

مركز زاي لبحوث اللغة العربية ZAI Arabic Language Research Center







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.
- <u>http://barec.camel-lab.com/</u>





المحلث المخادغ

والدَّجاجةُ الحَمْراء

ست مان

العُبْرَات

الصِّغَار الثَّلاث

هايدى

مجموعة القراءة المتدرجة

همي برنامجُ صدرَ بإشرافِ معهدِ تعليم القراءةِ البريطانيُّ، ووفقَ معاييرهِ، وشروطِهِ. تتألُّف مجموعةً القراءةِ المتدرَّجةِ مِن ثمانيةِ مُستوياتِ (أربِع مراحل)، لكلُّ منها لونُ يحدُّدُ درجة صعوبيَّهِ من حيثٌ مضمونِ القصَّةِ، ومستواها اللُّفويَّ، وعددِ كلماتِها، وتنوُّع مفرداتِها. جدولُ الألوانِ المرفقُ يوضَّحُ تدرُّج مستوى القراءةِ في هذهِ المجموعة. صُمَّمَتْ هذهِ الكبِّ ضمنَ مجموعاتِ تهدفُ إله، تعليم اللُفةِ العرييَّةِ بالقراءةِ المتَّرِعةِ المُتحَرِّةِ.

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

-

اليس بلاد العجائب

Level	Grade	Age	Examples
Level I	Grade 1	6	بَيْت، مَحَبَرَة، أَزْنَب، أَزْرَق، كَبِير، صَنَعَ، أَكُلَ، فَرِحَ، عَلى، لكِن house, tree, rabbit, blue, big, to make, to eat, to be happy, on, but
Level II	Grades 2-3	7-8	جَزِيرة، ذَهَب، سَنة، داكِن، أُسْطُوانِي، صَعْب، خَدَعَ، كافأً، قُرْبَ، إذا island, gold, year, dark, cylindrical, difficult, to cheat, to reward, near, if
Level III	Grade 4-5	9-10	رِئَة، مُنْحَف، مُعادَلة، ثُمْنَكِن، مُوَحَّد، أَغْرى، نَدُرَ، لَدى، كَيْ، ما إِنْحَتّى lung, museum, equation, possible, united, to entice, to be rare, with, for, no sooner than
Level IV	Grades 6-8	11-14	اِقْتِصاد، نُسْخ، طُمَّانينة، راقِي، مُثْبَت، نَكَثَ، أَغْضى، إِبَّانَ، إَمَّا، لَنَنْ economy, sap, tranquility, sophisticated, proven, to breach, to overlook, during, whereas, if (were)
Level V	Specialist	15 -	أَدَمَة، قَسْطَرة، هَيْضة، مِطْياف، لَوْذَع، شُمَتِي، لَحَا، طَعَنَ، لَدُنْ، أَتَى epidermis, catheterization, cholera, spectroscope, witty, bronchial, to denounce, to depart, with (≈ <i>chez</i> in French), wherever

دليل عربي21 دليل إلكتروني للكتب المصنّفة وفق معايير تصنيف عربي21 وهنادا طه



تم تصنيف أكثر من ٩٠٠٠ كتاب! 9,000+books classified!



classified! Hanada Taha Grade ACTFL SAMER Age Level مبتدئ أدنى 1 1 Ι 1 6 2 مبتدئ أدنى Ļ 6 1 Ι 3 مبتدئ متوسط Ι 6 1 ٤ مبتدئ متوسط 4 ۵ 1-2 I-II 6-7 5 مبتدئ أعلى I-II 6-7 1-2 4 مبتدئ أعلى 6 I-II 6-7 و 1-2 7 j 2 متوسط أدنى 7 Π 8 متوسط أدنى τ 2-3 Π 7-8 9 Ł 2-3 متو سط أو سط Π 7-8 متوسط أوسط 10 3 Π 7-8 ي متوسط أعلى 11 ک II-III 3-4 8-9 12 متقدم أدنى 5 4 III 9 متقدم أوسط 13 4-5 Ш 9-10 ? 14 متقدم أعلى 4-6 III-IV 10-11 ن 15 متقن أدنى 11-13 س 6-8 IV متقن أوسط 16 ٤ 6-8 IV 11-13 17 متقن أعلى ف 7-8 IV 12-14 18 V 15 9 متفوق ص

