RLRL 2023 *Spotlight Talk*
5 September 2023

بارق: المكنز العربي المتوازن لتقييم الانقرائية

BAREC: The Balanced Arabic Readability Evaluation Corpus

Nizar Habash
New York University Abu Dhabi

Hanada Taha
Zayed University

BAREC Project

- The project is a collaboration between **New York University Abu Dhabi's Camel Lab** and **Zayed University's ZAI Arabic Language Research Center**.

- The 2.5 years effort is funded by the Abu Dhabi Arabic Language Center (ALC), starting September 2023.

- The overarching objective of the BAREC project is to develop a comprehensive reference resource to facilitate the study and evaluation of Arabic readability across the Arab world.

- http://barec.camel-lab.com/

SAMER Project Lexicon (Al Khalil & Habash): 36,000 lexical entries

Level	Grade	Age	Examples
Level I	Grade 1	6	بَيْت، شَجَرَة، أَرْنَب، أَزْرَق، كَبير، صَنَعَ، أَكَلَ، فَرِحَ، عَلى، لكِن house, tree, rabbit, blue, big, to make, to eat, to be happy, on, but
Level II	Grades 2-3	7-8	جَزيرة، ذَهَب، سَنة، داكِن، أُسْطُواني، صَعْب، خَدَعَ، كافأَ، قُرْب، إذا island, gold, year, dark, cylindrical, difficult, to cheat, to reward, near, if
Level III	Grade 4-5	9-10	رِئة، مُتْحَف، مُعادَلة، مُمْكِن، مُوَحَّد، أَغْرى، نَدُرَ، لَدى، كَيْ، ما إنْ...حَتّى lung, museum, equation, possible, united, to entice, to be rare, with, for, no sooner... than...
Level IV	Grades 6-8	11-14	اِقْتِصاد، نُسْغ، طُمَأْنينة، راقٍ، مُثْبَت، نَكَثَ، أَغْضى، إبّانَ، إنَّما، لَدُنْ economy, sap, tranquility, sophisticated, proven, to breach, to overlook, during, whereas, if (were)
Level V	Specialist	15 -	أَدَمة، قَسْطَرة، هَيْضة، مِطْياف، لَوْذَعي، شُعَبيّ، لَحا، طَعَنَ، لَدُنْ، أَنّى epidermis, catheterization, cholera, spectroscope, witty, bronchial, to denounce, to depart, with (≈ *chez* in French), wherever

دليل عربي 21

دليل إلكتروني للكتب المصنّفة
وفق معايير تصنيف عربي 21
وهنادا طه

تم تصنيف أكثر
من ٩٠٠٠ كتاب!

9,000+ books
classified!

	Hanada Taha Level	Grade	ACTFL	SAMER	Age
1	أ	1	مبتدئ أدنى	I	6
2	ب	1	مبتدئ أدنى	I	6
3	ج	1	مبتدئ متوسط	I	6
4	د	1-2	مبتدئ متوسط	I-II	6-7
5	هـ	1-2	مبتدئ أعلى	I-II	6-7
6	و	1-2	مبتدئ أعلى	I-II	6-7
7	ز	2	متوسط أدنى	II	7
8	ح	2-3	متوسط أدنى	II	7-8
9	ط	2-3	متوسط أوسط	II	7-8
10	ي	3	متوسط أوسط	II	7-8
11	ك	3-4	متوسط أعلى	II-III	8-9
12	ل	4	متقدم أدنى	III	9
13	م	4-5	متقدم أوسط	III	9-10
14	ن	4-6	متقدم أعلى	III-IV	10-11
15	س	6-8	متقن أدنى	IV	11-13
16	ع	6-8	متقن أوسط	IV	11-13
17	ف	7-8	متقن أعلى	IV	12-14
18	ص	9	متفوق	V	15
19	ق	10-12	متميز	V	16-18

The Hanada Taha Arabic Readability Leveling Scale

- 19 levels of Arabic **book** readability
- Detailed guidelines including many linguistic and extralinguistic features

