# المبحث الثاني

## مدونات المتعلمين

**عبدالله بن يحيى الفيفي**

جامعة الإمام محمد بن سعود الإسلامية

الرياض، المملكة العربية السعودية

ayjfaifi@imamu.edu.sa

# المقدمة

أثبتت المدونات النصية (Corpora) منذ ظهورها قبل بضعة عقود قدرتها على إعادة الصياغة لمنهجية البحث في كثير من الجوانب اللغوية؛ وذلك لما توفره من إمكانات كبيرة جعلتها أساساً لمادة البحث اللغوي الحديث. هذا التطور الكبير ولَّد عدة أنواع من المدونات النصية، التي تختلف تبعاً لأغراضها وطريقة بنائها، ومن ذلك ما يعرف بمدونات المتعلمين (Learner Corpora). وقد أضحت هذه المدونات إحدى ركائز البحث في مجال اكتساب اللغة وتعليمها؛ لما تكشفه من دلالات على مستويات التعليم اللغوي، ومدى ملاءمة إستراتيجياته؛ ثم إن استخدامها انتشر بشكل كبير فبرزت عدة أنواع فرعية ذات أغراض مختلفة، مع تجاوز استخدامها للبحث في اكتساب وتعليم اللغة إلى مجالات أوسع كتأليف المعاجم، ومعالجة اللغة الطبيعية.

وقد ظهرت في الآونة الأخيرة بعض مدونات المتعلمين العربية، وبدأ استخدامها في مجال تعليم اللغة العربية للناطقين بغيرها؛ فبات من الضروري الحديث عن هذا النوع ووضعه في مكانه الملائم، وفي هذا المبحث نستعرض بعض الأسس النظرية التي وضعها الباحثون مثل تعريف مدونات المتعلمين، وبيان أنواعها، وما الذي يميزها عن غيرها، مع استعراض المدونات العربية منها، لينتقل الحديث بعد ذلك بشيء من التفصيل حول سبعة مجالات تمثل أهم النماذج التي استفادت - ولا زالت تستفيد - من مدونات المتعلمين، وهي (١) أبحاث اكتساب اللغة وتعليمها، (٢) التحليل التقابلي للغة المرحلية، (٣) تحليل الأخطاء بمساعدة الحاسب، (٤) دراسة التطور اللغوي لدى الطلاب، (٥) تأليف المعاجم، (٦) تدريس اللغة وتصميم المواد التعليمية، وأخيراً (٧) البحث في معالجة اللغة الطبيعية. عند الحديث عن المدونات اللغوية يبرز جلياً موضوع الوسم(١) (Tagging)، وقد اهتمت مدونات المتعلمين بوسم الأخطاء

(Error tagging) أكثر من غيرها؛ لذا خصّصنا لذلك مبحثاً يشرح هذا النوع، ويستعرض الدراسات التي عُنيت بوسم الأخطاء في اللغة العربية، مع الوقوف على بعض النماذج العملية لطريقة وسم الأخطاء. ونختم حديثنا في هذا المبحث بمجموعة من المصادر العلمية لمن أراد القراءة بتوسع حول مدونات المتعلمين.

## ما هي مدونات المتعلمين؟

لا فرق من حيث البنية بين مدونات المتعلمين والمدونات العامة، ولذا فإن تعريف قرانجر Granger (٢٠٠٢م) لمدونات المتعلمين - وهي أشهر من عرّف هذا النوع - كان مبنياً على تعريف سينكلير Sinclair (١٩٩٦م) للمدونات العامة. إلا أن ما يميز بين هذين النوعين أمران: الأول أن مصدر البيانات لمدونات المتعلمين ينحصر في تلك المواد التي يحررها متعلمو اللغة دون غيرهم، بينما تجمع نصوص المدونات العامة من مصادر متنوعة، والثاني أن الغاية من بناء مدونات المتعلمين تدور في الغالب حول اكتساب اللغة وتعليمها، في حين تستخدم المدونات العامة لأغراض عدة لا حصر لها. ومن هنا تعرف قرانجر مدونات المتعلمين بأنها:

مجموعة حاسوبية من البيانات النصية الواقعية للغة الثانية أو الأجنبية، والتي تم جمعها وفقاً لمعايير تصميم واضحة، ولغرض محدد في مجال اكتساب اللغة الثانية، أو تعليم اللغة الأجنبية، ويتم وسم هذه النصوص بطريقة معيارية ومتجانسة، مع توثيق أصلها ومصدر الحصول عليها (قرانجر، ٢٠٠٢: ٧)

وحتى يكون التعريف أكثر وضوحاً، فسنشرح بعض مفرداته فيما يلي:

**حاسوبية:** المقصود بكونها حاسوبية أن تكون محفوظة في الحاسب الآلي بطريقة تسمح بقراءة هذه النصوص آلياً، فالمستندات النصية المدخلة على شكل صور عن طريق الماسح الضوئي مثلاً لا تندرج تحت هذا التعريف.

**واقعية:** كونها واقعية يعني أن متعلمي اللغة أنتجوها في سياق طبيعي غير مصطنع، ولا بد هنا من التنويه أنه عندما يطلب من الطالب كتابة نص معين بغرض إدراجه في المدونة – أو الحديث عن موضوع معين سواءٌ أكان في مقابلة أم محادثة أم عرض تقديمي أم غير ذلك مع تسجيله صوتياً أو عن طريق الفيديو – فإن لغة الطالب لا تكون واقعية بشكل كامل كما لو كان يحادث أحد زملائه في الفصل مثلاً (قرانجر، ٢٠٠٢م). وبناءً على مقياس نسلهاف Nesselhauf (٢٠٠٤م: ١٢٨) الذي يتألف من أربع درجات: (١) نص طبيعي بشكل كامل، (٢) نص منتج من خلال عملية تعليمية، (٣) نص منتج من خلال مهام تتم مراقبتها والتحكم بها، (٤) وأخيراً برامج نصية ذات سيناريو محدد؛ ترى الباحثة أن مدونات المتعلمين لا تحوي نصوصاً واقعية يمكن تصنيفها تحت الدرجة الأولى، حيث إن منهجية تأليف المدونة تؤثر بشكل واضح في نوعية النصوص المنتجة من الطلاب، فهي في درجة أقل من تلك النصوص الطبيعية الخالصة.

**معايير تصميم واضحة:** يقصد بهذا تحديد العناصر الأساسية لبناء المدونة، والتي تتبع الهدف من إنشائها غالباً، وهذه العناصر كثيرة نذكر منها على سبيل المثال: (١) تحديد الطلاب الذين ستُضم نصوصُهم للمدونة، (٢) نوع المواد التي ستجمع منهم، (٣) تحديد نطاق المدونة (مؤسسة تعليمية، أو مدينة، أو دولة)، (٤) تحديد حجم المدونة (٥) مع منهجية جمعها (٦) ووسمها. ولقد عالج المبحث الثالث من هذا الكتاب موضوع تصميم المدونات بتفصيل أكثر، ولمزيد من المراجع انظر سينكلير (٢٠٠٥)، وقرانجر (٢٠٠٢، و٢٠٠٣ب).

**وسم النصوص بطريقة معيارية ومتجانسة:** يعد وسم النصوص أحد العناصر المرتبطة بشكل وثيق بالمدونات النصية، ومن هذا المنطلق لا بد أن يستند وسم النصوص في مدونات المتعلمين على إحدى المنهجيات المعيارية

المتبعة في مثل هذه العملية، إضافة إلى أهمية تطبيق هذا المنهجية بشكل متجانس على جميع أجزاء المدونة. (للمزيد حول معنى الوسم - ومثله التحشية - انظر «سابعاً: مصطلحات مهمة في مجال لسانيات المدونات» في المبحث الأول من هذا الكتاب).

**توثيق أصلها ومصدر الحصول عليها:** يقصد به المعلومات التي تؤخذ مع كل نص بغرض التوثيق وتسمى البيانات الوصفية للمدونة (Corpus metadata)، أو ترويسة المدونة (Corpus header)، ويعرفها برنارد Burnard (٢٠٠٥: ٤٠) بأنها «بيانات حول البيانات»، أي معلومات إضافية حول نصوص المدونة، وهي في الغالب قسمان: الأول بياناتٌ حول مؤلف النص مثل عمره، وجنسه، ولغته الأم، وجنسيته، ومستواه الدراسي أو اللغوي، ومدة دراسته للغة، إلى غير ذلك. والقسم الثاني معلوماتٌ حول النصّ نفسه، مثل نوعه الأدبي (مقال، أو قصة، أو رسالة أو بحث)، وشكله (مكتوب، أو منطوق)، ومكان وتاريخ تأليفه، وعدد كلماته، ونحو ذلك.

## بداية مدونات المتعلمين

بدأت مدونات المتعلمين مع ظهور «مدونة لونجمان للمتعلمين» Longman Learners' Corpus (شبكة مدونة لونجمان Longman Corpus Network، ٢٠١٢م) التي بدأ العمل عليها في نهاية ثمانينات القرن الماضي، ومع ظهور مدونات أخرى في ذلك الوقت مثل بنك بيانات اللغة الثانية من مؤسسة العلوم الأوروبية (European Science Foundation Second Language Databank) التي تضم نصوصاً مسجلة لبعض المهاجرين الناطقين بلغات مختلفة، إلا أن عدد العينة في هذه المدونة لم يتجاوز أربعة أشخاص، كما أن البيانات المستخرجة منهم لم تكن واقعية بل كان فيها درجة عالية من التوجيه؛ ولذا اعتُبرت مدونة غير معيارية مع أنها استوفت بقية معايير مدونات المتعلمين كإمكانية قراءة

النصوص آلياً (نسلهاف، ٢٠٠٤م). استمر بعد ذلك إنشاء مدونات المتعلمين على مدى العقدين الماضيين إلى أن فاقت المئة بحسب قائمة قرانجر لمدونات المتعلمين حول العالم (قرانجر، ٢٠١٢م)، إضافة إلى وجود عدد آخر منها لم يدرج بعد في هذه القائمة. وبعد هذا الانتشار لمدونات المتعلمين أصبح الكثير منها مفتوح المصدر، يمكن تنزيل نصوصها والبحث فيها باستخدام أي برنامج لتحليل المدونات، مثل المدونة اللغوية لمتعلمي اللغة العربية (الفيفي وآخرون Alfaifi et al.، ٢٠١٤م)، ومدونة المتعلمين العربية للنصوص المكتوبة (فروانة وتميمي Farwaneh and Tamimi، ٢٠١٢م)، وبعضها متاح للاستخدام من خلال مواقعها الخاصة على شبكة الإنترنت، مثل مدونة ميتشغان الأكاديمية للإنجليزية المنطوقة The Michigan Corpus of Academic Spoken English (سمبسون وآخرون Simpson et al.، ٢٠٠٢م)، والمدونة الروسية لمتعلمي الترجمة Russian Learner Translator Corpus (سوسنينا Sosnina، ٢٠١٤م)، ومنها ما هو تجاري مثل مدونة لونجمان السالفة الذكر، وكذلك مدونة كامبريدج للمتعلمين Cambridge Learner Corpus (٢٠١٢م)، وأيضاً المدونة الدولية لمتعلمي اللغة الإنجليزية International Corpus of Learner English (قرانجر، ٢٠٠٣م ب).

## حدود مدونات المتعلمين

مع اختلاف أنواع متعلمي اللغة فإنه قد يشكل على بعض الدارسين أحياناً ما يمكن أن يدخل تحت مسمى مدونات المتعلمين، فهل يشمل هذا النوعُ النصوصَ التي حررها متعلمو اللغة وإن كانوا من الناطقين بها (الذين يدرسون قواعدها الإملائية والنحوية مثلاً)؟ أو أنها مقصورة على مواد من إنتاج الناطقين بغيرها؟ وللإجابة على هذا السؤال سنستعرض الأنواع الثلاثة التالية:

الأول: مدونات لمتعلمي اللغة باعتبارها لغة ثانية أو أجنبية، وهذا النوع هو المقصود مباشرة بمدونات المتعلمين. ومثال هذا النوع مدونة المتعلمين التمهيدية للغة العربية (أبو حكيمة وآخرون، ٢٠٠٨م)، وكذلك مدونة متعلمي العربية المكتوبة (فروانة وتميمي، ٢٠١٢م).

الثاني: مدونات جمعت بين النوع الأول إضافة إلى مدونات متعلمي اللغة الناطقين بها باعتبارها لغتهم الأم؛ وذلك بغرض المقارنة بينهما. وهذا النوع يدخل أيضاً ضمن مدونات المتعلمين؛ وقد وجد الباحث عند دراسته لأكثر من مئة وخمسين مدونة من مدونات المتعلمين (الفيفي، لم ينشر بعد) أن عشرين في المئة منها تندرج تحت هذا النوع. ومثاله المدونة اللغوية لمتعلمي اللغة العربية (الفيفي وآخرون، ٢٠١٤م).

النوع الثالث: مدونات أُفرِدتْ لمتعلمي اللغة الناطقين بها باعتبارها لغتهم الأم، وذلك مثل مدونات المتعلمين العرب الذين يدرسون القواعد الإملائية والصرفية والنحوية للغة العربية، سواءٌ أكان ذلك في مراحل التعليم العام أم في التعليم الجامعي، بغرض دراسة السمات اللغوية لهذه الفئة مثلاً، وذلك مثل مدونة لوفين للمقالات الإنجليزية للناطقين بها (عبدالله ونور Abdullah and Noor، ٢٠١٣)؛ وهذا النوع من المدونات - استناداً على تعريف قرانجر (٢٠٠٢م) - لا يدخل ضمن مدونات المتعلمين التي تركز على اكتساب اللغة الثانية وتعليم اللغة الأجنبية (ثودي Thoday، ٢٠٠٧م)؛ لكن مع ملاحظة ازدياد عدد هذا النوع من المدونات، واشتراكها مع مدونات المتعلمين في كثير من الخصائص مثل البناء والأغراض، بل واشتراكهما أحياناً في مدونة واحدة كما هو الحال في النوع الثاني الآنف الذكر، فإنها أقرب ما تكون إلى مدونات المتعلمين منها إلى أي نوع آخر من المدونات.

