Some hypotheses on the formation of East Asian language families

Laurent Sagart
EHESS and CNRS, Paris
5 language families

- Sino-Tibetan
- Hmong-Mien (aka Miao-Yao)
- Tai-Kadai
- Austroasiatic
- Austronesian
The STAN hypothesis

- Sino-Tibetan-Austronesian
- Claims ST and AN are two branches of a macrophylum (most recent Sagart 2005)
- regards Tai-Kadai as a branch of Austronesian (most recent Sagart 2004)
- thus consolidating 3 of the 5 EA families

The Austronesian family
growing consensus among austronesianists

- passage to Taiwan c. 3500-3000 BCE of the pre-ANs by boat
- homeland and first diversification in Taiwan
Formosan homeland and Malayo-Polynesian

- All An languages outside of Taiwan form a subgroup: Malayo-Polynesian (MP)
- MP innovations identified by Blust 1977 and subsequent work.
- Support from archaeology: earliest ceramic cultures in Taiwan c. 3000-2500 BCE.
- Support form genetics: the precursor of the "polynesian motif" in mtDNA is in Taiwan and nowhere else (Trejaut et al).
Blust's subgrouping

Proto-Austronesian

n branches in Taiwan

Proto-Malayo-Polynesian
Reconstructing the PAn lexicon

- Blust reconstructed PAn vocabulary based on this phylogeny
- rice, Setaria, house, field, boat, bow-and-arrow, fishnet, pigs, dogs, weaving, basketry etc.
today

- The Austronesians in Taiwan grow rice and Setaria.
- Setaria is sacred, not rice.
cereals confirmed by archaeology

- Tsang in 2002 discovered a site with grains of setaria and rice in SW Taiwan, dated 2800-2500 BCE.
- confirming knowledge of these cereals by the PAn’s.
Where did the ANs come from?

- obviously from the other side of the straits,
- but the sites there are transient
- before that?
- from an area which had domesticated rice and Setaria italica
- and tooth evulsion
tooth evulsion
Dawenkou culture, south Shandong, late 4th mill. BCE

- rice
- Setaria italica
- houses
- tooth evulsion
- adaptation to coastal environment
migration hypothesis

- pre-Austronesians in Dawenkou, late 4th
- migrated south by boat
- fast
- serial founder effect
- in search of new land
- bypassing other neolithic cultures
Tainan, foxtail and millet, -2800 BCE

Middle Dawenkou c. -3500 BCE

Middle Yangshao = PST ?

Majiayao = PTB ?

Tainan, foxtail and millet, -2800 BCE
Austronesian and Sino-Tibetan

- pre-Austronesian = Middle Dawenkou, c. 3500 BCE
- proto-Sino-Tibetan= Middle Yangshao, Henan, c. 4000 BCE
- close geographically
- common agriculture
Middle Dawenkou, c. 3500 BCE

Middle Yangshao = PST?
Linguistic evidence for STAN

- Sagart 2005
- sound correspondences
- > 12% basic vocabulary
- shared derivational morphology
### a sound correspondence

<table>
<thead>
<tr>
<th></th>
<th>PAn -q</th>
<th>OC -ʔ</th>
<th>TB - k</th>
</tr>
</thead>
<tbody>
<tr>
<td>vomit / spit</td>
<td>*utaq ‘vomit’</td>
<td>吐 tʰaʔ 'spit'</td>
<td>Lushai tʃʰa:k &lt; tʰa:k 'spit'</td>
</tr>
<tr>
<td>earth</td>
<td>*taq</td>
<td>± tʰaʔ</td>
<td></td>
</tr>
<tr>
<td>brain</td>
<td>*punuq</td>
<td>腦 nʰ uʔ</td>
<td>TB *nuk</td>
</tr>
</tbody>
</table>
Prediction

- agricultural vocabulary should be shared
Setaria italica

- PAn *beCeŋ Setaria italica
- OC 稷 tsək Setaria italica
- Trung tɕəʔ55 millet
- Lhokpu cək Setaria italica
- Tibetan č’ag fodder
de-husked rice

