# Proving Dene-Yeniseian genealogical relatedness

Johanna Nichols, UC Berkeley
johanna@berkeley.edu

*Dene-Yeniseian Symposium*

Fairbanks and Anchorage, February 2008

***Stock*** = the oldest phylogenetic grouping that can be both demonstrated and reconstructed.

How can relatedness be proved (independent of reconstruction)?

***Stock*** = the oldest phylogenetic grouping that can be both demonstrated and reconstructed.

How can relatedness be proved (independent of reconstruction)?

Show that the number of resemblant elements significantly exceeds what would be expected by chance.

The statistics of searches

Specific, similar, and generic segments

Shared paradigms and subparadigms

Resemblant lexical items

Conclusion

## *The statistics of searches*

## *The statistics of searches*

The **individual-identifying** statistical threshold:

1/7000 or 0.000143

(since there are about 7000 languages on earth)

plus a conventional level of statistical significance:

| | | | | |
|---|---|---|---|---|
| 0.05 | 1/350,000 | or | 0.000 0029 | or 3 / 1,000,000 |
| 0.01 | 1/700,000 | | 0.000 0014 | or 1 / 1,000,000 |

(Rule of thumb: **5 zeroes** after the decimal point)

## *The statistics of searches*

This threshold can be met with shared morphological paradigms:

(1) Germanic suppletive paradigm for 'good' : 'better':

| | | |
|---|---|---|
| English | *good* | *better* |
| German | *gut* | *besser* |
| Swedish | *god* | *bättre* |

(2) Gender-number suffixes in Afroasiatic determiners (Greenberg 1960). Analysis (a) treats gender as neutralized in the plural; (b) treats it as syncretized.

| | (a) Sg. | Pl. | (b) Sg. | Pl. |
|---|---|---|---|---|
| Masc. | *-n* | } *-n* | *-n* | *-n* |
| Fem. | *-t* | | *-t* | *-n* |

(calculations to follow later)

## *The statistics of searches*

Sources of (statistical) freedom in defining lexical comparanda:

Resemblant consonants

## *The statistics of searches*

### Sources of freedom in identifying consonants

Two-consonant root:

- $C_1$ and $C_2$ (in that order)
- Each C is resemblant (*not* defined by regular correspondences or identity)
- Phonotactics (positioning of vowels, if any) irrelevant

So these represent the same CC root:

| | |
|---|---|
| *qof, geb, akpu, xpi* | (**similar** consonants) |
| plus: *hemi, ogw* | (**generic** consonants) |

## *The statistics of searches*

Other sources of freedom:

Semantics: same sense; a few senses' leeway; several senses' leeway

Form: strict parse; selective parse

Selective: *kep, kedep, dekp, pek* (all K-P)

## *The statistics of searches*

Calculation of probability: This is a *search* with several degrees of freedom.

What are the chances of tossing heads? 50%.

What if you get up to three attempts? 87.5%.

## *The statistics of searches*

Calculation of probability: This is a *search* with several degrees of freedom.

What are the chances of tossing heads? 50%.
What if you get up to three attempts? 87.5%.

Cumulative probability = $q_1 + q_2 + ... + q_i$

where $q_{i+1} = p\ (1 - q_i)$

p = event probability

q = cumulative probability;

$q_i$ = cumulative probability after the *i*-th trial

## *The statistics of searches*

Example:

Identical (particular) consonant: p = 0.05

(Average consonant inventory is about 20.)

Similar consonant: 3 distinctive features' leeway or about 1/7 of consonant inventory: p = 0.14

Generic consonant: 5 distinctive features' leeway or about 1/4 of consonant inventory: p = 0.23

Identical CC root: p = 0.0025

Similar CC root: p = 0.02

Generic CC root: p = 0.05

## *The statistics of searches*

Example:

Identical (particular) consonant: p = 0.05

(Average consonant inventory is about 20.)

