

# Quantifying Scripts

Defining metrics of characters for  
quantitative and descriptive  
analysis



University of  
St Andrews

---

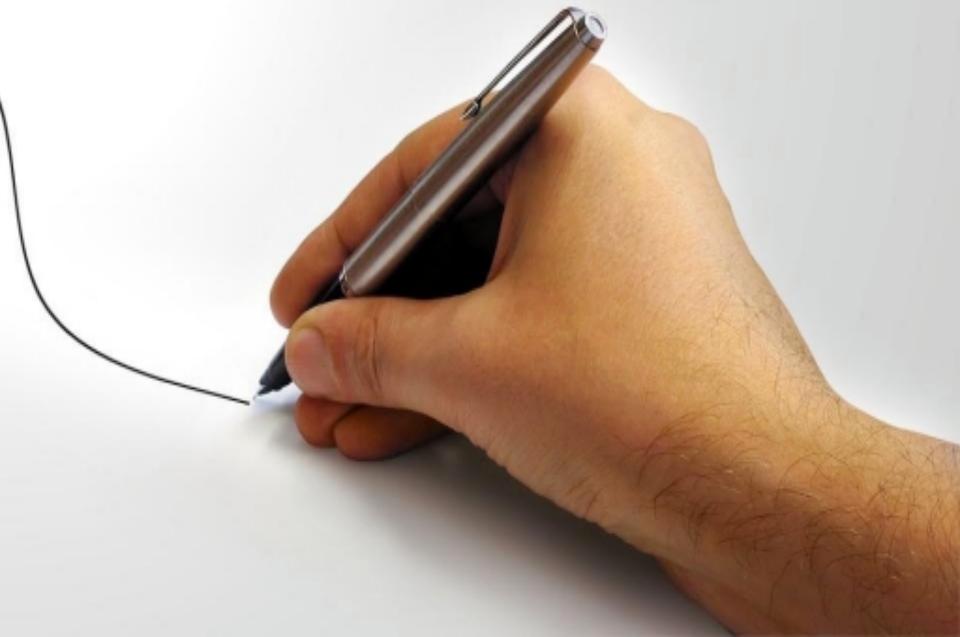
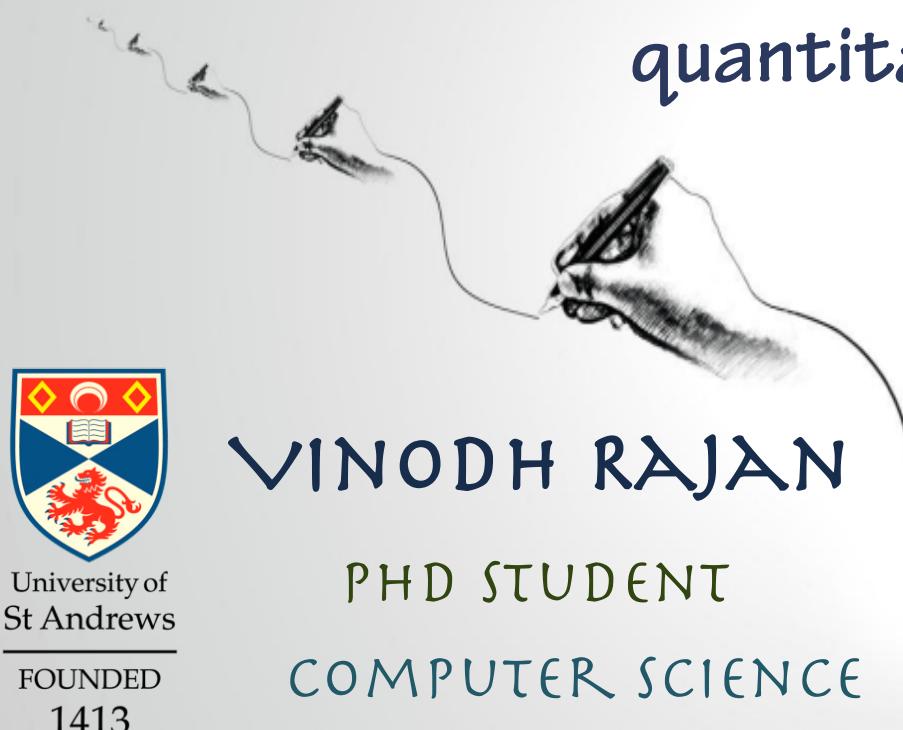
FOUNDED  
1413

