In: C. Percy, C.F. Meyer & I. Lancashire (eds.) *Synchronic Corpus Linguistics* (3-15). Amsterdam: Rodopi.

Johansson, S. & K. Hofland (1994) Towards an English-Norwegian parallel corpus. In: U. Fries, G. Tottie & P. Schneider (eds.) *Creating and using English language corpora* (25-37). Amsterdam: Rodopi.

Johansson, Stig, Geoffrey Leech & Helen Goodluck (1978) *Manual of Information to Accompany the Lancaster-Oslo/ Bergen Corpus of British English, for use with Digital Computers*. Department of English, University of Oslo.

Johns, T. (1997) Contexts: The background development, and trialing of a concordance-based CALL program. In: A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (eds.) *Teaching and Language Corpora* (100-115). London: Longman.

Jurafsky, Dan (2007) Lecture 6: Part of Speech Tagging (I). *LINGUIST 180: Introduction to Computational Linguistics*.

Kay, M. & M. Röscheisen (1988) Text-translation alignment. *Computational Linguistics* 19(1), pp. 121-142.

Kennedy, C. & T. Miceli (2001) An evaluation of intermediate students' approaches to corpus investigation, *LLT* 5(3), pp. 77-90.

Kenny, Dorothy (1998) Creatures of Habit? What Translators Usually Do With Words. *Meta* 43 (4), pp. 515-523.

Kettemann, B. & G. Marko (eds.)(2002) *Teaching and learning by doing corpus analysis*. Amsterdam/ New York: Rodopi.

Khoja, Shereen (2001)APT: Arabic part-of-speech tagger. In: *Proceedings of Student Workshop (NAACL 2001)*, Carnegie Mellon University, Pittsburgh, Pennsylvania, U.S.

Kjellmer, G (1986) The lesser man: Observations on the role of woman in modern English writings. In: J. Aarts & W. Meijjas (eds.) *Corpus Linguistics II* (163- 176). Amsterdam: Rodopi.

Kraif, Olivier (2001) Constitution et exploitation de bi-textes pour l'aide à la traduction. Thèse—Sciences du Langage. Université de Nice Sophia Antipolis.

Krieger, Daniel (2003) Corpus Linguistics: What it is and How it can be applied to teaching. *The International TESL Journal*, vol. IX, n° 3, March 2003. Viewable online at: <http://iteslj.org/Articles/Krieger-Corpus.html> (Retrieved: 21 March, 2005)

Laviosa-Braithwaite, Sara (2002) *Corpus-based translation studies. Theory, findings, applications*. Amsterdam/New York: Rodopi.

- (1998a) Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta* 43(4), pp. 557-570.
- (1998b) Universals of translation. In: Mona, Baker (ed.) *Routledge Encyclopaedia of Translation Studies* (288-291). London / New York: Routledge.

Laurence, Anthony (2005) AntConc: Design and Development of a Freeware Corpus Analysis Toolkit for the Technical Writing Classroom. *IEEE International Professional Communication Conference Proceedings*, pp. 729-735.

- (2004) AntConc: A Learner and Classroom Friendly Multi-Platform Corpus Analysis Toolkit. In: *IWLeL 2004, an Interactive Workshop on language e-Learning*, pp. 7-13.

Lee, Sherman (without a year) Corpus Linguistics: Developing a PolyU Language Bank. (Power Point)

Leech, Geoffrey (2005) Adding Linguistic Annotation. In: Martin, Wynne (ed.) *Developing Linguistic Corpora: a Guide to Good Practice*. Viewable online at: <http://www.ahds.ac.uk/creating/guide/linguistic-corpora/chapter 2.htm>

- (1977) *Semantics*. Penguin Books.

Léon, J. (2005) Claimed and unclaimed sources of Corpus Linguistics. *Henry Sweet Society Bulletin*, n° 44, pp. 36-50.

Le Petit Larousse Illustré (2006) Paris.

Louw, B (1997) The Role of Corpora in Critical Literary Appreciation. In: A. Wichmann, S., Fligelstone, T. McEnery & G. Knowles (eds.) *Teaching and Language Corpora*. London: Longman.

MacMullen, W. John (2003) Requirements Definition and Design Criteria for Text Corpora in Information Science. *SILS Technical Report TR 2003-03*. Chapel Hill, NC: University of North Carolina, School of Information and Library Science, Technical Report Series.

Magri-Mourgues, Véronique (2006) Corpus et Stylistique. *Corpus* n°5, Corpus et Stylistique – décembre 2006.

Maia, Belinda (2003) Some languages are more equal than others. Training translators in terminology and information retrieval using comparable and parallel corpora. In: Federico, Zanettin, Silvia Bernardini & Dominic Stewart (eds.) *Corpora in Translator Education* (43-53). Manchester: St. Jerome.

- (2000) Making corpora: A learning process. In: S. Bernardini & F. Zanettin (eds.) *I corpora nella didattica della traduzione* (47-60). Bologna: CLUEB.
- (2002) Do-it yourself disposable, specialised mini-corpora — Where next? Reflections on teaching translation and terminology through corpora. In: Tagnin, S.E.O pp. 219-234.
- (1998) Word order and the first person singular in Portuguese and English. *Meta* 43(4), pp. 589-601.

Malmkjaer, K. (1998) Love thy neighbour: will parallel corpora endear linguists to translators? *Meta* 43(4), pp. 534-541.

Marshman, Elizabeth (2003) Construction et gestion des corpus: Résumé et essai d'uniformisations du processus de la terminologie. Defining the norms for the ELECTIK research group. [En ligne] (page consultée le 24/07/2007).

<http://www.olst.umontreal.ca/pdf/terminotique/corpusenttermino.pdf>

Mauranen, A. (2002) Will 'translationese' ruin a contrastive study? *Languages in Contrast* 2(2), pp. 161-186.

Mayaffre, Damon (2005) Rôle et place des corpus en linguistique: Réflexions inductives. *Texto!* [En ligne], décembre 2005, vol. x, n. 4. Consultée le 10/11/2006.

http://www.revue-texto.net/Reperes/Themes/Mayaffre_Corpus.html

McArthur & McArthur (1992) *The Oxford Companion to the English Language*. Viewable online at: <http://www.engl.polyu.edu.hk/corpuslinguist/corpus.htm>

McEnery, A. & Z. Xiao (2007) Parallel and comparable corpora: What are they up to? In: Gunilla M. Anderman & Margaret Rogers (eds.) *Incorporating corpora: Translation and the linguist* (18-30). Translating Europe. Multilingual Matters, Clevedon, UK.

McEnery, Tony & Andrew Wilson (1996) *Corpus Linguistics*. 1st ed., Edinburgh: Edinburgh University Press.

- (1993) Corpora and Translation: Uses and future prospects. *UCREL*, Technical Papers vol.2, March 1993.

Mindt, Dieter (1991) Syntactic evidence for semantic distinctions in English. In: J. Aijmer & B. Altenberg (eds.) *Directions in English Corpus Linguistics*. Amsterdam: Rodopi.

Mohammadi, Dehcheshmeh Maryam (2007) Specialized monolingual corpora in translation. In: Gabe, Boker (ed.) *Translation Journal* 11(2), April 2007.

Mukherjee, Joybrato (2006) Corpus linguistics and English reference grammars. In: Antoinette Renouf & Andrew Kehoe (eds.) *Language and Computers. The Changing Face of Corpus Linguistics*, vol. 55 (337-354). Amsterdam/ New York: Rodopi.

Nikkhou, Mahtab & Khalid Choukri (2005) Report on Survey on Arabic Language Resources and Tools in Mediterranean Countries. *NEMLAR*, Center for Sprogteknologi, University of Copenhagen, Denmark.

Olohan, Maeve (2002a) Comparable Corpora in Translation Research: Overview of Recent Analyses Using the Translational English Corpus, *LREC Language Resources in Translation Work and Research Workshop Proceedings*, pp. 5-9.

- (2002b) Leave it out! Using a comparable corpus to Investigate Aspects of Explication in Translation. In: S.E.O., Tagnin (Org.) *Op. Cit.*, pp. 153-169.

Oostdijk, N. & P. de Haan (1994) Clause patterns in Modern British English. A Corpus-based (quantitative) study, *ICAME Journal* 18, pp. 41-80.

