

# How to compare Languages/Dialects quantitatively?

Sudheer Kolachina

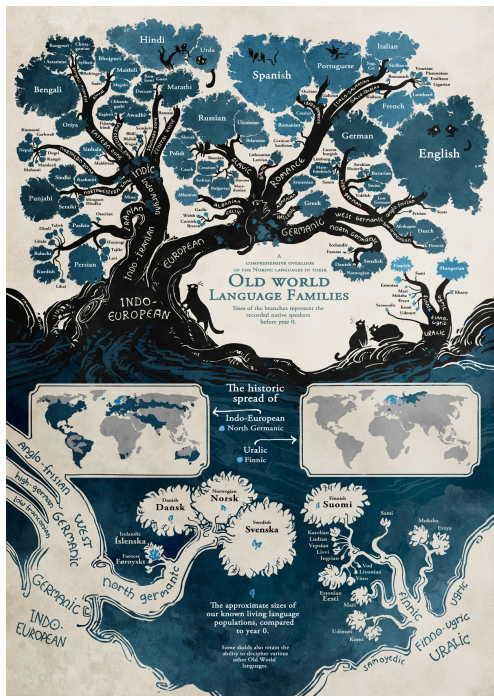
September 2, 2020

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.



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

## Systematic sound correspondences across languages

<b>Meaning</b>	<b><i>Sanskrit</i></b>	<b><i>Greek</i></b>	<b><i>Latin</i></b>	<b><i>Gothic</i></b>	<b><i>English</i></b>
'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	bropar	brother

Figure 1: Cognates in Indo-European languages

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



# Frame Title

## Centum versus Satem

Greek	Latin	Gothic	Sanskrit	Slavic
<b>k</b>	k	h	sh	s
<b>g</b>	g	k	j	z
<b>kh</b>	h/g/f	g	h	z
<b>p/t/k</b>	qu	wh	k	k
<b>b/d/g</b>	v/gu	q	g	g
<b>ph/th/kh</b>	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:  $\text{EditDist}(A,B) \neq \text{EditDist}(B,A)$

# Useful Links

[https://en.wikipedia.org/wiki/Edit\\_distance](https://en.wikipedia.org/wiki/Edit_distance)

[https:](https://www.britannica.com/topic/Indo-European-languages)

[//www.britannica.com/topic/Indo-European-languages](https://www.britannica.com/topic/Indo-European-languages)

[//https:](https://royalsocietypublishing.org/doi/10.1098/rsos.171504)

[//royalsocietypublishing.org/doi/10.1098/rsos.171504](https://royalsocietypublishing.org/doi/10.1098/rsos.171504)

[https://www.cambridge.org/core/books/](https://www.cambridge.org/core/books/dravidian-languages/CB6877BB2AFC237DA0B154E62F8DD898)

[dravidian-languages/CB6877BB2AFC237DA0B154E62F8DD898](https://www.cambridge.org/core/books/dravidian-languages/CB6877BB2AFC237DA0B154E62F8DD898)

