

# المدونات المتوازية، واستخداماتها في تعليم اللغات وتدريب المترجمين

د. هند بنت مطلق العتيبي  
أستاذ مساعد- كلية اللغات والترجمة  
جامعة الملك سعود  
[Hialotaibi@ksu.edu.sa](mailto:Hialotaibi@ksu.edu.sa)



# المحاور

- ما هي المدونة المتوازية (Parallel Corpus)؟
- استخدام المدونات المتوازية في تعليم اللغات
- استخدام المدونات المتوازية في تدريب المترجمين
- استعراض المدونات المتوازية العربية/ الإنجليزية
- مشروع المدونة العربية - الإنجليزية (AEPC)

## ما هي المدونة المتوازية (Parallel Corpus)؟

- يُعرّف سينكلير (Sinclair, 1995: 19) المدونة المتوازية بأنها "نصوص أصلية وترجماتها بلغة (أو بلغات) أخرى".
- كما تعرفها بيكر ( Baker , 1995 ) بأنها "مجموعة نصوص كتبت باللغة الأصل (أ) وترجماتها باللغة (ب)".

# استخدام المدونات المتوازية في تعليم اللغات

- بنهاية الثمانينات وبداية التسعينات العديد من الدراسات التي تدعم ذلك الاتجاه، ومنها جونز وستيفنز، Johns (1986 و Stevens, 1991)، وغيرها.
- ثمة منهجين لاستخدام المدونات في تعليم اللغة:

غير مباشر

مباشر

# المنهج المباشر

يعتمد المنهج المباشر على Data-Driven Learning

التعلم القائم على الشواهد

وهو المبدأ الذي قدمه Johns منذ ما يقارب 20 عاماً

(انظر: Johns, 1991, 2002)

مواجهة الطالب مباشرة بالبيانات قدر الإمكان؛ بحيث  
تخلق من المتعلم باحثاً لغوياً.

# Data-Driven Learning DDL

- A focus on the exploitation of authentic materials even when dealing with tasks such as the acquisition of grammatical structures and lexical items.
  - A focus on real, exploratory tasks and activities rather than traditional “drill and kill” exercises.
  - A focus on learner-centered activities.
  - A focus on the use and exploitation of tools rather than ready-made or off-the-shelf learnware.
- (Davies et al., 2012, p. 26)

# المنهج غير المباشر

- يركز على الدراسات البحثية؛ و البراهين التي يمكن استقاؤها من المدونات، وتأثيرها على تصميم المناهج الدراسية، واختيار المواد التعليمية التي تُقدَّم للطالب.
- كمثال: مدونة المتعلمين (Learners' Corpus) حيث تفيد هذه المدونات في تحليل أخطاء الدارسين من خلفيات ومستويات مختلفة، ثم تُسهم في دراسات اكتساب اللغة الثانية .

(انظر صالح 2012)

# مزايا استخدام المدونات في تعليم اللغات

- باستخدام الفعلي للغة (Real language use).
- يشجع استخدام المدونات في تعليم اللغة مبادئ التعلم الفعال.
- الاستقلالية في التعلم (autonomy).



# دراسات تناولت استخدام المدونات المتوازية

- John, E. S. (2001). A case for using a parallel corpus and concordancer for beginners of a foreign language. *Language learning & technology*, 5(3), 185-203.
- Chujo, K., Anthony, L., and Oghigian, K. (2009). DDL for the EFL classroom: Effective uses of a Japanese-English parallel corpus and the development of a learner-friendly, online parallel concordancer. In *Proceedings of the corpus linguistics conference, CL2009*. Liverpool University of Liverpool.
- SANGAWA, K. H. (2014). The learner as lexicographer: using monolingual and bilingual corpora to deepen vocabulary knowledge. *Acta Linguistica Asiatica*, 4(2), 53-65.
- Wong, T. S., and Lee, J. S. (2016). Corpus-based learning of Cantonese for Mandarin speakers. *ReCALL*, 28(02), 187-206.

# استخدام المدونات المتوازية في تدريب المترجمين

- ظهرت الفكرة في بداية التسعينات، على يد عدد من الباحثين، ومن أشهرهم: منى بيكر (انظر Baker (1995; 1993)، و جون سنكلير (أنظر Sinclair, 1995)، و (Aston 1999).

تطبيقي

نظري

# استخدام المدونات المتوازية في تدريب المترجمين

- يشير غوداك (Gouadec,2002) إلى أن تعليم المترجمين يتضمن "تدريب الطالب على تنفيذ وظائف محددة، ضمن أوساط محددة، بالاستعانة بأدوات ونظم محددة".