Oostdijk, N. (1988) A Corpus for studying linguistic variation. *ICAME Journal* 12.

Øverås, Linn (1998) In search of the third code: An investigation of norms in literary translation. *Meta* 43(4), pp. 571-588.

Parkinson, D. & S. Farwaneh (eds.) (2003 a) *Perspectives on Arabic Linguistics xv: Papers from the Fifteenth Annual Symposium on Arabic Linguistics*. Salt Lake City.

Pincemin, Bommier-Bénédicte (1999) Diffusion ciblée automatique d'informations: Conception et mise en œuvre d'une linguistique textuelle pour la caractérisation des destinataires et des documents. Chapitre IV, A, pp. 415-427. Université Paris IV, Sorbonnes.

Rastier, F. (2005a) Enjeux épistémologiques de la linguistique de corpus. In : G. Williams (éd.) *La linguistique de corpus*, Rennes, PUR, pp. 31-45. [En ligne sur] *Texto!* Consultée le 13/03/2006. Viewable online at: http://www.revue-texto.net/Inedits/Rastier_Enjeux.html

Reiss, K. (1981) Type, kind and individuality of text: decision making in translation. *Poetics today* 2(4), pp. 121-131.

Roe, P. (2000) The ASTCOVEA German Grammar in conText Project. In: B. Dodd (ed.) *Working with German Corpora* (199-216). Birmingham, England: University of Birmingham Press.

Roberts, A., L. Al-Sulaiti & E. Atwell (2005) aConCorde: Towards a proper concordancer for Arabic.

Rundell, Michael & Sylviane Granger (2007) From corpora to confidence. *MED Magazine*, Issue 46, August 2007. (Retrieved: 10/11/2007) viewable online at: <http://www.macmillandictionary.com/MED-Magazine/August2007/46-MEDMagazine-Contrib.htm>

Samy, Doaa, Sandoval Antonio Moreno, José M. Guirao & Enrique Alfonseca (2006) Building a Parallel Multilingual Corpus (Arabic-Spanish-English). *LREC*, pp. 2176-2181.

Schmied, J. (1993) Qualitative and quantitative research approaches to English relative constructions. In: C. Souter & E. Atwell (eds.) *Corpus-based computational linguistics* (85-86). Amsterdam: Rodopi.

Sharoff, Serge (2004) Towards basic categories for describing properties of texts in a corpus. In: M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, R. Silva, C. Pereira, F. Carvalho, M. Lopes, M. Catarino & S. Barros (eds.) *LREC04: Proceedings of Language Resources and Evaluation Conference*, vol. V, pp. 1743-1746. Portugal: Lisbon.

Shlesinger, Miriam (1998) Corpus- based interpreting studies as an offshoot of corpus-based translation studies. *Meta* 43(4), pp. 486-493.

- (1995) Stranger in paradigm: What Lies Ahead for simultaneance Interpreting Research? In: Daniel Giel (ed.) *Target* 7(1), special issue: Interpreting Research, pp 7-29.
- (1991) Interpreter latitude vs. Due process: Simultaneous and consecutive interpretation in multilingual trials. In: Sonja Tirkkonen-Condit (ed.) *Empirical research in translation and intercultural studies: Selected papers of the TRANS-SIF seminar* (147-55), Savonlinna 1988, Tübingen: Gunter Narr.
- (1989b) *Simultaneous interpretation as a factor in effecting shifts in the position of texts on the oral-literate continuum*. Unpublished MA thesis, Tel Aviv: Tel Aviv University.

Sinclair, John Mch.(2005 a) Corpus and Text-Basic Principles. In: Martin, Wynne (ed.) *Developing Linguistic Corpora: a Guide to Good Practice*. Oxbow Books. Available online at: <http://www.ahds.ac.uk/creating/guides/linguistic-corpora/chapter1.htm>

- (ed.) (2005b) *How to use corpora in language teaching*. Amsterdam: John Benjamins.
- Sinclair, J. McH., S. Jones & R. Daley (2004) *English Collocation Studies: The OSTI Report*. London/ New York: Continuum.
- (1996a) EAGLES Preliminary Recommendations on Corpus Typology. Technical report, EAGLES (Expert Advisory Group on Language Engineering Standards).
- (1996b) Preliminary Recommendations on text-typology. *EAGLES Document EAG-TCWG-TTYP/P*. Available at: <http://www.ilc.cnr.it/EAGLES96/texttype.textty.html>
- (ed.) (1987) *Looking Up—An account of the COBUILD Project in lexical computing*. London: Collins ELT.
- Sinclair, J. McH. et al. (1970) *English Lexical Studies: Report to OSTI on Project C/LP/08*. Department of English, University of Birmingham.

Soler, José (1993) Text Corpora: Meeting the Challenge of Information Excess. *Library & Linguistic Computing* 8 (4), pp. 224-226.

Stefanowitsch, Anatol (2003a) Corpus compilation. *Corpus linguistics handouts*. Viewable online at: <http://www-user.uni-bremen.de/~anatol/materials.html>

- (2003b) Corpus design. *Corpus Linguistics Handouts*.
- (2003c) Corpus linguistic methods. *Corpus Linguistics Handouts*.

Stubbs, Michael (2001) Texts, Corpora and Problems of Interpretation: A Response to Widdowson. *Applied Linguistics* 22 (2), pp. 149-172.

Tagnin, Stella E. O. (Org.) (2002a) *Cadernos de tradução: corpora e tradução*. Florianópolis: NUT, 2002, 1(9).

- (2002b) Corpora and the Innocent Translator: How can they help him? In: M. Thelen (ed.) *Translation ad Meaning*, Part 6, Proceedings of the Lodz Session of the 3rd Maastricht– Lodz Duo Colloquium on “Translation on Meaning” (489-496), Lodz, Poland, September 22-24, 2000, Maastricht: Universitaire Pers Maastricht.

Teich, E. (2002) System-oriented and text-oriented comparative linguistic research: Cross-linguistic variation in translation. *Languages in Contrast* 2(2), pp. 187-210.

Toury, Gide (1980) *In search of a theory of translation*. Tel Aviv: The Porter Institute for Poetics and Semiotics, Tel Aviv University.

- (1995) *Descriptive translation studies and beyond*. Amsterdam/Philadelphia: John Benjamins

Tribble, C. (1997) Improvising corpora for ELT: Quick-and-dirty ways of developing corpora for language teaching. In: J. Melia & B. Lewandowska-Tomaszczyk (ed.) *PALC' 97 Proceedings*. Lodz: Lodz University Press. Online at: <http://www.tribble.co.uk/text/Palc.htm>

Tymoczko, Maria (1998) Computerized Corpora and the Future of Translation Studies. *Meta* 43(4), pp. 652-659.

Vanderauwera, Ria (1985) *Dutch Novels Translated into English: The Transformation of a 'Minority' Literature*. Amsterdam: Rodopi.

van Eynde, Frank (2008) Corpus construction and annotation. Centre for Computational Linguistics. Katholieke Universiteit Leuven. Viewable online at: <http://www.ccl.kuleuven.be/Courses/LEA/corpusannotatie.ppt> (Consultée le 25/10/2007)

van Eynde, F. , J. Zavret & W. Daelemans (2000) Part-of-Speech Tagging and Lemmatization for the Spoken Dutch Corpus. In: M. Gravididou et al. (eds.) *Proceedings of the Second International Conference on Language Resources and Evaluation* (1427-1433). Athens.

van Mol, Mark (2003) Evolution of MSA, the Case of some Complementary Particles. In: D. Parkinson & S. Farwanah (eds.) *Op. Cit.*, pp. 135-147.

- (2000) Exploring annotated Arabic corpora: Preliminary results. http://www.ilt.kuleuven.ac.be/ilt/arabic_pdf/tunis.pdf. (Consultée le 12/10/2007).

van Mol, M & H. Paulussen (2001) AraLat: A relational database for the development of bilingual Arabic dictionaries. In: S. Lee (ed.) *Proceedings of Asialex 2001, Asian Bilingualism and the Dictionary* (206-211). Seoul, August 2001.

