

OxfordLanguages

Using *OED*'s dictionary data to
study language variation: **a demo**

UKLVC12 September 2019

Catherine Sangster

Gary Leicester

OXFORD

The *Oxford English Dictionary* illustrates the **meaning, history, and pronunciation of 600,000 words**, past and present, from across the English-speaking world.

OED is currently exploring ways in which its wealth of data can be **made more accessible** to academic researchers.

Oxford**Languages**

— **Blog post:**

<https://public.oed.com/the-oed-and-research/>

— **Survey:**

<https://public.oed.com/blog/the-oed-supporting-research-through-innovation>

OXFORD

The available *OED* data include the following:

phonemic **transcriptions**

- **British English:** 276057, in 228363 groups, in 204453 entries (revised and unrevised)
- **American English:** 208208, in 168866 groups, in 148412 entries (revised and c. $\frac{3}{4}$ of unrevised)
- **12 World Englishes**, transcribed according to contemporary phonological models, for relevant subsets of headwords

linked **sound files**

- studio-quality **human-voice recordings** of all the above

Oxford**Languages**

Transcriptions are in *OED*-style IPA but are **readily transformable** to any preferred phonological model or transcription system (as we presented at BAAP 2018)

		British English	American English
OED		 /'hɑ:tɪd/	 /'hɑrdəd/
ODO		/'hɑ:tɪd/ 	/'hɑrdəd/ 
ALD		/'hɑ:tɪd/ 	/'hɑ:rtɪd/ 

Our **searcher** allows you to use **multiple parameters** to create a subset of data to **fit your research requirements:**

- word containing / not containing (string)
- transcription containing / not containing (string)
- variants
- number of pronunciations given
- word frequency (with granularity)
- date ranges (earliest / latest quotation evidence)
- parts of speech
- status of word within entry
- presence of World English pronunciations

... what else would be useful?

OxfordLanguages

Also available:

Historical pronunciation transcriptions
from Murray's original
New English Dictionary (1888-) and its
supplements, for headwords whose
pronunciation has changed significantly,
and headwords now deemed obsolete.

Transcriptions can be input using SAMPA
or IPA (using onscreen keyboard tool)

regular expressions work in all text fields

`^ $.*? []`

Our **demo** was available in the **poster area** throughout UKLVC12.

Illustrative examples of search results are on the next two slides.

Please email catherine.sangster@oup.com if you would like to try out the searcher, or if you have any questions or feedback.

OED data dated Wed Aug 28 10:20:36 2019

Let us know your thoughts - click here to fill in a short survey

British with or without audio

word does contain du

transcription does contain dju

syllabic marker nasal g ʃ ʒ θ ð ŋ ʧ ɪ ɪ ɛ œ æ ɜ ɘ ʌ ʊ ʊ ɔ ɒ ɑ

and is any of a group containing 1 - 16 pronunciations

and the variants don't contain dʒu

syllabic marker nasal g ʃ ʒ θ ð ŋ ʧ ɪ ɪ ɛ œ æ ɜ ɘ ʌ ʊ ʊ ɔ ɒ ɑ

Word frequency range: 0 - 8

Date range: 1 - 2020

parts of speech: n. v. adj. suffix prefix comb._form adv. int. prep. pron. connective phr. conj.

section in entry main xref infl sub variant

comma separated batch values: comma separated news values:

