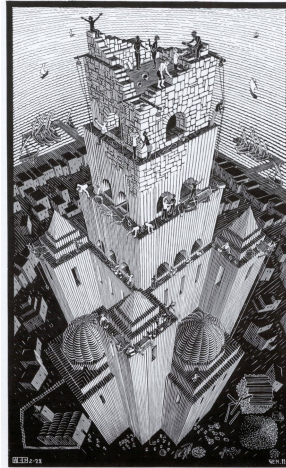


Null Models of Language Change

Dan Hruschka
Santa Fe Institute
dhrusch@santafe.edu

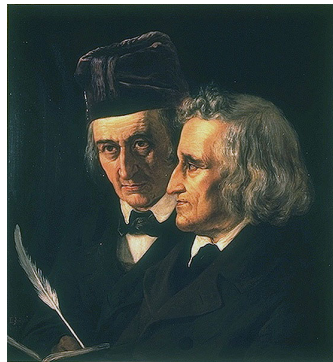


Questions in culture change

- Rates and kinds of change
- Systematicity of change
- Historical relationships & Contact processes
- Selection

Language as a cultural system

- Transmission is simple. Words are spoken, individuals hear them, and they are learned or reinforced
- High degree of fidelity
- Advantages
 - Long history of study
 - Common system for comparing sounds and grammar across languages
 - Has own discipline
- Disadvantages
 - Unique?
 - Has own discipline



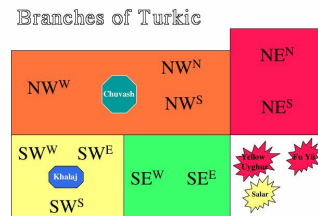
Jacob and Wilhelm Grimm (1855)

Opportunities

- A common mathematical platform for integrating diverse kinds of data
 - Linguistic
 - Genetic
 - Cultural Material

Questions in language change

- **Rates and kinds of change**
- Systematicity of change
- **Historical Relationships**
- **Selection**



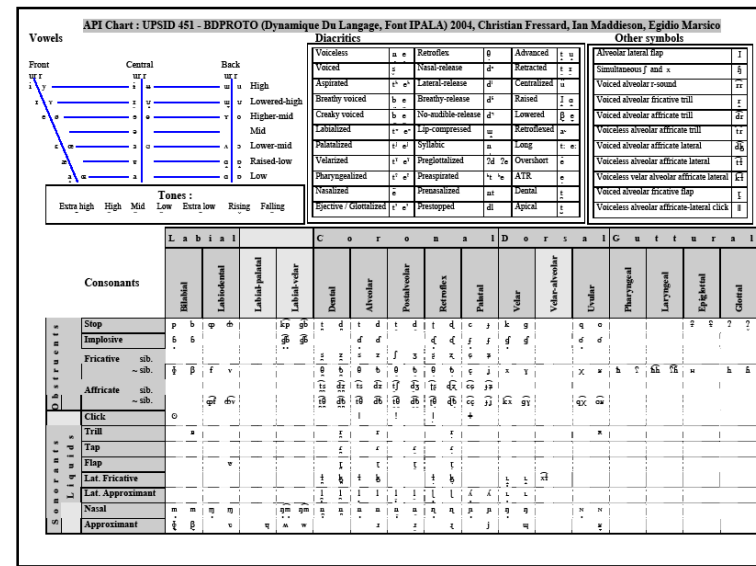
Lars Johanson (1998) The History of Turkic

Lecture 3 Layout

- Null models of phonological change
- Selection: homophony
- Testing historical relationships

Null models of phonological change

- “Although no comprehensive study of sound change that would allow us to distinguish common from uncommon innovations has ever been undertaken, historical linguists have acquired a sense of what kinds of change are likely to occur” (Blust 2004)