	Level	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
	مستوى	أ	ب	ج	د	هـ	و	ز	ح	ط	ي	ك	ل	م	ن	س	ع	ف	ص	ق
رسم الحركات	تشكيل	كامل	كامل	كامل	كامل	كامل	كامل	معظم	معظم	معظم	معظم	بعض	بعض	نادر	نادر	بدون	بدون	بدون	بدون	بدون
رسم الحركات	ترقيم	لا	لا	لا	لا	نعم	نعم	نعم	نعم	نعم	نعم	نعم	نعم	نعم	نعم	نعم	نعم	نعم	نعم	نعم
عدد الكلمات	كلمة/نص	0-20	15-30	30-40	40-100	40-100	80-120	8-120	8-120	100-300	300-400	300-800	300-800	450-1000	1000-1600	2k-5k	10k-25k	30k-50k	40k-60k	60k+
عدد الكلمات	كلمة/صفحة	0-1	2-4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
عدد الكلمات	كلمة/جملة	0-1	2-4	3-8	3-8	5-8	5-8	5-11	6-13	6-13	8-16	10-25	10-25	10-25	*-40	-	-	-	-	-
عدد الكلمات	سطر/صفحة	-	-	1-2	1-2	2-3	2-3	3-6	3-7	3-8	3-8	8-14	8-16	15-25	-	-	-	-	-	-
الأصالة		تربوي أصيل	تربوي أصيل	تربوي أصيل	تربوي أصيل	تربوي أصيل	تربوي أصيل	تربوي أصيل	تربوي أصيل	أصيل أو مترجم	أصيل أو مترجم	أصيل أو مترجم	أصيل أو مترجم	أصيل أو مترجم	أصيل أو مترجم	أصيل أو مترجم	أصيل أو مترجم	أصيل أو مترجم	أصيل أو مترجم	أصيل أو مترجم
الفصاحة/المفردات	ثلاثي/رباعي	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
الفصاحة/المفردات	مقطع/مقطعين	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
الفصاحة/المفردات	خماسي/سداسي	-	-	-	-	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-
الفصاحة/المفردات	كلمات بصرية	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	-	-
الفصاحة/المفردات	فصيحة / عامية	فصيحة قريبة من العامية	فصيحة قريبة من العامية	فصيحة قريبة من العامية	فصيحة قريبة من العامية	فصيحة قريبة من العامية	فصيحة قريبة من العامية	فصيحة قريبة من العامية	فصيحة قريبة من العامية	فصيحة قريبة من العامية	فصيحة	فصيحة	فصيحة	-	-	-	-	-	-	-
الفصاحة/المفردات	تكرار	عالي	عالي	عالي	عالي	بعض	بعض	بعض	بعض	-	-	-	-	-	-	-	-	-	-	-
الفصاحة/المفردات	كلمات جديدة/سياق ضروري	-	-	-	-	-	-	-	-	X	X	X	X	X	X	X	X	X	X	X
الرمزية		لا	لا	لا	لا	بسيط/متجاوز	بسيط/مدعوم	بسيط/مدعوم	بسيط/مدعوم	درجة/مستقل	درجة/مستقل	درجة/مستقل	درجة/مستقل	كثير/مستقل	كثير/مستقل	كثير/مستقل	كثير/مستقل	كثير/مستقل	كثير/مستقل	كثير/مستقل
النوع	مفاهيم	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
النوع	قصص	-	-	بسيطة	بسيطة	بسيطة	بسيطة	متنوعة	متنوعة	متنوعة	متنوعة	متنوعة	متنوعة	-	-	-	-	-	-	-
النوع	أناشيد	-	-	-	-	بسيطة	بسيطة	متنوعة	متنوعة	متنوعة	متنوعة	-	-	-	-	-	-	-	-	-
النوع	كتب معلومات	-	-	-	-	بسيطة	بسيطة	متنوعة	متنوعة	متنوعة	متنوعة	متنوعة	متنوعة							
النوع	قصص عائلية	-	-	-	-	-	-	متنوعة	متنوعة	متنوعة	متنوعة	-	-	-	-	-	-	-	-	-
النوع	مجلات	-	-	-	-	-	-	-	-	متنوعة	متنوعة	-	-	-	-	-	-	-	-	-
النوع	قصص ديني	-	-	-	-	-	-	-	-	متنوعة	متنوعة	-	-	-	-	-	-	-	-	-
النوع	قصص مصورة	-	-	-	-	-	-	-	-	متنوعة	متنوعة	-	-	-	-	-	-	-	-	-
النوع	قصص الغاز ومغامرات	-	-	-	-	-	-	-	-	-	-	متنوعة	متنوعة	-	-	-	-	-	-	-
النوع	كتب علمية	-	-	-	-	-	-	-	-	-	-	متنوعة	متنوعة	متنوعة	متنوعة	متنوعة	متنوعة	متنوعة	متنوعة	متنوعة
النوع	كتب أدبية	-	-	-	-	-	-	-	-	-	-	-	-	متنوعة	متنوعة	متنوعة	متنوعة	متنوعة	متنوعة	متنوعة
النوع	كتب تاريخية	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	متنوعة	متنوعة	متنوعة	متنوعة
النوع	كتب فلسفية	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	متنوعة
التراكيب	كلمة واحدة	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
التراكيب	تكرار	X	X	X	X	X	X	بعض	بعض	-	-	-	-	-	-	-	-	-	-	-
التراكيب	سجع	-	X	X	X	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-
التراكيب	حوار		-	-	بسيط	بسيط	بسيط	بسيط	بسيط	أكثر من واحد	أكثر من واحد	-	-	-	-	-	-	-	-	-
التراكيب	معلومة/مفهوم	-	-	-	-	-	-	أكثر من واحد	أكثر من واحد	أكثر من واحد	أكثر من واحد	متعدد	-	-	-	-	-	-	-	-
التراكيب	مبتدأ-خبر	-	X	X	X	X	X	X	X	X	X	بعض التعقيد	معقدة	-	-	-	-	-	-	-
التراكيب	اسم-صفة	-	-	X	X	X	X	X	X	X	X	بعض التعقيد	معقدة	-	-	-	-	-	-	-
التراكيب	فعل-فاعل-مفعول	-	-	X	X	X	X	X	X	X	X	بعض التعقيد	معقدة	-	-	-	-	-	-	-
التراكيب	جار-مجرور	-	-	-	X	X	X	X	X	X	X	بعض التعقيد	معقدة	-	-	-	-	-	-	-
التراكيب	ظروف/حال/إضافة	-	-	-	-	-	-	-	-	X	X	بعض التعقيد	معقدة	-	-	-	-	-	-	-
التراكيب	تشابيه/استعارات /مجازيات	-	-	-	-	-	-	-	-	X	X	X	X	X	X	X	X	X	X	X
التراكيب	تعقيد نحوي	-	-	-	-	-	-	-	-	-	-	-	-	X	X	X	X	X	X	X
التراكيب	تعقيد صرفي	-	-	-	-	-	-	-	-	-	-	-	-	X	X	X	X	X	X	X
التراكيب	تعقيد بلاغي	-	-	-	-	-	-	-	-	-	-	-	-	X	X	X	X	X	X	X

