Oxford Arabic Corpus

Pete Whitelock, Tressy Arts

Oxford University Press

Content

Source	Million words
Al-Ahram	37
Asharq Al-Awsat	184
Agence France Presse	42
Assabah	14
Al Hayat	209
An Nahar	253
Al-Quds Al-Arabi	19
Ummah Press	3
Xinhua News Agency	86
Arabic Writers Union of Damascus	10
Total	859

Analysis -MADA

- Nizar Habash et al., CADIM: Columbia's Arabic Dialect Modeling Group
- Buckwalter analyser
  + SRI Language Modeling
  + Support Vector Machine (ML)
- Functions
  - Tokenization
  - Vocalization
  - POS Tagging
  - Lemmatization

- ;;; SENTENCE AlnHr AlErAqy syAsy lA TA}fy !

- ;;WORD AlnHr

- ;;MADA: AlnHr asp:na cas:a enc0:0 gen:m mod:na num:s per:na pos:noun prc0:Al_det prc1:0 prc2:0 prc3:0 stt:d vox:na

- *1.004898 diac:Aln~aHora lex:naHor_1 bw:Al/DET+naHor/NOUN+a/CASE_DEF_ACC gloss:slaughtering;butchering pos:noun prc3:0 prc2:0 prc1:0 prc0:Al_det per:na asp:na vox:na mod:na gen:m num:s stt:d cas:a enc0:0 rat:y source:lex stem:naHor stemcat:N

- ^1.001650 diac:Aln~aHora lex:naHor_2 bw:Al/DET+naHor/NOUN+a/CASE_DEF_ACC gloss:throat pos:noun prc3:0 prc2:0 prc1:0 prc0:Al_det per:na asp:na vox:na mod:na gen:m num:s stt:d cas:a enc0:0 rat:y source:lex stem:naHor stemcat:N

- ;;WORD AlErAqy

- ;;MADA: AlErAqy asp:na cas:g enc0:0 gen:m mod:na num:s per:na pos:adj prc0:Al_det prc1:0 prc2:0 prc3:0 stt:d vox:na

- *1.006966 diac:AlEirAqiy~i lex:EirAqiy~_1 bw:Al/DET+EirAqiy~/ADJ+i/CASE_DEF_GEN gloss:Iraqi pos:adj prc3:0 prc2:0 prc1:0 prc0:Al_det per:na asp:na vox:na mod:na gen:m num:s stt:d cas:g enc0:0 rat:y source:lex stem:EirAqiy~ stemcat:Nall

Input to Sketch Engine

- <s>
- النحر   النَّحْرَ   النحر   nنَحْر-   NNAxCaGmMxNsPxSdVx   slaughtering;butchering
- العراقي   العِراقيِّ   aعِراقِيّ-   AJAxCgGmMxNsPxSdVx   Iraqi
- سياسيٌسِياسِيِّ   aسِياسِيّ-   AJAxCgGmMxNsPxSiVx   political
- لا   لا   pلا-   PNAxCxGxMxNxPxSxVx   no;not;non-
- طائِفِيٌ طائفي   aطائِفِيّ-   AJAxCgGmMxNsPxSiVx   sectarian;factional
- !   !   z-!   ZZAxCxGxMxNxPxSxVx   !
- </s>

- If you would like access to the corpus for research purposes, email me:

  pete.whitelock@oup.com

Appendix - Screenshots

(Oxford Arabic Corpus in
SketchEngine)

Sketch Engine

Search _____ in Help

User: tressy Corpus: Arabic Gigaword (plus AWU) Hits: 363954 [Search]_____ in Arabic Gigaword (plus AWU) ▾

Concordance
Word List
Word Sketch
Thesaurus
Sketch-Diff
? Help on main menu

? Help on Conc. menu
Save
View options
  KWIC/Sentence
Sort
  Left | Right
  Node
  References
  Shuffle
Sample
Filter
Frequency
  Node tags
  Node forms
  Doc IDs
Collocations

Page 1 of 18198 [Go] Next | Last

#20228 الانتقادات حدة تتعلق بحرية التعبير والاضطهاد المتواصل ل **كتاب**/NN **كتاب** لاهانتهم القومية التركية بموجب المادة 301 من قانون العقوبات

#29445 الذي بعث فينا رسولا من انفسنا يتلو علينا اياته ويعلمنا **كتاب**/NN **الكتاب** والحكمة . والحمد لله الذي ولي علينا من يقوم بالقسط ويحكم