# استخدام المدونات المتوازية في تدريب المترجمين

- طبيعة الاختلاف بين النص الأصلي وترجمته.
- تأثير اللغة الأم على الترجمة.
- المتلازمات اللفظية، والقوالب المألوفة، والتراكيب الشائعة في كلا اللغتين.
- التحقق من خيارات الترجمة المستمدة من مصادر أخرى.
- المقارنة بأعمال المترجمين المحترفين.
- زيادة الوعي بطبيعة النصوص التخصصية

# دراسات تناولت استخدام المدونات المتوازية

- Frankenberg-Garcia, A., and Santos, D. (2003). Introducing Compara: the Portuguese -English parallel corpus. *Corpora in translator education*, 71-87.
- Salhi, H. (2013). Investigating the Complementary Polysemy and the Arabic Translations of the Noun Destruction in EAPCOUNT. *Meta: Journal des aducteurs/Translators' Journal*, 58(1), 227-246.
- Rodríguez-Inés, P. (2014). Using corpora for awareness-raising purposes in translation, especially into a foreign language (Spanish–English). *Perspectives*, 22(2), 222-241.
- Haiyan, L. and Zhongxin, D. (2014). Effectiveness of self-built Chinese-English Corpus on Assisting Translation Teaching. *International Journal of Humanities and Social Science*, 4(7), 96-98.
- Verplaetse, H., and Heylen, K. (2015). Parallel Corpora for Medical Translation Training: An Analysis of Impact On Student Performance. *In proceedings of the 4th International Conference on Corpus Use and Learning to Translate (CULT)*. Alicante, Spain.

# المدونات المتوازية العربية/ الإنجليزية

- صعوبة معالجة اللغة العربية مقارنة باللغة الإنجليزية مثلاً؛ حيث تترابط الحروف مع بعضها كما تستخدم اللغة العربية أنواعاً متعددة من العناصر الخارجية، مثل: النقاط، والهمزات، والمد، وعلامات التشكيل.
- صعوبة العثور على نصوص مترجمة من مجالات متنوعة، وأوعية متعددة بشكل كافٍ لتجميع مدونة متوازنة.
- قصور البرمجيات الخاصة بمعالجة اللغة العربية، مثل برامج المحاذاة، والترميز، والتنقيح، والمحطات الصرفية، وغيرها.
- قلة الدعم المادي والمعنوي.
- القيود المتعلقة بحقوق النشر.

(النظر Al-Ajmi, 2011، العصيمي 2013، صالح 2012)

Corpus	Organization	Size	Purpose	Content	Availability
<b>The English–Arabic Parallel Corpus of United Nations Texts EAPCOUNT</b>	Carthage University	5,392,491 words	A research tool for applied and theoretical linguistic research	Resolutions and annual reports issued by different UN organizations and institutions	Available through: <a href="http://conferences.unite.un.org/UNCorpus">http://conferences.unite.un.org/UNCorpus</a>
<b>The Open Parallel Corpus (OPUS)</b>	Uppsala University		To support research in machine translation	Web-collected texts	Available through: <a href="http://opus.lingfil.uu.se/index.php">http://opus.lingfil.uu.se/index.php</a>
<b>MultiUN</b>	Language Technology Lab at DFKI, Germany	300 million words	To support research in machine translation	United Nations' websites	Available through: <a href="http://www.euromatrixplus.net/multi-un/">http://www.euromatrixplus.net/multi-un/</a>

Corpus	Organization	Size	Purpose	Content	Availability
<b>European Union corpus (EuroMatrix)</b>	European Union	51 million words in total, with 1.5 million Arabic words	To support research in machine translation	Proceedings of the European Parliament	Available through: <a href="http://www.euromatrix.net/">http://www.euromatrix.net/</a>
<b>Kuwait University English–Arabic Parallel Corpus</b>	Kuwait University	3 million words	Teaching, translation, and lexicography	Publications from Kuwait National Council	Only available for Kuwait University staff and students
<b>AMARA</b>	Qatar Computing Research Institute	2.6 million Arabic words; 3.9 million English	To support research in machine translation	Subtitles from educational videos	Available through: <a href="https://amara.org/en/">https://amara.org/en/</a>



Corpus	Organization	Size	Purpose	Content	Availability
Arabic–English Parallel corpus	Bangor University	27.8 million Arabic words; 30.8 million English words	To support and test machine translation systems	Translated texts from <i>Al-Hayat</i> newspaper and OPUS	Artificial Intelligence and Intelligent Agents research group, w.j.teahan@bangor.ac.uk.
Hadith Corpus	Leeds University	2 million words	Religious studies	Book of Sahih Albukhari	Not yet available

