**Language Breakdown**

Statscan provides a detailed breakdown of the census data for each electoral district. For the purpose of this analysis, we will look at ‘Mother tongue’ - ‘Single responses’ and group them in a way that will maximize information while minimizing complexity. We choose ‘Mother tongue’, not ‘most spoken at home’ because mother-tongue is the best indication for an ethnic breakdown

Categories will be:

English

French

Non-official languages

These non-official languages will be broken down further into:

* Aboriginal languages
* Afro-Asiatic languages (mainly semitic languages, includes Hebrew and Arabic)
* Austro-Asiatic languages (includes Khmer and Vietnamese); include Hmong-Mien languages in this group (computing required)
* Austronesian languages (includes Tagalog / Philipino)
* Dravidian languages (includes Tamil)
* The large group of Indo-European languages will be broken down (computing required):
  + European languages (Albanian, Armenian, Balt-Slavic, Celtic, Germanic, Greek), Italic, does not include English and French (doesn’t include it in Statscan either), also include Kartvelian languages (Georgian, computing required) and Uralic languages (computing required)
  + Indo-Aryan languages (includes Hindi, Punjabi, Gujarati, Marathi, Urdu…) - further breakdown may be useful.
  + Iranian languages (Kurdish, Farsi), include Indo-Iranian Languages n.i.e in this group (computing required)
* Japanese, Mongolic, Korean languages (computing required)
* Niger-Congo and Nilo-Saharan languages, include Creole languages (computing required)
* Sign languages and Other languages (computing required)
* The Sino-Tibetan languages group will be broken down:
  + Chinese languages - mainly Mandarin and Cantonese (computing required)
  + Tibeto-Burman languages
* Tai-Kadai languages (mainly Lao and Thai)
* Turkic languages