Similar consonant: 3 distinctive features' leeway or about 1/7 of consonant inventory: p = 0.14

Generic consonant: 5 distinctive features' leeway or about 1/4 of consonant inventory: p = 0.23

Identical CC root: p = 0.0025

**Also, cognate roots with C's showing regular correspondences.** Same as identical.

The number of resemblant two-consonant roots required in a binary comparison, with varying degrees of phonological and semantic leeway. Similar calculations for one-consonant roots. ($p_2$ = probability of two-consonant root; n = number of trials; entries are minimum numbers of words required to reach significance at < 0.05.)

| | 1 sense: | | | | 3 senses: | | | | 5 senses: | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | n = | 100 | 200 | 1000 | | 100 | 200 | 1000 | | 100 | 200 | 1000 |
| | | $p_2$ | | | | $p_2$ | | | | $p_2$ | | |
| Similar | 0.02 | 5 | 8 | 28 | 0.06 | 10 | 19 | 73 | 0.10 | 15 | 28 | 117 |
| " + select | 0.04 | 7 | 14 | 51 | 0.12 | 18 | 33 | 138 | 0.18 | 25 | 46 | 201 |
| Generic | 0.05 | 9 | 16 | 63 | 0.14 | 20 | 37 | 159 | 0.23 | 30 | 57 | 253 |
| " + select | 0.09 | 14 | 26 | 106 | 0.25 | 32 | 61 | 273 | 0.38 | 47 | 88 | 406 |
| One-consonant roots: | | | | | | | | | | | | |
| Generic | 0.14 | 20 | 37 | 159 | 0.37 | 45 | 86 | 396 | 0.54 | 64 | 120 | 567 |
| " + select | 0.27 | 34 | 65 | 294 | 0.54 | 64 | 120 | 567 | 0.72 | 80 | 155 | 744 |

The number of resemblant two-consonant roots required in a binary comparison, with varying degrees of phonological and semantic leeway. Similar calculations for one-consonant roots. ($p_2$ = probability of two-consonant root; n = number of trials; entries are minimum numbers of words required to reach significance at < 0.05.)

Red = best model of most actual long-range comparisons.

| | 1 sense: | | | | 3 senses: | | | | 5 senses: | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | n = | 100 | 200 | 1000 | | 100 | 200 | 1000 | | 100 | 200 | 1000 |
| | | $p_2$ | | | | $p_2$ | | | | $p_2$ | | |
| Similar | 0.02 | 5 | 8 | 28 | 0.06 | 10 | 19 | 73 | 0.10 | 15 | 28 | 117 |
| " + select | 0.04 | 7 | 14 | 51 | 0.12 | 18 | 33 | 138 | 0.18 | 25 | 46 | 201 |
| Generic | 0.05 | 9 | 16 | 63 | 0.14 | 20 | 37 | 159 | 0.23 | 30 | 57 | 253 |
| " + select | 0.09 | 14 | 26 | 106 | 0.25 | 32 | 61 | 273 | 0.38 | 47 | 88 | **406** |
| One-consonant roots: | | | | | | | | | | | | |
| Generic | 0.14 | 20 | 37 | 159 | 0.37 | 45 | 86 | 396 | 0.54 | 64 | 120 | 567 |
| " + select | 0.27 | 34 | 65 | 294 | 0.54 | 64 | 120 | 567 | 0.72 | 80 | 155 | 744 |

**An example of long-range comparison:**

**Nikolayev & Starostin's *North Caucasian Etymological Dictionary***

Nakh-Daghestanian (East Caucasian) root: (C)V(R)C

(C1 can be head gender marker)

West Caucasian root: C*(V)

C* = possibly complex

Matching strategy: Multiple selective parse

Match C1 or C2 of EC to any component of C*

If C1 of either language is unmatched it can be considered a gender prefix

Senses: Usually over 5 reported.

3600 reported cognates, 1800 of which have both WC and EC reflexes

No. trials: Wordlist = all available dictionaries for c. 40 languages.

## An example of long-range comparison:
## Nikolayev & Starostin's *North Caucasian Etymological Dictionary*

Model this search as a binary ND-WC comparison with these parameters:

Consonants: 1 similar (0.14), 1 arbitrary (0.5), total 0.07 for CC root

(Though in fact the possibility of calling C1 a gender marker makes this de facto not a root consonant, i.e. these are one-consonant roots.)

Selective parse (in addition to the arbitrary C1)

5 senses

Cumulative probability 0.35

Trials: ??? -- Estimate as 7200, twice the number of reported cognates

Successes: **1800** (cognates with WC representatives)

Needed: **2588** (a minimum, as the model above is very conservative)

## Another example:

## Ruhlen, *PNAS* 1998, Yeniseian - Na-Dene

Putative cognate sets for Proto-Yeniseian and Na-Dene from Ruhlen 1998, classified by phonological structure. All = Na-Dene forms from one or more of Haida, Tlingit, Eyak, Athabaskan. Ath. = Na-Dene forms from only (Proto-)Athabaskan.

