OxfordLanguages

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

UKLVC12 September 2019
Catherine Sangster
Gary Leicester

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.

——Blog post:

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

——Survey:

https://public.oed.com/blog/the-oedsupporting-research-through-innovation

OXFORD

The available OED data include the following:

phonemic transcriptions

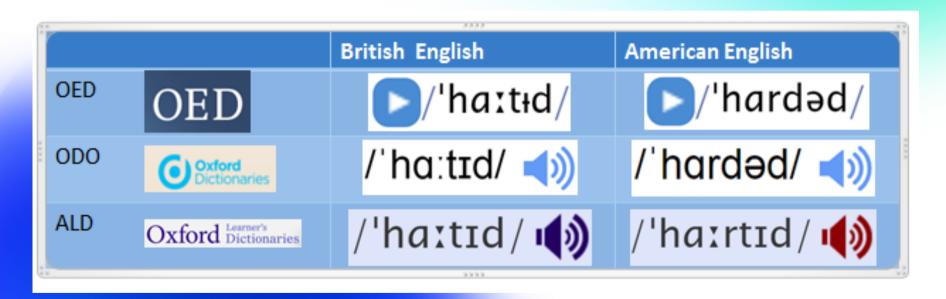
- British English: 276057, in 228363 groups, in 204453 entries (revised and unrevised)
- American English: 208208, in 168866 groups,
 in 148412 entries (revised and c. ¾ of unrevised)
- 12 World Englishes, transcribed according to contemporary phonological models, for relevant subsets of headwords

linked sound files

studio-quality human-voice recordings of all the above

OXFORD

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



OXFORD

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?

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

^ \$.*?[]

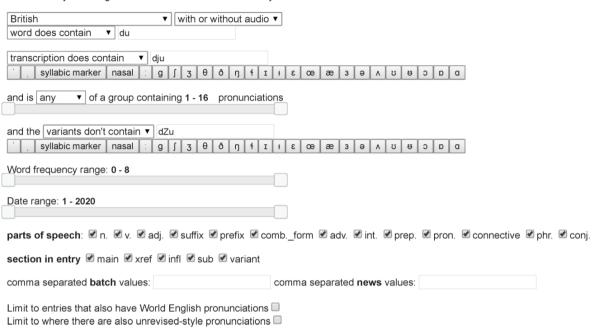
Our **demo** was available in the **poster area** throughout UKLVC12.

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

Please email <u>catherine.sangster@oup.com</u> 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



▼ ascending ▼ Search

#	entry	word	entry id
1	anthomodusa	anthomodusa	3608144

results sort order database order

#	entry	word	entry id	ph e:id	transcription	soundfile	position	variants
1	anthomedusa	anthomedusa	36081440	1174607030	ļanθəmɨˈdjuːzə	anthomedusa#_gb_1	1 of 2	,anθəmı dju∶sə
2	anthomedusa	anthomedusa	36081440	1174607040	ļanθəm₊'dju∶sə	anthomedusa#_gb_2	2 of 2	,anθəmı dju∶zə
3	Anthomedusae	Anthomedusae	45532832	1209526990	ุanθə(ʊ)mɨˈdjuːziː	anthomedusae#1_gb_1	1 of 1	-
4	cubomedusa	cubomedusa	268378	112498678	kju:bəʊmɨˈdjuːzə	cubomedusa#_gb_1	1 of 1	-
5	Cubomedusae	Cubomedusae	269398	117667538	,kjuːbəʊmɨˈdjuːziː	cubomedusae#1_gb_1	1 of 1	-
6	cubomedusan	cubomedusan	268380	116709342	kju:bอชm+ dju:zən	cubomedusan#_gb_1	1 of 1	-
7	dude	dude	58231	243190324	dju:d	dude#_gb_2	1 of 1	-
8	full due	full due	40229305	1209908140	ุfʊl ˈdjuː	full_due#1_gb_1	1 of 1	-
9	light duty	light duty	83535955	1278180680	ˈlʌɪt ˌdjuːti	light_duty#1_gb_3	1 of 1	-
10	May duke	May duke	115301	37914974	meı dju:k	may_duke#1_gb_1	1 of 1	-
11	medusiferous	medusiferous	115815	37570785	mɛdju: sɪf(ə)rəs	medusiferous#_gb_1	1 of 1	-
12	mesoduodenum	mesoduodenum	233691	13564227	mɛsə(ʊ)djuːəˈdiːnəm	mesoduodenum#_gb_1	1 of 1	-
13	nomen dubium	nomen dubium	256941	11200491	ุกอชmɛn ˈdjuːbɪəm	nomen_dubium#1_gb_2	1 of 2	nəʊmən dju:bɪəm
14	nomen dubium	nomen dubium	256941	11200492	nevmen 'dju:brem	nomen_dubium#1_gb_1	2 of 2	neumen 'dju:biem
15	nomen dubium	nomina dubia	256941	11200502	nomine dju:bie	nomina_dubia#1_gb_1	1 of 1	-
16	non-scheduled	non-scheduled	128090	34139590	npnˈʃɛdjuːld	non_scheduled#1_gb_2	2 of 6	npnˈʃɛdjʊld ▼
17	non-scheduled	non-scheduled	128090	34139593	npn skedju:ld	non_scheduled#1_gb_5	5 of 6	nonˈʃɛdjʊld ▼
18	orsedue	orsedue	132772	32965184	ˈɔːsɨdjuː	orsedue#_gb_1	1 of 1	-
19	Osler-Rendu-Weber	Osler-Rendu-Weber	262244	10596058	edaw ujbnar elaa	osler_rendu_weber#1_gb_2	1 of 2	psle rondju: verbe

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

