Common Voice

moz://a

Language Variants

Contents
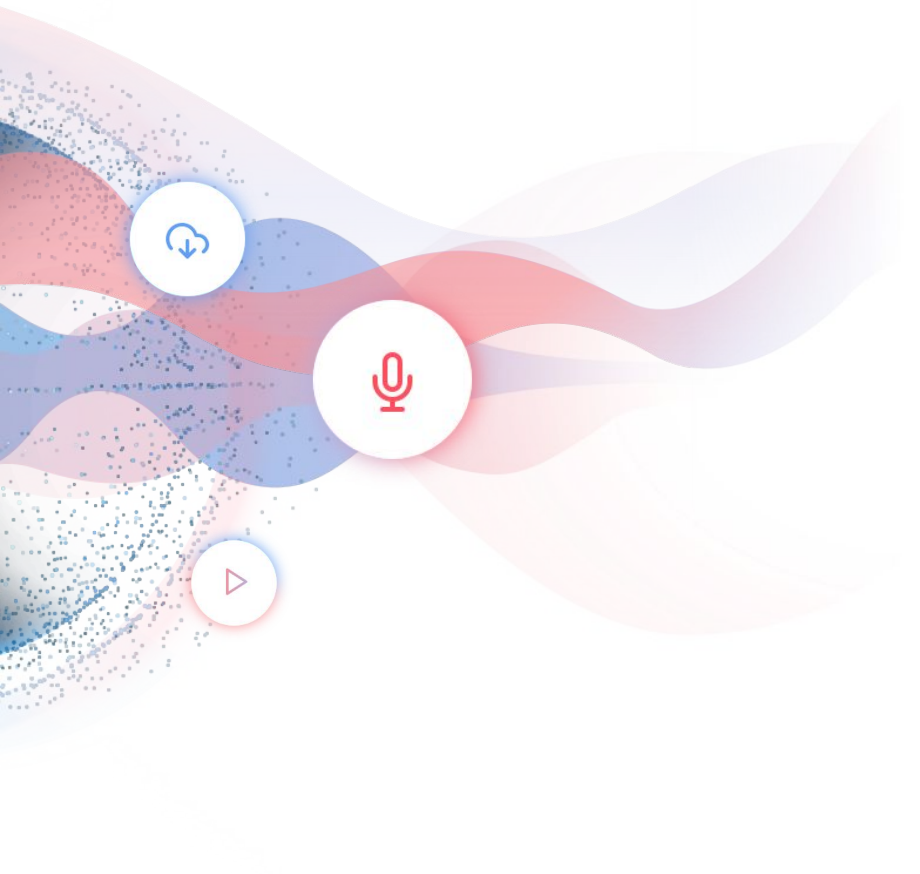
Mozilla Foundation stewards the Common Voice Dataset by providing product and software leadership for the projects goals.

Languages are such a personal aspect of our identities, it's vital that we engage and support the voices of our contributors.
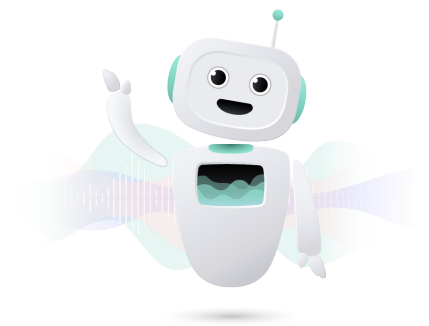
Common Voice

![moz://a]