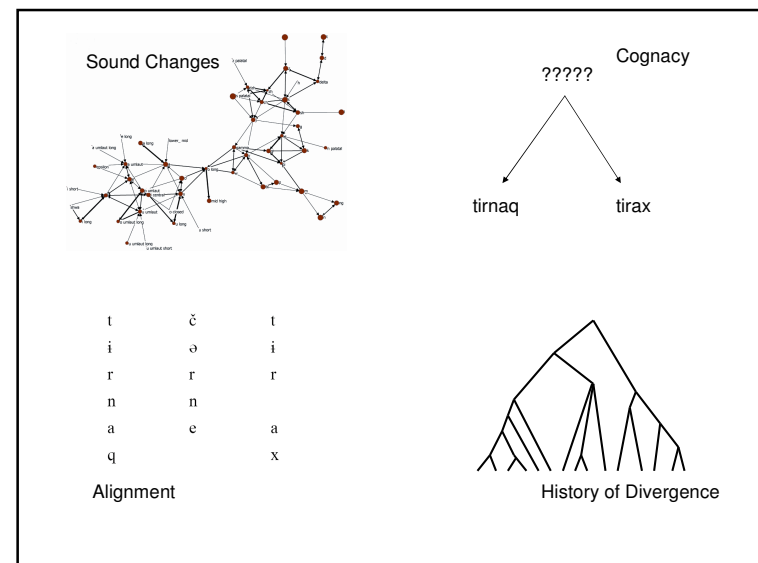
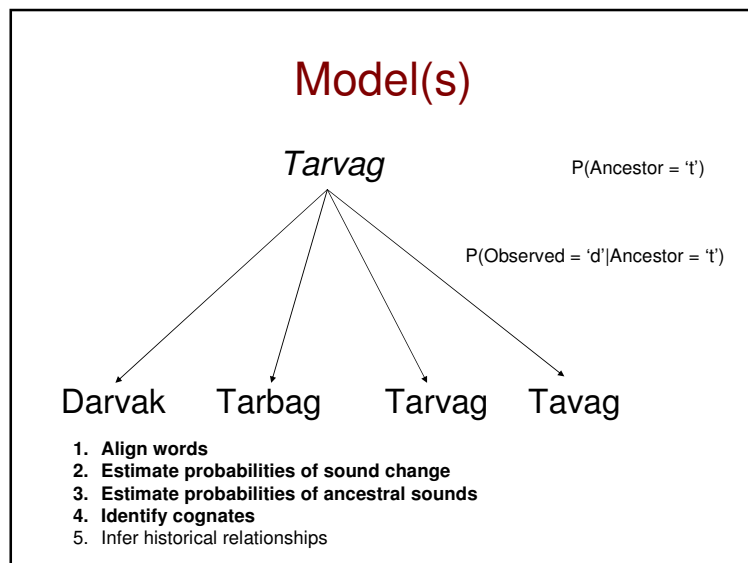
- Extant lexica (350 to 1400 words each) for 29 Turkic languages
- Phonetic transcriptions

Data

- Strengths
 - Data exists
 - Shallow time depth
 - CVCVC structure
- Weaknesses
 - Idealization
 - Synchronic

Screenshot of a browser window showing a table of Turkic lexica data. The table has columns for Proto-Turkic, Alt., Meaning, Karakhanid, Turkish, Tatar, Uzbek, Uighur, Sary-Yughur, and Azerbaidzhan. The data is organized into rows, with each row representing a specific word or morpheme and its variations across different Turkic languages and historical periods.

Proto-Turkic	Alt.	Meaning	Karakhanid	Turkish	Tatar	Uzbek	Uighur	Sary-Yughur	Azerbaidzhan
*köl	1140	ashes	köl (MK)	köl	köl	köl	köl	köl	köl
*küpük	988	1 bark (n.) 2 shell	qabiq (Tef.) 1, 2	kabuk 1, 2	qabiq 1, 2	qabiq 1, 2	qobuq (R, o-t) 1,		Gabik 1, 2
*karin	770	belly	qarin (MK, KB)	karin	qarin	qarin	qarin	qarin	Garin
*bedü-k	113	1 big 2 high	bedük 1 (MK, KB)	bujuk 1	bujuk 2	bujuk 1, 2	bujuk 1, 2	bezik 1	bojuk 1
*kül	1157	1 bird 2 duck	quš 1 (MK, KB)	kuš 1	qoš 1	quš 1	quš 1	Quš 1	Guš 1
*isä-	610	to bite	isä- (MK)	isä-					
*diš	2250	tooth	tiš (MK), tiš (KB)	diš	tiš	tiš	tiš, tiš	dis	diš
*kara	734	black	qara (MK, KB)	kara	qara	qara	qara, dial. qare	Gara	Gara
*kän	1068	blood	qan (MK)	kan	qan	qan	qan	qan	Gan
*kernäk	1055	1 bone 2 spongy bone		kernik 1	kinek 2	komik, gamik (dial.)			
*mäme (*beme)	1292	1 breast (fem.) 2 nipple		meme 1, Old Ormami, mama 1			mämi 1		mämi 2
*jak-	367	1 to burn (tr.) 2 light	jaq- 1 (MK)	jak- 1	jaq- 1, jaqtı 2	jaq- 1, jaydu 2	jaq- 1		jak- 1



$$1) \quad p(L_1, L_2, \dots, L_N) = \prod_{CognateClasses} \sum_{Ancestors} p(Ancestor) \prod_L p(Word_L | Ancestor)$$