## تصنيف مدونات المتعلمين

تشير قرانجر (٢٠٠٢م) إلى أن المدونات تصنف في الغالب وفق ثنائيات متقابلة، وسنأتي على أربع منها هي الأقرب إلى مدونات المتعلمين (الشكل ١)، مع ملاحظة أن القائمة اليمنى من هذه الثنائيات هي الغالب على مدونات المتعلمين الحالية.

| | | |
|---|---|---|
| أحادية اللغة<br>Monolingual | ⟷ | ثنائية اللغة<br>Bilingual |
| عامة<br>General | ⟷ | متخصصة<br>Technical |
| تزامنيّة (عرْضية)<br>Synchronic (Cross) | ⟷ | تعاقبية (طولية)<br>Diachronic (Longitudinal) |
| مكتوبة<br>Written | ⟷ | منطوقة<br>Spoken |

الشكل (١) أربع ثنائيات لتصنيف مدونات المتعلمين

إن أغلب مدونات المتعلمين تندرج تحت المدونات أحادية اللغة، مع وجود عدد قليل منها ثنائية اللغة (مترجمة) مثل مدونة سبنس Spence (١٩٩٨م). أما العموم والتخصيص هنا فيقصد به نوع المواد المضمنة في المدونة، لا نوع المشاركين فيها، والفرق بينهما أن مدونات المتعلمين تعد من المدونات المتخصصة مقارنة بالمدونات العامة؛ حيث تقتصر المشاركة فيها على متعلمي اللغة فقط كما ذكرنا سابقاً في تعريفها، ثم تتفرع بعد ذلك بحسب موادها إلى مدونة ذات لغة عامة، ومدونة ذات لغة خاصة، فالعامة تشمل نصوصاً

في عدة مجالات مختلفة دون تخصيص، بينما الخاصة تركز على النصوص المكتوبة لأغراض محددة، كالسياسة، أو التجارة، أو الدين، على أن يكون كتّابُ هذه النصوص من متعلمي اللغة، ومن هذا النوع الأخير تصنف مدونة المتعلمين «إنديانا» للأعمال (Indiana Business learner corpus)، التي جمعها كونر وآخرون Connor et al. (٢٠٠٢م)، وكانت نصوصها محصورة في الأغراض التجارية.

المدونة التزامنية هي تلك التي يتزامن جمع كل مادة أو جزء منها مع بقية المواد أو الأجزاء الأخرى، أي أنها تجمع تقريباً دفعة واحدة دون فواصل زمنية كبيرة، وقد تسمى كذلك مدونة عرَضية أو مقطعيّة. أما المدونة التعاقبية فتُجمع أجزاؤها على فترات زمنية متعاقبة، وتسمى أيضاً طولية(٢)؛ لأن موادها تُجمع بناء على خط زمني معين. أغلب مدونات المتعلمين من النوع الأول وذلك لصعوبة متابعة عينة محددة من متعلمي اللغة لعدة أشهر وربما لبضع سنوات، إلا أن هذا النوع الأخير مفيد لمراقبة وتحليل التطور الواقع بين كل فترة وأخرى. من المدونات التعاقبية المدونة السويدية لتطوير تعليم اللغة الثانية (The ASU corpus) التي جمعها هامربرغ Hammarberg(٢٠١٠م) لتكون مادة لدراسة التطور اللغوي لكل طالب على حدة، أو بالمقارنة مع بقية الطلاب؛ ومنها كذلك مدونة اللغة المرحلية للمتعلمين الصغار (The Corpus of Young Learner Interlanguage) التي جمعها هوسن Housen (٢٠٠٢م) بإجراء مقابلات مع ستة من متعلمي اللغة الإنجليزية، ثلاثة منهم ناطقون بالفرنسية وثلاثة ناطقون بالهولندية، وذلك لفترة زادت على ثلاث سنوات.

يرى كينيدي Kennedy (١٩٩٨م) أهمية اللغة المنطوقة لغلبة التواصل الشفهي بين الناس على التواصل الكتابي، ولكن عند مقارنة مدونات المتعلمين من حيث المحتوى المكتوب والمنطوق يظهر التفوق الكبير للمدونات ذات المواد المكتوبة، ومع أن ليتش Leech (١٩٩٧م) يقترح أن يكون الجزء المنطوق من

المدونة مساوياً على الأقل للجزء المكتوب، إلا أن المشقة الكبيرة في جمع المواد الصوتية وإدخالها للحاسب يفسر سبب التفاوت الكبير بين هذين النوعين.

## ماذا يميز مدونات المتعلمين عن غيرها؟

أشرنا في تعريف مدونات المتعلمين إلى اثنتين من السمات التي تميزها عن أنواع المدونات الأخرى، وسنفرد هنا مبحثاً خاصاً لبيانهما بشيء من التفصيل، مع إضافة سمتين أخريين أقل منهما درجة في التمييز بين مدونات المتعلمين وغيرها من المدونات.

مصدر البيانات: تُجمَع نصوص المدونات العامة من مصادر متنوعة، في حين تقتصر مدونات المتعلمين على المواد التي ينتجها متعلمو اللغة دون غيرهم، من نصوص مكتوبة أو أحاديث منطوقة؛ وقد تتنوع هذه المواد ما بين بحث، أو رسالة، أو قصة، أو مقال، أو كلمة تلقى مشافهة، أو إجابات في مقابلة مع معلم أو زميل دراسة، أو أداءً لأي مهمة لغوية تطلب من الطالب. وتجدر الإشارة هنا إلى أهمية أن تكون المواد المجموعة من الطلاب موحدة، أي تمثل أحد الأنواع السابقة، إما لكامل المدونة أو لكل مدونة فرعية منها، وإلا كانت النتيجة خليطاً من البيانات المختلفة التي لا يمكن الاستفادة منها أو تعميم النتائج عليها (قرانجر، ١٩٩٣م).

**الغرض:** يمكن ملاحظة أن اكتساب اللغة وتعليمها يمثلان الغرض الرئيس الذي تدور حوله مدونات المتعلمين (ثودي، ٢٠٠٧م)، ومع أن هناك أغراضاً أخرى قد تُستقى من هذا النوع مثل برامج تصحيح الأخطاء آلياً في مجال اللغويات الحاسوبية (انظر مثلاً زغواني وآخرين Zaghouani et al.، ٢٠١٤م)، إلا أن غالبية المدونات الحالية والأبحاث المبنية عليها تعكس الاهتمام الواضح بمجال تعليم اللغة واكتسابها.

**الوسم:** أشكال الوسم التي يمكن استخدامها في المدونات النصية كثيرة ومتنوعة، إلا أن مدونات المتعلمين تتميز بنوع خاص من الوسم، وهو وسم الأخطاء اللغوية، كونه أحد الأركان الأساسية التي تعتمد عليها بعض الأبحاث القائمة على مدونات المتعلمين. وسنورد بشيء من التفصيل بعض هذه الأبحاث عند حديثنا عن طرق الاستفادة من مدونات المتعلمين، كما سنفرد الحديث عن وسم الأخطاء في جزء آخر من هذا الفصل.

**الحجم:** مع أهمية حجم المدونة باعتباره أحد العناصر الرئيسة في تمثيل اللغة، إلا أن مدونات المتعلمين ذات حجم صغير في الغالب، وأكثرها يقل حجمه عن المليون كلمة، أما أكبرها فيحوي ما يقارب ٢٥ مليون كلمة مثل مدونة كامبريدج للمتعلمين (٢٠١٢م)، ومدونة جامعة هونج كونج للعلوم والتكنولوجيا (ميلتون ونانديني Milton and Nandini، ١٩٩٤م، وبرافيك Pravec، ٢٠٠٢م)، وكلاهما لمتعلمي اللغة الإنجليزية. وترى قرانجر (٢٠٠٣م ب) أن مدونة مكونة من مئتي ألف كلمة لاستخدامها في مجال اكتساب اللغة الثانية تعد كبيرة إذا ما قارنَّاها بالعينات البحثية الصغيرة التي يعتمد عليها الباحثون عادة في هذا المجال؛ لكنها في المقابل تعد صغيرة بالنظر إلى ما هو مستخدم في المجالات اللغوية الأخرى، حيث يعتمد الباحثون على مدونات تتكون من مئات الملايين وربما مليارات الكلمات، خصوصاً أن مثل هذه الأرقام الكبيرة أصبحت معياراً لمقارنة أحجام المدونات بعد أن كانت استثناءً في بداية نشأتها. كما أن هذا الرقم الذي ذكرته قرانجر - ٢٠٠,٠٠٠ كلمة - هو ما اعتمدته في المدونة الدولية لمتعلمي اللغة الإنجليزية كحد أدنى لكل واحدة من المدونات الفرعية التي تمثل مجموعة من الطلاب المتحدثين بلغة أم واحدة (انظر قرانجر، ٢٠٠٣م ب).

## مدونات المتعلمين العربية

إن المطّلع على المشاريع القائمة حالياً في مجال مدونات المتعلمين لا يمكن أن تخطئ عينه ما تحظى به اللغات الأجنبية، واللغة الإنجليزية على وجه

الخصوص (الفيفي، لم ينشر بعد)، من نصيب في هذه المدونات؛ فبعد دراسة لأكثر من مئة وخمسين مدونة في هذا المجال، وجد الباحث أن اللغة الإنجليزية تدخل في ٥٢٪ منها، بينما كان نصيب اللغة العربية ٣٪ فقط (خمس مدونات فقط). ولكن عند إمعان النظر في هذه المدونات الخمس نجد أن وتيرة الاهتمام بهذا النوع متسارعة وإن بدأت متأخرة، وهذا ينمُّ عن الحاجة الكبيرة لمدونات المتعلمين لما أثبتته الأبحاث القائمة عليها من فوائد علمية وعملية.

وقبل أن نورد معلومات حول مدونات المتعلمين العربية، فإنه من المستحسن أن نشير - خصوصاً لأولئك المهتمين بجانب التصميم وجمع البيانات لهذا النوع - إلى واحدة من أفضل الأمثلة على مدونات المتعلمين، وهي المدونة الدولية لمتعلمي اللغة الإنجليزية (International Corpus of Learner English)، التي أشرفت على بنائها رائدة البحث في هذا المجال سيلفيان فرانجر من جامعة لوفين ببلجيكا (انظر فرانجر، ٢٠٠٣م ب)، وتعتبر هذه المدونة في نظر كثير من الباحثين مدونة معيارية لما تتمتع به من تصميم دقيق، إضافة إلى المحتوى النصي الذي شارك في جمعه عدد كبير من الباحثين حول العالم، وقد تطورت النسخة الثانية من هذه المدونة لتضم ثلاثة ملايين وسبعمئة ألف كلمة لنصوص حررها متعلمو اللغة الإنجليزية في عدد من الدول، يمثلون ست عشرة لغة أم، ويجري العمل حالياً على النسخة الثالثة منها. أما بالنسبة للغة العربية، فهناك خمس مدونات لمتعلميها تتفاوت من حيث الحجم والخصائص، وفيما يلي نبذة موجزة حول كل منها.

### المدونة الأولى: مدونة غازي أبو حكيمة وآخرين

١. اسمها: مدونة المتعلمين التمهيدية للغة العربية - The Pilot Arabic Learner Corpus

٢. حجمها: ٩٠٠٠ كلمة

٣. المواد المشمولة فيها: نصوص مكتوبة

٤. الطلاب المشاركون فيها: متعلمو اللغة العربية باعتبارها لغة أجنبية في الولايات المتحدة الأمريكية

٥. الوسم: غير موسومة

٦. نشر المدونة للباحثين: المدونة لا زالت قيد التطوير وهي غير متاحة للاستخدام العام

١. المرجع: أبو حكيمة وآخرون Abuhakema et al. (٢٠٠٨م)