VINODH RAJAN

PHD STUDENT  
COMPUTER SCIENCE

sicsa\*

The Scottish Informatics &  
Computer Science Alliance



# Introduction

- Why quantify Scripts & Characters?
- Analysis of Characters/Glyphs have mostly been mostly qualitative
- Quantitative Analysis on Scripts can be extremely useful and informative.
  - Can learn more about human handwriting behavior



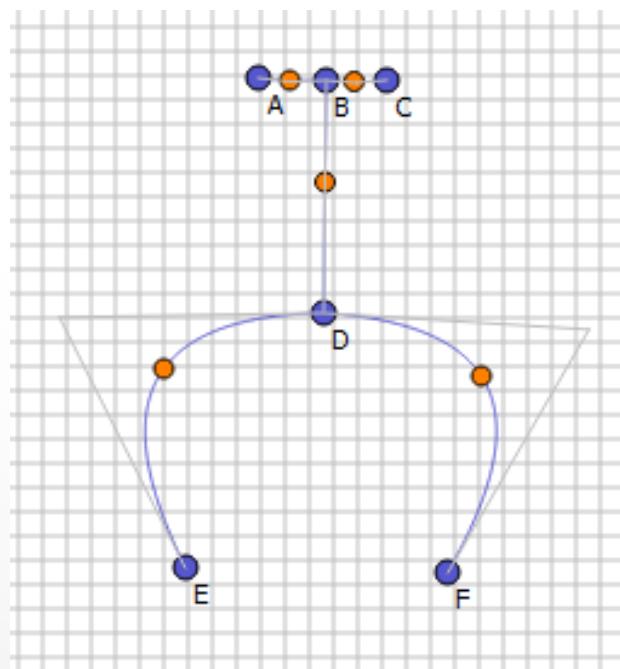
# What's in Characters?

- Visual Information
  - That relate to their appearance
- Production Information
  - That relate to the handwriting process that produce them
- Cognitive Information



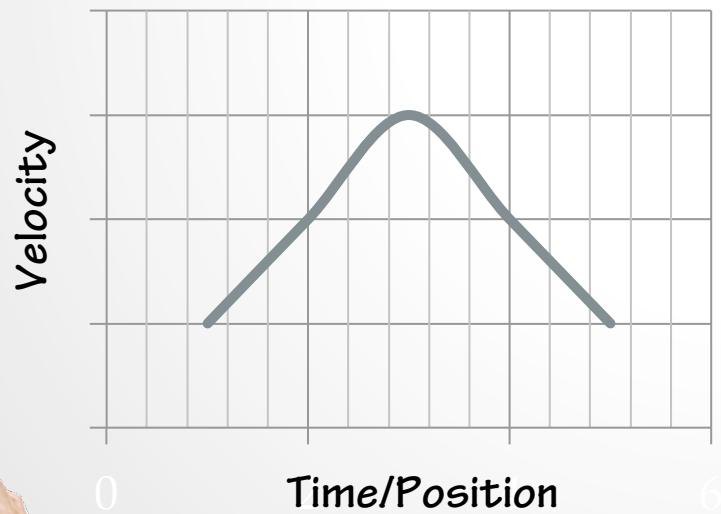
# Initial Representation

- Represent character as a set of B-splines
- B-Splines are efficient in representing the shape



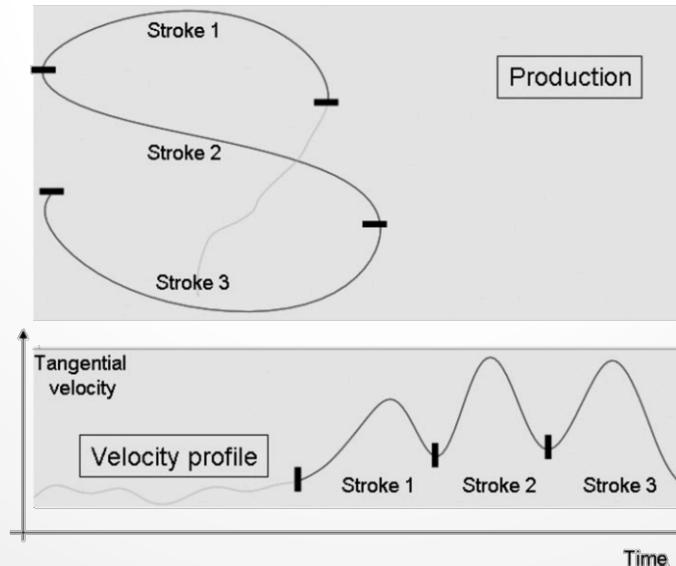
# Handwriting Kinematics

- Handwriting is a ballistic process
- A stroke creates a bell shaped velocity profile