Language
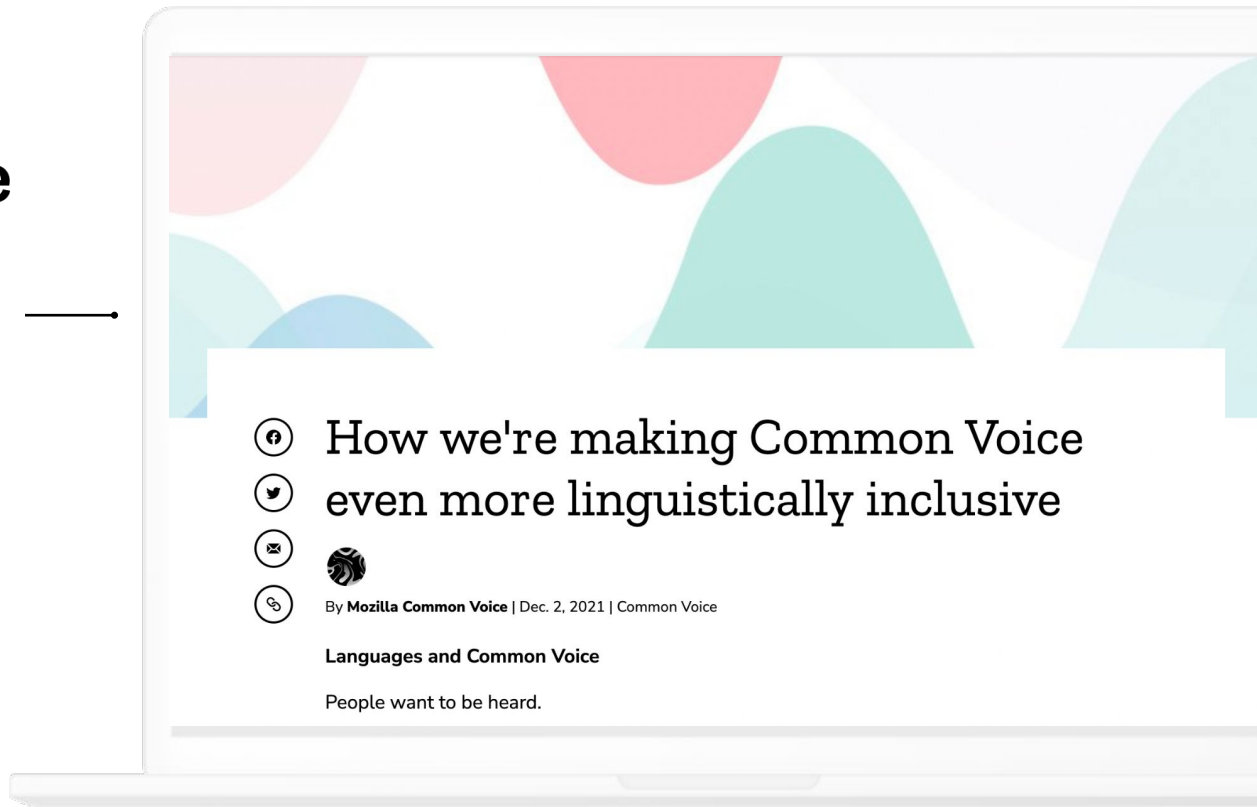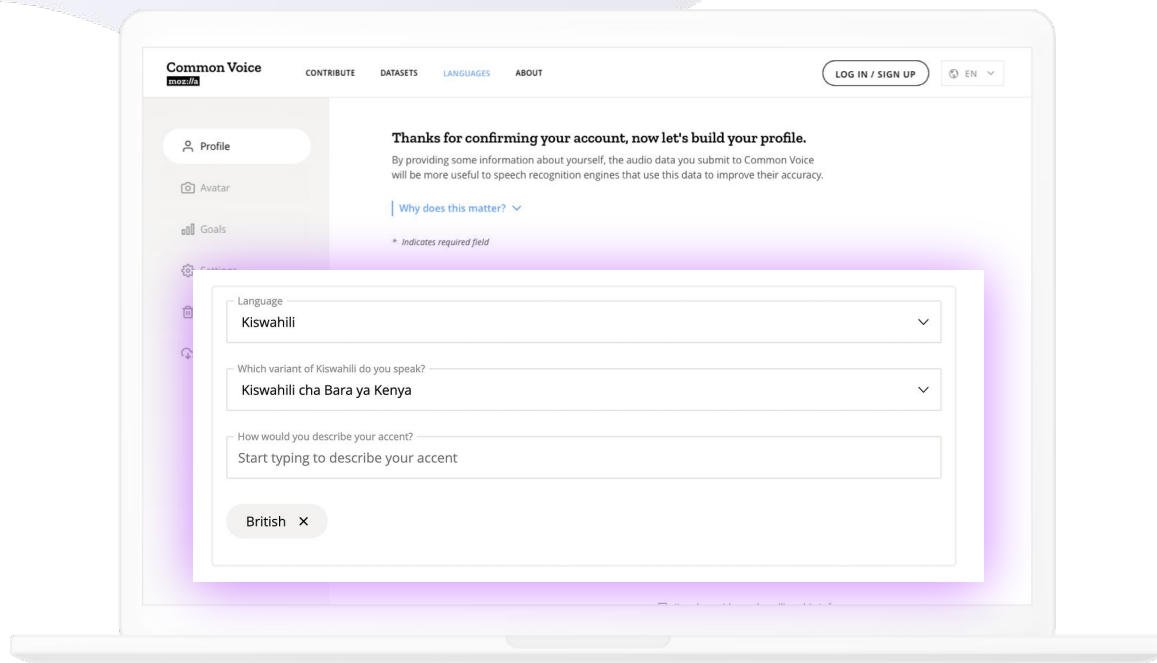## Variants
Accents

