YOUGLISH.COM: A PROMISING TOOL FOR PRONUNCIATION DICTIONARY LEXICOGRAPHY

DANIEL MCCARTHY

NEWCASTLE UNIVERSITY

Abstract

YouGlish.com allows the researcher to listen to hundreds of speakers pronouncing a specific word in a matter of minutes. The English language contains countless words whose pronunciation differs unpredictably from one individual to the next, such as *comparable* and *candidate*. Knowing the relative frequency of each variant pronunciation would allow lexicographers to order the variant pronunciations in a dictionary by frequency with much greater certainty than has been possible heretofore.

One existing methodology for identifying the most frequent pronunciation variant of a word is the pronunciation preference poll (Wells, 1999). Despite its advantages this method also comes with limitations. To address these limitations the present paper introduces the website YouGlish.com, highlighting its immense potential as a research tool for lexicographers of pronunciation and outlining its strengths and weaknesses. Investigation of 51 words using YouGlish indicates that the relative frequency of pronunciation variants in a number of the words (e.g. *applicable* in American English) has changed from that indicated by existing pronunciation dictionaries.

Keywords: lexicography, phonetics, pronunciation dictionaries, web-based research, accents of English.

Background

There are many words in the English language whose pronunciation differs unpredictably from one individual to the next. Is *comparable* stressed on the first syllable or the second? Is the final vowel of *candidate* /ə/ or /eɪ/? This is reflected in pronunciation dictionaries, which list more than one pronunciation for a large proportion of their entries. However, most of these entries have no information on the frequencies of the variants. Nevertheless, the relative frequency of the variants can be implied by the order in which they are listed, the most frequent variant appearing first. For example, for the word *candidate*, the *Cambridge English Pronouncing Dictionary* (Roach et al., 2011) lists the /ə/ variant first, whereas the *Longman Pronunciation Dictionary* (Wells, 2008) lists /eɪ/ first.

Observations such as this raise the following questions: How do writers of pronunciation dictionaries get their information? How do they decide what pronunciation of a word is the most frequent? Wells (1999: 33), himself a pronunciation lexicographer, writes that in most cases they rely on what he terms their 'informal impressions' of the language: whatever variant they have heard the most is the one they prioritize in the word's entry. This approach does seem reasonable for commonly used words. For example, anyone with substantial exposure to British English will sense that the most frequent pronunciation of the words *either* and *neither* has the vowel /aɪ/, whereas those acquainted with American English will sense that the /iː/ variant predominates. A poll on the matter reveals this to be true: Wells (2008: 265) reports the /aɪ/ variant as being preferred by 87% of British respondents, whereas among American respondents /iː/ is preferred by 84%.

Wells (1990; 1995; 1999; 2000; 2008) has pioneered the use of such pronunciation preference polls, which have been conducted by him for all three editions of his *Longman Pronunciation Dictionary*. Such polls have a number of advantages. The most important one is that they allow the lexicographer to check that their impression of the relative frequency of variants is accurate. The results of the polls also make pronunciation dictionaries more interesting for the reader.

The pronunciation preference polls do, however, come with limitations. Perhaps the greatest one is that only a small number of words can be investigated: Wells (2007) reports that the preference poll conducted for the first edition of his dictionary (Wells, 1990) contained about 90 words and the poll for the second edition (Wells, 2000) again contained only about 90 words. The number of words in the English language whose pronunciation differs unpredictably from one person to the next must be in the thousands, as evidenced by the entries in such dictionaries. Including all such words in a questionnaire would not be feasible since it would make the survey too long for participants to be bothered with. Furthermore, most of the

words on which English speakers differ as to their pronunciation are infrequent: many (if not most) speakers have never encountered such words, save perhaps in writing.

Another limitation of a pronunciation poll is that respondents may not be accurate in reporting their own usage. This is a problem that Wells has dealt with by asking respondents what pronunciation variant they *prefer* rather than what variant they *use* (1999: 33-34). Unfortunately this leaves us unsure of how much of a gap there is between what speakers prefer and what they actually use. We will see an example of this later.