- PAn *beRas ‘dehusked rice’
- OC 糯 *mə-rat-s ‘dehusked rice’
- Tibetan mbras ‘rice’
rice grain (ready to cook/as food)

- PAN *Semay ‘rice as food’
- OC 米 *C-mij? ‘grain’
- Gyarong smay ‘grain’
- Burmese hmai1 ‘mole’
- Garo mi? ‘rice’
- Karen me ‘cooked rice’
reap by cutting

- PAN *-Rit 'cut': Puyuma rit rit 'cut' (rice); Thao klhit 'cut something slender', ka-klhit 'scythe; harvesting knife'; Malay arit 'sickle; grass knife'
- OC 利 *rit-s 'profit, profitable' (< harvest)
- TB (Benedict) ri:t ‘reap’
pestle

- PAn (Sagart) *turtur 'to pound; pestle’
- OC 磯 t‘ur-s 'pestle’
- Trung ㄎㄨㄣ 'to pound; pestle’
Revision of higher AN phylogeny

- Sagart 2004
- based on a set of nested lexical innovations
- in the numerals 5-10
- quinary system in the process of changing to decimal
## Implicational Hierarchy of Numerals 5-10 in Formosan Languages and PMP

<table>
<thead>
<tr>
<th></th>
<th>Pitu '7'</th>
<th>Lima '5'</th>
<th>Enem '6'</th>
<th>Walu '8'</th>
<th>Siwa '9'</th>
<th>Puluq '10'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lui Lang</td>
<td>innai</td>
<td>(na)lup</td>
<td>(na)sulup</td>
<td>patulunai</td>
<td>satulunai</td>
<td>isit</td>
</tr>
<tr>
<td>Saisiat</td>
<td>saivuseaha</td>
<td>rasu</td>
<td>saivusa</td>
<td>makaspat</td>
<td>ra:ha</td>
<td>ranpon</td>
</tr>
<tr>
<td>Paze</td>
<td>xasebidusa</td>
<td>xasep</td>
<td>xasebuza</td>
<td>xasebaturu</td>
<td>xasebituru</td>
<td>xasebisupat</td>
</tr>
<tr>
<td>Favorlang</td>
<td>naito</td>
<td>achab</td>
<td>nataap</td>
<td>maaspat</td>
<td>tannacho</td>
<td>zchiett</td>
</tr>
<tr>
<td>Taokas</td>
<td>yweto</td>
<td>hasap</td>
<td>tahap</td>
<td>mahalpat</td>
<td>tanasu</td>
<td>(ta)isid</td>
</tr>
<tr>
<td>Atayal</td>
<td>pitu?</td>
<td>imagal</td>
<td>cziu?</td>
<td>spat</td>
<td>qeru?</td>
<td>lp uu</td>
</tr>
<tr>
<td>Sediq</td>
<td>pito</td>
<td>lima</td>
<td>mataro</td>
<td>maspat</td>
<td>manjali</td>
<td>maxal</td>
</tr>
<tr>
<td>Thao</td>
<td>pitu</td>
<td>rima</td>
<td>kaa-turu, makalhturu-turu</td>
<td>kahspat, maka(lh)-shpa-shpat</td>
<td>tanacu</td>
<td>maqcin</td>
</tr>
<tr>
<td>Siraya</td>
<td>pitu</td>
<td>rima</td>
<td>nem</td>
<td>kuixpa</td>
<td>matuda</td>
<td>saat kittian</td>
</tr>
<tr>
<td>Hoanya</td>
<td>pito</td>
<td>Lima</td>
<td>(mi)nun</td>
<td>(mi)alu</td>
<td>(a)sia</td>
<td>(miata)isi</td>
</tr>
<tr>
<td>Papora</td>
<td>pitu</td>
<td>nema</td>
<td>(ne)nem</td>
<td>mahal</td>
<td>(me)siya</td>
<td>(me)tsi</td>
</tr>
<tr>
<td>Tsou</td>
<td>pitu</td>
<td>eimo</td>
<td>nôme</td>
<td>vóeu</td>
<td>sis</td>
<td>máskë</td>
</tr>
<tr>
<td>Saaroa</td>
<td>(k)upito</td>
<td>(k)ulima</td>
<td>(k)enem</td>
<td>(k)ulo</td>
<td>(k)usia</td>
<td>(k)ma:fte</td>
</tr>
<tr>
<td>Kanabu</td>
<td>pitu</td>
<td>lima</td>
<td>nem</td>
<td>(h)a:ru</td>
<td>siya</td>
<td>ma:në</td>
</tr>
<tr>
<td>Bunun</td>
<td>pitu'</td>
<td>hima'</td>
<td>nuum</td>
<td>vau'</td>
<td>siva'</td>
<td>mas'an</td>
</tr>
<tr>
<td>Rukai</td>
<td>pitu</td>
<td>Lima</td>
<td>enem</td>
<td>vaLu</td>
<td>barātē</td>
<td>manjale</td>
</tr>
<tr>
<td>Paiwan</td>
<td>pitju</td>
<td>lima</td>
<td>enem, unem</td>
<td>alu</td>
<td>siva</td>
<td>pu'luq</td>
</tr>
<tr>
<td>Puyuma</td>
<td>pitu</td>
<td>Lima</td>
<td>nem</td>
<td>waLu</td>
<td>iwa</td>
<td>puLu</td>
</tr>
<tr>
<td>Amis</td>
<td>pitu</td>
<td>lima</td>
<td>'enem</td>
<td>falu</td>
<td>siwa</td>
<td>polo</td>
</tr>
<tr>
<td>Kavalan</td>
<td>pitu</td>
<td>rima</td>
<td>'nem</td>
<td>waru</td>
<td>siwa</td>
<td>betin</td>
</tr>
<tr>
<td>Ketagalan</td>
<td>pitu</td>
<td>tsjima</td>
<td>anem</td>
<td>wasu</td>
<td>siwa</td>
<td>labatan</td>
</tr>
<tr>
<td>PMP</td>
<td>*pitu</td>
<td>*lima</td>
<td>*enem</td>
<td>*walu</td>
<td>*siwa</td>
<td>*pu'luq</td>
</tr>
</tbody>
</table>
Higher AN phylogeny based on innovations among numerals
Stages in the AN settlement of Taiwan
seven
Position of Tai-Kadai