# Stroke Segmentation

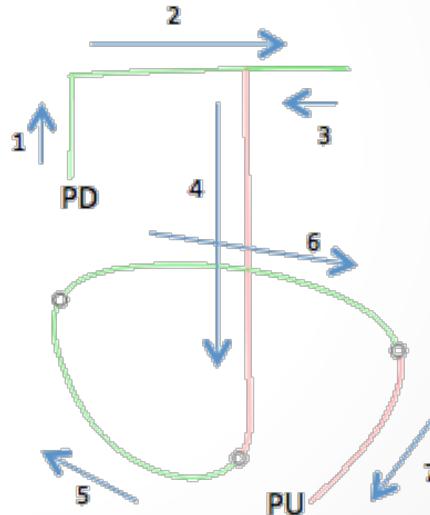
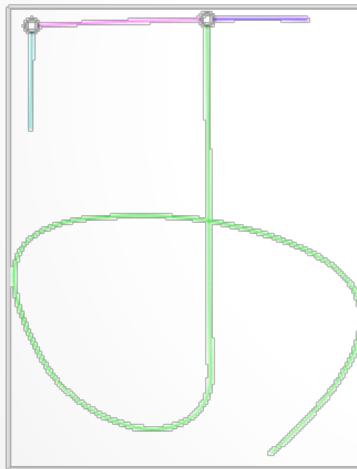
- User trajectory information
  - If not available, reconstruct
- Find points where Velocity slows down
- Segment strokes at those points



Kendel et al. (2010)

# Stroke-based Representation

- Character represented as strokes
- Natural representation



# Visual Information

- Length
- Divergence
- Length-Breadth Index
- Size
- Openness  
(= Divergence/Length)
- Compactness  
(= Length/Size)
- Average Curvature
- Distinctivity
- (=DTW Cost)
- Ascendance/  
Descendance
- Circularity/  
Rectangularity



# Production Information

- Stroke Counts
  - Primitive, Pen, Disjoined
- Changeability  
(= Upstrokes/Downstrokes)
- Dis-fluency
- Entropy
- N-gram Model
- Angle-Based metrics
  - Major, Initial, Divergence, Pen-Drag, Inter-stroke
- Pen-Drag Distance



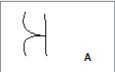
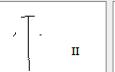
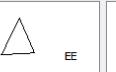
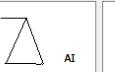
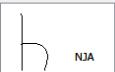
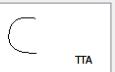
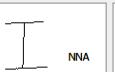
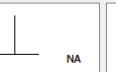
# Cognitive Information

- Handwriting Top-Down Process
- Information required to plan the trajectory of the character
- *No. of Landmark Points* correspond to that information
- *Ramer-Douglas-Peucker simplification algorithm*

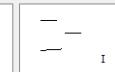
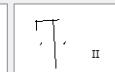
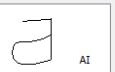
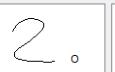
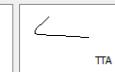
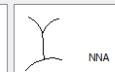
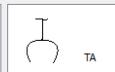
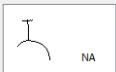
# Metrics of Scripts Vs Properties of Characters

- Most properties discussed where character properties
- How to define properties of Scripts?
  - Average?
  - Distribution?

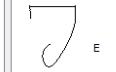
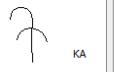
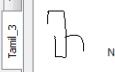
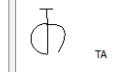
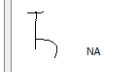
# Development of Medieval Tamil Script