How we're making Common Voice even more linguistically inclusive

By **Mozilla Common Voice** | Dec. 2, 2021 | Common Voice

Languages and Common Voice

People want to be heard.

| Language	Variant	Accent
A **shared system** of spoken, signed, or written symbols through which a group of people can understand one another. For the purposes of Common Voice and ASR, languages would be driven for the most part by **mutual intelligibility.** Example... Spanish	A specific form of a language or language cluster associated with a group of speakers, for example those living in a shared region or country, or who have a shared culture or heritage, and thus experience **vocabulary, grammar and norms that differentiate their speech from others.** *Note that we use the word variants rather than dialects because the latter has sometimes been used in derogatory ways, however variants are not limited to geographies. Example... Mexican Spanish	**Pronunciation differences** associated with a group or an individual speaker. This may take account of specific, complex or nuanced factors, such as other languages spoken, religious or tribal affiliation, or long periods of time spent in different places. Example... has lived in Mexico City, also speaks Nahuatl

How to add - Language Variants + Metadata

Profile Page

Common Voice
moz://a

Why is a variant important ?

No two people speak exactly the same, everyone has their own way of speaking.
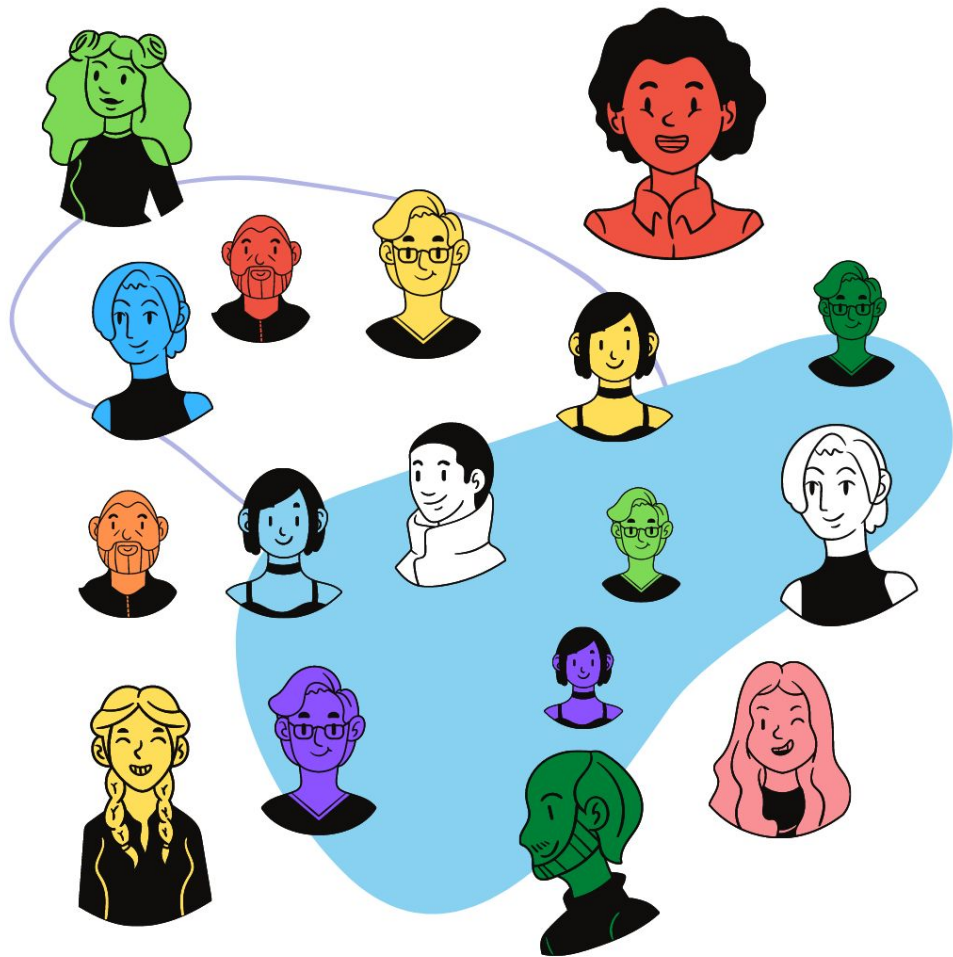
Choosing a variant

Step 1: Plan a gathering of language speakers together

Host the session in your native language

Checklist		
		Reach out to other contributors on Community Channels and linguists for your language. (Comms template will be shared)
		Host a community event or discuss asynchronously about steps 2 and 3 *Ps. Request support for funding via Mozilla [Resource Reps](#) for your community event*