$$2) \quad p(\text{Word}_L \mid \text{Ancestor}) = \sum_{\text{alignments}} p(\text{Word}_L, \text{alignment} \mid \text{Ancestor})$$

$$3) \prod_{positions} p(Word_L(position) | Ancestor, position) =$$

Likelihood we'll work with

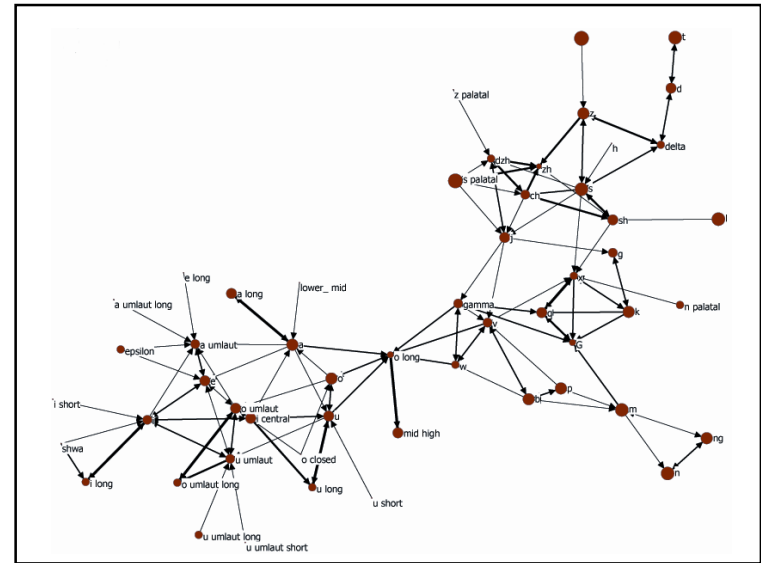
$$p(L_1, L_2, \dots, L_N \mid p(A), p(s \mid A), alignment) = \prod_{CognateClasses} \prod_{positions} \sum_{Ancestors} p(A) \prod_L p(s \mid A)$$