**المدونة الثانية: مدونة سميرة فروانة ومحمد تميمي**

١. اسمها: مدونة متعلمي العربية المكتوبة – Written Arabic The L2 Corpus

٢. حجمها: غير موضح، لكنه بحسب إحصاء تقديري للنصوص تضم حوالي ٣٥٠٠٠ كلمة

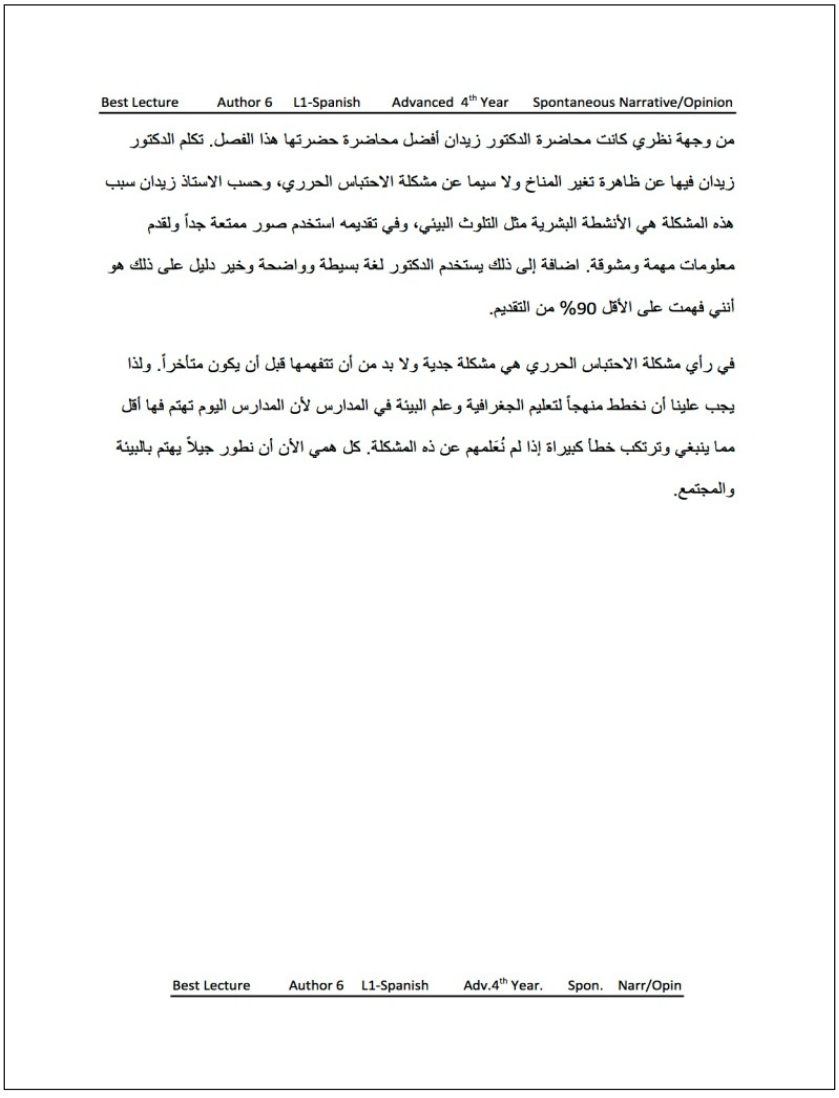
٣. المواد المشمولة فيها: نصوص مكتوبة

٤. الطلاب المشاركون فيها: متعلمو اللغة العربية باعتبارها لغة أجنبية في الولايات المتحدة الأمريكية

٥. الوسم: غير موسومة

٦. نشر المدونة للباحثين: المدونة متاحة على شكل ملفات PDF فقط، ويمكن تنزيلها من موقع المدونة http://l2arabiccorpus.cercll.arizona.edu

٧. المرجع: فروانة وتميمي (٢٠١٢م) (انظر الشكلين ٢، و٣)



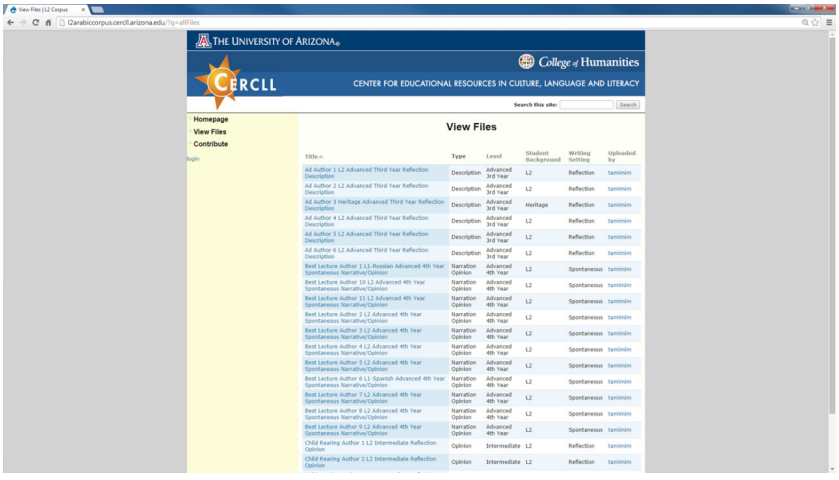
Best Lecture Author 6 L1-Spanish Advanced 4th Year Spontaneous Narrative/Opinion

من وجهة نظري كانت محاضرة الدكتور زيدان أفضل محاضرة حضرتها هذا الفصل. تكلم الدكتور زيدان فيها عن ظاهرة تغير المناخ ولا سيما عن مشكلة الاحتباس الحراري، وحسب الاستاذ زيدان سبب هذه المشكلة هي الأنشطة البشرية مثل التلوث البيني، وفي تقديمه استخدم صور ممتعة جداً ولقدم معلومات مهمة ومشوقة. اضافة إلى ذلك يستخدم الدكتور لغة بسيطة وواضحة وخير دليل على ذلك هو أنني فهمت على الأقل 90% من التقديم.

في رأي مشكلة الاحتباس الحراري هي مشكلة جدية ولا بد من أن تتفهمها قبل أن يكون متأخراً. ولذا يجب علينا أن نخطط منهجاً لتعليم الجغرافية وعلم البيئة في المدارس لأن المدارس اليوم تهتم فها أقل مما ينبغي وترتكب خطأ كبيراة إذا لم نُعلمهم عن ذه المشكلة. كل همي الأن أن نطور جيلاً يهتم بالبيئة والمجتمع.

Best Lecture Author 6 L1-Spanish Adv.4th Year. Spon. Narr/Opin

الشكل (٢) نموذج لملف نصي مع البيانات الوصفية من مدونة المتعلمين العربية للنصوص المكتوبة




الشكل (٣) موقع مدونة المتعلمين العربية للنصوص المكتوبة

## المدونة الثالثة: مدونة حسلينا حسان ومحمد فهام محمد غالب

١. اسمها: المدونة الماليزية لمتعلمي العربية – Malaysian Corpus of Arabic Learners

٢. حجمها: حوالي ٨٧٥٠٠ كلمة تقريباً

٣. المواد المشمولة فيها: مقالات وصفية ومقارنة، كتبت باستخدام الحاسب، يصل عددها إلى ٢٥٠ نصاً، متوسط طول النص الواحد ٣٥٠ كلمة

٤. الطلاب المشاركون فيها: متعلمو اللغة العربية الماليزيون في المستوى المتقدم

٥. الوسم: غير موسومة

٦. نشر المدونة للباحثين: هذه المدونة غير متاحة على شبكة الإنترنت، لكن المؤلفة تنوي إدراجها مستقبلاً ضمن مدونة أكاديمية أوسع يمكن

الوصول إليها عن طريق العنوان التالي: http://efolio.iium.edu.my/arabicconcordancer

٧. المرجع: حسان وغالب (٢٠١٣)

**المدونة الرابعة: مدونة محمد الكنهل وآخرين**

١. اسمها: مدونة تصحيح الأخطاء الإملائية

٢. حجمها: ٦٥٠٠٠ كلمة

٣. المواد المشمولة فيها: نصوص كتبت يدوياً من قبل الطلاب، ثم أدخلت بعد ذلك إلى الحاسب الآلي بواسطة نساخ مستقلين

٤. الطلاب المشاركون فيها: مجموعتان من الطلاب الجامعيين، كل مجموعة من جامعة مختلفة، دون تحديد للتخصصات أو المستويات

٥. الوسم: هناك نسختان لهذه المدونة، الأولى ملفات نصية (txt) وهي غير موسومة، والثانية قاعدة بيانات (مايكروسوفت أكسس – Microsoft Access) تم تصحيح الأخطاء فيها يدوياً.

٦. نشر المدونة للباحثين: يمكن تنزيل نصوص المدونة على شكل ملفات نصية من الرابط التالي: cri.kacst.edu.sa/Resources/TST_DB.rar

٧. المرجع: الكنهل وآخرون Alkanhal et al. (٢٠١٢م) (انظر الشكل ٤)

العنوان:أثر التدخين في الاماكن العامه وعيره نعم: لان التدخين ضار بي صحه المدخن، ولاذلك لضرر التدخين بي الناس وله أمراص كثيرة: سرطان الرئه وعيره من الأمراض. ولذلك منعة بعض الدول الكبير هذه الشائعه التي تثر على المجتمعات الكبير وتنثر على قوة بناء المجتمع . هل تعتقد أن مجتمع انتشار فيها التدخين تكون أمر سليمه وشبابه صالحين؟. أن من يدخن أمام صديق له قد يكون الضرر الصديق أكثر من المدخن وأثبت في بعض الت جارب على المدخن ومن يجلس أمام المدخن اكتشف أن أثر التدخين أكثر على الصديق الي يجلس دائما مع المدخن. لذلك منعه أكبر الدول التدخين في الأماكن العامه والمكاتب لأن: أكثر الأماكن التي يكون فيها التدخين السالبي، أو (السلبي) كما في الفقر التي قبلها وهو تحارب الكلم بين المدخن والشخص الغير مدخن. ومن سبب منع التدخين في الاماكن العامه والمكاتب، تشبه بعض الناس في الأقتداء في المدخنين وفعل مايفعلون( شرب الدخان) واذكر قصه بي اختصار يو جد رجل جا ئر مرض السرطان الرئه بسبب مصاحبه لأصدقاء يشربوا الدخان ف أصبح كأن يد خن معهم وهذا أخر منيجلس مع المدخنين لذلك منع الدول السليمه. وفي الختام اذكر ان كل مجتمع سليم يبي على الشباب فإذا فسد الشباب فسد المجتمع ولذلك ننصر إلى التدخين انه (الموت البطيء ) لان الشباب أكبر فئه يشربوا التدخين فإذا أصبح الشاب مريص فقد تصبح الأمه في عجز وكل مايضن بي الشباب والناس يمنع في كل مكان. s.w

الشكل (٤) نموذج لملف نصي من مدونة الكنهل

## المدونة الخامسة: مدونة عبدالله الفيفي وإريك أتويل

١. اسمها: المدونة اللغوية لمتعلمي اللغة العربية – Arabic Learner Corpus

٢. حجمها: ٢٨٢,٧٣٢ كلمة

٣. المواد المشمولة فيها: نصوص مكتوبة يدوياً، وأخرى باستخدام الحاسب، وكذلك مقابلات صوتية مسجلة، تدور جميعها حول موضوعين، الأول سردي والثاني يناقش الاهتمامات الدراسية

٤. الطلاب المشاركون فيها: متعلمو اللغة العربية الناطقون بغيرها ويمثلون ٤٧٪ من محتوى المدونة، إضافة إلى طلاب ناطقين بالعربية يمثلون ٥٣٪، وهم من كلا الجنسين (الذكور ٦٧٪، والإناث ٣٥٪)، يدرسون في عدد من المدارس الثانوية ومعاهد اللغة وجامعات المملكة العربية السعودية، وتتفاوت أعمارهم بين ١٦ و ٤٢ سنة.

٥. الوسم: المدونة المتاحة للتنزيل حالياً غير موسومة، وستتاح مستقبلاً نسخة تشمل وسم الأخطاء اللغوية. هناك نسخ إضافية من المدونة على بعض مواقع البحث، وهي مذكورة بالتفصيل على موقع المدونة المذكور أدناه.

٦. نشر المدونة للباحثين: يمكن تنزيل نصوص المدونة في ملفات نصية (txt) ، أو ملفات لغة الترميز الممتدة (xml)، إضافة إلى الأصول المكتوبة يدوياً في ملفات (pdf)، وكذلك التسجيلات الصوتية في ملفات (mp3) من خلال موقع المدونة على الرابط التالي: http://www.arabiclearnercorpus.com

٧. المرجع: الفيفي وآخرون (٢٠١٤م) (انظر الشكل ٥، و٦)

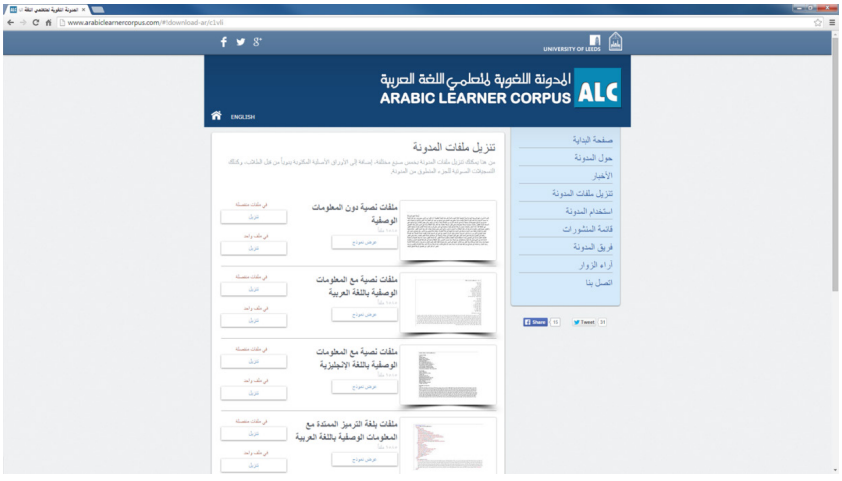
رمز النص :S319_T1_M_Pre_NNAS_W_C

معلومات الطالب
العمر :٢٠
الجنس :ذكر
الجنسية :نيجيري
اللغة الأم :اليوربا
أصالة اللغة :ناطق بغير العربية
عدد اللغات التي يتحدثها :٣
عدد سنوات تعلم اللغة العربية :١٠
عدد السنوات في بلدان عربية :١٠
المرحلة العامة :ما قبل الجامعة
المرحلة الدراسية :برنامج الدبلوم
السنة/المستوى :المستوى الثاني
المؤسسة التعليمية :معهد اللغة بجامعة الإمام

معلومات النص
النوع :سردي
مكان التأليف :في الصف
سنة التأليف :١٤٣٥
دولة التأليف :السعودية
مدينة التأليف :الرياض
محدد بوقت :نعم
استخدام مراجع :لا
استخدام كتاب قواعد :لا
استخدام معجم أحادي :لا
استخدام معجم ثنائي :لا
استخدام مراجع أخرى :لا
الشكل :مكتوب
الوسيط :مكتوب يدويا
الطول :١٧٨ كلمة

