How to compare Languages/Dialects quantitatively?
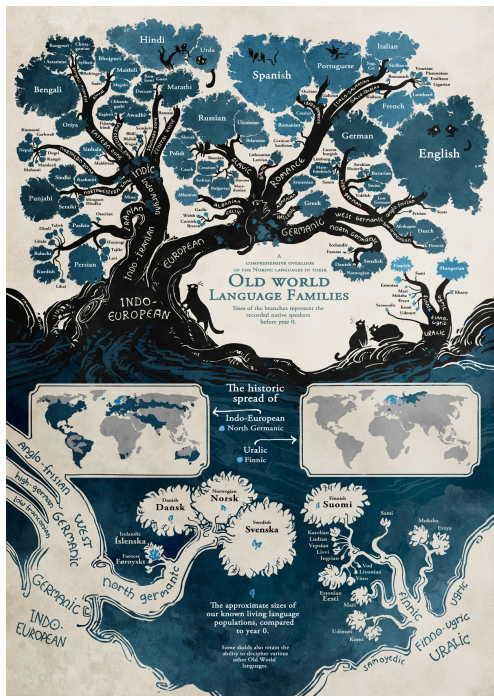
Sudheer Kolachina

September 2, 2020

Why are some languages more alike than others?

- ▶ Swedish is more similar to Danish than to Spanish.
- ▶ Hindi is more similar to Punjabi than to Thai.

- Languages can be similar either
  - due to descent from a common ancestor, Indo-European or Uralic - *genetic* relationship
  - due to convergence as a result of extensive language contact and multilingualism - *areal* relationship / notion of Linguistic area

COMPARATIVE-HISTORICAL LINGUISTICS refers to an area of linguistics that is concerned with comparing Languages to establish their genetic relatedness.

A COMPREHENSIVE OVERLOOK
OF THE NORDIC LANGUAGES IN THEIR

Old World
Language Families

Sizes of the branches represent the
recorded native speakers
before year 0.

- Comparative method
- Sound correspondences
- Reconstruction and Proto-language

Systematic sound correspondences across languages

Meaning	Sanskrit	Greek	Latin	Gothic	English
'Foot'	pad-	pod-	ped-	fo:tus	foot
'Father'	pitar	pate:r	pater	fadar	father
'You'	tvam	tu:	tu	þu	thou
'Dog'	Sva:n	kuo:n	canis	hunds	hound
'Hundred'	Satem	katon	centum	hunda	hundred
'Tribe/Family'	janas	genos	genus	kuni	kin
'Brother'	bhra:tar	phra:te:r	fra:ter	broþar	brother

Figure 1: Cognates in Indo-European languages

- ▶ Grimms's Law: evidence of systematic sound change
- ▶ Chain shift in Germanic:
  voiced aspirated > voiced > voiceless > fricative

Centum versus Satem

Greek	Latin	Gothic	Sanskrit	Slavic
k	k	h	sh	s
g	g	k	j	z
kh	h/g/f	g	h	z
p/t/k	qu	wh	k	k
b/d/g	v/gu	q	g	g
ph/th/kh	f/v/gu	w	gh	g

Figure 2: Velar and Palatal stops in Indo-European languages

Can we estimate inter-language similarity quantitatively?

- Edit Distance - Estimate similarity between strings by counting the number of edit operations required to transform one string into another
- http://www.let.rug.nl/~kleiweg/lev/
- Asymmetric: EditDist(A,B) != EditDist(B,A)

Useful Links

```
https://en.wikipedia.org/wiki/Edit_distance
https:
//www.britannica.com/topic/Indo-European-languages
//https:
//royalsocietypublishing.org/doi/10.1098/rsos.171504
https://www.cambridge.org/core/books/
dravidian-languages/CB6877BB2AFC237DA0B154E62F8DD898
```