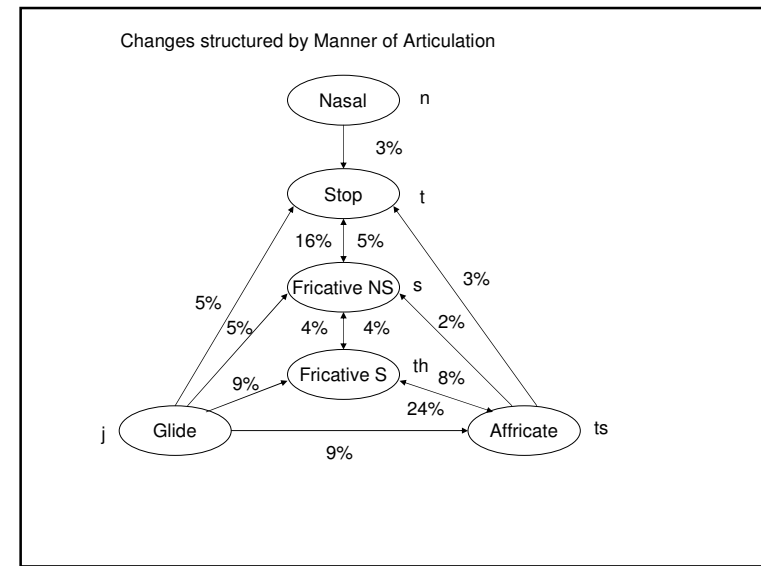
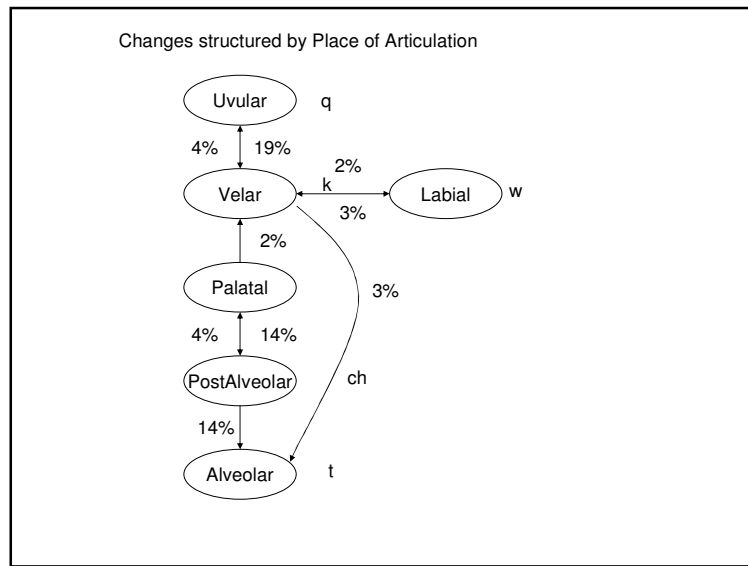
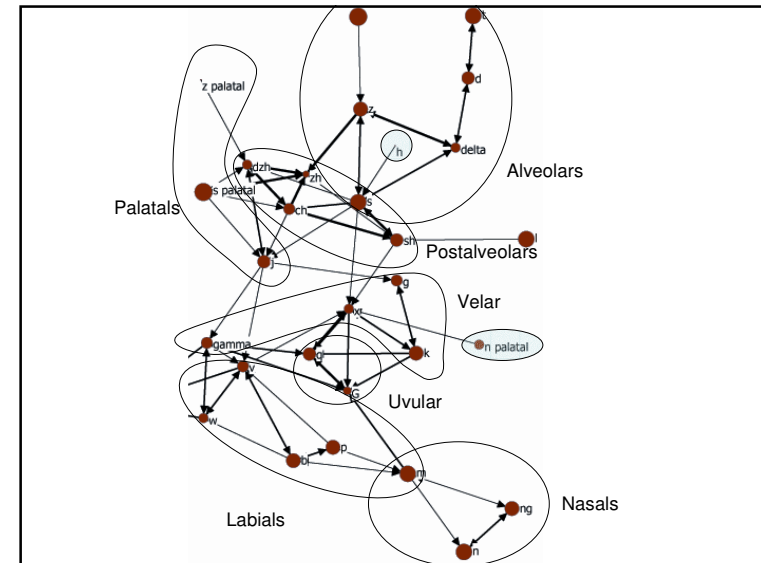
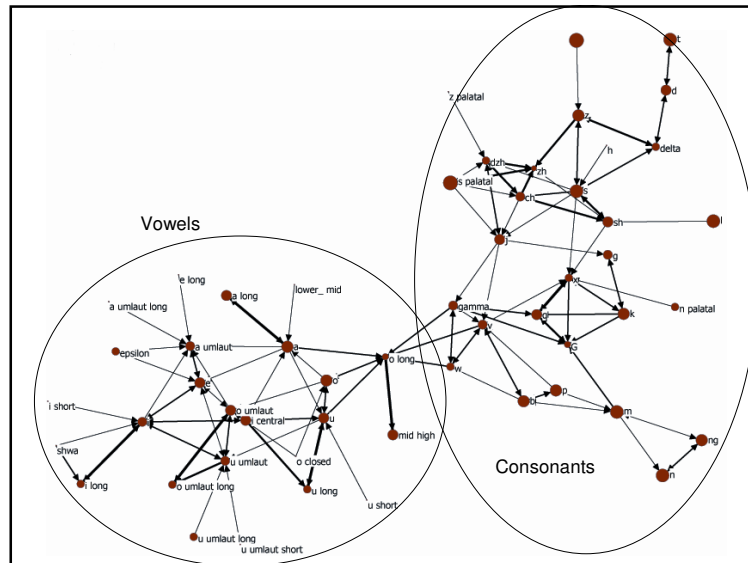
$P(s|A)$ = probability of observed sound given an ancestor
 $P(A)$ = probability of ancestor

Some assumptions were required to get here

Descriptives

- Estimating sound change probabilities
- Alignments
- Cognate judgments





Alignment: Etymology 14 “fingernail”

t t t t t t t d d d t t t t ċ t t d d t
i i i i i i i ə i i i i i i ə i i i i i
r
y y n n y n n m n n y y G n n y y n
a a a a a ɔ a a a a a a a e a a a a
q q k q q q q q G q x q q q x k q q q

Natural measure of cognacy

$$p(\text{alignment} \mid p(A), p(d \mid A)) =$$

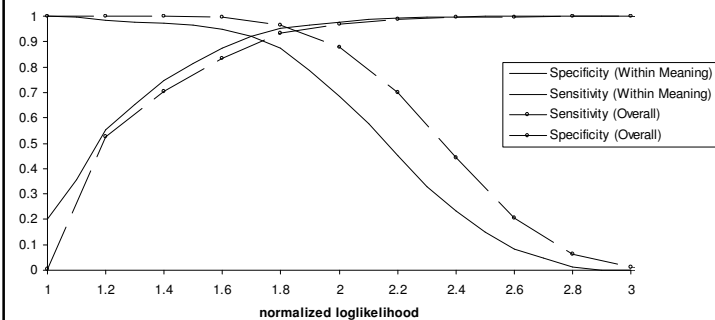
$$\prod_p \sum_A p(A) p(s_{1p} \mid A) p(s_{2p} \mid A)$$