Váradi, Tamás (2001) The linguistic relevance of corpus linguistics. In: P. Rayson, A. Wilson, T. McEnery, A. Hardie & S. Khoja (eds.) *Proceedings of the Corpus Linguistics 2001 Conference* (587-594). Lancaster: Lancaster University.

Varantola, Krista (2002) Disposable corpora as intelligent tools in translation. In Tagnin, S.E.O. (Org.) *Cadernos de tradução: Corpora e Tradução*. Florianópolis: NUT, 2002, vol.1, n°9, p.171-189.

- (2000) Translators, dictionaries and text corpora. In: S. Bernardini & F. Zanettin (eds.) *I corpora nella didattica della traduzione. Corpus Use and Learning to Translate* (117-136). Bologna: CLUEB.

Venuti, L. (2000) (ed.) *The Translation Studies Reader*. London/New York: Routledge.

Véronis, Jean (2000) Alignement de corpus multilingues. In: Jean-Marie Pierrel (ed.) *Ingénierie des Langues*, chapitre 6, pp.151-171. 1^{er} éd., Paris: Edition Hermès.

Véronis, J & Ph. Langlais (2000) Evaluation of parallel text alignment Systems: ARCADE. In: J. Véronis (ed.) *Parallel Text Processing*. Dordrecht: Kluwer.

Vicente, Christian (2006) Linguistique de corpus et traduction spécialisée: Eléments pour approche traductologique du lexique spécialisé. In: D. Gouadec (Dir.) *Traduire pour le Web* (147-156). Paris: La Maison du Dictionnaire.

Vinay, J.P. & J. Darbelnet 1995 [1958] comparative stylistics of French and English. A methodology for translation. Amsterdam / Philadelphia: John Benjamins Publishing Company

Williams, Geoffrey C. (2006) La linguistique et le corpus: Une affaire prépositionnelle. In: F. Rastier & M. Ballabriga (Dir.) *Corpus en Lettres et Sciences Sociales – Des documents numériques à l'interprétation, actes du XXVII colloque d'Albi, Langues et Signification*, pp. 151-158.

Willis, Jane (1998) Concordances in the classroom without a computer. In: Brian, Tomlinson (ed.) *Materials development in language teaching*. Cambridge.

Wilss, W. (1982) The Science of translation. Problems and methods. Tübingen: Gunter Narr Verlag.

Wollin, L. & H. Lindquist (eds.) (1985) *Translation Studies in Scandinavia*. Lund Studies in English 75. Lund: CWK Gleerup

Yoda, Lalbila Aristide (2005) *La traduction médicale du français vers le mooré et le bias*. Thèse.

Zanettin, Federico (2002a) Corpora in translation practice. *Proceedings from First International Workshop on Language Resources (LR) for Translation Work and Research*, Gran Canaria, 28 May 2002.

- (2001) Swimming in words: Corpora, translation and language learning. In: G. Aston (ed.) *Learning with corpora* (177-197). Bologna/ Houston, TX: CLUEB / Athelstan.
- (1998) Bilingual comparable corpora and the training of translators. *Meta* 43(4), pp. 616-630.

Zemanek, P. (2001) CLARA (Corpus Linguæ Arabic)/ An Overview in ELSNET (ed.) *Proceedings of ALC /EACL workshop on Arabic language processing*. Toulouse, France. Retrieved: April 23, 2007. Online at: <http://www.elsnet.org/acl2001-arabic.html>

Zimina-Poirot, Maria (2004) *Approches quantitatives de l'extraction de ressources traductionnelles à partir de corpus parallèles*. Ph.D Thesis, Paris Sorbonne University-Paris 3.

الفهرس

كلمة لا بد منها

أ- د	مقدمة
1	الفصل الأول: المدونة
2	I- ماهية المدونة
6	II- صناعة المدونة
7	II- 1. تصميم المدونة
9	II- 2. تجميع المدونة
9	II- 2- 1. الانتقال
15	II- 2- 2. مصدر النصوص
18	II- 3. إعداد النصوص
18	II- 3- 1. بيانات الوثيقة
18	II- 3- 2. التنقيح
19	II- 3- 3. الترميز
21	الترميز بأقسام الكلام
23	الإعراب
24	III- أنواع المدونة
28	VI- لسانيات المدونة: المصطلح و المنطلقات التاريخية، المنهج و الأدوات الإجرائية
28	VI- 1. المصطلح
30	VI- 2. المنطلقات التاريخية
33	VI- 3. المنهج و الأدوات الإجرائية
33	VI- 3- 1. المنهج
41	VI- 3- 2. آليات البحث و أدواته
41	VI- 3- 2- 1. الفهرسة الآلية
45	VI- 3- 2- 2. قائمة الوصيلات
46	VI- 3- 2- 3. جدول التكرار
48	الفصل الثاني: واقع البحث باعتماد المدونة عند الغرب و العرب
49	I - الواقع عند الغرب

55	1-I. المدونات في اللسانيات و الأبحاث اللغوية
56	1-1-I. المعجمية، المعجماتية و المصطلحية
61	2-1-I. المدونات، النحو و الصرف
64	3-1-I. المدونات، الداليات و الأسلوبيات
65	4-1-I. المدونات و المعالجة الآلية للغات
66	5-1-I. المدونات و اللسانيات الاجتماعية
67	2-I. المدونات و تعليم اللغات
73	3-I. المدونات و الترجمة الآلية
76	II- الواقع عند العرب
76	1-II. المدونات في خدمة العربية
82	2-II. المدونات العربية
82	1-2-II. مدونة كلارا
82	2-2-II. مدونة نيوزوير العربية
82	3-2-II. مدونة الحياة (2002)
83	4-2-II. مدونة النهار (2001)
83	5-2-II. مدونة العربية الفصحى
83	6-2-II. المدونة العامة للعربية العلمية
83	7-2-II. مدونة أرابيك جيغا وورد
83	8-2-II. المدونة المتوازية متعددة اللغات
85	9-2-II. مدونة اللغة العربية المعاصرة
88	10-2-II. مدونة اللغة العربية الدولية
105	الفصل الثالث: المدونة في الترجمة تنظيرا و تطبيقا
106	I- الترجمات باعتماد المدونات: الأصول و الاتجاهات
109	1-I. الشموليات الترجمية
110	1-1-I. التبسيط و تلافي التكرارات في النص المصدر
111	2-1-I. الإظهار
113	3-1-I. التطبيق
115	2-I. أنواع المدونات المعتمدة في الترجمة
115	1-2-I. مدونات أحادية اللغة
116	2-2-I. مدونات ثنائية/متعددة اللغة
122	3-2-I. تنظير المدونات المتوازية

126	4-2-I. استخدامات المدونات المتوازية و المقارنة.....
131	II- المدونات و تدريب المترجم
131	1-II. الترجمة و التدريب.....
133	2-II. تجربة الترجمة باعتماد المدونة.....
135	II 1-2 - الفرضيات
135	II 2-2- العينة المشاركة
135	II 3-2- العينة النصية
136	II 4-2- الأدوات و الموارد المتاحة.....
144	II 5-2- التجربة
144	II 6-2- نتائج الورشتين و تحليلها.....
161	II 7-2- تعاليق الطلبة و تساؤلاتهم.....
163	II 8-2- خلاصة التجربة.....
165	خاتمة
171	ملحق.....
172	الملحق رقم 1: المرمزات الآلية
173	الملحق رقم 2: سلسلة شارات ترميز مدونة بان
174	الملحق رقم 3: برنامجيات الفهرسة.....
175	الملحق رقم 4: مدونات عربية
176	الملحق رقم 5: النص و ترجمته
178	معجم البحث
182	مكتبة البحث
197	الفهرس
200	الملخصات.....
201	الملخص باللغة الانجليزية
214	الملخص باللغة الفرنسية

الملخصات

Abstract

"Gone are the days of the translator armed with pen and paper, or a typewriter, and dusty dictionaries." It is thus clear that, nowadays, translation schools and departments have to pay a great attention to translation technology and other modern electronic tools relevant for translators training, and integrate them in their translation curricula, i.e., considering and treating computer-aided translation technology as an integral part of the trainee's competence. The emerging discipline of corpus linguistics is one area which looks promising in this respect.

Bearing this in mind, we want through this study to explore the potential of corpora and corpus investigation tools in translator education.