عنوان النص :رحلتي إلى لاغوس.
النص :
عرفت أنني سأسافر إلى لاغوس .عاصمة نيجيريا سابقا، في رمضان لأمّ الناس في صلاة التراويح والتهجّد، ولكن الأعمال الكثيرة التي كانت على عاتقي والتي تلزمني التخلص منها تكاد تُنسيني وتمحو السفر عن ذاكرتي .كنت أدرّس بعض الطلاب في إحدى المدارس في مدينة إيوو ولاية أوشن ومما طلب مني إخراج الأسئلة لهم في الإختبار النهائي وفعلت كما طلبوا، وفجأة جاءت الأوراق للتصحيح .بدأت بتصحيح الأوراق لثلاث مواد درّسها، وقبيل المغرب تذكرت أنني سأسافر إلى لاغوس لأن ذلك اليوم هو التاسع والعشرون للشعبان .ماذا أفعل والأدوات مازالت كثيرة وكأنني لم أصحح منها شيئا أو إضافة إلى ذلك ما جهزت للسفر ولا استحممْتُ من كثرة الأوراق التي أصحّح جاءتني فكرة أن أعطي الأوراق لواحد من الإخوان ثم جهّزت للسفر وتوجهت إلى لاغوس .وصلت إلى المكان الذي أقصده الساعة العاشرة ولكن الحمد لله أن الهلال ما طلع في أي مكان ومعنى ذلك أن صلاة التراويح تبدأ في اليوم التالي .بدأنا التراويح في ليلة الثلاثين بعدد غفير من الناس والبهجة والسرور ما اختفيا في أوجههم لأنهم –فضل الله تعالى –شهدوا رمضان عام الفين وثلاثة عشر ميلاديا، أشهر الرحمة والمغفرة وعتق من النار

الشكل (٥) نموذج لملف نصي مع البيانات الوصفية من المدونة اللغوية لمتعلمي اللغة العربية




الشكل (٦) موقع المدونة اللغوية لمتعلمي اللغة العربية

## مجالات الإفادة من مدونات المتعلمين

تعد مدونات المتعلمين المصدر الرئيس للبيانات في كثير من الأبحاث اللغوية، حيث قدمت وسائل وأساليب جديدة للبحث أسهمت في الحصول على نتائج أكثر دقة وواقعية، إضافة إلى النتائج العلمية والعملية المستفادة من دراسة وتحليل المفردات والتراكيب اللغوية في مدونات المتعلمين (شابل، ١٤٢٨هـ)؛ وقد قدّم المبحث الأول من هذا الكتاب للقارئ بعض الأمثلة على استخدام المدونات في البحث اللغوي؛ لذا لن نبسط الكلام هنا عن الوسائل والأساليب بقدر ما سنركز حديثنا حول المجالات التي لعبت مدونات المتعلمين فيها دوراً ملموساً مما أدى لتطورها بشكل واضح عما كانت عليه في الماضي.

### أولاً: الأبحاث في مجال اكتساب اللغة (Language Acquisition Research)

بناء على الدراستين اللتين أجرتهما قرانجر (١٩٩٨م و ٢٠٠٢م) حول استخدام مدونات المتعلمين في أبحاث اكتساب اللغة فقد وجدت أن هذه الأبحاث

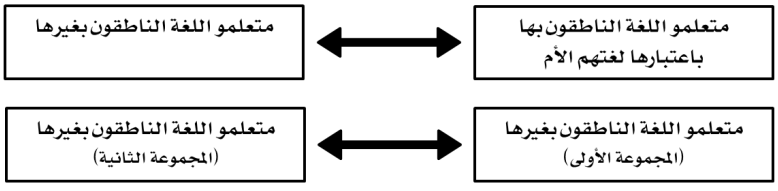
باتت تعتمد بشكل متزايد على هذا النوع من المدونات، وأن هذا التوجه يرجع إلى الأفق الواسع في الإمكانات والأفكار والرؤى التي توفرها مدونات المتعلمين لهذ الحقل الحيوي، الذي يهدف بشكل رئيس إلى فهم آلية اكتساب اللغة الثانية أو الأجنبية؛ إضافة إلى ذلك فإن هذه المدونات - رغم بعض العوامل التي قد تؤثر على نتائجها كما يرى ييب Yip (١٩٩٥م) - تعطي صورة أشمل لبحث جوانب لم يكن بمقدور الدراسات السابقة النظر فيها، ومن ذلك إمكانية دراسة الأشكال اللغوية التي قد يَشُقّ على الطلاب تعلُّمُها، ويصعب عليهم استخدامها، فيختارون تفاديها والالتفاف عليها من خلال أساليب لغوية أخرى، قد تسبب إطناباً واضحاً في غير موضعه.

إضافة إلى ذلك فإن مدونات المتعلمين تمثل أحد أنواع البيانات البحثية التي يمكن من خلالها دراسة اكتساب اللغة، حيث قسّم إليس Ellis (١٩٩٤م) هذه البيانات إلى ثلاثة أنواع، (١) الاستخدام اللغوي: وهي تلك البيانات والنصوص التي تعكس استخدام الطلاب للغة الثانية، سواءٌ أكانت من ناحية الفهم والاستيعاب أم من ناحية الإنتاج؛ (٢) الأحكام اللغوية: وهي الآراء التي يصدرها الطلاب حول اللغة الثانية، كأن يطلب منهم الحكم على جملة لغوية من ناحية التركيب النحوي؛ (٣) التقارير الذاتية: ويقصد بها ما يقدمه الطلاب من معلومات حول إستراتيجياتهم اللغوية عن طريق الاستبانات أو مهام التفكير المسموع ونحوهما. وبالرجوع إلى النوع الأول يقسم إليس «الاستخدام اللغوي» إلى صنفين، يطلق على الأول منهما **بيانات طبيعيّة** في حال لم يتم التحكم في أداء الطلاب عند الحصول عليها، وهذا ما تمثله إلى حد كبير مدونات المتعلمين، ويمثل الصنفَ الثاني **البياناتُ المستنطقة** إن كانت نتيجةً لمهام مقيدة أو أنشطة لغوية ذات رقابة. وفي حين أن الصنف الثاني يقدم معلومات أكثر تركيزاً للجانب المطلوب دراسته فإن مدونات المتعلمين تفتح المجال لدراسة كثير من الجوانب في نفس الوقت (نسلهاف، ٢٠٠٤م)، مع إمكانية دراسة العلاقة

بينها وأثر بعضها على بعض. ويرى إليس (١٩٩٤م) أن البحث الجيد هو ذلك الذي يعتمد على أكثر من نوع من هذه البيانات.

**ثانياً: التحليل التقابلي للغة المرحلية (Contrastive Interlanguage Analysis)**

التحليل التقابلي (Contrastive Analysis) واحد من المناهج المستخدمة في الأبحاث اللغوية، وتكمن أهميته فيما يكشفه من أوجه التشابه والاختلاف بين اللغات المقارنة؛ وفي مجال مدونات المتعلمين يركز الباحثون على التحليل التقابلي للغة المرحلية من خلال نوعين من المقارنات: النوع الأول بين الناطقين باللغة بوصفها لغتهم الأم ومتعلميها من الناطقين بغيرها، والنوع الثاني بين مجموعتين مختلفتين من متعلمي اللغة الناطقين بغيرها، بحيث تتميزان عن بعضهما في أحد العناصر التي قد تؤثر في عملية التعلم إيجاباً أو سلباً، مثل اللغة الأم للمتعلم، أو العمر، أو المرحلة الدراسية، أو مدة تعلمه للغة، أو عدد اللغات التي يتقنها، أو غير ذلك من العوامل (قرانجر، ١٩٩٨م)، انظر الشكل ٧.




الشّكل (٧) الأنواع الرئيسة للتحليل التقابلي باستخدام مدونات المتعلمين

يهدف النوع الأول إلى الكشف عن الفروق اللغوية بين الناطقين باللغة بوصفها لغتهم الأم مقارنة بمتعلميها الناطقين بغيرها، ويكشف لنا هذا النوع الخصائص المميزة للغة المرحلية للمتعلمين مع تزويد الباحثين بمعلومات دقيقة عنها، فعلى سبيل المثال يمكن معرفة إلى أي مدى يشيع استخدام كلمة أو تركيب

محدد بين كل فريق، ويمكن تبعاً لذلك استخلاص ما يسمى بالاستخدام الزائد، أو الناقص، أو الخاطئ لمفردات وتراكيب اللغة بين متعلميها الناطقين بغيرها؛ وهذا يساعد المختصين من باحثين ومعلمين ومصممي المناهج التعليمية في توجيه الطلاب إلى الاستخدام الصحيح والأمثل للغة، وذلك بتقريبهم ما أمكن إلى أساليب الاستخدام الواقعية للناطقين بها.

أما النوع الثاني من التحليل التقابلي للغة المرحلية فيكون بين مجموعتين مختلفتين من متعلمي اللغة الناطقين بغيرها، حيث تكون المقارنة على أساس أحد العناصر المؤثرة في عملية التعلم، فالمقارنة بين مجموعتين من الطلاب مع اختلاف اللغة الأم لكل منهما (كالمقارنة مثلاً بين مجموعة طلاب إنجليز ومجموعة طلاب صينيين يتعلمون العربية)، يساعدنا في معرفة أثر بعض اللغات على تعلم الطلاب للغة الهدف إيجاباً وسلباً. على سبيل المثال قد يكون تشابه أقسام الضمائر في بعض اللغات عاملاً مساعداً للطالب على إتقانها، بينما يجد غيره صعوبة في فهم الحدود الدقيقة لاستخدامها. ومثل هذا النوع من المقارنات بين مجموعات مختلفة من متعلمي اللغة الناطقين بغيرها، يمكننا من الإجابة على عدة تساؤلات نحو: هل يؤثر عدد اللغات التي يتحدث بها الطالب على مدى قدرته على تعلم اللغة؟ ما مدى أثر العمر على فهم اللغة، واكتساب المفردات، ودقة القواعد، ونحو ذلك؟ هل يؤثر مستوى الطالب لغوياً على شيوع استخدامه لبعض الكلمات أو العبارات؟

ولا بد هنا من التنبيه على دور التصميم الجيد للمدونة؛ ليتمكن الباحثون من إجراء التحليل التقابلي على أكبر عدد ممكن من العوامل، فجودة تصميم المدونة تؤثر بشكل كبير على جودة النتائج المستخرجة منها (سينكلير، ١٩٩١م، وقرانجر، ١٩٩٨م، والسليطي Al-Sulaiti، ٢٠٠٤م).

**ثالثاً: تحليل الأخطاء بمساعدة الحاسب (Computer-aided Error Analysis)**

بدأ العمل في مجال تحليل الأخطاء اللغوية في نهاية ستينات القرن الماضي - أي قبل بداية وجود مدونات المتعلمين التي بدأت في أواخر ثمانينات القرن نفسه - وقد جُمع كثير من العينات اللغوية للطلاب في ذلك الوقت، لكنها كانت تُستخدم كمستودع للأخطاء فقط، لا على أنها مدونات للمتعلمين، وبالتالي يمكن التفريق بينها وبين مدونات المتعلمين بأن هذه العينات لم تكن تُجمع وفق معايير تصميم واضحة، ولم يكن للحاسب الآلي دور في تحليل محتواها، إضافة إلى أنه كان يتم التخلص منها بمجرد استخراج الأخطاء ودراستها (نسلهاف، ٢٠٠٤م).

ومع استخدام أدوات تحليل المدونات على الحاسب الآلي في الوقت الراهن، أصبح لدى الباحثين إمكانات كبيرة للوصول إلى تحليل أعمق للأخطاء في مدونات المتعلمين، مع الحصول على نتائج أدق وبشكل أسرع. ويلعب وسم الأخطاء الدور الأكبر في هذا الجانب، فكلما كان تصنيف الأخطاء ووسمها يستند على معايير واضحة ودقيقة كانت نتائج البحث والتحليل أكثر جودة وبالتالي أكثر واقعية وفائدة.

ومن الأمثلة على ذلك تلك الدراسة التي أجراها لي Lee (٢٠٠٧م) في جامعة أسكس Essex، حيث قام ببناء مدونة تضم عشرين ألف كلمة، عبارة عن نصوص كتبها طلاب كوريون يدرسون اللغة الإنجليزية، أعمارهم بين ١٦ و١٧ عاماً، وبعد وسم جميع الأخطاء باستخدام أداة خاصة صممها هتشنسون Hutchinson (١٩٩٦م)، وهي أداة تعتمد على تصنيف الأخطاء اللغوية لدنيو وآخرين Dagneaux et al. (١٩٩٦م)، أمكن استنتاج عدد الأخطاء ونسبتها تحت كل فئة من فئات الخطأ العامة والفرعية بسرعة ودقة عاليتين. فعلى سبيل المثال وجد الباحث أن أخطاء القواعد اللغوية كانت الأكثر بنسبة ٣٧,٥٪، تلتها أخطاء علامات الترقيم بنسبة ١٤,٥٪. وفيما يخص التصنيفات الفرعية، وجد أن علامات التعريف والتنكير -

تحت تصنيف القواعد – كانت الأكثر عرضة للخطأ بنسبة ٨, ١٤٪. وهذا النوع من الأبحاث في استنتاج الأخطاء وتحليلها، يستخدم كأساس لكثير من الأبحاث المتقدمة، والأغراض التربوية، فهو يساعد الباحثين على تتبع مستويات الطلاب ومدى تطور قدراتهم اللغوية، إضافة إلى تقصي أوجه القصور في الإستراتيجيات الحاليّة لتعليم اللغة، ومن ثم العمل على تلافيها مستقبلاً، كما يساعد على تصميم كتب، ومواد تعليمية، ومعاجم طلابية أقدر على الوفاء باحتياجات المتعلمين وأهدافهم، وسيرد حديث أكثر عن بعض هذه الجوانب.

