Uralic lexical evolution
in basic and less basic vocabulary

Trees and networks compared

Jyri Lehtinen, Kaj Syrjänen, Terhi Honkola,
Kalle Korhonen, Outi Vesakoski and Niklas Wahlberg

Societas Linguistica Europaea 2012
29 Aug – 1 Sep 2012, Stockholm University
Beyond trees

- Trees only show splits, **no convergence**
- Why weak support for a split?
  - Which alternative connections make splitting ambiguous?
- Tree algorithms **always** produce **trees**
  - Does the data actually support a tree model?
- Alternative: **networks**
  - Show conflicting splits in a network-like pattern
Phylogenetic networks

In trees, divergence is clear-cut; if binary branching is not clearly supported, polytomous branchings possible.

Networks can display secondary connections caused by conflicts in the data; not "rooted", but show the distance of languages instead of historical divergence.

Lang 1

Lang 2

Lang 3

Lang 4

Lang 5
Phylogenetic networks in language history

• Suitability of networks where divergence from common ancestor not assumed
  ▫ e.g. Dunn et al. (2008): typological features of languages of Island Melanesia

• Effect of borrowing within language families
  ▫ e.g. Nelson-Sathi et al. (2010) on connections between branches of Indo-European

• Support of tree; showing unclear connections between branches (see Heggarty et al. 2010)

• Here, NeighborNet method used
Languages included

- Northern Saami
- Inari Saami
- Skolt Saami
- Kildin Saami
- Ume Saami
- Finnish
- Karelian
- Võro
- Komi
- Veps
- Mari
- Mansi
- Udmurt
- Selkup
- Nenets
- Hungarian
- Livonian
- Estonian
- Mordvin
Network of Ura100 data

Data as in Syrjänen et al. (accepted) with Inari and Kildin Sami and Võro added
Comparison of tree and network

Tree finds divergences well (phylogenetic signal); cause of polytomy nor weak support of branches not visible

Network: deep and unclear connections with other subgroups -> polytomies and weak branchings; Also, alternate connections for clear branches -> Earlier dialect continua or deep divergences with low data support
Role of basic vocabulary

- Vocabulary-based historical study mostly concentrates on 'basic vocabulary'
  - Words supposed to be maximally useful in showing tree-like divergence:
    1) resistant to borrowing; 2) stable through time;
    3) morphologically simple; 4) universally present

- However, just one aspect of family-internal connections

- No basic vocabulary lists immune to language contact
'Less Basic Vocabulary'

- Leipzig-Jakarta list (Tadmor 2009) formed in Loanword Typology Project (Haspelmath & Tadmor 2009a)
- Here, the impact of less basic vocabulary-like meanings on secondary connections studied
- Leipzig-Jakarta: meanings ranked 1-100
- ’Less Basic Vocabulary’ list: meanings 401-500 (World Loanword Database, Haspelmath & Tadmor 2009b)
The LBV list and data collection

- In Syrjänen et al. (accepted, Diachronica), divergence of Uralic studied with basic vocabulary (Swadesh, Leipzig-Jakarta)
- LBV:
  - easily borrowed: 'bake', 'clean'
  - often innovated: 'shake', 'only', 'piece'
  - formed with compounds, derivs.: 'deaf', 'rainbow'
  - seldom present: 'grandson', 'millet' ('ice' good in Uralic!)
- Same dictionaries, new languages: Kildin Sami, Inari Sami, Võro
Less Basic Vocabulary network

- Northern Sami
- Inari Sami
- Ume Sami
- Skolt Sami
- Kildin Sami
- Veps
- Karelian
- Finnish
- Võro
- Estonian
- Livonian
- Northern Mansi
- Tundra Nenets
- Selkup
- Udmurt
- Komi-Zyrian
- Eastern Khanty
- Hungarian
- Mari
- Erzya_Mordvin
Comparison: Ura100 vs. LBV

- Samoyed
  - Selkup
  - Tundra_Nenets
  - Eastern_Khanty
  - Northern_Mansi
  - Hungarian
  - Udmurt
  - Komi-Zyrian
  - Mari
  - Erzya_Mordvin
  - Ugric
  - Permic
  - Samoyed

- Finnic
  - Estonian
  - Livonian
  - Võro
  - Karelian
  - Veps
  - Ume_Sami
  - Northern_Sami
  - Inari_Sami
  - Skolt_Sami
  - Kildin_Sami
  - Sami
  - Northern_Sami
  - Inari_Sami
  - Skolt_Sami
  - Kildin_Sami
  - Ume_Sami
  - Komi-Zyrian
  - Udmurt
  - Selkup
  - Northern_Mansi
  - Eastern_Khanty
  - Hungarian
  - Mari
  - Erzya_Mordvin
  - Ugric
Tree from LBV
Conclusions

• Networks *do* show connections not visible in trees
• Trees find branches not easily detected from networks
• The two models *complement each other*
• Less basic vocabulary
  ▫ Shows effect of language contact because of borrowing
  ▫ Deep connections unclesarer because of faster replacement
• Adding LBV to basic vocabulary does not make subgroups clearer
  ▫ *Quality of data matters for trees* (cf. Syrjänen et al; accepted)
Thank you!