Our study is divided into three main chapters:

We have started with a preliminary chapter that introduces the pertinent concepts and lays the foundation of the rest of this work. We have defined 'corpora' as large collections of machine-readable texts (typically *.txt*, i.e. Plain Text format), either spoken or written, compiled on the basis of a set of well-defined criteria. Each text, in turn, is considered as a naturally occurring example of language. To put it in simple words, we can say that a text is an instance of language use or behavior.

We have mentioned that the contents of a corpus should be selected according to their "communicative function in the

community in which they arise" (external criteria), so that they represent, as far as possible, the language or variety of language from which they are chosen.

In a nutshell, 'representativeness' is one of the main features of a well-designed corpus. The search goals should also be paid attention to.

Once texts have been selected to be included in a corpus, they may be either left in their 'raw' state, or enriched with additional information. This process is known as 'corpus annotation'.

Basic mark-up may include information about the source of the text: information about its author(s)/speaker(s), publisher, copyright and contextual information, e.g. regional and social background of the author/speaker, their age and sex, place and time of publication/recording, etc. Moreover, some information about the original document structure may be added (e.g. headings, paragraphs). Higher level mark-up may involve meta-linguistic information, such as part-of-speech tagging (where every word in the corpus has a tag indicating its word class), syntactic parsing (information about the tree structure or the grammatical relations of the constituents), or even semantic annotation (where words, clauses, or phrases are tagged for semantic information such as semantic field, metaphor, and so on).

Studying language through corpora is known as 'Corpus linguistics'. Unlike Chomskyan Generative Grammar, corpus-based approaches to language study insist upon the primacy of

‘performance’, they do not rely on intuitive judgments about grammaticality supported by invented examples. All the texts in a corpus are authentic, attested instances of naturally-occurring linguistic data. Corpus linguistics thus is an empirical approach that involves studying examples of "what people have actually said, rather than hypothesizing about what they might or should say."

Depending on the purpose and the form, many types of corpora may be distinguished: written vs. spoken corpora, general vs. specialized, monolingual vs. bi-/multilingual, synchronic vs. diachronic, native speaker vs. learner, raw vs. annotated, parallel vs. comparable.

Corpora are not designed to be read linearly by a human reader. Rather the data is available to be processed by corpus analysis tools. Such corpus investigation software allows the user to manipulate and display the information in a corpus in various useful ways quickly and with great accuracy that would not be possible if undertaken manually.

Most corpus research software performs two basic functions. They reorder the corpus contents so that they can be observed and investigated by the user, and they calculate statistical information about the data. These functions can be assumed in three main ways: Concordances, collocate lists and frequency lists.

A concordance is an alphabetical listing of all occurrences of a particular search term, together with the contexts in which it appears. The context may be presented in a variety of ways; the

most commonly used one is called KWIC (Key Word In Context) format, in which the node is highlighted in the middle of each line, surrounded by a predefined extent of the context on either side. It should be remembered that each concordance line is independent of the line before and after, but they may be from the same text.

Furthermore, KWIC concordances can be sorted in order of appearance in the corpus, or alphabetically according to the word appearing directly to the left or the right of the key word. Most concordances allow many functions, like: case-sensitive/non-case sensitive searches, wildcard searches (e.g. *go** to retrieve *go*, *goes*, *going*, *gone*, etc.).

A collocate list is a list of all words occurring with the search item more often than simple chance would suggest. Usually, 'collocates' range from the fourth or the fifth word to the left to the fourth or the fifth word to the right. This measurement in words of the co-text of the node is often referred to as a 'span'. Furthermore, collocates are listed together with their frequencies of occurrence at that position. Collocate lists may be presented alphabetically or by frequency.

Users may also discover how many different words are in the corpus (*Types*) and how often each appears (*Tokens*). The list of all *types* together with the number of *tokens* is referred to as a 'frequency list'. Frequency lists can usually be arranged in order of ascending/descending frequency or in alphabetical order. The *type-token ratio* can be calculated: It can be said that the higher the ratio

the harder the text is to comprehend. Moreover, in some concordancers, there are the so-called 'Stop lists'; they are lists of words to be removed or ignored, e.g. common function words such as prepositions, conjunctions, etc.

Chapter two focuses on two axes. The first axis gives a survey of the importance of corpora in language research, e.g. lexicology, lexicography, grammar, stylistics, semantics, NLP, language teaching/learning and automatic translation.

The second axis demonstrates how the Arabic language lacks corpora and tools for corpus analysis. We have discussed the relevance of such resources to many Arabic language applications. Then, we have tried to show some trials that have been conducted to build Arabic corpora, which were planned to support various Arabic studies that depend on authentic data. We have come to the conclusion that most of these trials were unsuccessful and others were for commercial purposes. We have stressed on the increasing need of an Arabic corpus to represent the Arabic language.

Chapter three encompasses two parts. Part one deals with incorporation of the methods and tools of corpus linguistics into descriptive translation studies, which resulted in the development of a new methodology for inspecting translational-linguistic issues, i.e. Corpus(-based) Translation Studies.

CTS are "the application of corpus analysis techniques, both quantitative and qualitative, to the study of aspects of the product and process of translation." This new area of research within

translation studies is motivated by an interest in the study of translated texts as instances of language use in their own right.

One of the fundamental concepts in corpus-based translation studies has been the notion of ‘Translation Universals’, which resulted from the examination of the linguistic features characteristic of translational language. These features are universal to all translations, regardless of the source and target text languages. They include explicitation, simplification and normalization of the source text, as well as other related phenomena.

In that sense then, the emergence of corpus-based trend within the field of TS can be regarded as a direct consequence of a major change of perspective in TS that displaced the research focus from the source text to the target text.

In CTS, scholars have constructed a wide array of different types of corpora. They reflect the criteria according to which they are designed and the purpose for which they are created.

1- The multi-source language monolingual ‘comparable’ corpus: It consists of two sets of texts, one originally written in language A and one of similar texts translated into language A from a variety of different languages.

2- The bilingual/multilingual corpus: language pairs are put together either on the basis of parallelism or/and comparability.

Comparable bilingual corpora can be defined as a collection of texts composed independently in the languages involved, and put together on the basis of similarity of content, domain and

communicative function. Parallel bilingual corpora consist of texts originally written in a language A and their translation(s) into language B. source texts and target texts may be aligned on the level of words, patterns, sentences or even paragraphs or chapters.

Several studies showed that translation students with access to a comparable corpus, are able to enhance their understanding of the source texts and improve their ability to produce translations in the target language more fluently, since they are allowed to see how similar meanings were expressed in texts serving similar functions.

As regards parallel corpora, it has been observed that in the process of establishing equivalencies between parallel sets of texts, learners acquire information about the way in which an idea in one language is conveyed in another language. They can infer from them the translational strategies adopted by the translator(s).

It is worth noting that these studies are generally using corpora to translate out of the foreign language and into the native language; it is our belief that when translation is from a mother tongue into a foreign language, "the need for corpus tools grows exponentially". In the remaining part of this chapter, we tried to prove that fact through a pilot study carried out to see whether corpora could be a useful translation resource.

Our experiment is based on the assumption that a monolingual target-language corpus can be a useful resource for trainees translating from Arabic into English. We consider it will be more

useful than conventional resources in that it helps students improve the quality of their translations by:

- providing them with information about collocates and phraseological units;
- helping them to verify or reject decisions based on other tools as dictionaries;
- enabling them to confirm intuitive decisions;
- using target language patterns;
- learning how to use new words and expressions;
- extracting useful concepts and terms.

Then, participants were chosen for the experiment. Being at the fourth year of studies, students were native speakers of Arabic, because one of our objectives is to see if the corpus can help translators improve their foreign language competence. Moreover, they should be familiar with the basic principles of translation, and should have followed several training courses in translation from Arabic into English, dealing with press articles; but they had not dealt with the topic of the text to be-translated. Furthermore, they should be computer literate.

The text used for the experiment was a 215-word excerpt from a variety of articles taken from different agencies. It was selected according to two criteria:

- It should be short, because the experiment relied upon the good will of the students who volunteered their time to participate;

- Because participants were "novice" translators, i.e. students, the texts should be specialized, but not highly technical.