**رابعاً: دراسة التطور اللغوي لدى الطلاب**

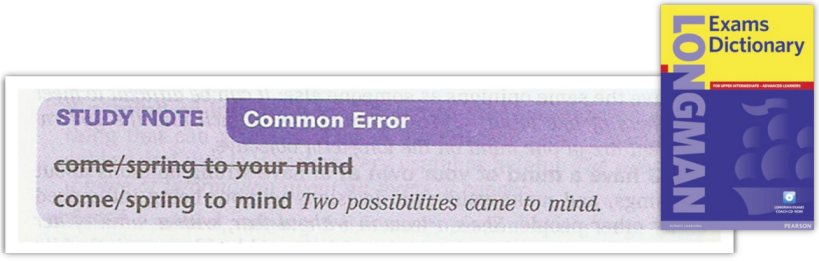
من الأبحاث القائمة على مدونات المتعلمين ما يختص بدراسة مستويات الطلاب وقدراتهم اللغوية وقياس التباين بين كل مستوى وآخر، مع دراسة العوامل المؤثرة في ذلك، ومن الأمثلة على ذلك تلك الدراسة التي قام بها باتيري وكاينز Buttery and Caines (٢٠١٢م) لبحث العلاقة بين مستوى الكفاية اللغوية مقارنة بطول الجمل التي يستخدمها الطلاب من ناحية، وبالتنوع والعدد في استخدام العبارات الظرفية من ناحية أخرى، وقد اعتمدوا في ذلك على وسم وتحليل الأخطاء الذي سبق أن ذكرنا أهميته باعتباره أساساً لمثل هذه الدراسات.

قد يكون هذا المجال هو الأساس الذي بدأ منه ذلك النوع المعروف بمدونات المتعلمين التعاقبية (Diachronic) أو الطولية (Longitudinal)، والتي من أهم أغراضها قياس مدى تطور الطلاب لغوياً سواءٌ أكان ذلك بالاعتماد على نسبة الأخطاء في كل مرحلة، أم من خلال تتبع ما يحدث من تحسن في استخدام بعض المفردات، أو العبارات، أو الأدوات اللغوية، أو غيرها، بالتزامن مع تطور مستوى الكفاية اللغوية لدى الطلاب.

### خامساً: تأليف معاجم الطلاب

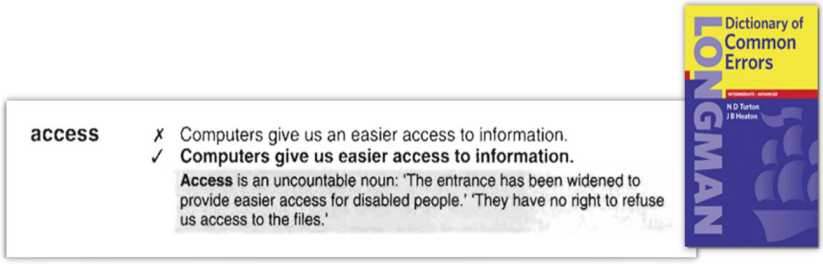
تعد صناعة المعاجم الموجهة لمتعلمي اللغة – الإنجليزية على وجه التحديد – من أوائل المواد التعليمية التي استفادت من الدراسات المعنية بتحليل مدونات المتعلمين، أما في الوقت الحالي فأغلب معاجم الطلاب تعتمد على هذا النوع من المدونات، خصوصاً المعاجم أحادية اللغة (Monolingual learners' dictionaries). ومن الصور الشائعة لاستخدام مدونات المتعلمين في بناء معاجم الطلاب، الاستفادة من مخرجات تحليل الأخطاء اللغوية على مستوى المفردات أو التراكيب، وتزويد قراء المعجم بتنبيهات حول هذه الأخطاء في المداخل المعجمية المرتبطة بها، مع بيان صور استخدامها الصحيح، ويعد معجم Longman Essential Activator (سمرز Summers، ١٩٩٧م) أول معجم يعتمد على مدونة متعلمين، حيث استُخدِمت «مدونة لونجمان» (شبكة مدونة لونجمان، ٢٠١٢م) في تأليف هذا المعجم (جيلارد وجاسبي Gillard and Gadsby، ١٩٩٨م، وقرانجر، ٢٠٠٣م ب، ونسلهاف، ٢٠٠٤م).

يعرض لنا الشكل ٨ مثالاً لخطأ شائع من «معجم لونجمان للاختبارات» Longman Exams Dictionary (بولن Bullon، ٢٠٠٦م: ٩٦٩)، وقد تم وضعه على شكل ملحوظة في نهاية المدخل المعجمي لكلمة «mind» للتحذير من الأساليب الخاطئة لاستخدام هذه الكلمة والتنبيه على البدائل الصحيحة لها.




الشّكل (٨) مثال لخطأ شائع من «معجم لونجمان للاختبارات»

وقد توسع البحث في تأليف معاجم الطلاب المبنية على مدونات المتعلمين حتى ظهرت معاجم مستقلة للتنبيه على أخطاء المتعلمين وتصويبها مثل «معجم لونجمان للأخطاء الشائعة»(٣) Longman Dictionary of Common Errors (تورتن وهيتن Turton and Heaton، ١٩٩٦: ٥)، انظر الشكل ٩.




الشّكل (٩) مثال لخطأ شائع من «معجم لونجمان للأخطاء الشائعة»

والمعاجم التي تعنى بالتنبيه على الأخطاء الشائعة قد وُجدت بكثرة في اللغة العربية، ومنها «معجم الأخطاء الشائعة» (العدناني، ١٩٨٣م)، و«المعجم الوجيز في الأخطاء الشائعة والإجازات اللغوية» (محمد، ١٤٢٦هـ)، و«معجم تقويم اللغة وتخليصها من الأخطاء الشائعة» (أَمُّون، التاريخ غير معروف)، وهي في الغالب موجهة للناطقين باللغة العربية، ولم تستفد من التطور الحاصل بين حقلي صناعة المعاجم ومدونات المتعلمين.

في مقابل المعاجم هناك بعض الدراسات التي اهتمت بدراسة شيوع الأخطاء لدى متعلمي اللغة، ومن ذلك دراسة العصيلي (١٤٠٥هـ) بعنوان «الأخطاء الشائعة في الكلام لدى طلاب اللغة العربية الناطقين بلغات أخرى»، وكذلك الشافعي وإبراهيم (١٤٠٨هـ) «الأخطاء الشائعة في الهجاء والإملاء بين تلاميذ المرحلة الابتدائية بمنطقة الرياض»، والحمد (١٤١٥هـ) «تحليل أخطاء التعبير الكتابي لدى المستوى المتقدم من دارسي العربية غير الناطقين بها في جامعة الملك سعود» وغيرها من الأبحاث ذات المنهجية الواضحة في

استخراج الأخطاء والحكم على شيوعها، لكنها استخدمت عينات بحثية خاصة عوضاً عن الاعتماد على مدونات المتعلمين، كما أن بناء معجم للطلاب لم يكن جزءاً من أي منها.

**سادساً: تدريس اللغة وتصميم المواد التعليمية**

من الطرائق المستحدثة في تعليم اللغة، والتي أجريت دراسات كثيرة حول فاعليتها في اكتساب اللغة، تلك التي تُستَخدم فيها المدونات داخل الفصل الدراسي فيما يعرف بالتعليم الموجه بالبيانات (Data-Driven Learning)، والذي يعرفه جونز وكينق Johns and King (١٩٩١م: iii) بأنه:

استخدام البيانات اللغوية التي يتم توليدها داخل الفصل الدراسي، عن طريق استخدام كشاف السياقات (Concordancer) على الحاسب الآلي، وذلك لمساعدة الطلاب على استكشاف العبارات والتراكيب اللغوية، وقياس مدى اطرادها في اللغة المستهدفة، ويشمل ذلك بناء أنشطة وتمارين لغوية اعتماداً على هذه البيانات المستخرجة من كشاف السياقات.

وتمثل لذلك قرانجر وتريبل Granger and Tribble (١٩٩٨م) بتدريب صفي تُستخدم فيه المدونات النصية العامة جنباً إلى جنب مع مدونات المتعلمين، حيث يطلب من الدارسين عمل قائمة بالكلمات التي ترد بعد لفظة محددة، ومقارنة أمثلة الناطقين باللغة (في المدونة العامة) مع أمثلة الناطقين بغيرها (في مدونة المتعلمين).

وسنورد هنا تدريباً مشابهاً لما ذُكر عند قرانجر وتريبل، لكنه مطبق على اللغة العربية، حيث استخدمنا المدونة اللغوية العربية لمدينة الملك عبدالعزيز للعلوم والتقنية(٤) (الثبيتي Al-Thubaity، ٢٠١٤م) باعتبارها مدونة عامة، والمدونة اللغوية لمتعلمي اللغة العربية (الفيفي وآخرون، ٢٠١٤م) باعتبارها مدونة متعلمين. في هذا التمرين يطلب من المتعلمين ما يلي:

١- إيجاد الكلمات التي تتلو عبارة «بالنسبة» في جميع الأمثلة المعطاة.

٢- فرز هذه الكلمات لإيجاد المفردات التي وردت فقط في مدونة المتعلمين ولم ترد في المدونة العامة، مع محاولة الحكم عليها بالسؤال عن مدى صحة استخدامها كما وردت في الأمثلة، وكيفية تصحيحها، ونحو ذلك.

**أولاً: نصوص الناطقين باللغة العربية (مدونة عامة)**

| | | |
|---|---|---|
| وأما العبد فقد يريد الشيء ويكون | بالنسبة | له شراً فيخرج من هذه الجهة عن كونه |
| وإن كانت مدته أو عمره طويل لكنه | بالنسبة | إلى الزمن بعامة –الزمن المطلق– لا شك أنه |
| وفق ما تيسر لمؤلفها، والترتيب ينفع المُتَلَقِّي لكن | بالنسبة | لنا سنجري على وفق ما جرى هو عليه |
| و أصبحت المخاطب المفضل | بالنسبة | للجهات المانحة للمساعدات و الممولة للبرامج |
| فرص الاختيار بين السلع والخدمات سواء | بالنسبة | للمنتجين أو المستهلكين كما إن المعرفة قد تلعب |
| وهو قلة كلام السلف وعظيم فقههم | بالنسبة | لمن بعدهم . فالصحابة رضي الله عنهم |
| ومن ثم اختيار القول الراجح في كل صورة . أما | بالنسبة | لتخريج الأحاديث فإن كان الحديث في الصحيحين |
| الذي هو بالنسبة إلى النحو كأصول الفقه | بالنسبة | إلى الفقه |
| مما أضعف عزيمة أفراده. أما | بالنسبة | إلى الجيش الإسلامي فقد كان قليل العدد من |
| وبالعادة يفضل هؤلاء السير على الأقدام . أما | بالنسبة | للحافلات السياحية فإن التحسن السياحي في |

**ثانياً: نصوص الناطقين بغير اللغة العربية (مدونة متعلمين)**

| | | |
|---|---|---|
| ولهذا أرى النجاح إلى الآن في هذا التخصص | بالنسبة | لي بعيدا، ولكن هدفي وإرادتي وعزمي |
| مشكلة مع العائلة التي كانت تأويني | بالنسبة | للطعام والأجرة وعدة مشاكل أخرى |
| اهتماماتي الدراسية هي كثيرة وعديدة ولكن | بالنسبة | إلى التخصص الذي اخترته فهو التخصص العلمي |
| وذلك أن الحاجة تدعو إليها | بالنسبة | أهل بلدي، معظمهم وأكثرهم محتاجون إلى الدعوة |
| بعد دراسة اللغة العربية | بالنسبة | للمعهد سألتحق بكلية أصول الدين بإذن الله |
| هذه هي قصة حياتي وأجمل قصة | بالنسبة | لي فقد قمت بوصفها لك واتمنى أن |
| وبإذن سيتحقق حلمي الذي اريده، أما | بالنسبة | في اختياري له فليس له سبب |
| فقد كانت من أكثر المواقف روحانيه | بالنسبة | الي، فلازلت استشعر ذلك الموقف |
| وايضاً في كلية الشريعة إستفادة كثيرة | بالنسبة | في كليات أخرى |
| أو يرغب بهذا التخصص. | بالنسبة | أسرتي هم يقولون أي التخصص أريد |

الشّكل (١٠) مثال لتدريبات التعليم الموجه بالبيانات

بالإجابة على الأسئلة أعلاه، سيلاحظ الطالب أن الناطقين باللغة العربية في جميع أمثلتهم استخدموا الحرفين «إلى» و«اللام» بعد عبارة «بالنسبة»، بينما استخدم متعلمو اللغة إضافة إلى هذين الحرفين، حرف الجر «في» والاسمين «أهل» و«أسرتي» مما يدفع الطالب للبحث حول صحة هذه التراكيب التي تفرد بها متعلمو اللغة دون الناطقين بها؛ ويظهر لنا من خلال هذا المثال القائم على المدونات النصية – ومنها مدونات المتعلمين – ما يمكن أن يكتسبه الطالب من استخدام لغوي صحيح لبعض العبارات، اعتماداً على بيانات واقعية غير مصطنعة، إضافة لما يكتسبه الطالب عند تفريقه بين الناطقين بالعربية والناطقين بغيرها من الوقوف على بعض الأخطاء المحتملة، والمبنية على بيانات واقعية أيضاً. ولا بد هنا من التنبيه على بعض المحاذير التي ذكرها الباحثون عند استخدام التعليم الموجه بالبيانات (انظر مثلاً تريبل و جونز Tribble and Jones، ١٩٩٧م، وجونز Johns، ١٩٩٤م، وباكرد Packard ، ١٩٩٤م)، ومن ذلك مثلاً الحاجة إلى تدريب الطلاب على استخدام المدونات، وبرامجها، وطرائق البحث فيها، مع الأخذ في الحسبان تفاوت الطلاب في القدرات الحاسوبية واللغوية، ومن المحاذير كذلك أهمية الاختيار الدقيق للأمثلة والسياقات التي تتوافق مع غرض التدريب، وكذلك تصميم التدريبات بما يناسب مستوى الطلاب، فإذا كان طلاب المستوى المتقدم قادرين على تمييز الخطأ في مدونات المتعلمين، فإن غيرهم من المستويات الأقل قد يظنه استخداماً صحيحاً إن لم يجد التوجيه المناسب، وبالتالي تكون المحصلة عكسية بأن يرسخ في ذهنه واحد أو أكثر من الأخطاء الواردة في مفردة أو تركيب لغوي.