Tamil 1

Tamil 1										
Tamil 2										
Tamil 3										

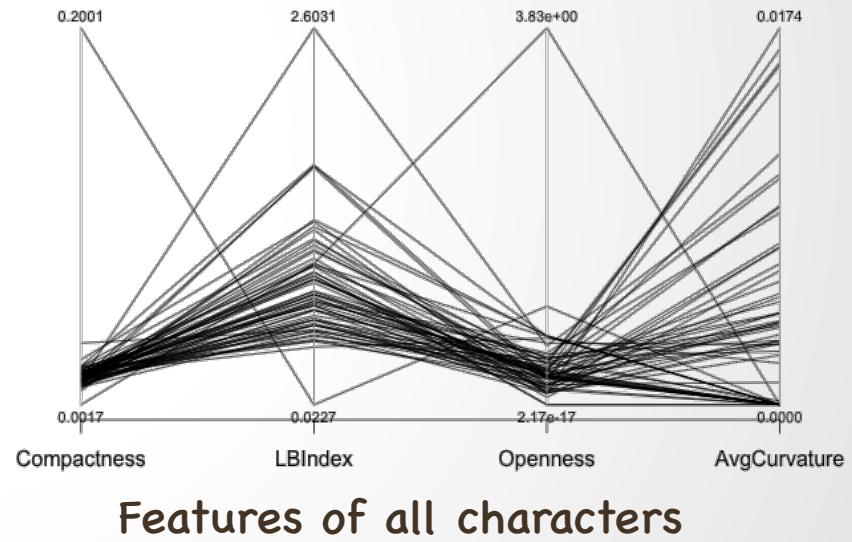
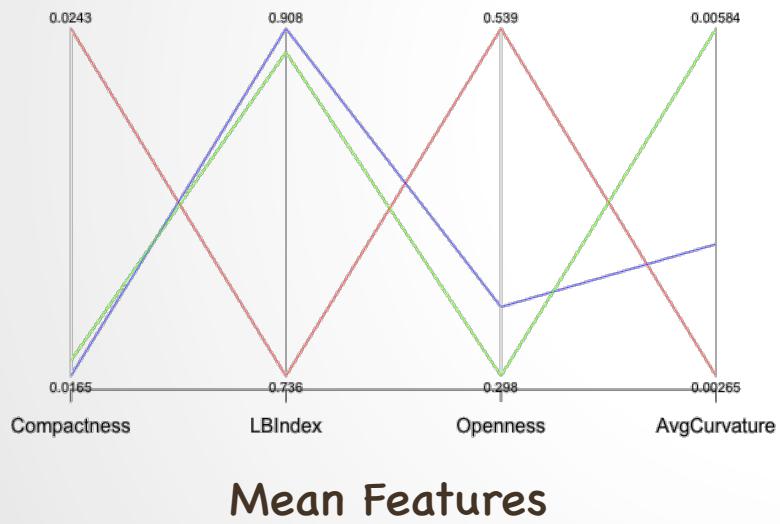
Tamil 2