# Linguistic Data Consortium (LDC)

المدونة	عدد الكلمات	الرسوم بالدولار
ISI Arabic/English automatically extracted parallel text	31 million	4,000
Uma Arabic/English parallel news text	2 million	3,000
Arabic/English parallel translation	42,000	3,000
Multiple translation Arabic (part 1)	23,000	1,000
Multiple translation Arabic (part 2)	15,000	1,000
Arabic newswire English translation collection	551,000	1,500
Arabic news translation text (part 1)	441,000	3,000
GALE phase 1 Arabic broadcast news parallel text (part 1)	90,000	1,500
IGALE phase 1 Arabic broadcast news parallel text (part 2)	56,000	1,500
UN bidirectional multilingual	1 million	4,000

# بعض أوجه القصور:

- غير متوفرة للطلاب والباحثين
- تتطلب رسوماً عالية
- بعضها يُعدّ متخصصاً، ولا يغطي موضوعات مختلفة،  
أو أوعية نشر متعددة
- اعتماد بعضها على الترجمة الآلية
- اعتماد بعضها على المعالجة الآلية بشكل كامل
- لم تصمم أساساً لهدف تعليمي

# مشروع المدونة العربية - الإنجليزية (AEPC)

## ● الهدف:

- تصميم مدونة متوازنة للنصوص الإنجليزية المترجمة للغة العربية للاستفادة منها كمصدر لتعليم اللغة وتدريب المترجمين

## ● الشروط:

- أن تكون ذات حجم ملائم (Size)
- الشمول (Representativeness)
- التوازن (Balance)
- الجودة (Quality)
- الإتاحة (Availability)

# المرحلة الأولى

## • إعداد النصوص :

- جمع النصوص المترجمة في مختلف المجالات، والمتوفرة؛ إما إلكترونياً على الشبكة العنكبوتية، أو المطبوعة
- تصنيفها وفق الفئات التالية:
  - اجتماعي - Social
  - سيرة ذاتية - Biography
  - أدبي - Literary
  - إداري - Administrative
  - طبي - Medical
  - قانوني - Legal
  - ديني - Religious
  - سياسي - Political
  - علمي - Scientific

# المرحلة الأولى

## • إعداد الملفات

تحتوي المدونة على ملفات بصيغة (XML)، تتضمن البيانات الوصفية (Metadata) التالية باللغتين العربية والإنجليزية:

- Title - العنوان
- Author - المؤلف
- Publisher - الناشر
- Year of Publication - سنة النشر
- Country – بلد النشر
- Author Gender – جنس المؤلف
- Medium - الوعاء
- Domain – المجال العام
- Topic – المجال الفرعي

# المرحلة الأولى

- **التجزئة (Segmentation):**  
تقسيم النصوص إلى مقاطع قصيرة على مستوى الجملة
- **المحاذاة (Alignment):**
  - ربط مقاطع النص الأصلي مع ترجماتها.

# الواجهة والأدوات:



Arabic-English Parallel Corpus

SEARCH  
ابحث

ABOUT  
عن المشروع

CONTACT US  
تواصل معنا

عدد الكلمات الكلي: 204117  
The total number of words : 204117

العدد الكلي للنصوص: 7877  
The total number of texts : 7877

Search results for " قال "

قال

Domain Selected domain Year Choose year Country Selected country

Topic Selected topic Medium Selected medium Author gender Choose gender

Search

نتائج بحث - (23) - search results

In English	In Arabic	From
If mental health can be defined, as author M. Scott Peck says, as a commitment to reality no matter what the cost	إذا عرفنا الصحة النفسية مثل ما قال الكاتب إم سكوت بيك "إن تعهد الحقيقة لا يهم ما يكلف"	Facing Addiction More info ?
Step 2: I wanted him to imagine that someone he loved said that same negative thing about himself or Herself.	الخطوة 2 : أردت منه أن يتخيل أن شخص ما يحبه قال عن نفسه ذات الأمور السلبية	The Life You Want More info ?
"So far, so good," he replied.	نناجحة حتى الآن" قال مجيباً	The Five Languages of



# المرحلة الثانية

- تحالف بحثي مع كلية الحاسب والمعلومات بجامعة الملك سعود
- تصميم أداة لعرض المتلازمات اللفظية (Colocations).
- تصميم أدوات التحليل الإحصائي الرسومي، graphical statistical analysis
- مزيد من خيارات التصفية لعرض نتائج أكثر دقة
- توسيم النصوص بحسب نوع الكلمة ( Part-of-speech tagging )