أما من ناحية تصميم المواد التعليمية، فقد قارنت بعض الدراسات (مثل لجنق Ljung، ١٩٩١م) بين تعليم اللغة باستخدام مواد تعليمية اعتمدت في تصميمها على المدونات العامة للناطقين باللغة، ومواد تعليمية أخرى مصممة بناء على اجتهاد وحدس المصمم أو المعلم نفسه، وأظهرت النتائج أن مخرجات

النوع الأول كانت أفضل بسبب اعتمادها على أمثلة لغوية حقيقية، وأن نقطة الضعف الكبرى في النوع الثاني تتمثل في إغفال بعض النواحي المهمة لمشاكل الطلاب اللغوية. لكن قرانجر Grangr، (١٩٩٨م) ترى أن المدونات العامة وحدَها لا تشير إلى درجة الصعوبة التي يواجهها متعلمو اللغة، سواءٌ أكان ذلك في المفردات أم في التراكيب، وهنا تأتي أهمية مدونات المتعلمين في تصميم المواد التعليمية، فلا شك أن اعتماد مصممي المواد التعليمية اللغوية على كلا النوعين - المدونات العامة للناطقين باللغة إضافة إلى مدونات المتعلمين الناطقين بغيرها - سيؤدي إلى تطور كبير، مع إمكانية توجيه هذه المواد لتلبي حاجات الطلاب اللغوية بشكل أكبر. ومن الناحية العملية فقد بدأ فعلياً إنتاج بعض الكتب التعليمية المبنية على نتائج أبحاث أجريت على مدونات المتعلمين، وهي نماذج إيجابية مشجعة (قرانجر، ٢٠٠٣م ب).

**سابعاً: البحث في معالجة اللغة الطبيعية (Natural Language Processing)**

يستفيد مجال معالجة اللغة الطبيعية من مدونات المتعلمين في بعض المجالات المهمة، منها على سبيل المثال التدقيق الإملائي، والتدقيق النحوي، المبنيان على نتائج تحليل الأخطاء، مما يعطي دقة أكبر في اكتشاف الأخطاء الإملائية والنحوية، وكذلك في اقتراح البدائل الصحيحة لها، وترتيبها بحسب أولويتها (انظر نورفيج Norvig، ٢٠٠٧م، و جارفسكي ومارتن Jurafsky and Martin، ٢٠٠٩م، والكنهل وآخرون، ٢٠١٢م)؛ من ذلك أيضاً بناء تطبيقات لتعليم اللغة حاسوبياً، وتستطيع هذه البرامج الإفادة بشكل مباشر من مدونات المتعلمين، لتوليد تدريبات تساعد الطلاب على تجاوز المشاكل الفعلية التي تواجههم؛ وكذلك المساعدة على اختيار نصوص القراءة المناسبة لمستوى الطلاب آلياً، وغير ذلك من التطبيقات الحاسوبية، انظر مثلاً شانق وشانق Chang and Chang (٢٠٠٤م)، والنسخة الإلكترونية من دراسة موريرس Meurers (لما

تطبع بعد) بعنوان: مدونات المتعلمين ومعالجة اللغة الطبيعية (Learner Corpora and Natural Language Processing).

## وسم مدونات المتعلمين

كما ذكرنا في بداية الفصل، فإن وسم الأخطاء اللغوية هو أكثر أنواع الوسم شيوعاً في مدونات المتعلمين، حيث تعتمد كثير من الأبحاث في مادتها على نتائج تحليل هذه الأخطاء، كأبحاث اكتساب وتعليم اللغة، وكذلك تأليف المعاجم، وتصميم الكتب التعليمية، ومنها أيضاً بعض تطبيقات اللغويات الحاسوبية. واستناداً على هذا الأساس سنذكر بعض الجوانب المتعلقة بوسم الأخطاء اللغوية في مدونات المتعلمين.

### أولاً: ما هو وسم الأخطاء؟

وسم الأخطاء يعني وضع رمز خاص لكل نوع من أنواع الخطأ اللغوي ليسهل تمييزه عن الأنواع الأخرى عند البحث في المدونة وتحليل مادتها، ومن هذه الأنواع مثلاً الخطأ في الهمزة، والخطأ في حروف الكلمة (كزيادة حرف أو إسقاطه أو استبداله بحرف آخر)، ومنها كذلك الخطأ في الإعراب، والخطأ في علامات الترقيم (زيادة علامة ترقيم، أو إسقاطها، أو استبدالها بأخرى خطأً)، وهكذا. وفي الغالب يتم حصر الأخطاء المراد وسمها، ثم تُصنّف بناء على مجالات أو فئات عامة؛ ليسهل فهمُها واستخدامُها من قبل القائمين على عملية الوسم وكذلك التحليل.

### ثانياً: الدراسات التي عنيت بتصنيف الأخطاء

منذ عقود مضت ظهرت كثير من الدراسات المهتمة بتحليل الأخطاء اللغوية التي يقع فيها الطلبة في كتاباتهم، ورغم أن هذه الدراسات قد صنَّفتْ هذه الأخطاء إلى عدة أنواع (انظر مثلاً العصيلي، ١٤٠٥هـ، والعتيق، ١٤١٢هـ،

والحمد، ١٤١٤هـ، والعقيلي، ١٤١٥هـ)، إلا أن تركيزها ظل محصوراً في العينات البحثية التي تعالجها، متبعة ما يعرف بطريقة التصنيف «من أسفل لأعلى» (Bottom-up approach)، أي أن تكون البداية بالبحث عن الأخطاء في عينة الدراسة وتحليلها، ومن ثم يوضعُ تصنيف لها، ولم يمتد اهتمامها لبناء تصنيفٍ شاملٍ بطريقة البحث «من أعلى لأسفل» (Top-down approach)، حيث يكون التصنيف هو الموضوع أولاً من خلال دراسة أو تصور عام للأخطاء، ومن ثم تتم عملية البحث عن هذه الأخطاء اللغوية وتحليلها في النصوص والمدونات وغيرها. هذه التقسيمات الشاملة للأخطاء لا تقتصر فائدتها على وسم مدونات المتعلمين، بل تتعدى ذلك إلى فوائد أخرى، مثل استخدامها في تعليم اللغة وتصحيح النصوص للطلاب من قِبَل معلميهم، حيث تمنح المعلمَ والطالبَ وضوحاً في تحديد نوع الخطأ بدقة يسهل معها التعرف على أسبابه وطرائق علاجه.

من المحاولات القائمة لبناء تصنيف لأنواع الأخطاء في اللغة العربية، ما قام به غازي أبو حكيمة وآخرون (٢٠٠٨م) من ترجمةٍ لأحد تقسيمات الأخطاء في اللغة الفرنسية (قرانجر، ٢٠٠٣م أ) ومحاولة تطبيقه على اللغة العربية. ويشتمل هذا التصنيف على ثلاث طبقات، الأولى مجالات الخطأ وهي: الشكل، والصرف، والقواعد، والمفردات، والنحو، وازدواجية اللغة، والأسلوب، وعلامات الترقيم، والأخطاء المطبعية؛ أما الطبقة الثانية فحوت فئات الأخطاء الفرعية المندرجة تحت المجالات التسعة السابقة، والطبقة الثالثة جاءت تحت اسم الفئات القواعدية وفيها أربعة أقسام: الاسم، والفعل، وعلامات الترقيم، والرابع لم يسمّه ولعله الحرف، وهذه الطبقة الثالثة عُنيت بالبنية الصرفية أكثر من عنايتها بالأخطاء، انظر أبو حكيمة وآخرين (٢٠٠٨م)، وكذلك الفيفي وأتويل (٢٠١٢م) لمزيد من الإيضاح حول هذا التصنيف.

التصنيف الآخر قام به الباحث كجزء من مشروع المدونة اللغوية لمتعلمي اللغة العربية (الفيفي وآخرون، ٢٠١٣م)، وفيه طبقتان، الأولى تشمل مجالات الخطأ، وقد رُتبتْ تبعاً لتسلسل المستويات اللغوية كما يلي: خطأ إملائي، خطأ صرفي، خطأ نحوي، خطأ دلالي، وبعدها يأتي الخطأ في علامات الترقيم. تحت كل مجال من هذه المجالات الخمسة مجموعة من الفئات الفرعية اعتمد فيها الباحث على عدة دراسات اهتمت بتصنيف الأخطاء وطرق وسمها، وبهذا جمع بين طريقتي البحث السالف ذكرهما، الأولى المسماة «من أسفل لأعلى»، حيث تم بناء تصنيف الأخطاء المقترح اعتماداً على نتائج مجموعة من الدراسات العلمية التي درست عينات كبيرة من أخطاء متعلمي اللغة (انظر الفيفي وأتويل، ٢٠١٢م)، ومن ثم انتقل إلى الطريقة الثانية المسماة «من أعلى لأسفل» من خلال اختبار التصنيف وتطبيقه على عينات من مدونات المتعلمين، ومنها المدونة اللغوية لمتعلمي اللغة العربية. كما قام الباحث بتقويم هذا التصنيف عدة مرات مستقلة لتحسينه وتنقيحه، وفي كل مرة يطبق على عينة مختلفة من النصوص؛ للتأكد من دقته وشموله ووضوح تصنيف الأخطاء فيه. كما قام الباحث بدراسة حول استخدام هذا التصنيف بالمقارنة مع تصنيف أبو حكيمة، وذلك بوسم عينة من النصوص بواسطة باحثَيْن مستقلَّيْن، وتم وسم هذه العينة مرتين باستخدام أحد التصنيفين في كل مرة، مع قياس مدى التطابق بين الباحثَين عند استخدام كل واحد من تصنيفات الأخطاء؛ وقد أظهرت النتائج أن التطابق بين المشاركَين عند استخدام التصنيف المقترح كان أكثر منه عند استخدام تصنيف أبو حكيمة، وللاطلاع على نتائج هذه الدراسة انظر الفيفي وآخرين (٢٠١٣م).

**ثالثاً: الأخطاء التي توسم**

يعتمد ذلك على الغرض من دراسة الأخطاء، فعلى الرغم من استخدام أغلب الباحثين للتصنيفات الموجودة بسبب شمولها لأهم الأخطاء اللغوية، إلا أن هناك من يدرس نقاطاً محددة يعمد فيها إلى استخدام تصنيفه الخاص

الذي يُحقّقُ أهداف بحثه بشكل مباشر، ومن ذلك مثلاً أن يكتفي بوسم الأخطاء التي يجري عليها دراسته مع إضافة بعض التفاصيل عليها. فمن يدرس أخطاء الهمزة على سبيل المثال، قد يتعمق أكثر في تصنيفها لتشمل كل الأخطاء المحتملة، كأن يقسمها مثلاً إلى أخطاء الهمزة في أول الكلمة، والخطأ في الهمزة المتوسطة، والخطأ في الهمزة المتطرفة، وكل واحد منها له أقسامه الفرعية، ففي أول الكلمة قد يتعلق الخطأ بكتابة همزة الوصل أو إسقاط همزة القطع، وفي وسطها قد تُقسّم الأخطاء بناء على الحرف الذي تكتب عليه الهمزة، وهكذا. أما مدونات المتعلمين - موضوع حديثنا في هذا المبحث - فكثير منها يعتمد على تصنيفات الخطأ العامة كتلك التي أوردناها سابقاً (أبو حكيمة وآخرون، ٢٠٠٨م، والفيفي وآخرون، ٢٠١٣م)؛ وسبب ذلك يعود إلى عموم الهدف من إنشاء مثل هذه المدونات، فناسب أن يكون وسم الأخطاء فيها عاماً كذلك.

**رابعاً: كيفية وسم الأخطاء**

يستخدم الباحثون عدة طرق لوسم الأخطاء تعتمد على نوع الملفات المستخدمة لحفظ النصوص (txt أو xml مثلاً)، وكذلك برامج تحليل هذه النصوص، إضافة إلى جدول تصنيف الأخطاء نفسه الذي يفرض أحياناً بعض القيود على طريقة وضع الوسوم. من هذه الطرائق أن تُدرِج وسم الخطأ بعد الكلمة أو العبارة الخاطئة مباشرة، على أن يكون الوسم محصوراً بعلامتين تميزهما عن النص، وهي في الغالب الأقواس المثلثة <>، كما يضيف بعض الباحثين إلى وسم الخطأ الشكل الصحيح للكلمة أو العبارة الخاطئة (انظر الشكل ١١). ويمكن لبعض البرامج المستخدمة في تحليل المدونات - مثل ورد سميث تولز(٥) WordSmith Tools (سكوتّ Scott، ٢٠١٢م)- إدراكُ أن ما بين هذه الأقواس وسوم إضافية على النص وليست من أصله، فلا يتم عرضها مع النص الأصلي، مع تمكين المستخدم من البحث في هذه النصوص وتحليلها بناء على تلك الوسوم.