In the first workshop, the 10 trainees were divided into 5 groups. They were presented with the text to be translated, and were not informed of the objective of the experiment. Each group was given 2 hours to translate the text using their conventional resources as necessary: monolingual/bilingual dictionaries, some press articles relating to the subject field, as well as traditional search engines.

Translations were collected without giving any immediate feedback.

The second workshop took place a longer time delay after the first one. First, students were presented with WebCorp, the corpus analysis tool used in our experiment.

WebCorp is an online search tool that uses the Web as a corpus, producing concordance lines based on results from the search engines Google, AltaVista, Yahoo, LiveSearch, Ask, MetaCrawler and Open Directory.

After have been given some training how to use this tool, students became familiar with its basic features, i.e. a concordancer that finds and displays all occurrences of a search term in a KWIC format, and a collocation viewer that allows them to see which words "go together", as well as word list viewer.

The corpus used was a variety of newspaper articles downloaded from Times online, Telegraph, Guardian, Observer, Independent, Mirror, The Sun, News of the World, Daily Star, NY

Times, LA Times, Sun Times, Washington Post, and from BBC website.

Participants were then asked to translate the same text using WebCorp.

Finally, they were asked to comment on the usefulness of different types of resources, their attitudes to working with them, and their observations and impressions.

To put it in a nutshell, the examination of translations produced with the aid of conventional resources and trainees' comments led us to find that in spite of repeated warnings from translator trainers to 'beware' of dictionaries, students are still very dependent on them. In other words, we can say that our student have a blind faith in their dictionaries to such an extent that they incorporate almost anything they encounter there into their translations, because they believe that it must be acceptable for inclusion.

Furthermore, when they cannot find an answer in a dictionary, they feel obliged to resort to their intuition to make a decision about a given choice. Inevitably, such choices that could not be found in dictionaries were present in other types of resources such as press articles. However, none of the trainees used them, because consultation of such resources is too time-consuming.

By contrast, the analysis of translations produced with the help of WebCorp, revealed that users have a lot of 'faith' in corpus

because its contents were authentic instances of language that had been produced by native speakers.

Above all, this experiment supports our hypothesis that a corpus can be a useful performance-enhancing tool in translating. Trainees could ensure a native-like and natural translation. Such "naturalness" is probably more difficult to achieve if undertaken otherwise, particularly when translating into the foreign language.

Corpora also generally allow unpredictable, incidental learning in that the user may notice somewhat unexpected or 'serendipitous' discoveries. Broadly speaking, creative searching results in a wealth of unfamiliar or unknown uses, and leads to answers to questions that the translator did not even think of asking in the first place.

It is worthy of notice that corpora reveal regularities, not rules; i.e. no corpus can tell us that something cannot be said— but it can provide some evidence as to how likely a word or expression is to occur, i.e. how it is commonly used.

The evidence that emerges from data may help avoid a literal translation which, even if possible in principle, may have the effect of sounding awkward and non-native-like. Furthermore, this evidence makes it easier for translators to take radical decisions, and helps them "be less bound to the source material and feel much more confident when deviating from the way things are expressed in the source material if they feel that the changes are justified." (Varantola, 2003, p. 67)

In this respect, researchers and trainers in the field of translation have recommended to incorporate corpora into the academic setting, and regard them as an essential part of modern translational competence. In practical terms, this means that translator training should include courses in corpus use, corpus compilation and corpus analysis software.

However, many of the students may be resistant to using the corpus, and they often claim to prefer working with the conventional resources. In fact, "whenever a new method of working is introduced, there is always a learning curve involved." Thus, it is up to the trainer to reduce this resistance, and change that way of thinking.

With this end in view,

- We exhort trainers to teach their students that "Exposure is training". To put it in simple words, we learn to behave, or how to behave in particular situations, because we are exposed to the behavior of others and, as far as translation is concerned, this principle is equally valid;

- Students should design and use their own 'mini', 'do-it-yourself', or disposable corpora as a means of translating, terminology extraction and genre analysis.

- "Skills for vertical reading of concordance lines ... are different from those required for horizontal reading of a continuous text. This aspect should be explicitly taught in the classroom."

- Trainers should encourage their students to become active participants in the teaching/learning process (compiling corpora, selecting texts, checking, etc.).

- They should be aware that concordances can be printed out and examined in the translation class as an aid to the analysis of source and target text features. Corpora can also help them evaluate students' translations and provide more objective feedback.

- Trainees should realize that computer skills are an absolute must. Besides learning to surf the Internet and mastering Word and Excel, they are also required to be able to use specific search tools like WebCorp.

- They should learn they have to observe how language behaves at all levels, from the lexicon to the text.

- They should realize that the additional time required by creating and consulting a corpus is compensated for by saving in other translation-related tasks, such as dictionary consultation. Moreover, the effort is rewarded by improving the quality of their translations.

On the whole, the present study, even if being conducted on a very small scale, is hoped to be of benefit to those interested in teaching translation and training would-be translators. It is also recognized that there is still much work to be done; however, we believe that this study can be seen as a small first step in this area.

Résumé

A l'heure actuelle, le marché de la traduction connaît une évolution constante et rapide. En parallèle, de nombreux outils et logiciels propres à l'industrie de la traduction ont vu le jour, principalement en raison de la généralisation des technologies de l'information.

Une connaissance approfondie de technologies et des outils d'aide à la traduction constitue désormais une nécessité incontournable pour tout traducteur, puisqu'ils font partie intégrante de l'environnement de son travail, permettant d'améliorer la qualité et d'augmenter la productivité.

En terme de formation pour ce marché de plus en plus complexe, il dès lors impératif non seulement de développer les compétences linguistiques et de traduction —base fondamentale de cet apprentissage— mais également de forger des compétences solides en technologies.

Nombreux sont les didacticiens de la traduction qui ont revendiqué la redéfinition des compétences du traducteur qui s'avère nécessaire pour s'adapter à ce marché émergent et le conquérir. Néanmoins, la situation n'a guère changé et encore aujourd'hui, "les formules pédagogiques utilisées dans de nombreux cours pratiques de traduction se limitent aux méthodologies traditionnelles (performance magistrale, traductions collectives et la méthodologie de 'lisez et traduisez')."

Probablement, la revendication devrait commencer par l'établissement du pont manquant entre l'enseignement de la traduction et la composante technologique. Et c'est dans ce sens, il nous semble, que l'on peut dire que la découverte (ou la création) du nouveau continent, miraculeusement surgi, qui est la '**Linguistique de corpus**', semble être prometteuse à cet égard.

Compte tenu de cela, la présente étude porte sur les corpus et le rôle qui sont appelés à jouer dans l'enseignement de la traduction. En vue d'un premier déblayage du terrain, nous avons commencé par un chapitre préliminaire qui constitue la charpente théorique de tout mémoire. Il s'agit d'un cadre conceptuel présentant l'ensemble des notions auxquelles nous avons fait appel tout au long de la recherche.

Ainsi, nous avons d'abord tenté de cerner le concept de "**corpus**".

Dans les différentes définitions de "corpus" élaborées par différents spécialistes de la question, peu importe l'application pour laquelle il est construit, on trouve les notions fondamentales de "collection", d'"échantillon", et de "représentativité", ainsi que les notions de "critères explicites" et de "format électronique".

D'une manière générale, le corpus constitue une *collection* de textes écrits ou parlés –disponibles ou convertis en *format électronique* (en général, format .txt)- portant sur un domaine donné. De plus, la sélection des textes doit reposer sur des *critères* de rassemblement bien précis, en fonction des objectifs de la recherche. Pour parler clair, le linguiste de corpus, au lieu de forger

lui-même des exemples de ce qu'il croit être l'**usage**, se fonde sur des textes attestés, authentiques d'utilisation de la langue.

Par texte *authentique*, on entend un texte crée dans une situation communicative ordinaire.

Ces *échantillons* textuels doivent être en effet *représentatifs* de la langue et du phénomène à décrire ou à étudier. En d'autres mots, un corpus est dit *représentatif* s'il reflète le plus fidèlement possible un usage dans une langue et plus précisément des faits de langue réels que l'on souhaite étudier. Il est également important de rappeler qu'il faut diversifier les auteurs ainsi que les collectivités d'appartenance, pour assurer une meilleure représentativité des usages.