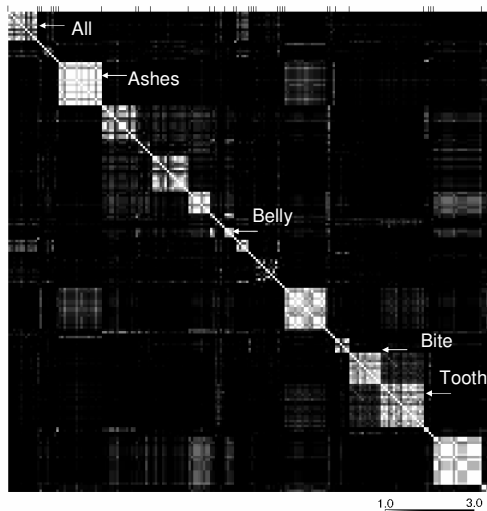
Divergence between words

$$NL = \ln(\text{ML}(X, Y)) - 0.5(\ln(\text{ML}(X, X)) + \ln(\text{ML}(Y, Y)))$$

Distance	Word1	Mean1	Word2	Mean2
0.35	a:t	name	at	horse
0.75	jat-	to sleep	jad-	to spread
1.5	Guš	bird	qoš	pair
2.0	čum-	to dive	Jum-	to close
3.0	qus-	to vomit	qoš-	to join
4.0	köj-	to burn	köz	eye
5.0	ben	we	mun	defect, sin
6.0	ini	brother	eki	two

Cognate Identification





2. Selection against homophony?

- Shrimb and Shrimp
- 50 and 15
- Alternative accounts
 - Phonetic reflection
 - Lexical replacement

Homophones in Salar

jel: wind, year

jaš: age, tear

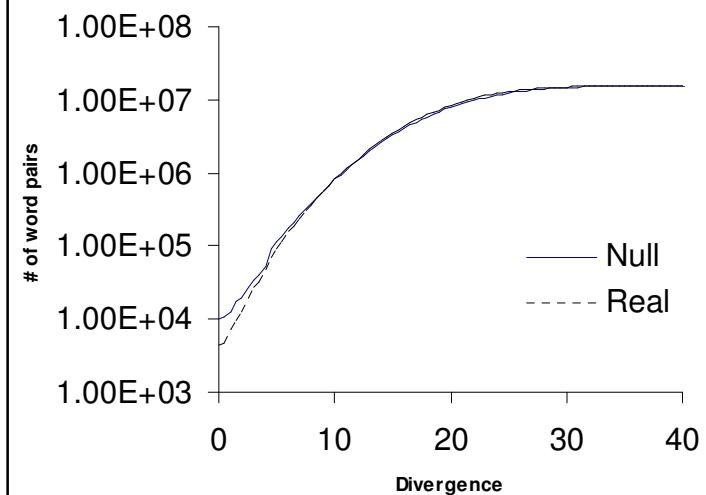
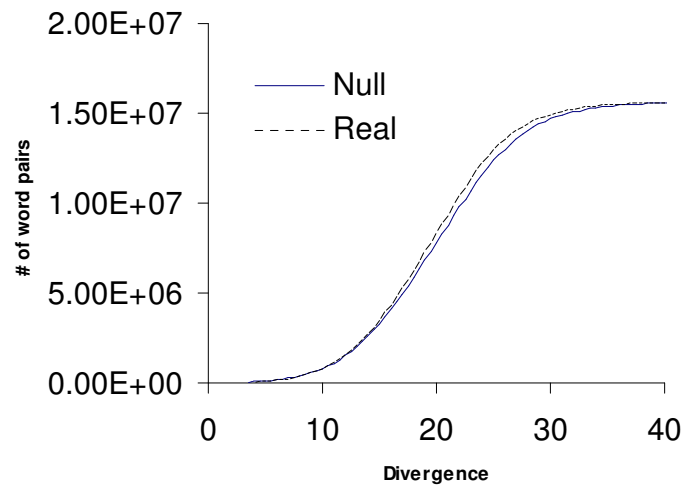
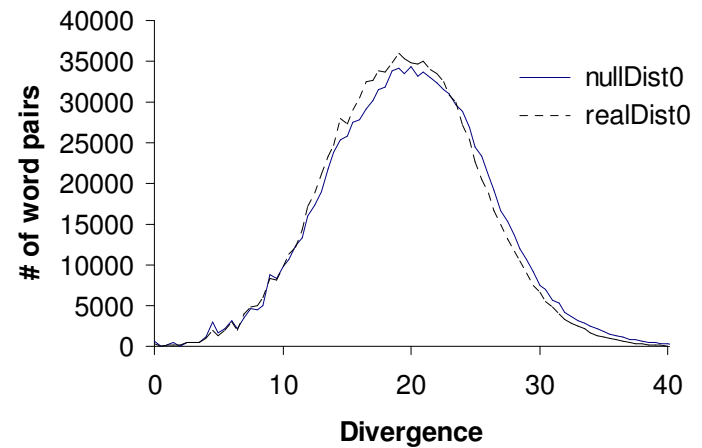
jiz: face, hundred

Selection against homophony?