A New Research Tool: YouGlish

The net result of these limitations is that we would like to have an alternative way of investigating the frequency of pronunciation variants, one that compensates for the limitations of preference polling. Ideally this tool would also contain enough material on it that we could investigate the pronunciation of words that are rare. The obvious answer, of course, is the internet. In the 19 years since Wells wrote in 1999 the number of videos of recorded speech on the internet has exploded. Therefore one might imagine that pronunciation lexicography has become easier with the arrival of such material. For example, if one were interested in determining the current pronunciations of an obscure item such as *reconnoitre*, one could search for videos containing the word. However, a video containing a word in its title does not guarantee that it will contain that word being uttered. Even if it does contain the word being uttered, if the video is long the researcher can end up wasting a lot of time waiting for it to be uttered.

These difficulties can be overcome with YouGlish.com (YouGlish, 2018). This website was designed for learners of English as a foreign language but in this paper I am concerned with highlighting its value for researchers.

The website consists of a database of videos. Crucially, however, all the videos have subtitles. This means that when the user types a word into the search bar, YouGlish jumps to

instances in its database of that word being uttered. The user can listen to each clip of the word being uttered by clicking the skip button underneath the video. Thus listening to a word being pronounced by a dozen speakers takes only a minute.

It is this rapidity that makes YouGlish a powerful tool for the pronunciation lexicographer. For example establishing the relative frequency of the two pronunciations of *applicable* can be done in a few minutes. Or maybe the researcher is interested in what percentage of speakers pronounce a /t/ in *often*, and whether the percentage is higher in British English than in American English. Again, this can be investigated within minutes. The researcher listens to each clip containing the word and enumerates how many speakers have the /t/ and how many do not.

The sample size (i.e. number of speakers listened to) can be increased depending on how much data the researcher desires for establishing the pronunciation variation in a given word with confidence: if one pronunciation greatly predominates over another, then only a modest number of video clips need be listened to, whereas if usage is more evenly split then a larger number of clips can be listened to.

The videos on YouGlish come from YouTube (hence the name of the website, a portmanteau of 'YouTube' and 'English'). As of September 2017 the website contained over 20 million videos, which have been subtitled using automatic speech recognition software. By July 2018 this number had risen to over 30 million. Figure 1 shows the layout of the website's home page. British, American, and Australian English can be examined separately, using the buttons for each.