#47984 الخيون الاربعاء 10 شوال 1427 ه 1 نوفمبر 2006 العدد 10200 <p> مصائر **كتاب**/NN **الكتب** والمكتبات هو المشهد غير المرئي من الماساة العراقية . لقد

#48022 مثل العراق لم يهدا منذ خمسينات القرن الماضي . لقد تماهت **كتاب**/NN **الكتب** مع مصائر اصحابها ، وكانت موضع ريبة واستهداف دائمين . ول

#48052 ذلك ، كانت عوائل المغضوب عليهم تلجا ، اول ما تلجا الي حرق **كتاب**/NN **الكتب** او دفنها عميقا في الارض . <p></p> وفي المنفي العراقي ، الذي امتد

#48076 الذي امتد طويلا ، نشات مشكلة اخري ، هي كيفية المحافظة علي **كتاب**/NN **الكتب** والمكتبات حين يضطر المنفي الي اختيار منفي اخر . في الوطن

#48103 المؤرخ عباس العزاوي ، وهو علي فراش الموت ، العمال يرمون ب **كتاب**/NN **كتب** من الشبابيك الي بطن سيارة مكشوفة ، وكانهم يرمون طابوقا

#48121 بطن سيارة مكشوفة ، وكانهم يرمون طابوقا (الاجر). <p></p> وبيعت **كتاب**/NN **كتب** الكاتب واللغوي هاشم الطعان الي جامعة الكوفة . وسريعا ما

#48146 باللوريات تحت نظرات افراد الاسرة الحزينة . وكانت المفاجاة ان **كتاب**/NN **الكتب** " فرهدت " قبل دخولها الي الجامعة ، وقسمت بين الاساتذة والمسؤولين

#48257 الدلي ، 93 عاما ، فما زال يبحث عن طريقة لحفظ ما في حوزته من **كتاب**/NN **كتب** ، ومن امانيه ان يعود الي العراق ، ويؤسس مكتبة في مسقط راسه

#48280 العراق ، ويؤسس مكتبة في مسقط راسه سوق الشيوخ ، يجمع فيها شتات **كتاب**/NN **كتب** . <p></p> وفي لندن ايضا ، يعيش المعاناة نفسها المعماري محمد

#48324 اربعينات القرن الماضي ، طريقُها الي مخازن باردة ، وفيها **كتاب**/NN **كتب** تعود الي القرن السابع عشر ، ومخطوطات عربية وفارسية ، ومجموعة

#51737 فيصل بن مشعل بن سعود بن عبد العزيز نائب امير منطقة القصيم ، **كتاب**/NN **كتابا** بعنوان " سر دوام النعم ". <p></p> ذ يشارك الدكتور احمد محمد علي

#51804 القران الكريم والسنة التابعة لرابطة العالم الاسلامي اخيرا **كتاب**/NN **كتابا** جديدا في موضوع الاعجاز العلمي بعنوان " المنح الالهية في

#54342 الوطنية للملكة . <p></p> ولم يغب عن ابناء وبنات نجران تفعيل دور **كتاب**/NN **الكتاب** وبيان اهميته في تحقيق التواصل الحضاري والتعريف بالامكانات

#60161 البلد " و " ليلة سقوط بغداد "، و " فرحان ملازم ادم "، و " ملك و **كتاب**/NN **كتابة** " السفارة في العمارة " و " عمارة يعقوبيان " و " دم الغزال

#60358 علي السبع وفاطمة الحسني ومحمد صيرفي . السيناريو والحوار **كتاب**/NN **كتب** المصري بلال فضل الذي يشارك في كتابة القصة ايضا مع الناقد

#61700 زعماء الثورة ، وجميع الوثائق والمخطوطات ، ومكتبة تضم ابرز **كتاب**/NN **الكتب** والبحوث والدراسات التي كتبت عن الثورة ، والافلام التسجيلية

#62366 ان ينصرف المدعوون الي تناول الاطباق التركية الشهيرة . <p></p> **كتاب**/NN **كتاب** يفضح " مطبخ " لجان تحكيم الجوائز الادبية الفرنسية هدايا

#62542 غراسيه "، عرض عليه سيارة علي سبيل الهدية ووعده باعادة طبع **كتاب**/NN **كتب** القديمة في حال اعطي صوته للرواية الصادرة عن هذه الدار ذاتها

