

# Mysterious Ms ?

- Metrics
- Maximum Parsimony (MP)
- Maximum Likelihood (ML)
- Markov Chain Monte Carlo (MCMC)

# Data structures

- Distance data (metrics)
  - approach used in linguistics to analyze Swadesh lists
- Character data (splits)
  - approach used in many biologically-inspired recent lexicostatistic analyses



# Innovations of Gray & Atkinson (2003)

- Use MCMC instead of simple metrics
- Underlying process is replacement of lexemes used for specified meanings: meaning change
- Dating without a strict “glottoclock”

# Problems

- Root is hand-picked
- Resolution of deeper ancestral nodes is still not clear
- Anatolian-hypothesis dates depend on details of early grouping and branching

# Simple methods are not necessarily bad

- Text automatically selected from internet
- Classified by language through 3-gram vectors of normal orthography
- Similarity of 3-gram vectors can be used as metric
- Major IE subgroups are reasonably well attested