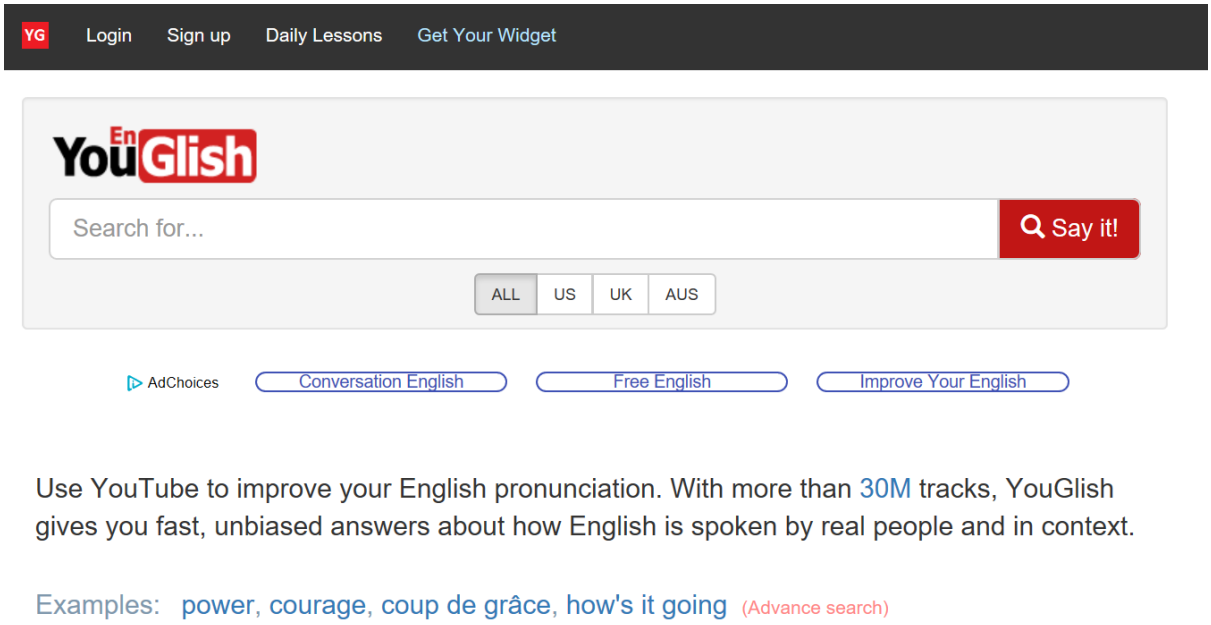

Figure 1: The front page of YouGlish.com. (Retrieved 6 December 2018.)

Regarding the question of how Americans pronounce the word *applicable*, Wells (2008: 38, citing Vaux's online questionnaire) found that 64% of American respondents stress the word on the first syllable, 36% on the second. Yet my own listening of 400 American speakers on YouGlish found that only 42% stressed the word on the first syllable, 58% on the second. One likely explanation for the discrepancy is that the language has changed from when Wells cited Vaux's questionnaire (2007) and the present day.

Earlier I noted that pronunciation preference polls leave open the question of whether speakers' preferences differ from their actual usage. Take the word *ordinary*. How is this pronounced in British English? Wells's (2008: 568) preference poll found that 34% of his British respondents preferred the variant ending in /ɛri/ (as opposed to /əri/ or /ri/). But when 100 speakers of British English were examined on YouGlish I found that only 10% of them used this variant. Admittedly my sample size was much smaller than that of Wells, but it does raise the suspicion at least that speakers' preferences can depart considerably from their actual usage. This would not be surprising in the case of *ordinary* since the /ri/ variant deviates from the spelling more than the other two options in that it has one less syllable, and laypeople are

probably more likely than linguists to be swayed by spelling when introspecting on pronunciation preferences. The proportion of the British YouGlish sample using the /ri/ variant was 79%, over double the 34% found by Wells. With a larger sample of YouGlish speakers this apparent discrepancy between preference and actual usage could be investigated with still greater precision.

What kind of subject matter is found in YouGlish videos? The videos appear not to be a random sample of YouTube videos, but rather are limited to material that is deemed suitable for the learner. Some material is sourced from television programmes, while much of it comes from public speaking such as academic lectures and biblical sermons. The result is that the sample seems to be skewed towards speakers with above-average educational attainment. However, this skewing of samples towards educated usage is also found to a similar (if not greater) degree in pronunciation polling: Wells (1999: 33, 35) describes the 1,932 speakers he recruited for his poll as 'speech-conscious', that is, the kind of person who is interested enough in language to take nearly an hour of their time to fill out a survey.

In sum, the results of YouGlish research should not be assumed to be representative of the *entire* population, be it of American or British speakers, but rather tilted towards middle-class and/or educated usage.

One might wonder to what extent the pronunciation of individual words like *ordinary* or *applicable* vary **within** British and American English. One possible criticism of YouGlish as a research tool is that videos cannot be searched according to the speaker's regional background: it is not possible to only display search results for, say, Mancunian. However, in the context of pronunciation dictionary lexicography this is not a major limitation: Wells (1999: 33) remarks the following about variation in the pronunciation of *schedule* and other words: 'Uncertainties such as these are often not a matter of social or local accent, since the same variability appears to exist in RP as in regional accents'.

Challenges with Using YouGlish

What are the difficulties in using YouGlish as a research tool? One annoyance is that Australian or American English videos may appear in the results for British English videos. (Less frequently, British English may appear in American and Australian English videos.) To some extent this is understandable: if material consists of an interaction between a British and American speaker, there is no easy way for YouGlish to separate the parts with the British speaker and the parts with the American speaker. More concerning is the appearance of *entire* videos of Australian English in the British English search results, which I have noted. This is just one of several reasons I will outline for why it is necessary to pay close attention when using YouGlish for surveying the frequency of pronunciation variants. That is, the linguist cannot assume that because a video has been put in the 'UK' category, the speaker is British.

A related difficulty with using YouGlish for polling pronunciation is that not all the results come from native speakers. Thus if one wants to exclude non-native speakers from usage surveys, it is important to listen closely to the clip for signs of a non-native accent. Fortunately it is usually easy to do so from a short clip, which means that this issue does not in my experience slow down the use of YouGlish for research.