Concordance
Word List
Word Sketch
Thesaurus
Sketch-Diff

? Help on main menu

Save
Change options
Turn on clustering
More data
Less data
Switch menu position

كِتاب *(noun)*   Arabic Gigaword (plus AWU) freq = 363954

Na* V	22101	1.5	Nn* V	14957	1.4	V Nn*	3861	2.1	Ng* N	91590	1.6	Ng N*	45146	0.8
أَصْدَرَ	1206	7.08	صَدَر	1419	7.78	صَدَر	280	5.51	هَذا	9641	5.52	تاريخ	2675	7.17
قَرَأَ	1185	8.75	كَانَ	1280	3.27	قَال	143	1.29	مَعْرِض	3316	8.35	طِفْل	748	5.92
وَجَّه	717	6.62	وَقَع	616	5.8	لَيْسَ	115	2.55	نَشْر	2045	7.36	واحِد	688	5.09
تَضَمَّن	675	6.66	قَرَأَ	608	7.91	ضَمَّ	106	3.61	صُدُور	1900	8.12	اللّه	651	5.45
كَانَ	600	2.17	قَال	450	2.93	حَمَل	90	3.21	دار	1789	7.29	سِيرَة	549	7.97
ضَمَّ	588	5.99	أَصْدَرَ	319	5.19	تَنَاوَل	71	3.93	عُنْوان	1708	7.48	اِسْتِقالَة	482	6.52
نَشَر	564	6.22	تَضَمَّن	318	5.62	أَصْدَرَ	70	3.06	مُؤَلَّف	1534	8.41	تُراث	439	6.66
قَدَّم	515	4.77	جَاء	252	4.26	تَضَمَّن	59	3.27	مُراجَعَة	1502	7.95	قِصَّة	426	6.47
وَضَع	493	5.4	تَنَاوَل	223	5.48	قَرَأَ	58	4.73	قِراءة	1418	7.79	رَقْم	388	4.97
أَرْسَل	443	6.6	أَتَى	188	4.38	قَدَّم	46	1.33	بَعْض	1322	4.89	تَكْلِيف	385	6.1
كَتَب	397	6.23	أَشَار	185	3.94	جَمَع	39	3.52	تَوْقِيع	1283	6.78	قِيمَة	367	4.31
وَقَع	385	5.09	بَدَأ	173	3.12	كَتَب	39	3.04	تَأْلِيف	1184	7.9	نَفْس	335	3.16
أَلَّف	358	8.3	عَرَض	172	4.77	نَشَر	37	2.41	صاحِب	1183	6.23	مُراسِل	330	6.02
تَنَاوَل	341	6.03	ضَمَّ	158	4.13	وَضَع	36	1.69	مُقَدَّم	1175	7.64	وَزِير	309	1.66
حَمَل	296	4.85	تَأَلَّف	141	6.7	اِعْتَبَر	35	1.07	أَلْف	1157	4.66	رَئِيس	295	0.78
اِعْتَبَر	291	4.07	قَدَّم	139	2.9	تَحَدَّث	34	1.92	كُلّ	1135	3.89	أَدَب	293	6.06
عَرَض	238	5.19	ذَكَر	132	2.87	مَثَّل	32	2.07	إصْدار	1100	7.0	كُلّ	290	1.95
تَلَقَّى	234	5.02	نَشَر	131	4.16	شَكَّل	29	1.18	أَحَد	1094	4.42	رِواية	271	5.7
لَيْسَ	225	3.48	تَحَدَّث	113	3.6	أَثَار	27	2.13	صَفْحَة	842	6.91	مُتَرْجِم	270	7.13
صَدَر	210	4.97	تَرْجَم	109	6.48	أَمْكَن	27	0.69	ثَلاث	780	4.11	ثَلاث	251	2.52
اِشْتَرَى	207	6.63	كَشَف	105	4.26	رَوَى	25	4.11	أَهْل	762	6.15	غَيْر	227	2.28

خَوْل-p N	1490	2.9
نَدْوَة	422	6.66
لِقاء	100	2.07
جِوار	36	1.13
حُضُور	35	1.36
نِقاش	23	3.12
مُلاحَظَة	17	3.26
مُناقَشَة	17	1.07
سُؤال	17	0.68
دِراسَة	15	0.13
مُحاضَرَة	14	3.52
مُداخَلَة	13	3.83
حَلْقَة	11	1.95
جَدَل	11	1.9
مُسْتَدِير	9	5.02