«وما كنتو < OG > كنت > وحيد < OM > وحيدًا > في هاذا < OD > هذا > الأمر < PM > الأمر، > بل كانت < XG > كان > معي زملاءي < OH > زملائي > الآخرين < XC > الآخرون >»

الشّكل (١١) مثال لوسم الأخطاء وتصويبها ضمن النص

من طرق وسم الأخطاء كذلك فصل الكلمات بحيث تكون كل كلمة – أو وحدة صرفية أو معجمية – في سطر مستقل، على أن يُجعل وسم الخطأ مقابلاً للكلمة الخاطئة، وتَفصِل بينهما مسافة جدولة (tab)، كذلك يكون التصويب بعد الوسم مع وجود نفس الفاصل (انظر الشكل ١٢)، وتستطيع أداة البحث والتحليل سكتش إنجن(٦) Sketch Engine (كيلقارف Kilgarriff، ٢٠٠٤م) قراءة هذا النوع من التنسيق.

| | | |
|---|---|---|
| وما | | |
| كنتو | OG | كنت |
| وحيد | OM | وحيداً |
| في | | |
| هذا | OD | هذا |
| الأمر | PM | الأمر، |
| بل | | |
| كانت | XG | كان |
| معي | | |
| زملاءي | OH | زملائي |
| الآخرين | XC | الآخرون |

الشّكل (١٢) مثال لوسم الأخطاء وتصويبها على شكل أعمدة

وفيما يلي (الشكل١٣) نموذج تم توليده آلياً بتنسيق XML، من نصوص المدونة اللغوية لمتعلمي اللغة العربية، وذلك بعد وسم وتصحيح الأخطاء فيه.

```
<?xml version="1.0"?>
<doc ID="S037_T2_M_Pre_NNAS_W_C">
    <text>
        <title>
            <t id=1>تخصص</t>
            <t id=2>في</t>
            <t id=3>المستقبل</t>
        </title>
        <p id=1>
            <t id=6>لما</t>
            <t id=7>ننظر</t>
            <t id=8>إلى</t>
            <t id=9>العالم</t>
            <t id=10 ErrTag="XG" ErrForm="الإسلامية" CorrForm="الإسلامي">الإسلامية</t>
            <t id=11>اليوم,</t>
            <t id=12>نرى</t>
            <t id=13>كثيرا</t>
            <t id=14>من</t>
            <t id=15>المشكلات</t>
            <t id=16>وقع</t>
            <t id=17>المسلمون</t>
            <t id=18>فيها,</t>
            <t id=19>ونتساءل</t>
            <t id=20>ما</t>
            <t id=21 ErrTag="PC" ErrForm="الحل," CorrForm="الحل؟">الحل,</t>
            <t id=22>وماذا</t>
            <t id=23>نفعل</t>
            <t id=24>حتى</t>
            <t id=25>نساهم</t>
            <t id=26>في</t>
            <t id=27>إصلاح</t>
            <t id=28 ErrTag="PC" ErrForm="أمتنا." CorrForm="أمتنا؟">أمتنا.</t>
            <t id=29>منذ</t>
            <t id=30>أن</t>
            <t id=31>بدأت</t>
            <t id=32 ErrTag="OM" ErrForm="بالالتزام" CorrForm="بالالتزام">بالالتزام</t>
            <t id=33>بالدين</t>
            <t id=34>وتركت</t>
            <t id=35>الجاهلية,</t>
            <t id=36>بدأت</t>
            <t id=37>بالتعرف</t>
            <t id=38>على</t>
            <t id=39>أحوال</t>
            <t id=40>المسلمين</t>
            <t id=41>اليوم</t>
            <t id=42>وفي</t>
            <t id=43 ErrTag="SW" ErrForm="التاريخ," CorrForm="الماضي،">التاريخ,</t>
            <t id=44>فعرفت</t>
            <t id=45>أنا</t>
            <t id=46>قد</t>
            <t id=47>كذبنا</t>
```

الشّكل (١٣) ملف بتنسيق XML من المدونة اللغوية لمتعلمي اللغة العربية بعد وسم الأخطاء وتصويبها

وقبل أن نختم حديثنا هنا، نود التنبيه على أن مدونات المتعلمين لم تقتصر على وسم الأخطاء، وإنما فصلنا الحديث حوله لأهميته الخاصة في هذا النوع من المدونات، مع التأكيد على أن مدونات المتعلمين قابلة - كغيرها من المدونات - لوسم الكلمات من الناحية الصرفية، والنحوية، والدلالية، والبلاغية، وغير ذلك من أشكال الوسم التي لا يمكن استقصاؤها هنا، وعلى قدر غنى المدونات من ناحية الوسوم تزداد الفائدة من مادتها. ولنضرب لذلك أمثلة، فعندما تكون الأخطاء موسومة، إضافة إلى وسم الكلمات صرفياً، يمكننا البحث عن الأخطاء في بنية صرفية معينة، مثل أخطاء الهمزة في الأفعال المضارعة. ولو أضفنا لذلك الوسم النحوي لأمكننا البحث مثلاً عن أخطاء المطابقة في الإعراب، أو الجنس، أو العدد، أو التعريف والتنكير، ما بين الأسماء والصفات، وهكذا.

## الخاتمة

استعرضنا في هذا المبحث مدونات المتعلمين من عدة جوانب، ووقفنا على مجالات استخدامها، ورأينا ما تزخر به من إمكانات، وما يمكن أن تسهم به في سبيل الرقي بالبحث اللغوي لآفاق أرحب، بدايةً من اكتساب وتعليم اللغة، وتأليف المعاجم، وتصميم المواد التعليمية، وصولاً لاستخدامها في مجال اللغويات الحاسوبية، وتعليم اللغة بمساعدة الحاسب. ومع تعدد هذه الأبحاث والموضوعات فلا زال المجال واسعاً، والطريق ممتدة للاستفادة من مدونات المتعلمين في مجالات جديدة وباستخدام طرائق شتى.

كما أتينا على أهم أنواع الوسم في مدونات المتعلمين، وهو وسم الأخطاء، وأكدنا على أن هذه المدونات يمكنها الاستفادة من أنواع الوسوم الأخرى مما يوسع إمكاناتها ويضاعف طرق البحث فيها. وبما أننا قد أتينا على مجموعة من مدونات المتعلمين العربية في إشارة إلى بداية الاهتمام بها، فإن مما

ينبغي التنبيه عليه في هذا المقام أن العمل في مجال مدونات المتعلمين العربية، ووسمها، والاستفادة منها في مجالات البحث اللغوي - أو غيرها من المجالات - لا يزال في أوله، ويحتاج للكثير من الدراسات ليصل إلى مستويات أفضل، ولعل هذا الفصل قد مهَّد الطريق لذلك، وكوَّن نقطة انطلاق لكثير من الباحثين بإذن الله تعالى.

## شكر وعرفان

للزملاء الكرام الذين تفضلوا بمراجعة هذا المبحث، وأبدوا ملاحظات قيمة ساعدتْ على تنقيحه وتهذيبه.

## الحواشي

(١) من المصطلحات المستخدمة للدلالة على عملية الوسم «التحشية» أو «الترميز»، وفي هذا الفصل اعتمدت مصطلح الوسم، انظر «قائمة مصطلحات لسانيات ا لمدونات اللغوية» للدكتور محمود إسماعيل صالح:

صالح، محمود إسماعيل: قائمة مصطلحات لسانيات المدونات اللغوية corpus-/01/2014/http://dr-mahmoud-ismail-saleh.blogspot.co.uk linguistics.html#more

(٢) يمكن تسميتها كذلك «تراكمية»، أو «تتبعية»، لكني فضلت مصطلح مدونة طولية ليكون مقابلاً مناسباً للمدونة العرضية.

(٣) توجد نسخة معربة من هذا القاموس بعنوان «معجم لونجمان للأخطاء الشائعة في اللغة الإنجليزية (إنجليزي - إنجليزي - عربي)»، نقله إلى العربية نبيل راغب (٢٠٠٧م):

راغب، نبيل: معجم لونجمان للأخطاء الشائعة في اللغة الإنجليزية، الشركة المصرية العالمية للنشر – لونجمان، الجيزة – مصر، ٢٠٠٧م.

(٤) يمكن الوصول إليها عن طريق الرابط التالي: http://www.kacstac.org.sa

(٥) يمكن تنزيله من الرابط التالي: http://www.lexically.net/wordsmith/

(٦) يمكن الوصول إليها عن طريق الرابط التالي: http://www.sketchengine.co.uk

## مصادر إضافية حول مدونات المتعلمين

من الملاحظ وجود قصور كبير في المراجع العربية التي تحدثت عن مدونات المتعلمين، ولعل البحث الذي نشر للباحث (الفيفي، ٢٠١٢م) في مؤتمر هندسة وعلوم الحاسب باللغة العربية، من أوائل المراجع العربية التي تناولت هذا النوع من المدونات بشيء من التفصيل، وهو مذكور ضمن قائمة المراجع لهذا الفصل، وسأشير هنا إلى أهم المصادر باللغة الإنجليزية:

**مدونات المتعلمين الحاسوبية واكتساب اللغة الثانية وتعليم اللغة الأجنبية**

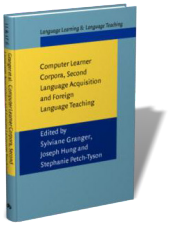
**Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching**

يتناول هذا الكتاب مجموعة من الأبحاث مقسمة على ثلاثة محاور رئيسة:

دور مدونات المتعلمين الحاسوبية في اكتساب اللغة الثانية وتعليم اللغة الأجنبية

المناهج القائمة على المدونات في بحث اللغة المرحلية

وسائل تدريس اللغة الأجنبية المبنية على المدونات




الكتاب من إصدار دار النشر الألمانية

John Benjamins Publishing Company

في ٢٠٠٢م

**عشرون عاماً من البحث في مدونات المتعلمين: نظرة على الماضي للمضي قدماً**

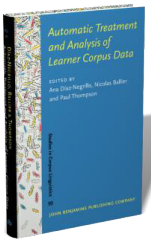
**Twenty Years of Learner Corpus Research: looking back, moving ahead**

يضم هذا الكتاب أكثر من أربعين بحثاً تمثل السجل العلمي لأول مؤتمر يقام حول مدونات المتعلمين في سبتمبر ٢٠١١ في جامعة لوفاين ببلجيكا، وقد طبعت الأبحاث المقدمة للمؤتمر في هذا الكتاب متسلسلةً دون تصنيف، لكنها شملت عدة موضوعات مثل تصميم مدونات المتعلمين، وجمع بياناتها، ووسمها، وكذلك تحليلها صوتياً، ونحوياً، ودلالياً، وبلاغياً، وغيرها من الجوانب التي تعكس الاهتمام بهذا النوع من المدونات.

**التحليل والمعالجة الآلية للبيانات في مدونات المتعلمين**

**Automatic Treatment and Analysis of Learner Corpus Data**

يتناول هذا الكتاب مجموعة من جوانب البحث المتعلقة بمدونات المتعلمين، وقد ضم – بعد مقدمة حول مستقبل هذا النوع من المدونات – مجموعة من الأبحاث توزعت على ثلاثة فصول:



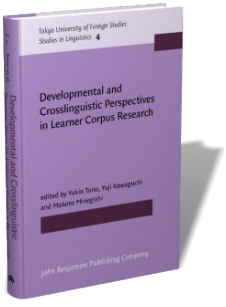

الأول: جمع مدونات المتعلمين، ووسمها، وتبادل موادها

الثاني: وسائل آلية للتعرف على خصائص لغة المتعلمين

الثالث: تحليل البيانات في مدونات المتعلمين

**رؤى لغوية تطويرية حول أبحاث مدونات المتعلمين**

**Developmental and Crosslinguistic Perspectives in Learner Corpus Research**




يشتمل هذا الكتاب على فصلين، يركز الفصل الأول على مشروع المدونة النصية الدولية للغة المرحلية (The International Corpus of Crosslinguistic Interlanguage)، وما أجري عليها من أبحاث، بينما يضم الفصل الثاني مجموعة متنوعة من مدونات المتعلمين، ذات البيانات المنطوقة (Spoken learner corpora)، وما تم عليها من أبحاث تتعلق ببنائها، ووسمها، وتحليلها.

**المجلة الدولية لأبحاث مدونات المتعلمين**

**International Journal of Learner Corpus Research**

وهي دورية حديثة يبدأ صدور العدد الأول منها مطلع عام ٢٠١٥م من دار النشر الألمانية John Benjamins Publishing Company. تهتم هذه المجلة بنشر الأبحاث التي تتناول بناء مدونات المتعلمين، ووسمها، وتحليلها، واستخدامها في أي مجال من مجالات البحث اللغوي. ورابطها على الشبكة العنكبوتية:

https://benjamins.com/#catalog/journals/ijlcr/main

# المراجع

## المراجع العربية

**أَمُّون، هلا:** معجم تقويم اللغة وتخليصها من الأخطاء الشائعة، دار القلم للطباعة والنشر والتوزيع، بيروت، تاريخ الطبع غير معروف.

**حسان، حسلينا و غالب، محمد فهام محمد:** مشروع جمع المدونات النصية الخاصة بالنصوص الأكاديمية في اللغة العربية، مجلة مجمع اللغة العربية الأردني، العدد الخامس والثمانون، ٢٠١٣م، ٥٧-٧٦.

**الحمد، محمد ماجد:** تحليل أخطاء التعبير الكتابي لدى المستوى المتقدم من دارسي العربية غير الناطقين بها في جامعة الملك سعود، رسالة علمية غير منشورة، جامعة الإمام محمد بن سعود الإسلامية، الرياض ، ١٤١٤هـ.

**شابل، كارول:** تطبيقات الحاسب الآلي في اكتساب اللغة الثانية: أسس للتعليم والقياس والبحث العلمي، ترجمة سعد بن علي وهف القحطاني، جامعة الملك سعود، الرياض، ١٤٢٨هـ.