En ce qui concerne la taille d'un corpus, il n'existe pas de consensus sur la **taille** idéale qu'il devrait avoir. Pour certains, avec qui l'on abonde dans le même sens, elle devrait être déterminée en fonction des objectifs de l'étude, de même qu'en fonction du caractère des textes ainsi que des domaines auxquels ils appartiennent. Fort est pourtant de reconnaître que la taille de corpus est aussi souvent limitée par la disponibilité des textes, d'une part, et la question des droits d'auteurs, d'autre part.

Les textes constituant un corpus peuvent être accompagnés d'informations spécifiques. Ces enrichissements sont connus sous le nom d'"**annotation de corpus**". On distingue:

- L'annotation de structure : les attributs de formatage (paragraphes, sections, titres, etc.) ;

- L'information textuelle : information sur la source du texte (auteur/locuteur, âge, sexe, niveau d'études, appartenance sociale, type de texte, registre, date et lieu de publication/d'enregistrement, etc.).
- L'annotation linguistique : elle comprend l'analyse linguistique du contenu de chaque texte, telle que : étiquetage morphosyntaxique (tagging), lemmatisation, analyse syntaxique, marquage prosodique des corpus oraux transcrits, annotation des erreurs dans les corpus des apprenants.

Un type d'annotation bien connu est l'*étiquetage grammatical* qui spécifie la catégorie grammaticale de chaque mot dans le texte selon son contexte. Ces enrichissements du corpus brut permettent au chercheur de formuler des requêtes bien plus ciblées qu'auparavant. Il pourra obtenir, par exemple, les *occurrences* de telle ou telle séquence de catégorie grammaticale (par exemple : adverbe+ adjectif+ nom), d'un certain lemme (par exemple, le verbe *boire*) quelle que soit sa forme fléchie (*bois, boit, buvons, boive, etc.*), ou d'un certain mode verbal (par exemple, le subjonctif).

L'annotation proprement syntaxique est plus sophistiquée: ici on cherche à fournir pour chaque énoncé son arborescence. Ceci permet d'effectuer une analyse syntaxique superficielle qui fournit les constituants (groupe nominal, verbal, prépositionnel, etc.).

Il va de soi que même si les corpus non-annotés (ou bruts) constituent des ressources intéressantes, la présence des annotations les rend plus utiles dans la recherche linguistique.

L'approche de la langue fondée sur les corpus s'appelle "**la linguistique de corpus**".

A la différence de la linguistique chomskyenne, qui se limite avant tout à l'étude de la '*compétence linguistique*' considérée comme le véritable objet d'analyse, la linguistique de corpus s'est surtout intéressée à la parole, et déplace le centre de gravité de la '*compétence*' vers la '*performance*'. Autrement dit, cette approche envisage la langue selon une perspective *empirique*, dans la mesure où elle conçoit que le corpus est un observable nécessaire en linguistique. Or, au moins deux approches se font face pour peut-être se compléter. Pour les uns, le corpus est un "*observatoire* d'une théorie *a priori*" (l'approche appliquée au corpus/ *corpus-based approach*), pour les autres, le corpus est un "*observé dynamique* qui permet de décrire puis d'élaborer des modèles *a posteriori*" (l'approche motivée par le corpus/ *corpus-driven approach*).

Pour sa part, le texte est "l'unité minimale, et le corpus l'ensemble dans lequel cette unité prend son sens."

Etant donné qu'un corpus est un objet construit en fonction des objectifs définis au préalable, on distingue, entre autres, des corpus écrits vs. corpus parlés, corpus (à objectifs) généraux vs. corpus spécialisés, diachroniques vs. synchroniques, ouverts (constamment enrichis) vs. fermes, corpus d'apprenants vs. corpus

de locuteurs natifs, corpus monolingues vs. corpus bi-/multilingues, corpus annotés vs. non-annotés.

Les corpus ne sont pas faits pour être consultés de manière séquentielle (comme un livre), mais pour être interrogés à l'aide des outils de manipulation des données. La plupart de ces logiciels d'interrogation de corpus accomplit deux fonctions de base :

- Réorganiser les contenus du corpus d'une manière qu'il est permis au chercheur de les exploiter avec clarté et précision ;
- Traduire ces données en langue des chiffres (des calculs statistiques).

Pour ce faire, trois caractéristiques de ces logiciels sont distinguées : les concordances, les listes de collocatifs et les listes de fréquences.

Une **concordance** est la liste de toutes les occurrences d'un **pivot**, alignées verticalement en colonne, entourées de part et d'autre par leur **contexte**.

En pratique, une concordance se calcule à partir de trois éléments :

- Un pivot*, typiquement un mot (mais cela peut être généralisé à toutes sortes d'items dont les occurrences sont repérables par le **concordancier** : lemme, expression, etc.), aux contextes duquel on s'intéresse ;
- Une taille de contexte* (éventuellement détaillée en contexte gauche et contexte droit, qui respectivement précède et suit le pivot). La quantité de contexte du **mot clé** peut être modifiée.

- iii. *Un ordre de présentation des contextes sélectionnés*, typiquement ordre de présence dans le corpus, ou tri alphabétique sur le mot qui précède le pivot (tri gauche) ou sur celui qui le suit (tri droit).

Il est important de mentionner que chaque ligne est indépendante de la ligne avant et après, mais il se peut que deux lignes appartiennent au même texte.

Pour ce qui de la *forme à chercher*, il peut s'agir d'une simple séquence de caractères. Le concordancier montre tous les passages où apparaît cette séquence.

A la place de la séquence de caractères, on peut également préciser une expression régulière. Il s'agit d'une formule qui couvre plusieurs séquences de caractères.

Les '*Recherches KWIC*' (mot clé en contexte) permettent de chercher des chaînes de caractères ou des expressions régulières dans les textes du corpus.

Dans le champ '*Mot ou expression*', l'utilisateur introduit le(s) élément(s) à chercher : un mot (entre deux blancs), plusieurs mots (également entourés de blancs), une partie d'un mot (sans blancs), une expression régulière. Puis, il détermine la '*largeur du contexte*', et le '*nombre maximale de réponses*' souhaitées. Enfin, il choisit le(s) *corpus* dans le(s)quel(s) il veut effectuer une quête. Les différentes occurrences de l'élément cherché sont affichées en rouge dans certains concordanciers.

La liste des **collocatifs** est la liste de tous les mots qui tendent à apparaître fréquemment avec le mot clé. Ces collocatifs s'arrangent à partir du 4^{ème} ou 5^{ème} mot à gauche au 4^{ème} ou 5^{ème} mot à droite. De plus, ils sont répertoriés avec leurs fréquences d'apparition à cette position. Les collocatifs peuvent être triés alphabétiquement ou par ordre de fréquence.

A partir de la liste des **fréquences**, l'utilisateur peut délimiter le nombre de **types** (c'est-à-dire, toute unité lexicale figurant dans le corpus) et la fréquence d'apparition de chaque type (le **token**).

De même, ce genre de listes peut être présenté par ordre alphabétique, ou par ordre croissant/décroissant de fréquences. En outre, dans certains concordanciers, il y a ce que l'on appelle '**Stop listes**' qui sont des listes de mots à ignorer, tels que les prépositions, les conjonctions.

Après avoir présenté tout cet appareillage qui appartient au cheptel notionnel de l'approche linguistique basée sur corpus, et dressé un bref aperçu historique, nous avons vu qu'il était temps d'apprécier l'apport de la linguistique de corpus aux différents domaines de recherche, à l'instar de la grammaire, la lexicologie, la lexicographie, la terminologie, la terminographie, la stylistique, la sémantique, le TAL, l'enseignement/l'apprentissage de langues, ainsi que la traduction automatique.

C'est au premier volet du chapitre II que nous avons tenté d'apporter un meilleur éclairage sur la place qu'occupent les corpus dans chaque domaine mentionné.

Dans le second volet, nous avons cru bon faire une présentation assez détaillée de nombreuses applications des corpus en Arabe, avant de faire un bref survol des différents travaux menés dans le cadre de quelques études arabes basées sur corpus. Les constatations faites dénotent un manque flagrant de ressources et d'outils.