خَوْل-p N*	1710	3.3
دِفاع	125	2.61
حال	68	1.66
مَوْضُوع	62	1.65
تارِيخ	50	1.53
حُكْم	41	0.66
سِياسَة	29	0.18
إسْلام	18	1.7
سِيرَة	16	4.22
تَجْرِبَة	16	0.69
ثَقافَة	14	0.71
تَخْفِيف	12	1.62
فَنّ	12	1.24
أَدَب	11	1.84
مُساواة	9	2.57
كَيْفِيَّة	9	1.37

عَن-p N*	9490	4.0
دار	524	5.87
ةُ	354	0.92
تارِيخ	276	3.98
حَياة	226	3.3
سِيرَة	181	7.37
مَرْكَز	108	1.63
مَوْضُوع	104	2.38
تَجْرِبَة	101	3.31
حَرْب	99	1.37
فَنّ	81	3.93
إسْلام	80	3.79
عَلاقَة	64	0.6
ذٰلِك	60	0.03
مُؤَسَّسَة	53	1.1
رِحْلَة	47	2.69
وِزارَة	46	0.12
أَدَب	45	3.77
قَضِيَّة	43	0.52
شاعِر	42	3.38
أَمِير	40	2.57
مَنْشُور	37	5.4
سِياسَة	36	0.48
مَكْتَبَة	34	3.93
كِتاب	34	1.3

تَحْتَ-p N*	730	4.5
عُنْوان	425	5.96
رَقْم	19	0.79
شَمْس	16	3.41
طَبْع	13	2.85
إشْراف	11	1.28
شِعار	10	1.53

Adj N*	57927	1.5
جَدِيد	5519	7.58
مُقَدَّس	3478	10.21
مَدْرَسِيّ	2650	10.16
أَبْيَض	1716	8.14
أَوَّل	1557	5.96
عَرَبِيّ	1436	5.04
مَفْتُوح	1422	8.73
أَخِير	1367	6.51
صادِر	1362	8.38
آخَر	904	6.05
قَدِيم	834	7.48
دِينِيّ	732	7.13
شَهِير	619	8.05
رَسْمِيّ	520	5.44
مُهِمّ	493	6.4
كَثِير	453	5.95
سابِق	451	4.92
مِصْرِيّ	449	5.02
أَكْثَر	443	5.72
عِلْمِيّ	442	6.45
خاصّ	419	4.6
مَذْكُور	417	7.03
إلِكْتُرُونِيّ	414	6.86
ضَخْم	359	6.87

Concordance
Word List
Word Sketch
Thesaurus
Sketch-Diff

? Help on main menu

Change options

Switch menu position

كِتاب /رِوَايَة

Arabic Gigaword (plus AWU) freq = 363954/108385

| كِتاب | 6.0 | 4.0 | 2.0 | 0 | -2.0 | -4.0 | -6.0 | رِوَايَة |

V Nn*	3861	1905	2.1	3.7
دازَ	9	34	1.8	3.7
حَكَىَ	14	31	4.7	6.0
كَتَّبَ	39	51	3.0	3.4
رَوَىَ	25	20	4.1	3.9
رَصَدَ	16	12	4.4	4.1
تَرْجَمَ	18	9	4.3	3.4
حَمَلَ	90	25	3.2	1.4
قَرَأَ	58	14	4.7	2.7
تَناوَلَ	71	14	3.9	1.6
رَسَمَ	12	0	3.1	0.0
صَدَرَ	280	32	5.5	2.4
شَرَحَ	10	0	3.2	0.0
تَضَمَّنَ	59	0	3.3	0.0
جَمَعَ	39	0	3.5	0.0
ضَمَّ	106	0	3.6	0.0
عالَجَ	17	0	3.7	0.0
بيع	9	0	3.8	0.0
إحْتَوَىَ	21	0	4.0	0.0
حَوَىَ	10	0	4.0	0.0
طَبَعَ	9	0	4.0	0.0
أهْدَىَ	9	0	4.2	0.0