| | All | Ath. only |
|---|---|---|
| 2 consonants, strict parse | 16 | 11 |
| 2 consonants, selective parse | 9 | 9 |
| 1 consonant, strict parse | 6 | 5 |
| 1 consonant, selective parse | 4 | 2 |
| 0 consonants | 1 | 1 |
| Total | 36 | 28 |
| Total using selective parse | 14 (39%) | 11 (39%) |
| Total with 2 consonants | 25 | 20 |

## Another example:
## Ruhlen, *PNAS* 1998, Yeniseian - Na-Dene

Parameters of Yeniseian-Athabaskan search:

- 3 senses (most sets contain 2 or 3 different glosses)
- Generic consonants
- 2 consonants (2-cons. sets extracted from the larger corpus)
- Selective parse (used especially for glottal stop, 39% of sets)
- 200-word Proto-Yeniseian wordlist (Athabaskan 1000+)

| | Found | Needed/200 |
|---|---|---|
| Total sets | 28 | |
| Total using selective parse | 11 (39%) | |
| Total with 2 generic consonants | 20 | 37 |
| (needed for selective parse) | | 61 |
| (Needed /1000 | | 273 ) |

## Another example:

## Ruhlen, *PNAS* 1998, Yeniseian - Na-Dene

Parameters of Ruhlen's Yeniseian-Athabaskan search:

- 3 senses (most sets contain 2 or 3 different glosses)
- Generic consonants
- 2 consonants (2-cons. sets extracted from the larger corpus)
- Selective parse (used especially for glottal stop, 39% of sets)
- 200-word Proto-Yeniseian wordlist (Athabaskan 1000+)

Additional complicating factor: both compared wordlists are reconstructed protolanguages.

| | Found | Needed |
|---|---|---|
| Total sets | 28 | |
| Total using selective parse | 11 (39%) | |
| Total with 2 generic consonants | 20 | 37 |
| (same, plus selective parse) | 20 | 61 |

(Needed /1000 273 )

## *The statistics of searches:*

## ***Previous long-range comparisons***

Offer *only* lexical evidence in support of relatedness.

Generous degrees of freedom (phonological, semantic, phonotactic).

Far fewer proposed cognates than needed.

Multilateral comparison also has many degrees of freedom in the choice of languages.

## *The statistics of searches:*
## *Previous long-range comparisons* ***and Vajda's***

Offer *only* lexical evidence in support of relatedness.

Paradigms

Generous degrees of freedom (phonological, semantic, phonotactic).

Regular correspondences (identity)

Far fewer proposed cognates than needed.

(No count available)

Multilateral comparison also has many degrees of freedom in the choice of languages.

Protolanguage to protolanguage; or PY/Ket to any Athabaskan (i.e. some freedom)

*The statistics of searches:*

# *Morphological paradigms*

**Algic pronominal affixes.** I, II = Wiyot allomorph sets.

| | Proto-Algonquian | I | II | Yurok |
|---|---|---|---|---|
| 1st person | * ne- | | du(÷)-d- < *n- | ÷ne- |
| 2nd | * ke- | | khu(÷)- | kh- k'e- |
| 3rd | * we- | u(÷)- | w- | ÷we- / ÷u- |
| Indefinite | * me- | | | b- < *m- me- |

Probability, calculated as 4 identical consonants in a 4-member paradigm:

## *Morphological paradigms*

Germanic *good : better*

| | | |
|---|---|---|
| English | *good* | *better* |
| German | *gut* | *besser* |
| Swedish | *god* | *bättre* |

*good*: g = 0.05 or 0.14
V = 0.5
d = 0.05 or 0.14
positive = 0.5

*bett-*: b = 0.05 or 0.14
V = 0.5
t = 0.05 or 0.14
comparative/superlative = 0.5

Overall probability if taken as 4 identical consonants:
0.000 000 39 (4 / 10,000,000)
If taken as 4 similar consonants (p = 0.14 each):
0.000024 (2 / 100,000)
If taken as two similar two-consonant roots:
0.000096 (9.6 / 100,000 or about 1 / 10,000)

# *Morphological paradigms*

**Gender-number suffixes in Afroasiatic determiners** (Greenberg 1960). Analysis (a) treats gender as neutralized in the plural; (b) treats it as syncretized.

| (a) | Sg. | Pl. | (b) | Sg. | Pl. |
|---|---|---|---|---|---|
| Masc. | -n | } -n | | -n | -n |
| Fem. | -t | | | -t | -n |