More difficult cases are those involving native speakers of English who have lived in more than one part of the English-speaking world. For example, there are several videos on YouGlish featuring the entrepreneur Elon Musk, who spent the first 17 years of his life in South Africa, then moved to Canada, and has settled down in the United States. His accent is a mix of South African and North American so I have excluded him from surveys of American English. But it is possible that there are other cases of speakers with mixed accents that I failed to detect. I do not expect such cases to be large enough in number to result in them affecting the results appreciably, but it is important to be aware of the possibility nonetheless.

One final point should be noted: I have not attempted to exclude Canadian speech from my examination of American English: it is not normally possible to determine from a short clip whether the person is speaking American English or Canadian English, since most of the phonetic differences between the two varieties (Labov, Ash, and Boberg, 2006: 216-224) are too modest to be noticed in such a small time frame. Thus the results for 'American' English in this paper should be understood to be North American English. In contrast, speakers of Irish English appearing in the British English videos have been excluded since the phonetic features of such accents are immediately noticeable, making the removal of such speakers straightforward.

All the difficulties discussed heretofore are ones which I believe to be fundamentally tractable, especially for a phonetically-trained native speaker of English. At worst such difficulties introduce a modicum of measurement noise into the data as opposed to fundamentally skewing the results. There is, however, a difficulty with using YouGlish which, if not avoided, *can* fundamentally skew the data: the search results contain *all* instances in the database of a given speaker uttering the search term. For example, I determined that the 299 American English search results for the word *cyclic* were uttered by just 63 speakers. How is this possible? If a word is used frequently in a video (because of it being a technical term in a lecture, for example), it can easily occur many times. Thus if the researcher does not pay attention to whether a clip is coming from a speaker already counted in the poll, the data are at risk of containing multiple counts of the same speaker.

The word *cyclic* is an extreme example because it happens to be used heavily in chemistry lectures as a constituent of various technical terms, with the result that the lecturers often end up uttering the word over and over. Most other words are less liable to this repetition problem, but it is nevertheless something that has to be guarded against when using each clip as a data point. This is particularly true of words where usage is fairly evenly split: *cyclic* itself

is an example, being pronounced with /aɪ/ by 30 out of 63 of the American speakers and with /ɪ/ by the remaining 33 of them. If all 299 instances of the word in the search results had been included in the data, then the sample would have been skewed in favour of those speakers who happened to utter the word the most.

One final limitation of YouGlish is that the number of tokens of a word varies massively from one word to the next. For the most part this is inescapable: a corpus based on real-life recordings will tend to reflect the power-law frequency of words (Zipf, 1935, cited by Pinker, 2011: 212-213) in real life. However it is also clear that coverage of British English lags behind the coverage of American English. Given that the population of Great Britain is approximately 63 million whereas the combined population of the United States and Canada is approximately 361 million, we would expect the number of search results for 'American English' to be roughly five to six times greater than the number of results for 'British English'. Although there are some words in which the British results exceed this ratio (e.g. *portrait*, for which there are 786 'British English' results as against 1,660 'American English' results), there appear to be far more cases in which the 'American English' results vastly outnumber the 'British English' results, sometimes by a factor of over 20 to 1.

Beneficiaries of YouGlish-Based Research

There are two groups who could benefit from YouGlish-based research: lexicographers of pronunciation and speech technologists.

How might pronunciation dictionaries benefit from YouGlish-based research? Let us take a concrete example: the word *status*. The *Longman Pronunciation Dictionary* (2008) recommends the pronunciation /ˈsteɪtəs/, both for learners of British English and American English. But is the word *status* in fact pronounced this way by most speakers? The word does not appear in any of Wells's pronunciation preference polls, which means we must turn to YouGlish. A sample of 50 British speakers reveals that /ˈsteɪtəs/ is the almost universal form,

with 95% of the speakers using this variant. Among the 100 American English speakers sampled, however, just 6% used this form! Instead, /ˈstætəs/ is the predominant form, used by all of the remaining 94%. Another example of a word where using YouGlish can correct pronunciation dictionaries is *nascent*, for which *Longman* recommends the variant with /æ/: in a sample of 100 American English speakers on YouGlish this variant was found among just five of the speakers, the other 95 going for the /eɪ/ variant.