Ng* N	91590	27320	1.6	1.7
إخْتِلاق	0	75	0.0	6.3
بَطَل	45	1288	1.9	7.0
كِتابَة	107	959	4.0	7.6
حَسَب	244	1046	4.2	6.6
قِراءة	1418	404	7.8	6.5
غِلاف	328	49	6.6	5.2
تَرْجَمة	648	132	7.0	5.4
نَشْر	2045	517	7.4	5.6
صاحِب	1183	301	6.2	4.4
مُؤلِّف	1534	213	8.4	6.4
طَبْع	417	53	6.5	4.3
صَفْحَة	842	127	6.9	4.7
عُنْوان	1708	269	7.5	5.1
طِباعَة	283	16	6.5	3.7
مَضْمُون	485	41	6.6	3.8
صُدُور	1900	164	8.1	5.1
تَأْلِيف	1184	74	7.9	4.6
إصْدار	1100	57	7.0	3.1
بَيْع	727	28	6.3	1.9
مُقَدَّم	1175	37	7.6	3.2
تَأْقِيع	1283	29	6.8	1.6

Ng N*	45146	10182	0.8	0.6
وَلِيمَة	0	31	0.0	6.2
كاتِب	55	371	3.1	6.1
آيَة	42	58	4.3	5.8
كلاسيكيّ	82	78	5.5	6.6
خَيال	81	86	5.3	6.4
أَدِيب	85	70	5.3	5.9
قِصَّة	426	234	6.5	6.0
تَوْراة	137	25	6.4	5.6
مُتَرْجِم	270	49	7.1	5.8
سِيرَة	549	89	8.0	6.3
جَيْب	126	11	5.7	3.1
رواية	271	30	5.7	2.9
أَدَب	293	13	6.1	2.0
تاريخ	2675	103	7.2	2.6
طِفْل	748	20	5.9	0.8
اللّٰه	651	0	5.5	0.0
عتاريخ	63	0	5.5	0.0
طَبْخ	77	0	5.7	0.0
نَبيّ	117	0	5.7	0.0
شُكُر	148	0	5.8	0.0
مُراسِل	330	0	6.0	0.0

N بَيْنَ-p N*	840	224	0.5	0.4
دين	9	0	0.4	0.0
يَد	86	0	2.6	0.0
دَفَّة	40	0	7.4	0.0

N بَعْد-p N*	555	173	0.7	0.7
إِنْتِهاء	10	0	0.3	0.0
كِتاب	19	0	0.5	0.0
صُدُور	10	0	1.3	0.0
وَفاة	14	0	1.5	0.0

N إِلَى-p N*	9491	1136	2.0	0.8
سِينَما	0	17	0.0	3.9
أُخْرَى	0	16	0.0	3.7
إِنْكِلِيزِيّ	34	16	5.0	4.4
فَرَنْسِيّ	33	18	3.8	3.1
عَرَبِيّ	80	29	3.0	1.6
لُغَة	154	32	4.0	1.7
فَخامَة	9	0	3.0	0.0
حَرِيرِيّ	21	0	3.1	0.0
بَطْرِيَرْك	24	0	3.2	0.0
تِلْمِيذ	23	0	3.2	0.0
نَقِيب	25	0	3.6	0.0
أَمانَة	44	0	3.7	0.0
نِيابَة	54	0	3.7	0.0
أَنان	16	0	3.9	0.0
لَحُود	20	0	3.9	0.0
مُدَّعِي	20	0	4.2	0.0
بابا	22	0	4.3	0.0
مُسْتَأْجِر	12	0	4.4	0.0

Adj N*	57927	19083	1.5	1.8
بُولِيسِيّ	13	338	2.8	8.9
تَوْراتِيّ	15	84	3.0	7.1
واقِعِيّ	16	125	2.8	6.8
فائِز	21	87	3.4	6.8
تارِيخِيّ	316	732	5.9	7.5
أَدَبِيّ	276	91	6.8	6.1
شَهِير	619	171	8.1	7.2
مُصَوَّر	199	23	6.7	5.0
مُؤَرَّخ	192	16	6.7	4.6
مَذْكُور	417	60	7.0	4.9
جَدِيد	5519	1056	7.6	5.3
ضَخْم	359	34	6.9	4.2
قَدِيم	834	93	7.5	4.7
صادِر	1362	112	8.4	5.3
دِينِيّ	732	33	7.1	3.0
مَفْتُوح	1422	11	8.7	2.3
نادِر	244	0	6.8	0.0
إِلِكْترُونِيّ	414	0	6.9	0.0
مُوَجَّه	329	0	6.9	0.0
سَماوِيّ	254	0	7.0	0.0
مَطْبُوع	256	0	7.1	0.0
دِراسِيّ	358	0	7.2	0.0
أَبْيَض	1716	0	8.1	0.0
	2650	0	10.2	0.0