متميز

V

16-18

19

ق

10-12

The Hanada Taha Arabic Readability Leveling Scale

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

	Level مستوی	1	2 ب	3 E	4 3	5	6 و	7 ز	8 2	9 ط	10 ي	11 ك	12 J	13 P	14 ن	15 س	16 E	17 ف	18 ص	19 ق
		کامل	کامل	کامل	كانل	كامل														
رسم الحركات رسم الحركات		الحاصل لا	کامل لا	حاصل لا	ي لا	تعم	کامل تعم	معظم تعم	معظم نعم	معظم نعم	معظم تعم	يعشن تعم	يعلس نعم	ٽابر نعم	ٽاڊر نعم	بدون تعم	يدون نعم	يدون تعم	يدون نعم	ون هم
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-	2k-5k	10k-25k	30k- 50k	40k- 60k	60k
عدد الكلمات	كلمة/منفحة	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	100	1-2	1-2	2-3	2-3	3-6	3-7	3-8	3-8	8-14	8-16	15-25			<u>141</u>	2	194	4
الأميالة		تريوي أصيل	تريوي اسيل	ٽريوي آمسيل	تريوي أمىيل	تريوي امىيل	تريوي أصيل	تريوي اسيل	ٽربوي اصيل	اصيل أو مترجم	اصيل او مثرجم	اصپل او مثرجم	اصیل او مثرجم	اصيل أو مترجم	اصيل او مترجم	اصيل أو مثرجم	اصيل او مترجم	اصیل او مترجم	اصيل او مترجم	يل أو رجم
الفصاحة/ المقردات	ثلاثي/رياعي	х	x	x	x	- 2	<u>.</u>	S.	12	2	2	1	- 19 -	<u>_</u>	<u></u>	S.		12	- 22	2
القصباحة/									100	-										
اللغردات القصاحة/	مقطع/مقطعين خماسمي/	x	х	x	x	•				-	•		-	•			•			
المقردات	سداستي	(2)	<u>i</u>	1(2)	- 23	х	x	х	х	32	<u>\$</u>	20	1997	-	192	- 25	<u>,</u>	8	-34	1
الفصاحة/ اللغردات	كلبات بصرية	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x			×
القصاحة/	فميمة /	فصيحة قريبة من				فصيحة قريبة من				فصيحة قريبة من										
المقردات القصاحة/	عامية	العامية	العامية	العامية	العامية	العامية	العامية	العامية	العامية	العامية	فمبيحة	فميحة	فميعة	.*.			*	*	4	
المقردات الفصاحة/	تکرار	مالي	مالي	عالي	عالي	يعض	يعض	يعض	يعض	8		÷.		1		3	(*)	8	28	1
القصاحة/ المقردات	کلمات جدیدة/ سیاق ضروری			1.00					1.04	x	x	x	x	x	x	x	x	x	x	x
		-				11		11	0	15		15	6	1.00	15	1.5	LK	1.15		
الرمزية		×	¥	¥	۲	بسيط/ متجاوز	بسيط/ مدعوم	بسيط/ مدعوم	يسيط/ مدعوم	درجة/ مستقل	درجة/ مىنتقل	درجة/ مستقل	درجة/ مستقل	کٹیر / مستقل	کٹیر/ مستقل	کٹیر/ مستقل	کٹیر/ مستقل	کٹیر/ مستقل	کٹیر/ مستقل	ير/ ستقل
النوع	مقاهيم	х	Х		- 81					10	÷.	÷		*		÷		8	10	
التوع	تمنمن			يسيطة	يسيطة	يسيطة	يسيطة	متتوعة	متتوعة	متتوعة	مثنوعة	متتوعة	متنوعة			-		÷.		
التوع	أناشيد	÷.	- 64			بسيطة	بسيطة	متنوعة	متتوعة	متتوعة	متنوعة						*	÷.	34	14
النوع	كآتب معلومات	17	12	1.00	59	بسيطة	بسيطة	متتوعة	متتوعة	متتوعة	متنوعة	متتوعة	متنوعة							
التوع	قصحن عائلية		- 24	- 68 S	- 20	-		مثنوعة	متتوعة	ملتوعة	متنوعة		20		(2)			×.	- 34	
النوع	مجلات									متتوعة	متنوعة	*	0.00			*			.*	1
التوع النوع	قصص ديڏي قصص مصورة		1.1		-					ملتوعة ملتوعة	مثنوعة مثنوعة	-		-		-				
	قصبص الغاز										- albert									
النوع	ومغامرات	18	- CR	1.00			18	- CP			- 81	متتوعة	متنوعة	10	340	1		8	34	1.28
التوع	كتب علمية		4		1			-	•	-		متتوعة	متنوعة	ملتوعة	متنوعة	متتوعة	متنوعة	متتوعة	متنوعة	ترمة
النوع	كنب أدبية	18	3 9	1.60	81	(4)	78	- 24	1.00		(4)	4.		ملتوعة	متنوعة	متتوعة	متنوعة	متتوعة	ملتوعة	وعة
التوع	كتب تاريخية	13	1.4	. •													متنوعة	متتوعة	متثوعة	فوعة
التوع	كتب فلسفية	18	1.4		÷.		18				-		(#)			÷			1	ics
التراكيب	كلمة واحدة	Х	24	1.1	1.12	1.2	12	10		<u>_</u>		- 21	1423						12	
التراكيب	تكرار	x	x	x	x	x	x	بعض	يعفن									*1		
التراكيب	سجع	12	x	x	x	x	x	x	x	12	- A	22	127	1	121			1.1	1	1
	53 - 56									اکثر من	أكثرمن									
التراكيب	حوار		- 24	1.40	pint	يسيط	بسيط	يسيط	يسيط	واحد	واحد	-2	1.00	+		*	90	×2.		-
التراكيب	معلومة/مقهوم	24	a		8	×	28	أكثر من واحد	اکثر من واحد	اکثار من واحد	أكثر من واحد	متعدد	200			æ	8		34	
التراكيب	میتدا-خبر		x	x	x	x	x	x	x	x	x	بعض التعليد	معقدة		1283	*				
التراكيب	اسمحمقة	17		x	x	x	x	x	x	x	x	بعض التعقيد	معقدة					-		
التراكيب	فعل فاعل مقعول	- S.	14	x	x	x	x	x	x	x	x	بعش التعقيد	معقدة	3	(e);	8	1	13	14	3
التراكيب	جار -مجرور	10	5.		x	x	x	x	x	x	x	بعض التعقيد	معقدة	4	828	ц.		29	<u>نة</u>	4
	ظروف/حال/إضنافة	52	5	1943	21	*	- 52	5	1.43	x	x	بعض التعقيد	معقوة		1942	- 12			-	
التراكيب	تشابيه/استعارات			1040	-		14			x	x	x	x	x	x	x	x	x	x	x
	-dollars/				11 22		25		1.50	•	~									
التراكيب	/مجازيات ثعقبد تحدي			-		-					-			x	x	x	x	X	X	
	/مجازيات تعقيد تحوي تعقيد مىرفى	1	1		5	-	1	1		2	-	1		x	x	x	x	x	x	x