Le troisième chapitre, quant à lui, est subdivisé en deux parties. Dans la première partie, nous nous sommes plus particulièrement sur l'exploitation des corpus en vue de la recherche portant sur la traduction. Cette exploitation a permis l'essor d'une nouvelle approche qui envisage les diverses questions traductionnelles.

La traduction basée sur corpus est motivée par la conviction que les textes traduits sont des exemples attestés, authentiques d'usage de la langue des traductions. L'observation minutieuse de cette langue a permis de mettre en lumière des caractéristiques linguistiques universelles qui sont propres à toutes les traductions, peu importe la langue source ou la langue cible. Il s'agit des "**universaux de la traductions**"; parmi lesquels, on distingue l'**explicitation**, la **simplification** et la **normalisation** du texte source, pour ne pas citer que quelques exemples.

Ainsi, nous pourrions dire que l'essor de *la traductologie basée sur corpus*, branche du domaine-parent 'la traductologie', indique un changement dans l'intérêt et la perspective qui a déplacé

le centre de gravité de recherche en traduction du texte source vers le texte cible.

Divers types de corpus sont déjà construits. Ils reflètent les critères selon lesquels ils sont conçus et la finalité pour laquelle ils sont créés.

On appelle **corpus parallèle** un ensemble de couples de textes tel que, pour un couple, un des textes est la traduction de l'autre. Il est intéressant d'**aligner** ces corpus, c'est-à-dire de faire correspondre chaque unité du texte en langue source avec chaque unité du texte en langue cible (au niveau des paragraphes, phrases et mots) pour disposer d'un jeu de donnée bilingue, en particulier dans les domaines spécialisés où le vocabulaire et l'usage des mots et des expressions évoluent rapidement.

Un **corpus comparable** est composé de textes dans des langues différentes mais partageant une partie du vocabulaire employé, ce qui implique généralement que les textes parlent d'un même sujet, à la même époque et dans un registre comparable.

Plusieurs études ont montré que l'utilisation d'un corpus comparable lors d'un exercice de traduction, permet aux apprenants d'appréhender le texte source, et d'améliorer leur capacité à produire des traductions acceptables dans la langue cible, dans la mesure où ils sont autorisés de voir comment un même sens a été exprimé dans des textes servant à des fonctions similaires.

En ce qui concerne le corpus parallèle, il a été observé que dans le processus d'établir des équivalences entre des ensembles de

textes, les apprenants acquièrent des informations sur la façon dont une idée dans une langue est transmise dans une autre langue. De plus, ils peuvent en déduire les diverses stratégies traductionnelles adoptées par le(s) traducteur(s).

En regardant de près ces études, on constate qu'elles ont généralement recours aux corpus pour traduire vers la langue maternelle. Bien entendu, nous croyons que, lorsque la traduction est de la langue maternelle vers une langue étrangère, le recours à ces ressources devient une nécessité incontournable. Dans ce qui a suivi de notre mémoire, nous avons tenté de démontrer comment l'enseignement de la traduction va un tirer un plein profit du travail sur corpus, d'où se manifeste l'originalité de notre travail.

Notre expérimentation reposait sur l'hypothèse qu'un corpus monolingue, composé de textes écrits en anglais (langue cible), pourrait être utile pour les apprenants traduisant de l'arabe en anglais. Nous estimions qu'il serait plus utile que les ressources conventionnelles dans la mesure où il les aide à perfectionner la qualité de leurs traductions pour les raisons suivantes :

- le corpus permet de repérer les collocations ;
- la consultation des données authentiques aide à vérifier des décisions tenues à partir d'autres outils (par exemple, les dictionnaires) ;
- le corpus permet de confirmer les décisions basées sur l'intuition ;

- les apprentis-traducteurs se trouvent confrontés à un éventail des patrons de la langue cible ;
- ils rendent compte de l'usage réel et répandu de nouveaux mots et expressions ;
- ils extraient des termes et des concepts utiles.

Ensuite, les participants ont été choisis pour l'expérimentation. L'échantillon était composé de dix sujets qui ont travaillé deux par deux. Ils étaient tous inscrits en 4ème année au moment de la passation du test. De plus, ils étaient des locuteurs natifs de la langue arabe, vu que l'un de nos objectifs est de voir si les corpus peuvent aider les apprentis-traducteurs, en général, à perfectionner leur compétence en langue étrangère. En outre, ces sujets participants à notre expérimentation et maîtrisant les ABC de l'outil informatique, devraient être familiarisés avec tant de principes fondamentales de la traduction, et avoir suivi plusieurs cours de formation en traduction journalistique de l'arabe en anglais, sans avoir forcément abordé le sujet du texte à traduire.

Il nous semble important de noter que ces apprenants ont participé à l'étude sur une base volontaire sans avoir aucune information relative à leur compétence en traduction (arabe-anglais).

Le texte choisi, quant à lui, était extrait d'un ensemble d'articles de diverses agences de presse. Il était d'une longueur de 215 mots. Ce texte reflète deux critères selon lesquels nous l'ayons choisi :

- Il a été court car, comme nous l'avons mentionné plus haut, les étudiants ont participé à l'expérimentation de façon volontaire ;
- Puisque les sujets étaient des 'apprentis-traducteurs', le texte devrait être spécialisé, mais pas très technique.

Pour le premier atelier, les cinq groupes de participants devaient traduire le texte pendant deux heures, en ayant la possibilité de recourir aux diverses ressources traditionnelles, telles que : les dictionnaires monolingues et bilingues, des articles de presse relatifs au sujet, ainsi que les moteurs de recherche connus. Bien sûr, l'objectif de cette expérimentation n'a pas été mentionné.

A la fin de cet atelier, les traductions ont été ramassées, et aucune explication ou commentaire n'a été fourni aux sujets.

Le deuxième atelier a eu lieu une semaine après le premier. Les trois premières heures ont été consacrées la présentation de WebCorp, le logiciel d'interrogation de corpus utilisé dans notre expérimentation.

WebCorp est un système de traitement de corpus qui permet d'exploiter les textes sur la toile. Ce logiciel permet, en utilisant des moteurs de recherche du Web (Google, AltaVista, Ask, Yahoo, LiveSearch, Meta Crawler et Open Directory), de générer des concordances de mots ou de structures en utilisant le Web comme corpus. Il est possible de restreindre la recherche à un domaine particulier, voire à des types de journaux.

Après avoir reçu quelques instructions sur l'utilisation de WebCorp, les apprenants se sont familiarisés avec ses principales composantes : le concordancier, l'afficheur des collocatifs et des listes de mots.

Le corpus utilisé, pour sa part, est une variété d'articles de journaux en ligne téléchargés à partir des web sites de : Times, Telegraph, Guardian, Observer, Independent, Mirror, The Sun, News of the World, Daily Star, NY Times, LA Times, Sun Times, Washington Post et BBC.

Dans un premier lieu, les sujets ont été invités à traduire le texte précédent à l'aide de WebCorp. Ils ont aussi été invités à noter leurs observations et à faire des commentaires sur l'utilité des différents types de ressources, ainsi qu'à mentionner leurs attitudes à l'égard de ces ressources.

Les résultats obtenus à la suite de l'analyse des cinq traductions, les commentaires émis par les sujets ainsi que l'observation de la façon dont ils se sont servis des outils de référence, nous ont permis de constater que :

- L'ensemble de nos apprenants se fient exclusivement aux dictionnaires bilingues ;
- Cette constatation est peu surprenante tenant compte du fait qu'ils ne savent pas assez bien se servir d'un dictionnaire ;
- Ils jugent comme acceptable tout mot apparaissant dans l'un ou l'autre de leurs dictionnaires ;

- Force est de constater que pour les problèmes de collocations, l'emploi du dictionnaire s'est révélé peu efficace dans la mesure où les sujets ont souvent cherché à valider une collocation douteuse en se référant à l'article du collocatif ;
- Notons par ailleurs qu'ils se fient à leurs propres intuition (!) plutôt que de se référer systématiquement à d'autres ressources, telles que les dictionnaires monolingues (Oxford English Dictionary) ou les articles mis à leur disposition.