Tamil 1										
Tamil 2										
Tamil 3										

Tamil 3

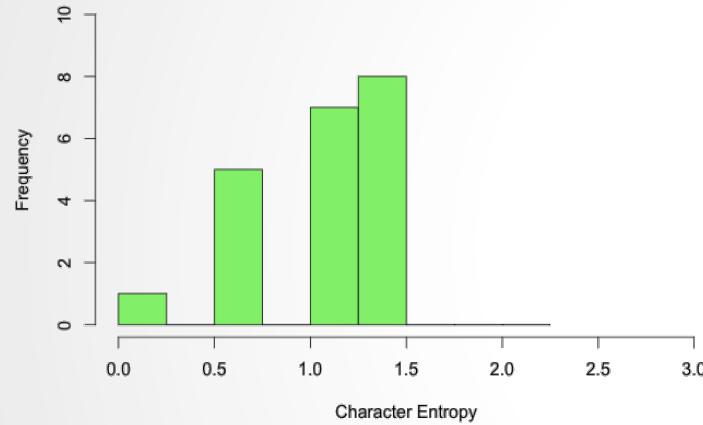
# Visual Features

- Parallel Co-ordinate plots of various visual features

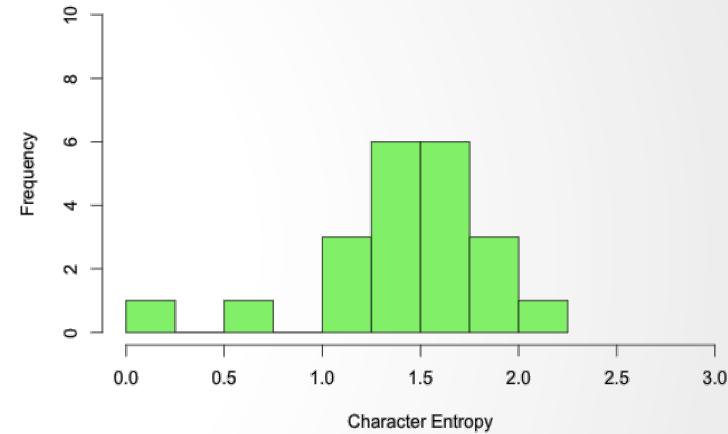


# Distribution of Entropy

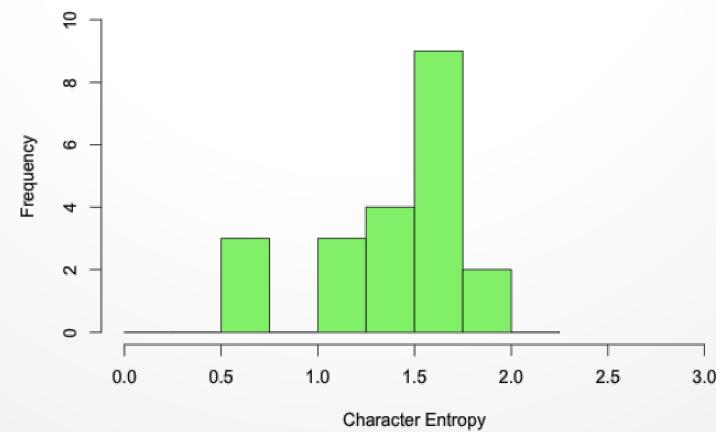
- Histogram of Entropy across the 3 scripts