الناشر: ميتاء الخياط الرسام: فاحي فاضل السنة: 2020



مُشْكِلَةُ قَطْقوط

جدی وجد تی



Ь

ميرفت البلتاجي

الرسام: ندى بكر السلة: 2022

فبالى

مشكلة قطقوط

الناشر: شركة أنف للاستشارات التعليميّة

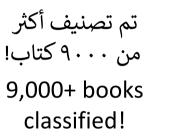
Ļ

واقعى

الناشر: دار کلمات الرسام: ريما الكوسا الستة: 2018

جدّي و جدّتي







	Hanada Taha Level	Grade	ACTFL	SAMER	Age 6		
1	i	1	مبتدئ أدنى	Ι			
2	ų	1	مبتدئ أدنى	Ι	6		
3	٦	1	مبتدئ متوسط	Ι	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	j	2	متوسط أدنى	II	7		
8	2	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	শ্র	3-4	متوسط أعلى	II-III	8-9		
12	ل	4	متقدم أدنى	III	9		
13	٩	4-5	متقدم أوسط	III	9-10		
14	ů	4-6	متقدم أعلى	III-IV	10-11		
15	س	6-8	متقن أدنى	IV	11-13		
16	٤	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		

دلیل عربی21

دليل إلكتروني للكتب المصنّفة

وفق معايير تصنيف عربي21



في العالم بيوت كثيرة ي عبد العال Gadig الناشرة دار كلفات الرسام: خريستل الحلّدل السنة: 2013



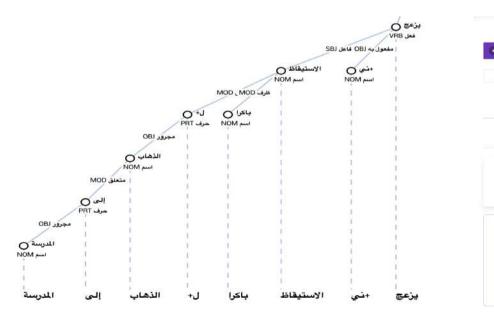
5

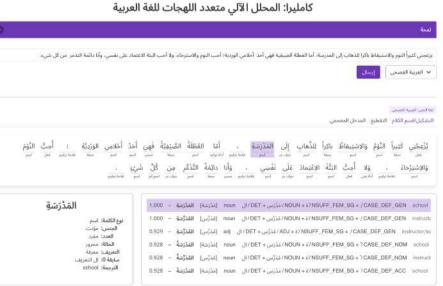






https://github.com/CAMeL-Lab/camel_tools





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

Camel Treebank Style Analysis

BAREC Planned Deliverables

- A 10M word corpus of text excerpts distributed across readability levels, genres, topics, and text origins.
 - Fair-use of text samples.
 - $_{\odot}~$ 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.
 - \circ $\,$ 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...
 - $_{\circ}$ Any feedback on the plan?
 - Pointers to resources?
 - Possible collaborations?
 - General advice from previous and similar experiences?
- $_{\circ}~$ Any questions you may have?