دليل عربي21

دليل إلكتروني للكتب المصنّفة
وفق معايير تصنيف عربي21
وهنادا طه

تم تصنيف أكثر
من ٩٠٠٠ كتاب!

9,000+ books
classified!

	Hanada Taha Level	Grade	ACTFL	SAMER	Age
1	أ	1	مبتدئ أدنى	I	6
2	ب	1	مبتدئ أدنى	I	6
3	ج	1	مبتدئ متوسط	I	6
4	د	1-2	مبتدئ متوسط	I-II	6-7
5	هـ	1-2	مبتدئ أعلى	I-II	6-7
6	و	1-2	مبتدئ أعلى	I-II	6-7
7	ز	2	متوسط أدنى	II	7
8	ح	2-3	متوسط أدنى	II	7-8
9	ط	2-3	متوسط أوسط	II	7-8
10	ي	3	متوسط أوسط	II	7-8
11	ك	3-4	متوسط أعلى	II-III	8-9
12	ل	4	متقدم أدنى	III	9
13	م	4-5	متقدم أوسط	III	9-10
14	ن	4-6	متقدم أعلى	III-IV	10-11
15	س	6-8	متقن أدنى	IV	11-13
16	ع	6-8	متقن أوسط	IV	11-13
17	ف	7-8	متقن أعلى	IV	12-14
18	ص	9	متفوق	V	15
19	ق	10-12	متميز	V	16-18

أ
نشر الملابس
واقعي
الناشر: دار كلمات
الرسام: هادي فاضل
السنة: 2018

ب
جدّي و جدّتي
ملاك عبيد
واقعي
الناشر: دار كلمات
الرسام: ريما الكوسا
السنة: 2018

ج
لا أستطيع النّومَ!
دار كلمات
واقعي
الناشر: ميثاء الخياط
الرسام: هادي فاضل
السنة: 2020

ه
أخجل من قراءتي
حنان طريق
واقعي
الناشر: دار أبجديات
الرسام: ندى بكر
السنة: 2022

ز
في العالم بيوت كثيرة
ري عبد العال
واقعي
الناشر: دار كلمات
الرسام: كريستل الحلّال
السنة: 2013

ط
مشكلة قطقوط
ميرفت البلّاجي
خيالي
الناشر: شركة ألف للاستشارات التعليميّة
الرسام: ندى بكر
السنة: 2022

ق
العصفور الأزرق و حكايات أخرى
ماري- كاثرين دونوغ - ترجمة: ماري طوق
خيالي
الناشر: دار الهدهد
الرسام: لا رسوم
السنة: 2017