- What would a lexicon “like” Karakalpak look like?
- Need a null model
- Build the lexicon from the ground up based on certain assumptions (word lengths, phonotactics)

Some neutral models

- Bag of sounds
- Position dependent probabilities
- 1 step phonotactic probabilities
- 2 step phonotactic probabilities
- Hierarchical construction based on syllables
- Above with vowel harmony



Is there selection against homophones?

- Don't know yet
- Maybe our neutral model is missing something else?
- What's going on with longer range interactions? Syllabic structure and vowel harmony.

- ## Is there selection against homophones?
- Don't know yet
 - Maybe our neutral model is missing something else?
 - What's going on with longer range interactions? Syllabic structure and vowel harmony.

Recap

- Modeling extant data to ask questions about
 - **Rates and Kinds of Change:** They reflect physical constraints on articulation
 - **Selection:** Tentative evidence for homophone avoidance, but need better null models
 - Historical Relationships and Contact
 - Regularity

- ## Recap
- Modeling extant data to ask questions about
 - **Rates and Kinds of Change:** They reflect physical constraints on articulation
 - **Selection:** Tentative evidence for homophone avoidance, but need better null models
 - Historical Relationships and Contact
 - Regularity

Historical Relationships & Contact

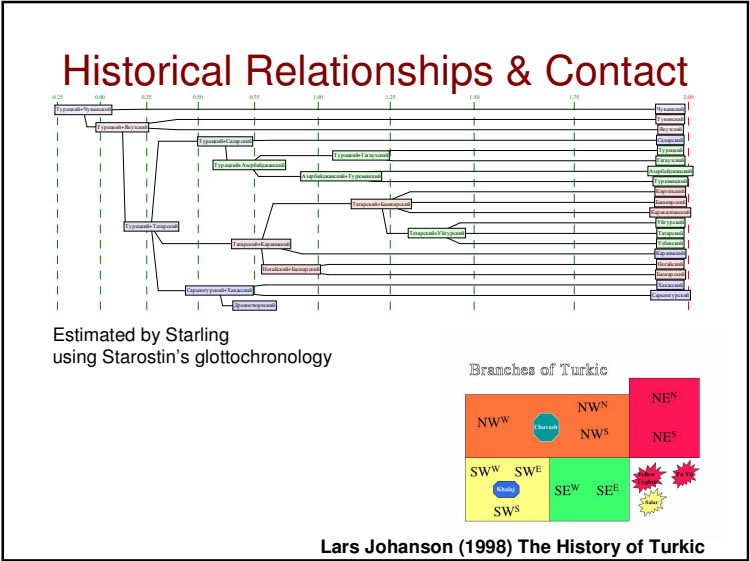
Estimated by Starling
using Starostin's glottochronology

Branches of Turkic

Branches of Turkic

NW ^W	NW ^N	NE ^N
SW ^W	NW ^S	
SW ^S	SE ^W	SE ^E

Lars Johanson (1998) The History of Turkic

[illegible]

Historical Relationships & Contact

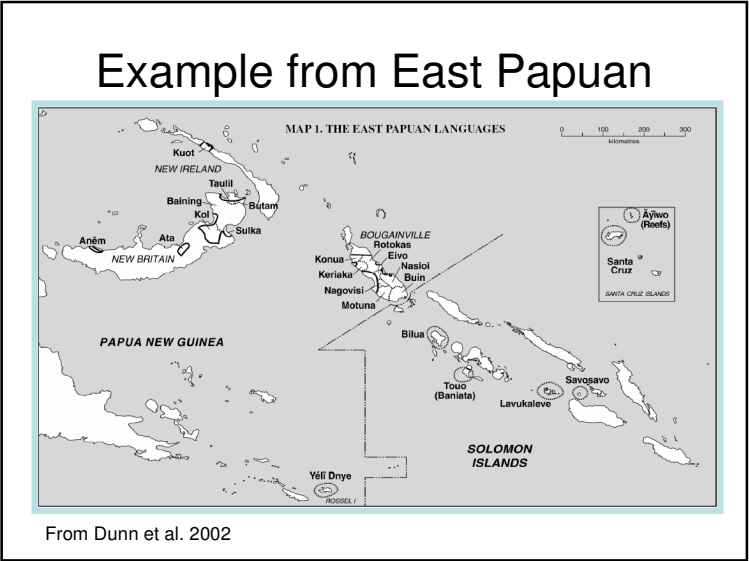
Estimated by Starling
using Starostin's glottochronology