Limit to entries that also have World English pronunciations

Limit to where there are also unrevised-style pronunciations

results sort order database order ascending Search

#	entry	word	entry id	ph e:id	transcription	soundfile	position	variants
1	anthomedusa	anthomedusa	36081440	1174607030	,anθəmɪˈdjuːzə	anthomedusa#_gb_1	1 of 2	,anθəmɪˈdjuːzə
2	anthomedusa	anthomedusa	36081440	1174607040	,anθəmɪˈdjuːsə	anthomedusa#_gb_2	2 of 2	,anθəmɪˈdjuːzə
3	Anthomedusae	Anthomedusae	45532832	1209526990	,anθə(ʊ)mɪˈdjuːziː	anthomedusae#1_gb_1	1 of 1	-
4	cubomedusa	cubomedusa	268378	112498678	,kjuːbəʊmɪˈdjuːzə	cubomedusa#_gb_1	1 of 1	-
5	Cubomedusae	Cubomedusae	269398	117667538	,kjuːbəʊmɪˈdjuːziː	cubomedusae#1_gb_1	1 of 1	-
6	cubomedusan	cubomedusan	268380	116709342	,kjuːbəʊmɪˈdjuːzən	cubomedusan#_gb_1	1 of 1	-
7	dude	dude	58231	243190324	djuːd	dude#_gb_2	1 of 1	-
8	full due	full due	40229305	1209908140	,fʊlˈdjuː	full_due#1_gb_1	1 of 1	-
9	light duty	light duty	83535955	1278180680	ˈlaɪtˌdjuːti	light_duty#1_gb_3	1 of 1	-
10	May duke	May duke	115301	37914974	ˈmeɪˌdjuːk	may_duke#1_gb_1	1 of 1	-
11	medusiferous	medusiferous	115815	37570785	,medjuːˈsɪf(ə)rəs	medusiferous#_gb_1	1 of 1	-
12	mesoduodenum	mesoduodenum	233691	13564227	,mesə(ʊ)djuːəˈdiːnəm	mesoduodenum#_gb_1	1 of 1	-
13	nomen dubium	nomen dubium	256941	11200491	,nəʊmənˈdjuːbɪəm	nomen_dubium#1_gb_2	1 of 2	,nəʊmənˈdjuːbɪəm
14	nomen dubium	nomen dubium	256941	11200492	,nəʊmənˈdjuːbɪəm	nomen_dubium#1_gb_1	2 of 2	,nəʊmənˈdjuːbɪəm
15	nomen dubium	nomina dubia	256941	11200502	,nɒmɪnəˈdjuːbɪə	nomina_dubia#1_gb_1	1 of 1	-
16	non-scheduled	non-scheduled	128090	34139590	nɒnˈʃedjuːld	non_scheduled#1_gb_2	2 of 6	nɒnˈʃedjəld ▼
17	non-scheduled	non-scheduled	128090	34139593	nɒnˈskedjuːld	non_scheduled#1_gb_5	5 of 6	nɒnˈʃedjəld ▼
18	orsedue	orsedue	132772	32965184	ˈɔːsɪdjuː	orsedue#_gb_1	1 of 1	-
19	Osler–Rendu–Weber	Osler–Rendu–Weber	262244	10596058	,ɒsləˌrɒndjuːˈwebə	osler_rendu_weber#1_gb_2	1 of 2	,ɒsləˌrɒndjuːˈveɪbə

OED data dated Wed Aug 28 10:20:36 2019

American with or without audio

word does contain ^ob

transcription does contain ^əb

syllabic marker nasal g ʃ ʒ θ ð ŋ ʈ ɪ ɛ œ æ ɜ ə ʌ ʊ ɘ ɔ ɒ ɑ

and is any of a group containing 1 - 16 pronunciations

and the variants don't contain ab

syllabic marker nasal g ʃ ʒ θ ð ŋ ʈ ɪ ɛ œ æ ɜ ə ʌ ʊ ɘ ɔ ɒ ɑ

Word frequency range: 0 - 8

Date range: 1 - 2020

parts of speech: n. v. adj. suffix prefix comb._form adv. int. prep. pron. connective phr. conj.

section in entry main xref infl sub variant

comma separated batch values: comma separated news values:

Limit to entries that also have World English pronunciations

Limit to where there are also unrevised-style pronunciations

results sort order database order ascending Search

#	entry	word	entry id	ph e:id	transcription	soundfile	position	variants	pos	freq	section	dates
1	obduct	obduct	129527	34048069	əb'dæk(t)	obduct#_us_1	1 of 1	-	v.	3	main	1623 - 1
2	obduction	obduction	129528	34048196	əb'dækʃ(ə)n	obduction#_us_1	1 of 1	-	n.	3	main	1578 - 2
3	objectable	objectable	129616	34074482	əb'dʒɛktəb(ə)l	objectable#_us_1	1 of 1	-	adj.	2	main	1656 - 2
4	objectant	objectant	129617	34074636	əb'dʒɛkt(ə)nt	objectant#_us_1	1 of 1	-	n.	2	main	1635 - 1
5	objecter	objecter	238302	13074449	əb'dʒɛktər	objecter#_us_1	1 of 1	-	n.	2	main	1609 - 2
6	objectified	objectified	257525	11061990	əb'dʒɛktəˌfaɪd	objectified#_us_3	1 of 1	-	adj.	4	main	1854 - 1
7	objector	objector	129649	34081771	əb'dʒɛktər	objector#_us_1	1 of 1	-	n.	5	main	1617 - 1
8	oblata	oblata	246430	12236075	əb'lədə	oblata#_us_1	1 of 1	-	n.	2	main	1721 - 1
9	obnoxious	obnoxious	129779	34107520	əb'nɒkʃəs	abnoxious#_us_2	1 of 1	-	adj.	5	main	1572 - 2
10	obnoxiously	obnoxiously	129780	34108161	əb'nɒkʃəsli	obnoxiously#_us_1	1 of 1	-	adv.	3	main	1625 - 2
11	obnoxiousness	obnoxiousness	129781	34108293	əb'nɒkʃəsənəs	obnoxiousness#_us_2	1 of 1	-	n.	3	main	1610 - 2
12	obnoxiously	obnoxiously	129782	34108515	əb'nɒksədi	obnoxiously#_us_1	1 of 1	-	n.	1	main	1924 - 2
13	obscene	obscene	129823	34113159	əb'siːn	obscene#_us_1	1 of 1	-	adj.	5	main	1571 - 1
14	obsceneness	obsceneness	129825	34113772	əb'si(n)nɪs	obsceneness#_us_1	1 of 1	-	n.	2	main	1637 - 1
15	obsequience	obsequience	257399	11049060	əb'sikwiəns	obsequience#_us_1	1 of 1	-	n.	2	main	1830 - 2