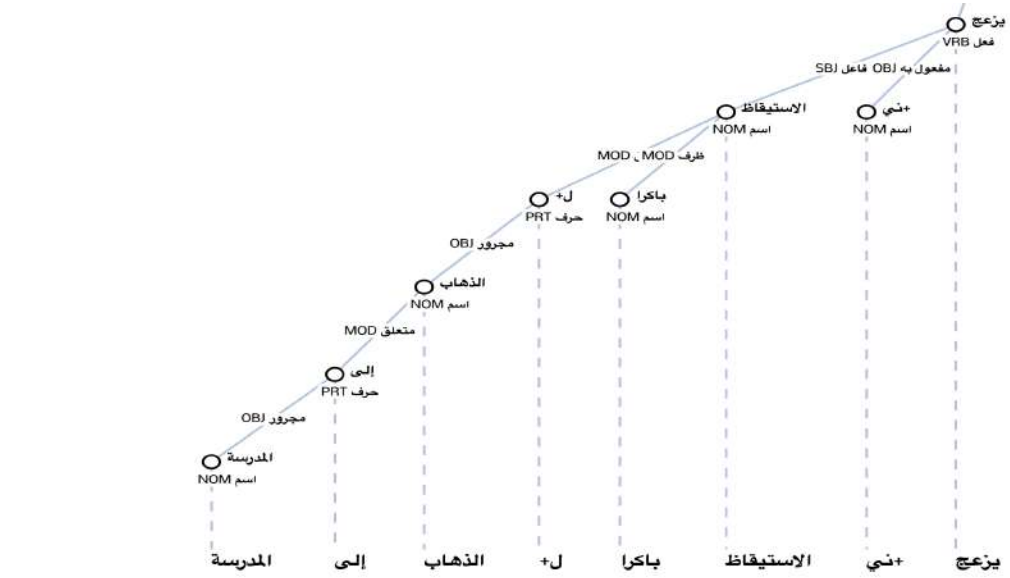
جامعة نيويورك أبوظبي

NYU | ABU DHABI

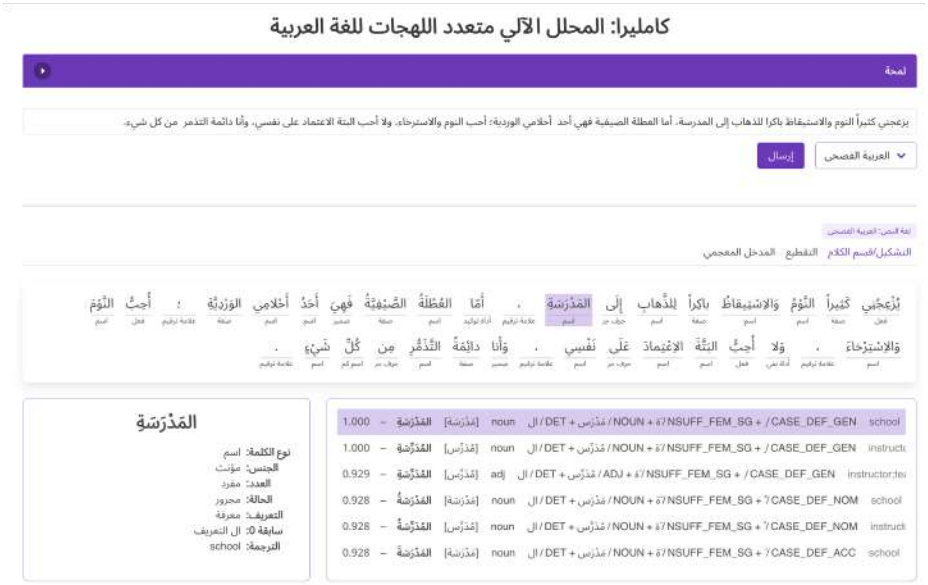
أدوات كامل

CAMeL Tools

https://github.com/CAMeL-Lab/camel_tools



Camel Treebank Style Analysis



http//:camelira.camel-lab.com/

BAREC Planned Deliverables

- A **10M word corpus** of text excerpts distributed across readability levels, genres, topics, and text origins.
  - Fair-use of text samples.
  - Simulate the wide range of read materials in the world.
- **Human annotations** for sub-samples of the 10M corpus
  - 1M words - sentence segmentations & sentence level readability leveling in the Taha system.
    - Guidelines to map the system from book-level to sentence-level.
  - 100K words – syntactic annotations in CATiB dependency tree style.
  - 40K lemmas – annotating a morphological lexicon with Taha's 19 levels.
- **Open-source AI tools**
  - Models for automatic readability using the annotated data.
  - Automatically annotate the 10M corpus.
  - Automatic readability reporting on new texts to help authors make informed decisions.

Q+A

- As this is a new project, may we ask you for…
  - Any feedback on the plan?
  - Pointers to resources?
  - Possible collaborations?
  - General advice from previous and similar experiences?
- Any questions you may have?