Branches of Turkic

NW ^W	(Central)	NW ^N	NE ^N
		NW ^S	
SW ^W	(South)	SE ^W	SE ^E
SW ^S		SE ^E	

Lars Johanson (1998) The History of Turkic

Example from East Papuan

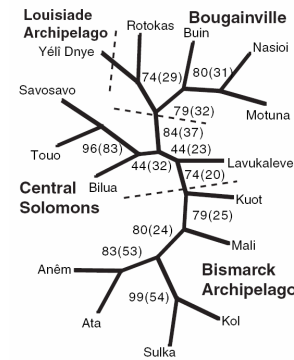


Example from East Papuan

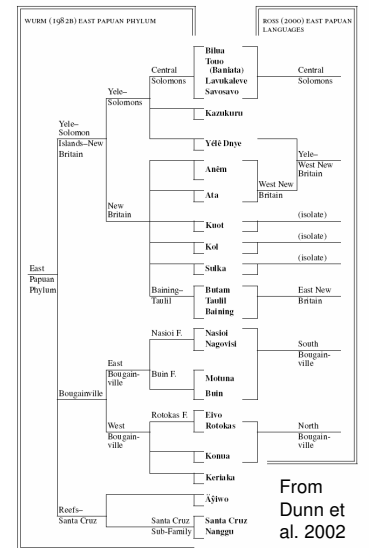
Setting

- Region of great linguistic diversity (~1000 languages)
- Overlapping settlement of Western Oceanic language groups (~3500ybp) and 'East Papuan' language groups (unknown arrival dates)

Debated Classifications



From Dunn et al. 2005



From
Dunn et
al. 2002

Competing Theories

- Long range divergence (Dunn et al. 2005).
"Most plausible...divergence of Papuan languages from a common ancestral stock, as part of late Pleistocene dispersals."
- Recent contact (Terrell et al. 1997)
 - Recent West Oceanic dispersal
 - Region of endemic bilingualism
 - Between-language transfer of syntactic features

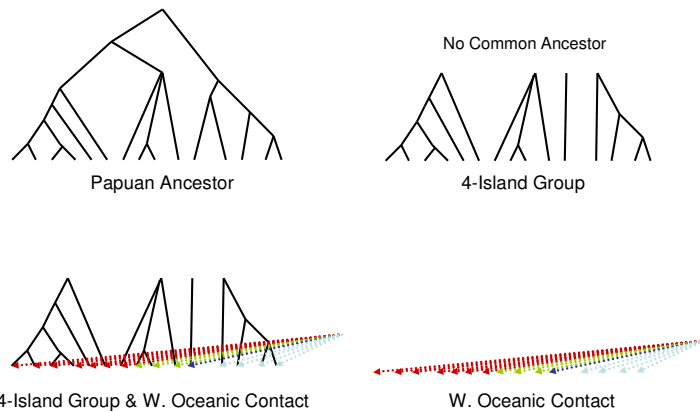
Data

- 15 "East Papuan" languages
- 16 West Oceanic languages
- 11 sound system & 114 syntactic features
- Dichotomous (presence/absence)
- Example of feature: presence of copula

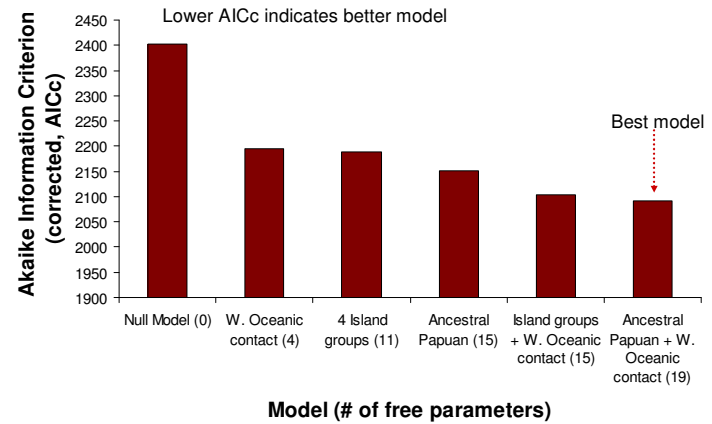
"I am a person" vs. "Ya chelovek"

Data from Dunn et al. (Science 2005)