# خاتمة



Arabic-English Parallel Corpus

# المراجع

- Aijmer, K. (Ed.). (2009). *Corpora and language teaching* (Vol. 33). Amsterdam and Philadelphia: John Benjamins Publishing.
- Al-Ajmi, H. (2003). Compiling an English-Arabic parallel text corpus. In M. Murata, S. Yamada & Y. Tono (Eds), *Proceedings of Asian Association for Lexicography* (pp.51–54).
- Al-Ajmi, H. (2004). A new English–Arabic parallel text corpus for lexicographic applications. *Lexikos*, 14(1).
- Alkahtani, S., and Teahan, W. J. (2015, December). A New Parallel Corpus of Arabic/English. In *Proceedings of the Eighth Saudi Students Conference in the UK* (p. 279). World Scientific.
- Aston, G. (1999). Corpus use and learning to translate. *Textus*, 12(2), pp. 289–314.
- Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, and E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair*. Amsterdam and Philadelphia: John Benjamins, pp. 233-250.
- Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. *Target*, 7(2), 223-243.
- Baker, M. (1998). *Routledge encyclopedia of translation studies*. London: Routledge: 74-76, 244-248.
- Baker, M. (1999). The role of corpora in investigating the linguistic behaviour of professional translators. *International Journal of Corpus Linguistics*, 4(2), 281-298.
- Barlow, M. (1996). Parallel texts in language teaching. In S. Botley, J. Glass, T. McEnery, & A. Wilson (Eds.), *Proceedings of teaching and language corpora 1996* (pp. 45-56). Lancaster, UK: UCREL Technical Papers Volume 9.
- Bernardini, S. (2004). Corpus-aided language pedagogy for translation education. In K. Malmkjær (Ed.), *Translation in undergraduate degree programmes* (pp. 97-111). Amsterdam and Philadelphia: John Benjamins Publishing.
- Bernardini, S. and Castagnoli, S. (2008). Corpora for translator education and translation practice. In E. Yuste-Rodrigo (Ed.), *Topics in language resources for translation and localization* (pp. 39-55). Amsterdam and Philadelphia: Benjamins Publishing.
- Biber, D. (1993). Representativeness in corpus design. *Literary and linguistic computing*, 8(4), 243-257.
- Botley, S., Glass, J., McEnery, A. and Wilson, A. (1996). *Proceedings of TALC 96*, Lancaster, UK: UCREL Technical Papers Volume 9.
- Botley, S. P., McEnery, A. M., and Wilson, A. (Eds.). (2000). *Multilingual corpora in teaching and research*. Amsterdam-Atlanta, G.A.: Rodopi
- Bowker, L., and Pearson, J. (2002). *Working with specialized language: a practical guide to using corpora*. London: Routledge.
- Brown, P. F., Pietra, V. J. D., Pietra, S. A. D., and Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. *Computational linguistics*, 19(2), 263-311.
- Brown, P. F., Lai, J. C., and Mercer, R. L. (1991, June). Aligning sentences in parallel corpora. In *Proceedings of the 29th annual meeting on Association for Computational Linguistics* (pp. 169-176). Association for Computational Linguistics.
- Chujo, K., Anthony, L., and Oghigian, K. (2009). DDL for the EFL classroom: Effective uses of a Japanese-English parallel corpus and the development of a learner-friendly, online parallel concordancer. In *Proceedings of the corpus linguistics conference, CL2009. Liverpool University of Liverpool*.
- Cohn, T. and Lapata, M., (2007). Machine translation by triangulation: Making effective use of multi-parallel corpora. In *Proceedings of 45th Annual Meeting of the Association for Computational Linguistics*, pages 828-735.
- Davies, G., Otto, S. and Rüschoff, B. (2013). Historical perspectives on CALL. In: M. Thomas, H. Reinders & M. Warschauer (a c. di.), *Contemporary computer-assisted language learning* (pp. 19-38). Londra: Bloomsbury.
- Eisele, A., and Chen, Y. (2010). MultiUN: A Multilingual Corpus from United Nation Documents. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC)*. European Language Resources Association (ELRA), pp. 2868–2872.
- Flowerdew, L. (2015). Data-driven learning and language learning theories: Whither the twain shall meet. In A. Leńko-Szymańska and A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 15–36). Amsterdam: John Benjamins.
- Frankenberg-García, A. (2015). Training translators to use corpora hands-on: challenges and reactions by a group of thirteen students at a UK university. *Corpora*, 10(3),