As for speech technology, the results of YouGlish-based research could be used to refine the prior probabilities of pronunciation variants. For example, automatic speech recognition software specializing in American English would have lower prior probabilities for the /ˈsteɪtəs/ variant of *status* than software for British English, which would have the by-product of reducing the confusion probability with *stasis* in the former variety but not the latter. However, given speech technologists' preference for bottom-up data-driven automated solutions over top-down knowledge-driven solutions it is unlikely that YouGlish-based research would be utilized by ASR in practice.

Nevertheless one might wonder whether ASR could somehow be used to automate the extraction of pronunciation variants from YouGlish. This would vastly enlarge the amount of pronunciation information harvested from YouGlish relative to a linguist listening to video clips one at a time. Unfortunately there are a number of barriers that would have to be surmounted before YouGlish could be used in this way. We have already seen one factor: a fully automated extraction of the variants of *cyclic* on YouGlish would have yielded statistics on the frequency of /ˈsɪk-/ and /ˈsaɪk-/ skewed in favour of those speakers who happened to utter the word the most. A possible solution to this is to only take one pronunciation of a given word from a given YouGlish video. However, certain famous speakers such as Barack Obama, Theresa May, and David Cameron appear in multiple YouGlish videos. Thus even if one only picked a single pronunciation of a given word from a given video, one could still end up with

multiple pronunciations from a single speaker. A human being watching the videos can get around this problem since they can see with their own eyeballs that they have already counted Theresa May's pronunciation of the word *status* and do not need to count it twice. This is the major advantage of manual compilation of usage statistics over ASR approaches.

Another hurdle for someone trying to compile statistics on pronunciation usage with YouGlish automatically is that sometimes the subtitles are mistimed with the video, such that the word of interest occurs a second or two before the clip begins. This requires the viewer to manually rewind the video to find the desired word. It is not at present clear how an automated procedure would deal with this.

Some Findings using YouGlish

The aim of this section is to give the linguist a flavour of the kind of research that can be done with YouGlish. Table 1 below shows the results for a selection of words that have been investigated by the present author using the website. The words in question were specifically chosen because examination of existing dictionaries (Roach et al. 2011; Wells 2008) indicated that more than one pronunciation variant is in use among speakers of English for each of them. YouGlish allows the researcher to take the additional step of identifying which of these variants is the most frequent one.

The sample size varies considerably from one word to another. For example for the word *portrait* I examined 100 British speakers but only 50 American speakers. This is because after listening to just 50 Americans it was clear that the /-ət/ variant is the most frequent variant by far (94%). In contrast I listened to a larger number of British speakers as usage is more divided: 38% /-ət/, 62% /-eɪt/. Note that for infrequent words such as *esoteric* or *worryingly*, the reason the sample size is small is that there are not many instances of the words on YouGlish.

Although YouGlish can be used to establish the percentages of particular pronunciations, it is important to bear in mind that there is a margin of error associated with these percentages, the precise magnitude of which will vary depending on the sample size for that word. One might wonder how large the margin of error is for the frequencies given. This is indeed an open question but it is interesting to note that Greg Jenner's online survey of the word *Brexit* (N = 6200) found 73% preferring the /-ks-/ variant, 27% the /-gz-/ one (Hall, 2017), which is strikingly similar to results found below for *exit* (N = 50), namely that 74% reported /-ks-/, 26% /-gz-/.

One final use of YouGlish is that it can be used to identify new variants: though not listed below I have discovered that most Americans do not have /j/ in *spurious* (a possibility that is not even listed by *Longman*) and that *de jure* and *a priori* have far more variant pronunciations than listed by any existing dictionary.