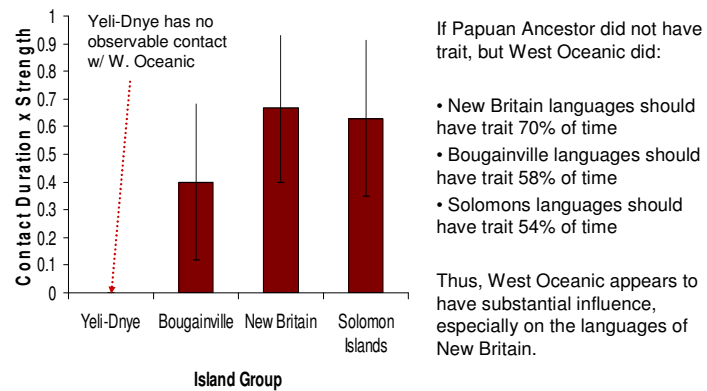
Model Variants



Model Comparison

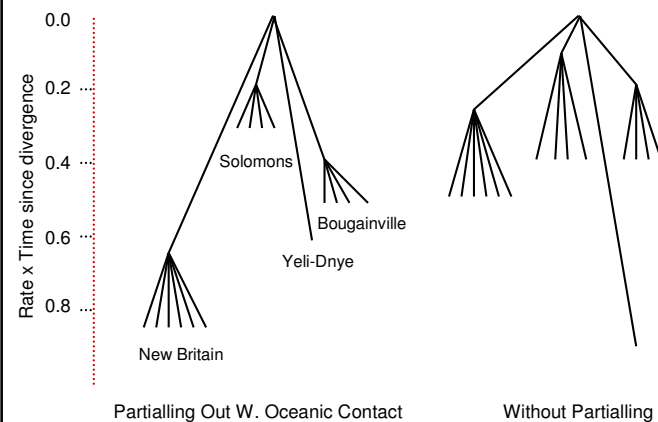


Estimated contact & influence



Partialling Out Contact:

Island Groups become more distant from "Common Ancestor"



Conclusions

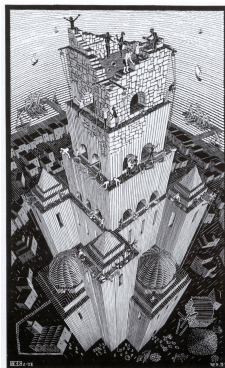
- An “East Papuan Ancestor” and West Oceanic contact both account for some similarity between East Papuan languages
- A “tree structure” might arise from a combination of processes, including more recent contact
- External data (in this case West Oceanic language data) is useful for assessing contact

Lessons

- Time depths of historical processes may be much more shallow than originally expected (Rather than 15,000 y divergence, it may be convergence over last 3000 y).
- The importance of using external data (historical, archeological, linguistic) to develop testable models.

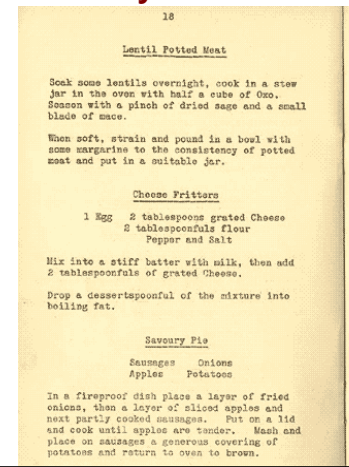
Concluding Questions

- What if anything can language change tell us about culture change in general?



Other kinds of cultural systems?

- Does such data exist?
- Is there sufficient data to test hypotheses about rates of change hybridization?
- An IPA, a common metric, for other cultural systems? For cuisine?
- Are systems with an “IPA” different from others?



Recap

- Lecture 1. Models of cultural diversity
 - Human communities vary greatly in social organization, behaviors, and knowledge
 - Many theories to explain diversity
 - » Darwinian processes
 - » Neutral theory
- Lecture 2. Simple models of social learning
 - Use of mathematical models to hone intuitions
 - Comparing predictions of models against real observational data
 - Experimental evidence for how people learn from others

Recap

- Today. Null Models of Language Change.
 - Null models of phonological innovation
 - Is there selection against homophony?

Wrapping up

- Diversity in human communities leaves many open questions
- We have reviewed some tools available to devise and test and compare theories of cultural change and diversity
- Data, data, data
- In addition: Careful case studies of change

Data

- Cross-cultural datasets
 - Human Relations Area Files (Yale University)
 - Ethnography
 - Archeology
 - Standard Cross-cultural sample (Murdock and White 1966)
 - Western North American Indians (Jorgensen 1981)
- Linguistic datasets
 - Evolution of Human Languages Project
 - University of Auckland (Austronesian, Bantu languages)
 - Language corpuses (British National Corpus)
- Forager Toolkits (Collard, U. of British Colombia)
- Greek City-States