Quant à l'analyse des traductions faites à l'aide de WebCorp, nous a révélé que la consultation du corpus a été extrêmement bénéfique pour les apprentis-traducteurs :

- Comme il était évident, le corpus a suggéré des usages qui sont plus idiomatiques, naturels et acceptables ;
- Il a aussi aidé les sujets à voir les divers stratégies déployées par le(s) traducteur(s) pour la résolution des différents types de problèmes rencontrés ;
- La constatation la plus frappante est que le corpus a permis de mettre en lumière un certain nombre d'usages inattendus ou des découvertes "fortuites", qui conduisent à des réponses à des questions que l'utilisateur n'a même pas pensé à poser en premier lieu, favorisant ainsi l'apprentissage imprévisible.

L'ensemble de ces constatations intéressantes nous permet de conclure que le corpus s'avère un atout précieux dans la traduction (arabe-anglais), favorisant l'éveil au langage (Language Awareness –pour emprunter le terme cher à Lynne Bowker) et permettant aux apprenants de trouver réponses à toutes leurs questions.

Toutefois, il est important de mettre en exergue que les corpus révèlent des régularités, et non pas des règles.

L'évidence qui résulte de l'observation des données réelles peut contribuer à l'évitement d'un mot-à-mot qui, même s'il est possible en principe, peut conduire à une traduction maladroite et dénaturisée. De plus, cette évidence permet au traducteur de tenir une décision catégorique concernant un choix donné, et lui libère des servitudes du texte source, dans la mesure où le traducteur se sent plus confiant quand il "s'écarte" de la façon selon laquelle les réalités sont exprimées dans la langue source, notamment si cet 'écart' est justifié.

A cet égard, chercheurs et formateurs dans le domaine de la traduction ont recommandé l'intégration des corpus dans le cursus universitaire en les considérant comme partie intégrante des compétences traductionnelles modernes. En pratique, cela signifie que la formation des traducteurs devrait comprendre des cours portant sur l'utilisation des corpus, leur compilation et les divers logiciels d'interrogation des corpus. Tout ceci dénote la nécessité de revoir certaines pratiques relatives à l'enseignement de la traduction.

Dans cette perspective,

- Les apprenants doivent bien saisir l'idée que l'on apprend à se comporter, ou comment se comporter dans des situations particulières, car on est exposé au comportement des autres et, autant que la traduction est en cause, ce principe est également valable ;
- Il est primordial d'enseigner aux étudiants à se servir de corpus, de les familiariser avec la structure des concordances et de leur faire comprendre la nature des informations y sont présentées. Leur faire comparer certaines ressources permettrait aussi de développer des réflexes de recherche d'information ;
- " Les compétences nécessaires pour lire les lignes de concordances verticalement...sont différentes de celles requises pour une lecture séquentielle d'un texte." Cet aspect devrait être explicitement enseigné dans la classe ;
- Les étudiants doivent percevoir les corpus comme un tout cohérent possédant une certaine nature, structure et finalité, et non comme un regroupement désordonné d'unités lexicales (à l'image des dictionnaires qui présentent le lexique comme une suite linéaire d'unités lexicales n'ayant aucun lien entre elle sinon celui de "commencer par la même lettre") ;
- La présentation et l'exploration adéquates d'un corpus contribueraient, entre autres, à la perception plus

cohérente de sa nature en mettant en lumière les relations qui servent à tisser ce réseau complexe de textes ;

- La consultation d'un corpus en classe peut constituer un exercice fort bénéfique, puisqu'il permet aux apprenants de constater la variété de propriétés inhérentes à la langue, d'acquérir de nouvelles stratégies de résoudre de problèmes traductionnels, et d'apprendre à douter de certains aspects langagiers (le lexique, la terminologie, les collocations, le caractère figé des locutions, etc.) qu'ils maîtrisent moins ;
- Il faut que les apprenants eux-mêmes prennent conscience de leur responsabilité dans le déroulement des cours. Autrement dit, chacun doit participer activement au processus enseignement/apprentissage (la compilation des corpus, la sélection des textes, la consultation, etc.) ;
- Les concordances peuvent être imprimées et examinées en classe afin d'analyser les caractéristiques de textes source et cible ;
- Il est essentiel que les apprenants rendent compte du fait que la composante informatique est exigée dans le profil moderne de quiconque désire être 'traducteur' ;
- Les corpus peuvent également être utilisés pour évaluer les traductions des apprenants avec tant de précision et d'objectivité ;

- Mais, plus que tout, le corpus permet aux étudiants d'apprendre à raffiner davantage leur intuition, à laquelle ils se fient en cas d'absence de justifications possibles de certains choix.

Nous voilà donc au terme de notre recherche sur les corpus et leur rôle dans la formation des (futurs) traducteurs. Evidemment, tout travail de recherche présente certaines limites, et ceci particulièrement vrai dans le cadre d'un mémoire de maîtrise pour lequel le temps et les moyens sont eux-mêmes assez limités. Il est donc important de bien délimiter notre champ d'investigation et de reconnaître les limites découlant des choix effectués. Voici quelques remarques qui méritent d'ailleurs d'être faites à propos de notre travail:

- Le nombre de sujets (10) est très limité et ne permet pas de faire aucune généralisation. Comme nous l'avons souligné à quelques reprises, compte tenu du fait que les sujets ont participé à l'étude de façon volontaire, nous croyons tout de même que l'observation de cinq dyades nous a permis de dresser un portrait assez juste de l'importance que revêtent les corpus dans l'enseignement de traduction. Il pourrait être intéressant de refaire une étude semblable avec plus des étudiants, peut-être en tenant compte, cette fois, de leur niveau en traduction (arabe-anglais) ;

- Pourquoi nous intéresser au WebCorp plutôt qu'à procéder à la construction de notre propre corpus pour mener à bien cette recherche ? La raison est simple : comme nous l'avons mentionné plus haut, la construction d'un corpus serait une tâche largement au-dessus des limites d'un mémoire de maîtrise vu que le temps et les moyens sont assez limités ;
- Si notre recherche présente certaines limites au niveau des échantillons, pour ne pas citer qu'un seul exemple, elle a par contre le mérite de s'intéresser à une question sur laquelle ne portent que très peu de travaux menés dans des universités européennes. De plus, les universités du monde arabe entier n'ont pas encore adopté cette approche, à l'exception de l'université de Koweït qui a construit un corpus parallèle (arabe-anglais) pour enseigner la traduction et la terminologie.

Finalement, nous espérons que, en dépit des limites de la zone explorée, la présente recherche aura permis de défricher une certaine partie de la question, de procurer quelques réponses et d'ouvrir surtout la voie à des travaux futurs de plus grande envergure.

Remerciement

J'aimerais d'entrée de jeu témoigner ma profonde gratitude à mon directeur de recherche, pour l'intérêt qu'il a porté à mon travail. Je tiens également à offrir mes sincères remerciements à ma codirectrice de recherche. Ils m'ont permis, par leur grande générosité, leur infinie patience et leurs conseils inestimables, de mener à terme ce mémoire.

Je veux également souligner ma reconnaissance envers Issam ALMASHHARAWI et M. Benyoucef FERHAT qui ont généreusement contribué à la révision de notre mémoire.

Je tiens à exprimer ma sincère gratitude envers Messieurs Mohammed DAOUD, Wissam TOUHAMI et Abderrahmane ZAOUÏ pour avoir bien voulu faire partie de mon jury et avoir lu attentivement mon mémoire. Merci à eux pour leur patience.

Ce travail n'aurait sans doute jamais été entrepris sans mes parents, mes frères, ma sœur et toute ma famille, qui m'ont vivement encouragée à poursuivre mes études aux cycles supérieurs et qui m'ont donné leur soutien sans faille. Ils ont toujours cru en moi. Je ne serais pas rendue si loin sans votre présence dans ma vie.

Je ne pourrais passer sous silence le soutien précieux que m'ont apporté Mlle Ahlem SEGHOÛR, Mme BENTAYEB Nacera, Bakhta, Fatima et Houaria. A elles aussi je dois ce travail.

*La dernière mais non la moindre ; je désire remercier ma chère amie **Amira** et sa famille pour leur soutien inconditionnel. Je ne saurais oublier toute la générosité que m'a témoignée **Amira** dans les moments les plus difficiles et qui m'a fait prendre conscience de toute la valeur de l'amitié.*

Je vous dédie ce travail...

Fatima Zohra TOUTAOUÏ