Word	**Variety**	**Variant 1**	**Frequency**	**Variant 2**	**Frequency**	**Sample Size**
administrative	Br	VVVəV	.96	VVVeɪV	.04	50
amateur	Br	ˈVVtʃə	.318	ˈVVtə	.671	
		VVˈtɜː	.012			85
amenable	Am	iː	.72	ɛ	.28	50
	Br	iː	.2	ɛ	.8	10
anti-	Am	-i	.37	-aɪ	.63	100
antidote	Am	-ə-	.95	-i-	.05	22
	Br	-ə-	.43	-i-	.28	
		-ɪ-	.28			21
apparatus	Am	-eɪ-	.02	-æ-	.98	50
	Br	-eɪ-	.87	-æ-	.13	23
applicable	Am	ˈVVVV	.422	VˈVVV	.578	400
aristocrat	Br	ˈærəstVV	.762	əˈrɪstVV	.238	21
candelabra	Am	ɑː	1	æ	0	10
	Br	ɑː	.6	æ	.4	5
candidate	Am	-ət	.517	-eɪt	.483	300
	Br	-ət	.7	-eɪt	.3	100
comparable	Am	ˈVVV	.8	VˈVVV	.2	100
	Br	ˈVVV	.478	VˈVVV	.522	67
compass	Am	ʌ	.71	ɒ	.29	80
consortium	Am	-ʃəm	.72	-tiəm	.10	
		-ʃiəm	.18			50
	Br	-ʃəm	.05	-tiəm	.95	20

dastardly	Br	ɑː	.67	æ	.33	6
de facto	Am	də	.67	di	.22	
		deɪ	.11			100
	Br	də	.5	deɪ	.35	
		diː	.15			20
esoteric	Br	iːs-	.73	iːz-	.04	
		ɛsə-	.19	ɛzə-	.04	26
exit	Am	ks	.46	gz	54	150
	Br	ks	.26	gz	.74	50
formidable	Am	V́VVV	.57	VV́VV	.43	100
	Br	V́VVV	.33	fɔːˈmɪd-	.32	60
		fəˈmɪd-	.35			
geography	Br	ˈdʒɒg-	.663	dʒiːˈɒg-	.337	86
hurricane	Br	-ən	.59	-eɪn	.41	27
hydraulic	Br	ɒ	.94	ɔː	.06	18
identification	Am	aɪ-	.86	ə-	.14	50
identify	Am	aɪ-	.76	ə-	.24	50
illusory	Am	z	.7	s	.25	
		ʒ	.05			20
	Br	z	.7	s	.3	10
immediate	Br	-diə-	.52	-dʒiə-	.02	
		-djə-	.23	-dʒə-	.23	100
immersion	Br	-ʃ-	.85	-ʒ-	.15	20
innovative	Br	V́VəV	.85	V́VeɪV	.12	
		V́ɪVV	.02	V́ɒVV	.01	100
interest	Br	-ɛst	.72	-əst	.28	100
legislative	Br	VVəV	.98	VVeɪV	.02	50
legislature	Br	VVəV	1	VVeɪV	0	13
liaison	Am	liːˈeɪzɒn	.75	ˈliːəzɒn	.14	
		ləˈjeɪzɒn	.04	ˈleɪzɒn	.05	
		ˈljeɪzɔ̃	.01	ˈleɪəzɒn	.01	100
	Br	-ɒn	.93	-ən	.07	28
mandatory	Br	ˈVətri	.89	ˈVətəri	.04	
		VˈdeɪtriVˈdætri	.04		.04	27
multi	Am	-i	.67	-aɪ	.31	
		-ə	.02			45
Muslim	Am	-ʊz-	.15	-ʊs-	.36	
		-ʌz-	.43	-ʌs-	.06	100
	Note: about half of -ʊs- users appear to be Muslim Americans.					
nascent	Am	eɪ	.95	æ	.05	100
	Br	eɪ	1	æ	0	6
November	Am	ə	.2	oʊ	.8	50
	Br	ə	.46	oʊ	.54	50
often	Am	-f-	.84	-ft-	.16	50
	Br	-f-	.62	-ft-	.38	50
omit	Am	ə-	.3	oʊ-	.7	50
	Br	ə-	.83	oʊ-	.17	12
partisan	Br	ˈVVzæn	.577	VVˈzæn	.308	