Probability calculated with **specific** consonants (p = 0.05):

| | | | |
|---|---|---|---|
| (a) | p = 0.000 0045 (4.5 / 1,000,000) | (b) | p = 0.000 0020 (2 / 1,000,000) |

Probability calculated with **similar** consonants (p = 0.14):

| | | | |
|---|---|---|---|
| (a) | p = 0.000099 (9.9 / 100,000) | (b) | p = 0.000043 (4 / 100,000) |

## *Morphological paradigms*

**Insufficient evidence:** *n : m* personal pronoun systems in the Americas (*n* in 1sg, *m* in 2sg, same paradigmatic positions)

Calculated as 2 identical consonants in a 2-member paradigm:

0.000625 (6 in 10,000)

Same, as 2 identical consonants in particular places in a 6-member paradigm:

0.00007 (7 in 100,000)

## *Morphological paradigms:*
## ***Dene-Yeniseian***

TAM prefixes (Vajda 2008: Table 12).

*çi telic *Ga non-telic

2 consonants; similar? identical? (0.05 * 0.05 or 0.14 * 0.14)
same functions (0.5)
same position (no search)
exhaustive paradigm (no search)

0.00125 or 0.0098 **(insufficient)**

## *Dene-Yeniseian morphological paradigms*

Yeniseic stative-resultative *ẏ, ND perfective/stative *yi
following these TAM prefixes

C: 0.05 or 0.14
function: ??
position: (no search)
exhaustive? (no search)

## *Dene-Yeniseian morphological paradigms*

TAM affixes (Vajda 2008: Table 13). Yeniseic prefixes (former suffixes in bipartite), EAT suffixes

*¬ progressive *n perfect

2 consonants; similar? identical? (0.05 * 0.05 or 0.14 * 0.14)
same functions (0.5)
arguably same position (no search)
exhaustive paradigm in Yeniseian, search in EAT
0.00125 + (0.00125*.5 = 0.000625) = 0.001875 or more
or 0.0098 + (0.0098 * .5) = 0.0147 or more **(insufficient)**

## *Dene-Yeniseian morphological paradigms*

Three pieces of insufficient evidence
Between the three of them, exhausting the TAM-like slots
Same ordering of the three slots

Increases the diagnostic value, possibly as far as:

| | Table 12 | Stat.-Perf. | Table 13 | Product |
|---|---|---|---|---|
| Identical C's | 0.00125 | 0.05 | 0.0098 | 0.000 000 6125 |
| Similar C's | 0.001875 | 0.14 | 0.0147 | 0.000 001 378 |

**Sufficient** (5 or 6 zeroes after the decimal point)

# *Dene-Yeniseian morphological paradigms*

Spatial and shape prefixes (Vajda 2008:§2.2.3):

**n-* 'round'; 'around'
**d-* 'long'; 'along'
**qu-* (Ath.) / *hu-* (Yen.) 'flat; area'

Exhaustive in subslot? (if so, no search)

| Consonants | Forms | Functions | Cum. prob. |
|---|---|---|---|
| Identical: | $0.05^3 = 0.000125$ | $0.33^3 = 0.037$ | 0.000 005 |
| Similar: | $0.14^3 = 0.002744$ | $0.33^3 = 0.037$ | 0.000 102 |

## *Dene-Yeniseian morphological paradigms*

Conclusion so far:

Two separate pieces of near-sufficient or even sufficient evidence

Outcome: **Sufficient evidence** from the morphology; Yeniseian and Na-Dene are related

## *Lexical comparanda*

Vajda 2008 contains about 60+ proposed cognate sets

Similar or identical consonants
Usually two consonants
No selective parsing
Wordlist size unknown
Semantic range 3 senses?

| Needed for wordlist: | 100 | 200 | 1000 |
|---|---|---|---|
| 2 similar C's | 10 | 19 | 73 |

## *Conclusions*

Morphological evidence:  Sufficient to establish relatedness

Lexical evidence:  Respectable showing

***Based on:***

Nichols, Johanna. 1996. The comparative method as heuristic. Mark Durie and Malcolm Ross, eds., *The Comparative Method Revisited,* 39-71. Sydney: Oxford UP.

---. 1998. The first four discoveries of America. The First Americans, AAA Annual Meeting, Philadelphia.

---. 2000. Linguistic evidence on the peopling of the New World. Symposium: The Initial Peopling of the New World. AAAS annual meeting, Washington, DC, February.

---. 2007. Typology in the service of classification. Workshop on Alternative Approaches to Language Classification, Stanford, August.