The Tai-Kadai family according to W. Ostapirat
## Austronesian and Buyang numerals

<table>
<thead>
<tr>
<th></th>
<th>Buyang</th>
<th>PMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>ça(^{54})</td>
<td>duha</td>
</tr>
<tr>
<td>3</td>
<td>tu(^{54})</td>
<td>telu</td>
</tr>
<tr>
<td>4</td>
<td>pa(^{54})</td>
<td>e(m)pat</td>
</tr>
<tr>
<td>5</td>
<td>ma(^{312})</td>
<td>lima</td>
</tr>
<tr>
<td>6</td>
<td>nam(^{54})</td>
<td>enem</td>
</tr>
<tr>
<td>7</td>
<td>tu(^{312})</td>
<td>pitu</td>
</tr>
<tr>
<td>8</td>
<td>ma(^{0}) du(^{312})</td>
<td>walu</td>
</tr>
<tr>
<td>9</td>
<td>va(^{11})</td>
<td>siwa</td>
</tr>
<tr>
<td>10</td>
<td>put(^{54})</td>
<td>puluq</td>
</tr>
</tbody>
</table>
The twin MP and TK migrations
Once on the mainland

- Tai-Kadai was relexified twice
- once by a local language
- once by Chinese
- with which it aligned typologically
  - word order
  - 3+1 tones
  - monosyllables
  - loss of morphology
thank you for your attention