		ˈVVzən	.077	ˈVVsæn	.038	26
penchant	Br	ˈpɒ̃ʃɒ̃	.8	ˈpɛntʃənt	.2	5
phenomenon	Br	-ən	1	-ɒn	0	24
portrait	Am	-ət	.94	-eɪt	.06	50
	Br	-ət	.38	-eɪt	.62	100
precedent	Am	prɛs-	1	priːs-	0	22
	Br	prɛs-	1	priːs-	0	22
princess	Br	V́V	.59	VV́	.41	49
project	Br	ɒ	1	əʊ	0	25
research	Br	VˈV	.62	ˈVV	.38	100
route	Am	uː	.44	aʊ	.56	200
second	Am	-d	.5	-t	.5	22
status	Am	-eɪ-	.06	-æ-	.94	100
	Br	-eɪ-	.95	-æ-	.05	50
version	Br	-ʃ-	.467	-ʒ-	.533	152
worryingly	Am	VVVV	.75	VVV	.25	8
	Br	VVVV	.375	VVV	.625	16

Table 1: Summary of results using YouGlish.

Conclusion

The aim of this paper has been to stimulate interest in using YouGlish when making decisions in lexicography. This tool has enormous potential to improve the quality of pronunciation dictionaries by helping the lexicographer decide in a principled way the order in which to list variant pronunciations, identify new variants, whether to omit obsolescent variants, and provide quantitative data on the frequencies of variants.

Summary

To summarize, these are the strengths of using YouGlish as a research tool:

1. Avoids relying on polling people's 'preferences', which can differ from actual usage.
2. Avoids relying on lexicographer's own impressions, which can be inaccurate.
3. Can listen to hundreds of speakers pronouncing a specific word in a matter of minutes.

And these are YouGlish's disadvantages:

1. The 'American English' and 'British English' search results can contain other accents.
2. The search results can contain more than one token from the same speaker.
3. The sample size for infrequent words can be small, especially for British English.

References

Forvo (2018). Website (retrieved 6 December 2018: https://forvo.com/).

Hall, D. (2017). ‘‘Breksit’ or ‘bregzit’? The question that divides a nation.’ Online article, (retrieved 6 December 2018: https://theconversation.com/breksit-or-bregzit-the-question-that-divides-a-nation-82278).

Labov, W., Ash, S., Boberg, C. (2006). *The Atlas of North American English: Phonetics, Phonology and Sound Change.* Berlin and New York: Mouton de Gruyter.

Pinker, S. (2011). *The Better Angels of Our Nature: The Decline of Violence in History and Its Causes.* London: Penguin.

Roach, P., Setter, J., and Esling, J. (2011). *Cambridge English Pronouncing Dictionary.* 18th Edition. Cambridge University Press.

Vaux, B. (2003). Online questionnaire (retrieved 6 December 2018: http://www.tekstlab.uio.no/cambridge_survey/maps).

Wells, J. C. (1995). Age grading in English pronunciation preferences. In Elenius, K. and Branderus, P. (eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences.* Stockholm. 3, pp. 696-699.

Wells, J. C. (1999). ‘British English pronunciation preferences: a changing scene.’ *Journal of the International Phonetic Association*, Vol. 29(1), pp. 33-50. Available at: http://www.phon.ucl.ac.uk/home/wells/changingscene.pdf).

Wells, J. C. (2007). ‘Pronunciation preference survey 2007.’ Blog entry, 11 May 2007 (retrieved 6 December 2018: http://www.phon.ucl.ac.uk/home/wells/blog0705a.htm).

Wells, J. C. (2008). *Longman Pronunciation Dictionary.* Third Edition (First Edition 1990, Second Edition 2000). Harlow, UK: Pearson Education.

YouGlish (2018). Website (retrieved 6 December 2018: http://youglish.com/).

Zipf, G. K. (1935). *The psycho-biology of language: An introduction to dynamic philology.* Boston: Houghton Mifflin.

About the author:

Daniel McCarthy is a PhD candidate in Speech Sciences at the School of Education, Communication and Language Sciences in Newcastle University where he is completing a thesis on the acoustics of place of articulation in English plosive consonants. His research interests include acoustic phonetics, auditory science, and dialectology. Over the last four years, he has been the teaching assistant on the Phonetics II module for students in the BSc Speech and Language Sciences and MSc Speech and Language Pathology courses. He has presented his research at the ICPhS, BAAP, and PaPE conferences.

E-mail: d.mccarthy1@newcastle.ac.uk