Siraya
language in 10 mins

Chu-Cheng
Demographics

- Siraya was an Austronesian language spoken in Southern Taiwan.

- no known native speakers since late 19th century
in the Central Pacific subgroup (34°) or the lowering of high vowels in morphemes identifying Central-Eastern Malayo-Polynesian (35°). The trees support 26 of the 34 putative Austronesian language subgroups and linkages discussed in (28). Of the remaining seven unsupported groups, two are linkages that lack exclusively shared innovations (Central and Western Malayo-Polynesian), and one is only supported by a single sound change (East Formosan). The remaining five (Western Oceanic, Malayo-Chamic, Greater Central Philippines, Greater Barito, and Barrier Islands/North Sumatra) may be obscured in our analyses because of conflicting signals caused by undetected borrowing between neighboring languages. Our results place the Formosan languages of Taiwan at the base of the trees immediately after the outgroups (Fig. 1). Following these are the languages of the Philippines, Borneo/Sulawesi, Central Malayo-Polynesian, South Halmahera/West New Guinea, and the Oceanic languages. This chained topology is precisely the structure predicted by the pulse-pause scenario.

One potential problem with this analysis is that Old Chinese may be too distantly related to Austronesian to reliably root the tree, whereas...
in the Central Pacific subgroup (34) or the lowering of high vowels in morphemes identifying Central-Eastern Malayo-Polynesian (35). The trees support 26 of the 34 putative Austronesian language subgroups and linkages discussed in (28).

Of the remaining seven unsupported groups, two are linkages that lack exclusively shared innovations (Central and Western Malayo-Polynesian), and one is only supported by a single sound change (East Formosan). The remaining five (Western Oceanic, Malayo-Chamic, Greater Central Philippines, Greater Barito, and Barrier Islands/North Sumatra) may be obscured in our analyses because of conflicting signals caused by undetected borrowing between neighboring languages. Our results place the Formosan languages of Taiwan at the base of the trees immediately after the outgroups (Fig. 1). Following these are the languages of the Philippines, Borneo/Sulawesi, Central Malayo-Polynesia, South Halmahera/West New Guinea, and the Oceanic languages. This chained topology is precisely the structure predicted by the pulse-pause scenario.

One potential problem with this analysis is that Old Chinese may be too distantly related to Austronesian to reliably root the tree, whereas...
History: since 17th century

• Romanized writing system introduced by the Dutch for missionary purposes; usage continued for 200 years

• eventually fell out of use due to Chinese colonization
  • last attested text: 1830s
  • last known speakers: probably 1890s

• probably a few remnant words in Taiwanese

• attempts of revival started in late 20th century
what do we know about Siraya?

- main source: 17th century Siraya-Dutch bitexts
  - the gospel of Matthew
  - the Heidelberg catechism
  - 4 short dialogues between natives
- land contracts (~200 of them)
- wordlists collected in late 19th century ~ 20th century
Het H. Evangelium na [de beschrijvinge]

MATTHEI.
Het eerste Capittel.

1 Het Boek des Gelachts Jesu Christi, des Loons Davids / des Loons Abrahams.
2 Abraham gewan Isaac, ende Isaac gewan Jacob, ende Jacob ghewan Judam / ende sijn broeders.
3 Ende Judas ghewan Phares ende Zara by Thamar, ende Phares ghewan Elrom, ende Elrom ghewan Aram.
4 Ende Aram ghewan Aminadab, ende Aminadab ghewan Paallon, ende Paallon ghewan Salmon.
5 Ende Salmon ghewan Booz by Rachab, ende Booz ghewan Obed by Ruth, ende Obed ghewan Jesse.
6 Ende Jesse ghewan David den Konich, ende David de Konings ghewan Solomon by de ghene die Urias

Hagnau ka D'illigh
Matiktik ka na safoulat ti

MATTHEUS.
Naunamou ki lbægb ki loul.

1 Oulat ki kavoytan ti Jezus Christus, ka na alak ti Abraham.
6 Ti Jesse ta ni-pou-alak ti David-an ka na Mei-faou ka Si bavau. Ti David ka na Mei-faou na ni-pou-alak ti Salmon-an p'ouhkoua
phonemics

- written using Roman alphabets
- distinction between long and short vowels.
- there was probably a schwa (interchangeable uses of e and i)
- distinction between voiced/voiceless stops
- b/ v are allophones
- we know nothing about stress, tone etc..
word classes: nouns

- reduplication for plurality e.g. *tau* ‘person’ *tau-tau* ‘people’

- nominalization through Ca-Reduplication e.g. *dilux* ‘lead’ *da-dilo* ‘priest’

- agentives can be formed with the prefix *tama-* e.g. *kuting* ‘to read’ *tama-k’ting* ‘reader’
word classes: verbs

- Tense-Aspect-Mood marker suffixes; can also have pronoun enclitics
  - *pako-lulux-en=au=kaw*
  - v4.invoke-praise-UO=1S.GEN=2S.NOM
  - “I thank you”
- distinguish between actor-oriented (AO) and undergoer-oriented (UO) voices
- bound verb construction very productive
  - *mattaey-:* ‘talking’
    - +*duma* ‘opposite’: *mattaey-duma-duma*: talk against
    - +*ma-riang* ‘good’: *mattaey-riang*: bless
syntax

• word order is typically VSO

• if the actor is not subject, it comes after the verb: VASX

• word order not strict; also cleft construction exists

• 3 cases for nouns
  • NOM: subjects
  • GEN: ‘secondary case’: actor when subject is not an actor, patient, possessor
  • OBL: locative, means
Give us this day our daily bread
and they were baptized by him in the river Jordan.
And whatever you will bind on earth will be bound in heaven.
MT Systems / NLP
Resources I know of
(speculative) MT difficulties

• morphological complications: Siraya has rich prefixes/suffixes/infixes

• marker on the verb directly decides which NP is the agent
  • no distinction in word order / case marking
  • *ni-ha tin ta malituk ki Mairang tin*
    ‘he hid his master’s money.’
  • *ni<*>ha tin ta malituk ki Mairang tin*
    ‘his master’s money hid him.’

• lack of native speakers