		Report back your findings in a blog posted on discourse/element or via email to commonvoice@mozilla.org

Common Voice
moz://a

Choosing a variant

Step 2: Does your language require variants ?

Checklist	
	Using ISO-639-3 to inform your decision about whether you need your own language code?
	What if some countries, communities and regions have very different vocabulary?
	Are words in the language pronounced differently in different countries or regions?
	Read through the Language Variant Guidance

Common Voice
moz://a

Choosing a variant

Step 3: Choose BCP-47 code for each variant you are choosing

Checklist	
	Find out if there is an existing BCP-47 Code using this https://r12a.github.io/app-subtags/
	Create a BCP-47 Code if none exists using; a language tag, an optional script tag, an optional region tag, an optional variant tag. Learn more on the playbook
	Share your thoughts on what should be featured Submit via email to Commonvoice@mozilla.org or in this form

Common Voice
moz://a

Example Variants

- es-MX: Spanish, as spoken in Mexico in Latin script.
- ca-valencia: The Valencian variant of Catalan.
- sr-Latn-ijekavsk: Serbian written in Latin script, spoken in Serbia and with the ijekavian variant (e.g. writing and pronouncing *rijeka* "river" instead of *reka*). This variant is spoken both in Serbia and Bosnia-Herzegovina but with few differences so the region tag is left out.
- rm-vallader: The Vallader variant of Romansch.
- pa-Guru: Punjabi written in Gurmukhi script. This variant is mostly used in India,

Deadlines for Submission

- Variants to be added by the September Dataset release need to be submitted by the end of July
- Variants to be added by the December Dataset release need to be submitted by the end of October

Submit via email to Commonvoice@mozilla.org or in this [form](#)

Common Voice
moz://a