**الشافعي، إبراهيم محمد و إبراهيم، عبدالحميد صفوت:** الأخطاء الشائعة في الهجاء والإملاء بين تلاميذ المرحلة الابتدائية بمنطقة الرياض، مركز بحوث كلية التربية في جامعة الملك سعود، ١٤٠٨هـ.

**العتيق، زايد مهلهل:** تحليل الأخطاء الدلالية لدى دارسي اللغة العربية من غير الناطقين بها في مادة التعبير الكتابي، رسالة علمية غير منشورة، جامعة الإمام محمد بن سعود الإسلامية، الرياض، ١٤١٢هـ.

**العدناني، محمد:** معجم الأخطاء الشائعة، مكتبة لبنان، بيروت، الطبعة الثانية، ١٩٨٣م.

**العصيلي، عبدالعزيز إبراهيم:** الأخطاء الشائعة في الكلام لدى طلاب اللغة العربية الناطقين بلغات أخرى: دراسة وصفية تحليلية، رسالة علمية غير منشورة، جامعة الإمام محمد بن سعود الإسلامية، الرياض، ١٤٠٥هـ.

**العقيلي، عبدالمحسن سالم:** تحليل الأخطاء في بعض أنماط الجملة الفعلية للغة العربية في الأداء الكتابي لدى دارسي المستوى المتقدم، رسالة علمية غير منشورة، جامعة الإمام محمد بن سعود الإسلامية، الرياض، ١٤١٥هـ.

**الفيفي، عبدالله وأتويل، إيريك:** المدونات اللغوية لمتعلمي اللغة العربية: نظام لتصنيف وترميز الأخطاء اللغوية، المؤتمر الدولي الثامن لهندسة وعلوم الحاسب الآلي باللغة العربية، ٢٠١٢م، القاهرة، مصر.

**محمد، جُودة مبروك:** المعجم الوجيز في الأخطاء الشائعة والإجازات اللغوية، مكتبة الآداب، القاهرة، ١٤٢٦هـ - ٢٠٠٥م

**المدونة اللغوية العربية لمدينة الملك عبدالعزيز للعلوم والتقنية:** مدينة الملك عبدالعزيز للعلوم والتقنية، ٢٠١٣م، http://www.kacstac.org.sa

## المراجع الأجنبية

**Abdullah, Shazila and Noor, Noorzan Mohd:** Contrastive Analysis of the Use of Lexical Verbs and Verb-noun Collocations in Two Learner Corpora: WECMEL vs. LOCNESS. In Ishikawa, Shin (Ed.), *Learner corpus studies in Asia and the world.* (Vol. 1), Papers from LCSAW2013. 2013, pp. 139–160.

**Abuhakema, G., Faraj, R., Feldman, A. and Fitzpatrick, E:** Annotating an Arabic Learner Corpus for Error. In: proceeding of *the*

*International Conference on Language Resources and Evaluation, LREC 26 May - 1 June 2008. Marrakech, Morocco, 2008.*

**Alfaifi, Abdullah, Atwell, Eric and Abuhakema, Ghazi:** Error Annotation of the Arabic Learner Corpus: A New Error Tagset. In: *Language Processing and Knowledge in the Web, Lecture Notes in Computer Science. 25th International Conference, GSCL 2013, 25-27 September 2013. Darmstadt, Germany, Springer,* 14 - 22 (9), *2013.*

**Alfaifi, Abdullah, Atwell, Eric and Hedaya, Ibraheem:** Arabic Learner Corpus (ALC) v2: A New Written and Spoken Corpus of Arabic Learners. In the proceedings of the *Learner Corpus Studies in Asia and the World (LCSAW 2014), 31 May - 01 Jun 2014,* Kobe, Japan. <http://www.arabiclearnercorpus.com>, 2014.

**Alkanhal, Mohamed, Al-Badrashiny, Mohamed, Alghamdi, Mansour and Al-Qabbany, Abdulaziz:** Automatic Stochastic Arabic Spelling Correction With Emphasis on Space Insertions and Deletions. *IEEE Transactions on Audio, Speech, and Language Processing,* 20(7), 2012, pp. 2111–2122.

**Al-Sulaiti, Latifa:** *Designing and Developing a Corpus of Contemporary Arabic*. Unpublished thesis, University of Leeds, 2004.

**Al-Thubaity, Abdulmohsen. O.:** A 700M+ Arabic corpus: KACST Arabic corpus design and construction. *Language resources and evaluation*. 2014. DOI 10.1007/s10579-014-9284-1.

**Bullon, Stephen:** *Longman Exams Dictionary.* Pearson Longman. England, 2006.

**Burnard, Lou.** Metadata for corpus work. In M. Wynne (Ed.), *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford, Oxbow Books, 2005, pp. 30-46.

**Cambridge Learner Corpus:** *Cambridge University Press.* 2012, Retrieved from: http://www.cambridge.org/gb/elt/catalogue/subject/custom/item3646603/Cambridge-English-Corpus-Cambridge-Learner-Corpus/?site_locale=en_GB

**Chang, Jason S. and Chang, Yu-Chia:** *Computer Assisted Language Learning Based on Corpora and Natural Language Processing: The Experience of Project CANDLE.* IWLeL 2004, Waseda University, Tokyo. 2004, Retrieved from: https://dspace.wul.waseda.ac.jp/dspace/bitstream/2065/1391/1/02.pdf

**Connor, Ulla, Precht, Kristen and Upton, Thomas:** Business English: learner data from Belgium, Finland and the U.S. In S. Granger, J. Hung and S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam & Philadelphia, Benjamins, 2002, pp. 175-194.

**Dagneaux, Estelle, Denness, Sharon., Granger, Sylviane, and Meunier, Fanny:** *Error tagging manual*. version 1.1, 1996.

**Ellis, Rod:** The Study of Second Language Acquisition. Oxford University Press, Oxford, 1994.

**Farwaneh, Samira and Tamimi, Mohammed:** *Arabic Learners Written Corpus: A Resource for Research and Learning*. 2012, Retrieved from the University of Arizona, the Center for Educational Resources

in Culture, Language and Literacy Web site: http://l2arabiccorpus.cercll.arizona.edu/?q=homepage

**Gillard, Patrick and Gadsby, Adam:** Using a learners› corpus in compiling ELT dictionaries. In Granger (Ed.) *Learner English on Computer.* Longman, London and New York, 1998, pp. 159-171.

**Granger, Sylviane and Tribble, Chris:** Learner corpus data in the foreign language classroom: form-focused instruction and data-driven learning. In Granger (Ed.) *Learner English on Computer.* Longman, London and New York, 1998, pp. 159-171.

**Granger, Sylviane:** A Bird›s-eye View of Computer Learner Corpus Research. In S. Granger, J. Hung and S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam & Philadelphia, Benjamins, 2002, pp. 3-33.

**Granger, Sylviane:** Error-tagged learner corpora and CALL: a promising synergy. *CALICO Journal,* 20(3), 2003 أ, pp. 465-480.

**Granger, Sylviane:** *Learner corpora around the world.* 2012, Retrieved from the Université Catholique de Louvain, Centre for English Corpus Linguistics Web site: http://www.uclouvain.be/en-cecl-lcworld.html

**Granger, Sylviane:** The computer learner corpus: a versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on Computer.* London/New York, Addison Wesley Longman, 1998, pp. 3 - 18.

**Granger, Sylviane:** The International Corpus of Learner English. In J. Aarts, P. de Haan and N. Oostdijk (Eds.), *English Language Corpora: Design, Analysis and Exploitation*. Amsterdam, Rodopi, 1993, pp. 57-69.

**Granger, Sylviane:** The International Corpus of Learner English: A New Resource for Foreign Language Learning and Teaching and Second Language Acquisition Research. *TESOL Quarterly*, 37(3), 2003 ب, pp. 538-546.

**Hammarberg, Björn:** *Introduction to the ASU Corpus, a Longitudinal Oral and Written Text Corpus of Adult Learners' Swedish with a Corresponding Part from Native Swedes*. Stockholm University, Department of Linguistics, 2010.

**Housen, Alex:** A corpus-based study of the L2-acquisition of the English verb system. In: Granger, S., Hung, J. and Petch-Tyson, S. (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam, John Benjamins, 2002, pp. 77-116.

**Hutchinson, John:** *UCL Error Editor.* Louvain-la-Neuve: Centre for English Corpus Linguistics, Université Catholique de Louvain, 1996.

**Johns, Tim and King, Philip:** Classroom Concordancing, *ELR-Journal*. 4, 1991, pp. 1-13.

**Johns, Tim:** From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.),

*Perspectives on Pedagogical Grammar.* Cambridge, Cambridge University Press, 1994, pp. 293-313.

**Jurafsky, Daniel, and Martin, James:** *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. 2nd edition, Prentice-Hall, 2009.

**Kennedy, Graeme:** *An Introduction to Corpus Linguistics*. London, Longman, 1998.

**Kilgarriff, Adam, Rychly, Pavel, Smrz, Pavel, & Tugwell, David:** The Sketch Engine. In the proceedings of *the Euralex, 6-10 July 2004*. Lorient, France, 2004.

**Lee, Dong Ju:** *Corpora and the classroom: a computer-aided error analysis of Korean students' writing and the design and evaluation of data-driven learning materials*. Unpublished thesis, University of Essex, 2007.

**Leech, Geoffrey:** Teaching and Language Corpora: a Convergence. In A. Wichmann, S. Fligelstone, T. McEnery and G. Knowles (Eds.), *Teaching and language corpora*. London, Longman, 1997, pp. 1-23.

**Ljung, Magnus:** Swedish TEFL Meets Reality. In Johansson, S. and Stenstrom, A-B. (Eds.) *English Computer Corpora: Selected Papers and Research Guide*. Mouton de Gruyter, Berlin and New York, 1991, pp. 245-256.

**Longman Corpus Network:** *The Longman Learners› Corpus*. 2012, Retrieved from: http://www.pearsonlongman.com/dictionaries/corpus/learners.html.

**Meurers, Detmar:** Learner Corpora and Natural Language Processing, *Cambridge Handbook of Learner Corpus Research*. edited by Sylviane Granger, Gaëtanelle Gilquin and Fanny Meunier. Cambridge University Press, Forthcoming, Retrieved from: http://www.sfs.uni-tuebingen.de/~dm/papers/Meurers-LCNLP-draft.pdf

**Milton, John, and Nandini, Chowdhury:** Tagging the interlanguage of Chinese learners of English. In L. Flowerdew and A. K. K. Tong (Eds.), *Entering Text*. Hong Kong, The Hong Kong University of Science and Technology, 1994.

**Nesselhauf, Nadja**: Learner Corpora and Their Potential in Language Teaching. In: Sinclair, J. (Ed.), *How to Use Corpora in Language Teaching*. Amsterdam & Philadelphia, Benjamins, 2004, pp. 125-152.

**Norvig, Peter:** *How to Write a Spelling Corrector.* 2007, Retrieved from: http://norvig.com/spell-correct.html.

**Packard, V. :** Producing a concordance-based self-access vocabulary package: some problems and solutions. In: Flowerdew, Lynne & Tong, Anthony. (Eds.), *Entering Text*. Language Centre, University of Science and Technology, 1994, pp. 215-226.

**Pravec, N. A. :** Survey of learner corpora. *ICAME Journal*, 26, 2002, pp. 81-114, Retrieved from: http://icame.uib.no/ij26/pravec.pdf.

**Scott, Mike:** *WordSmith Tools version 6*. Liverpool: Lexical Analysis Software, 2012, http://www.lexically.net/wordsmith.

**Simpson, R. C., Briggs, S. L., Ovens, J. and Swales, J. M. :** *The Michigan Corpus of Academic Spoken English*. Ann Arbor, MI, The Regents of the University of Michigan, 2002, Retrieved from the University of Michigan, English Language Institute Web site: http://micase.elicorpora.info/about-micase

**Sinclair, John:** Corpus and Text - Basic Principles. In M. Wynne (Ed.), Developing Linguistic Corpora: a Guide to Good Practice. Oxford, Oxbow Books, 2005, pp. 1-16.

**Sinclair, John:** *Corpus, Concordance, Collocation*. Oxford, Oxford University Press, 1991.

**Sinclair, John:** *EAGLES. Preliminary recommendations on Corpus Typology,* 1996, Retrieved from: http://www.ilc.cnr.it/EAGLES/corpustyp/corpustyp.html

**Sosnina, Ekaterina:** *Russian Learner Translator Corpus (RusLTC),* 2014, Retrieved from: http://rus-ltc.org

**Spence, Robert:** A corpus of student L1–L2 translations. In S. Granger & J. Hung (Eds.), *Proceedings of the International Symposium on Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Hong-Kong, The Chinese University of Hong Kong, 1998, pp. 110–112.

**Summers, Della:** *Longman Essential Activator.* Longman. England, 1997.

**Thoday, Elizabeth:** Issues in building learner corpora: an investigation into the acquisition of German passive constructions. In *the 2nd Newcastle Postgraduate Conference in Theoretical and Applied Linguistics (25 June 2007)*. Newcastle, UK, 2007.

**Tribble, Chris and Jones, Glyn:** *Concordances in the Classroom: A Resource Book for Teachers*. Houston, TX, Athelstan, 1997.

**Turton, Nigel D. and Heaton, J.B. :** *Longman dictionary of common errors,* Pearson Longman. England, 2nd edition, 1996.

**Zaghouani, Wajdi, Mohit, Behrang, Habash, Nizar, Obeid, Ossama, Tomeh, Nadi and Oflazer, Kemal:** (Large-scale Arabic Error Annotation: Guidelines and Framework. In proceedings of the International Conference on *Language Resources and Evaluation (LREC 26-31 May 2014). Rejkavik, Iceland,* 2014.