Tamil 1



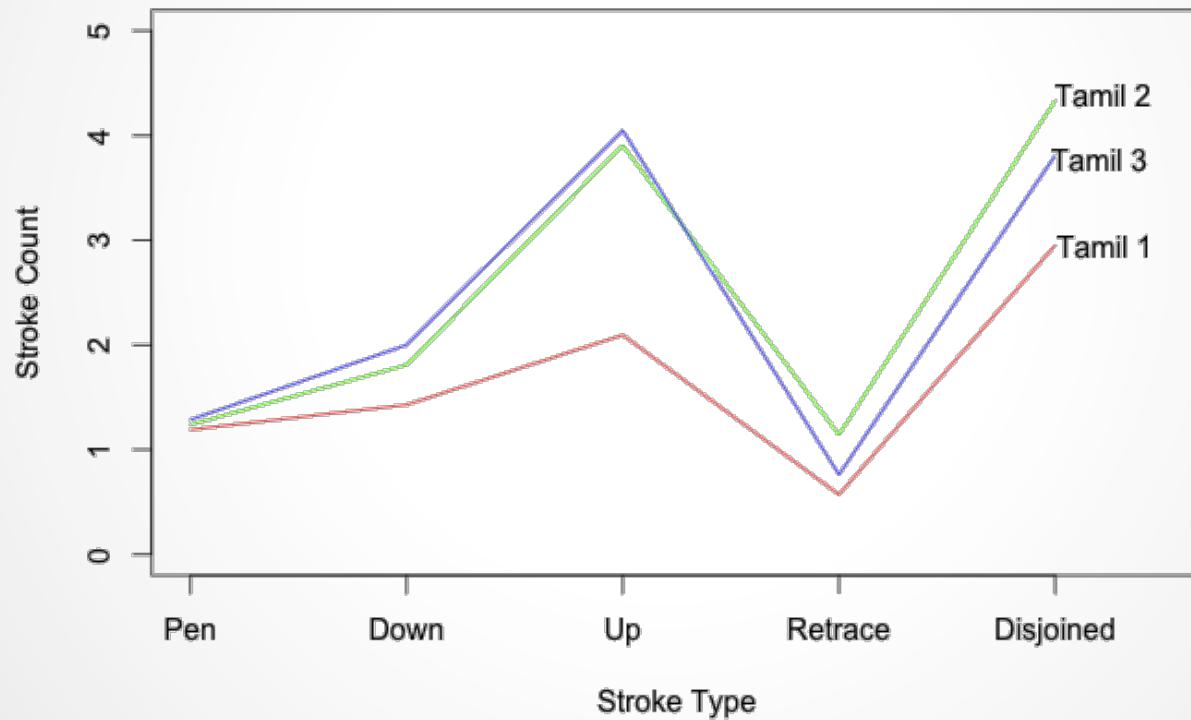
Tamil 2



Tamil 3

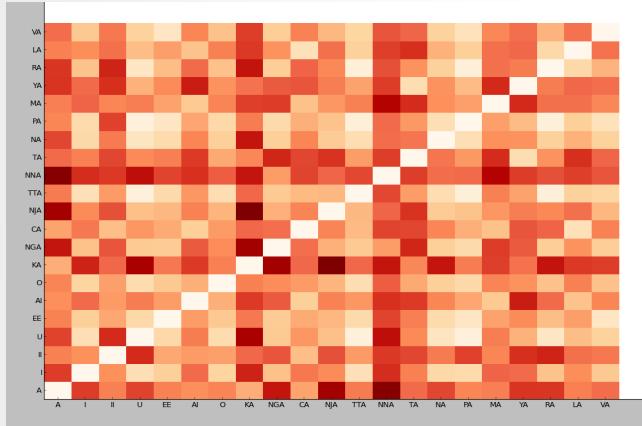
# Stroke Counts

- Change in different types of stroke counts

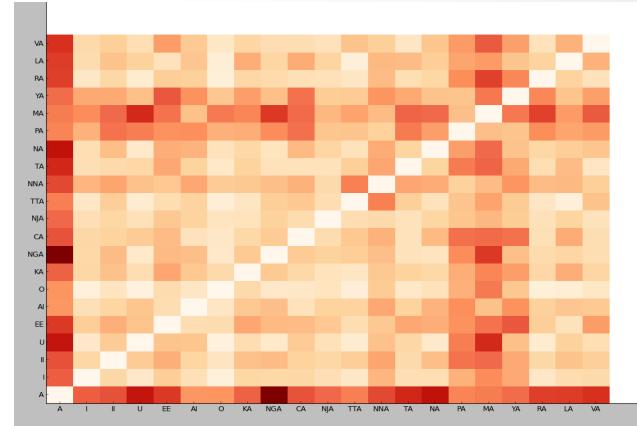


# Distinctivity Change

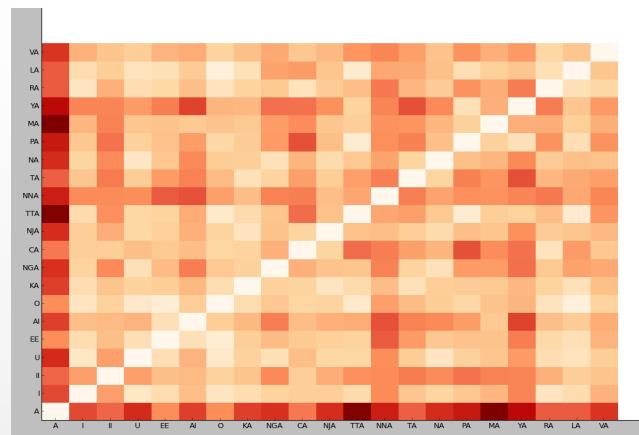
- Heatmaps for Distinctivity



Tamil 1



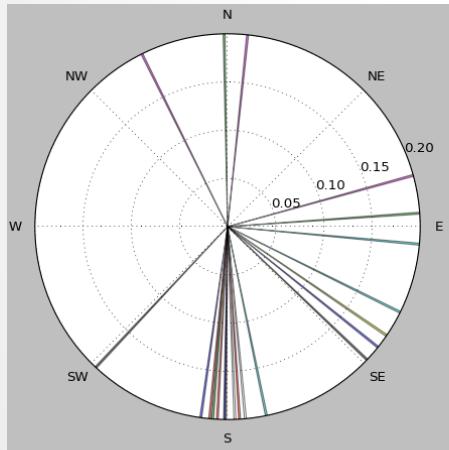
Tamil 2



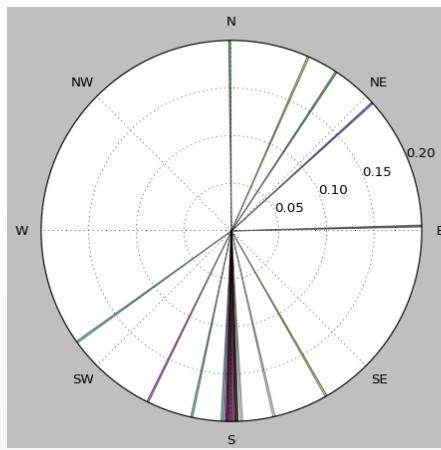
Tamil 3

# Major Pen Direction

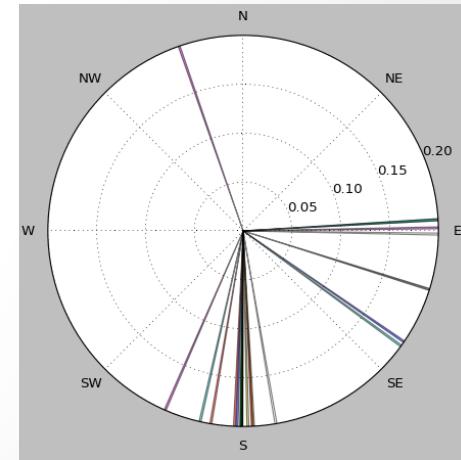
- Major stroke direction seems to have been constant – Downwards



Tamil 1



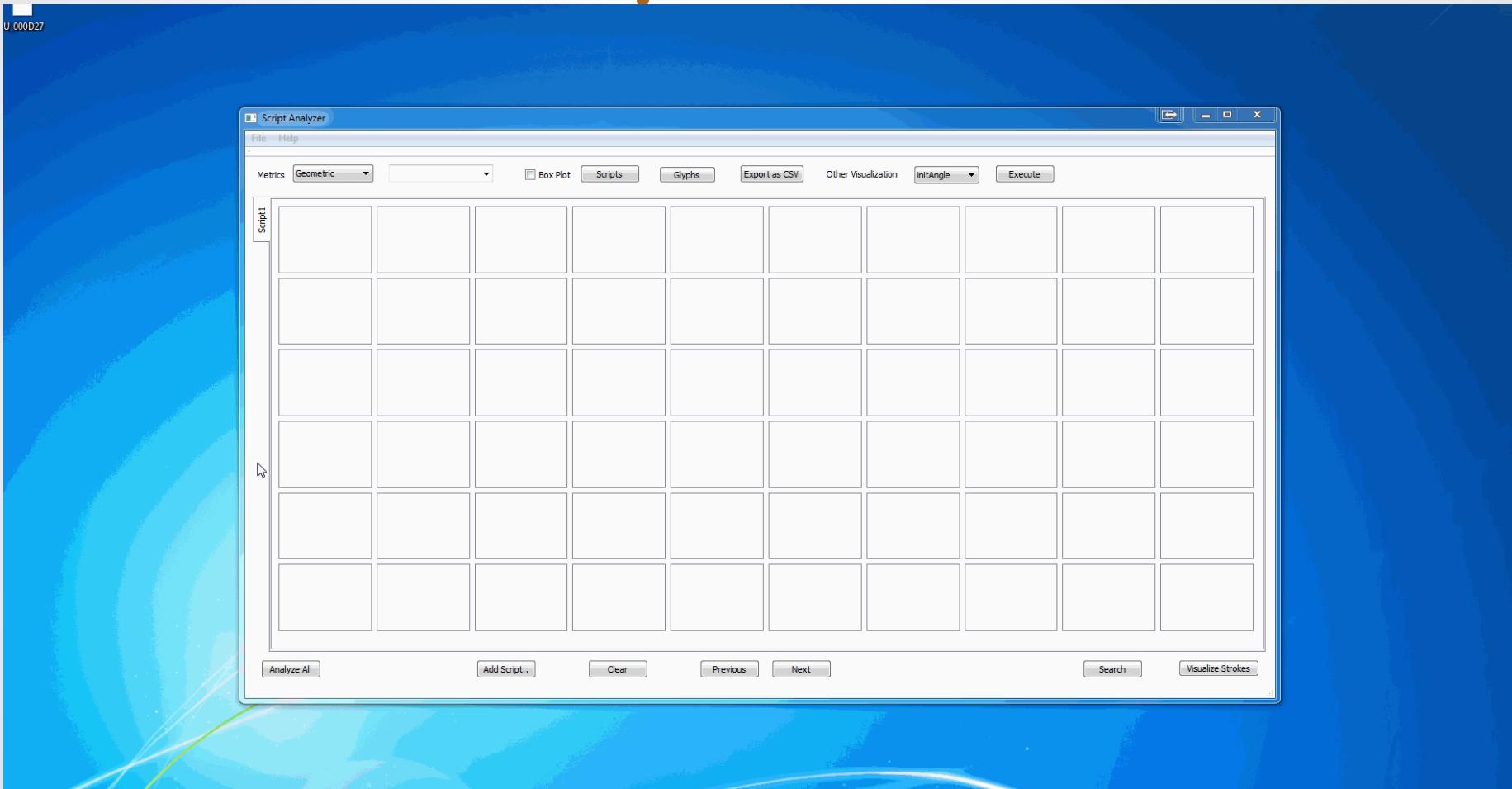
Tamil 2



Tamil 3

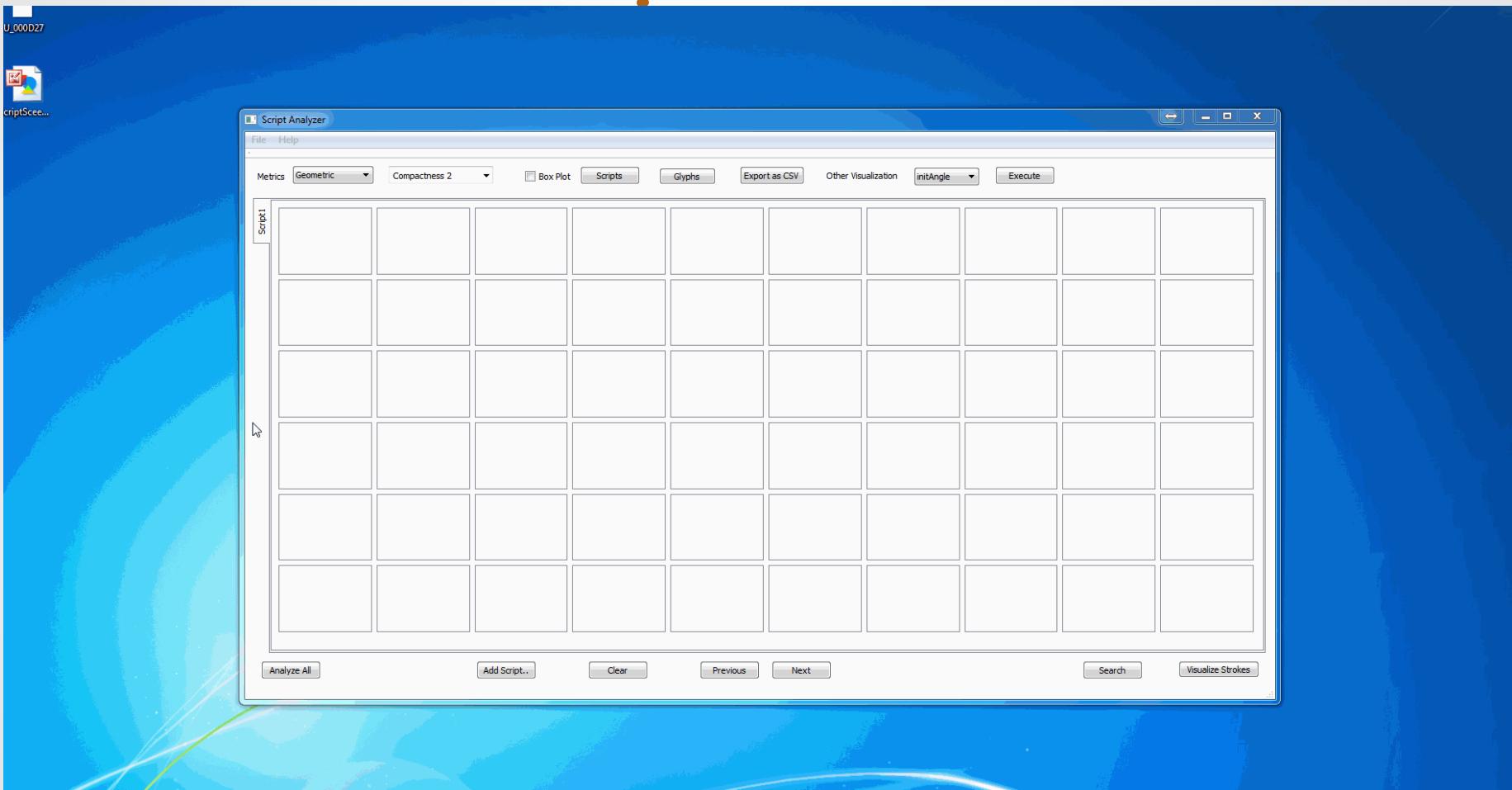


# Prototype Demo! - 1





# Prototype Demo! - 2



# Future Work

- More Metrics ?
  - Was anything important that captures a salient feature got missed out?
- Most metrics are aimed for descriptive analysis
  - How well do this perform for classification work?



*atha bodhisattvo\* [...] viśvāmitram ācāryam evam āha - katamāṁ me bho upādhyāya lipiṁ śikṣāpayasi ? brāhmī kharoṣṭī puṣkarasāriṁ aṅgalipīṁ vaṅgalipīṁ magadhalipīṁ [...] drāviḍalipīṁ [...] cīnalipīṁ lūnalipīṁ hūṇalipīṁ [...]. āsāṁ bho upādhyāya catuṣṣaṣṭīlipīnāṁ katamāṁ tvam śiṣyāpayiṣyasi ?*



University of  
St Andrews

FOUNDED  
1413