

「言語における系統・変異・多様性とその数理」研究発表会
2017年12月5日15:40-16:20 国立国語研究所・多目的室

オーストロネシア諸言語における 系統・変異・多様性 と数理分析の可能性

国立民族学博物館／総合研究大学院大学


菊澤 律子

- はじめに
1 目的／オーストロネシア語族とは
- 比較（歴史）言語学の方法
2 比較方法と数理処理
- 統計処理の可能性
3 系統分類と数理処理
- 応用編
4 統語構造の比較再建と数理処理
- まとめにかえて（予告編）
5 垂直伝搬と水平伝搬



本報告の目的


- 本発表では、オーストロネシア諸言語を例にとり、これまで手作業で行われてきた言語の歴史研究の手法のうち、計量的アプローチによる効率化と客観化をはかれる可能性があるのかどうか、あるとしたらどのような内容があり得るのか、検討するための素材を提供する。



背景

- オーストロネシア語族は、歴史言語学における比較方法（the Comparative Method）を用いて言語間の系統関係を特定することができた成功例としてよく知られている（Ross 1996a: 184）。
- 文献による記録がほとんど残っていないため、系統関係を含む史の変遷の解明には、現在話されている言語の比較が不可欠である。具体的には、音対応に基づく同源語の特定と音および語彙の再建、共有変化に基づく系統関係の特定、音対応の階層に基づく間接継承語（同じ系統の言語からの借用語）の特定、などといった手続きによる。






背景

• 一方で、オーストロネシア諸言語を対象とした計量言語学的な研究では、

- (1) 言語学者により同源性が特定されたデータを利用した系統分析 (e.g., Russel et al., Greenhill, et al., and others) か、
- (2) 同源・借用を区別せず、言語の相対としての類型論的な近似性を見る研究 (e.g., Wichmann to appear, Greenhill 2016?)

のいずれかとなっており、比較方法の手法を機械処理してより広い言語に適用するという取り組みはない (cf. Grime 2007, List and Hill 2017)。

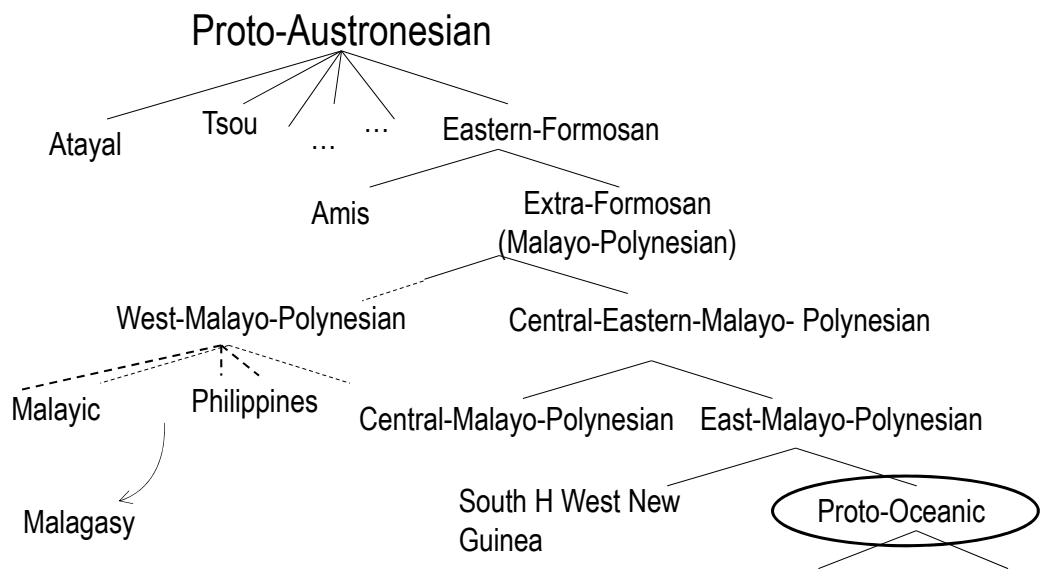


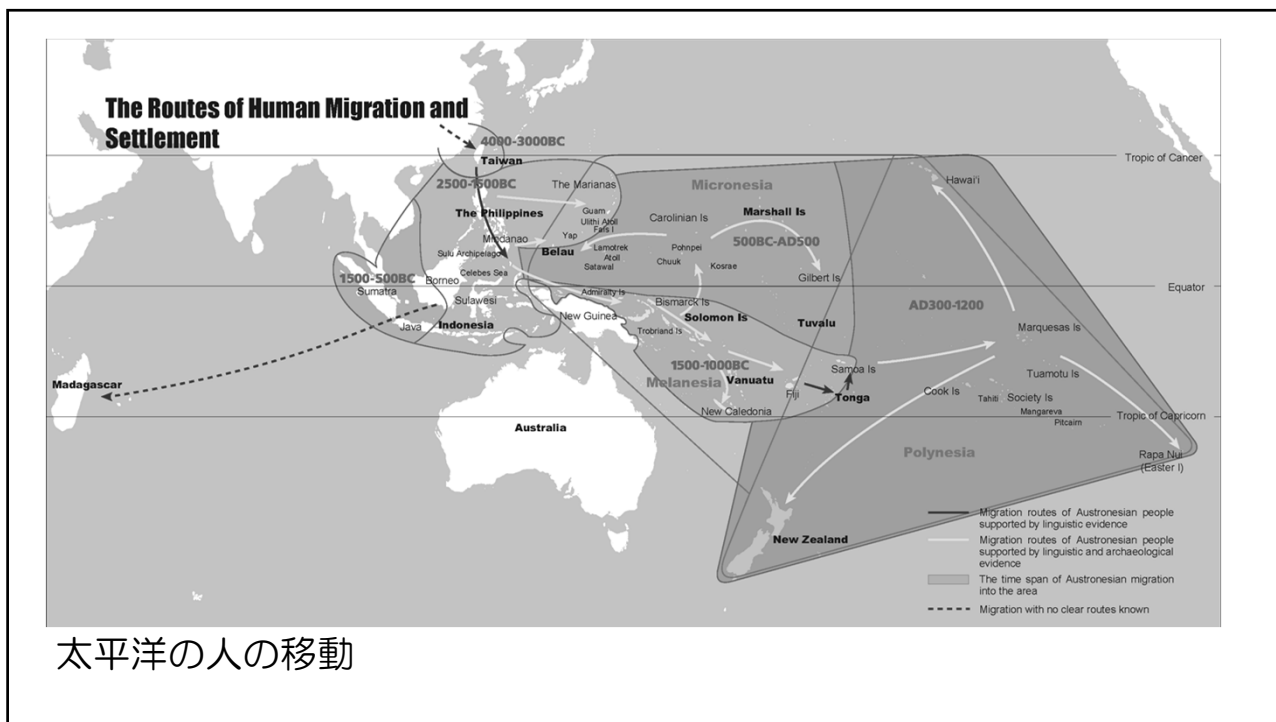
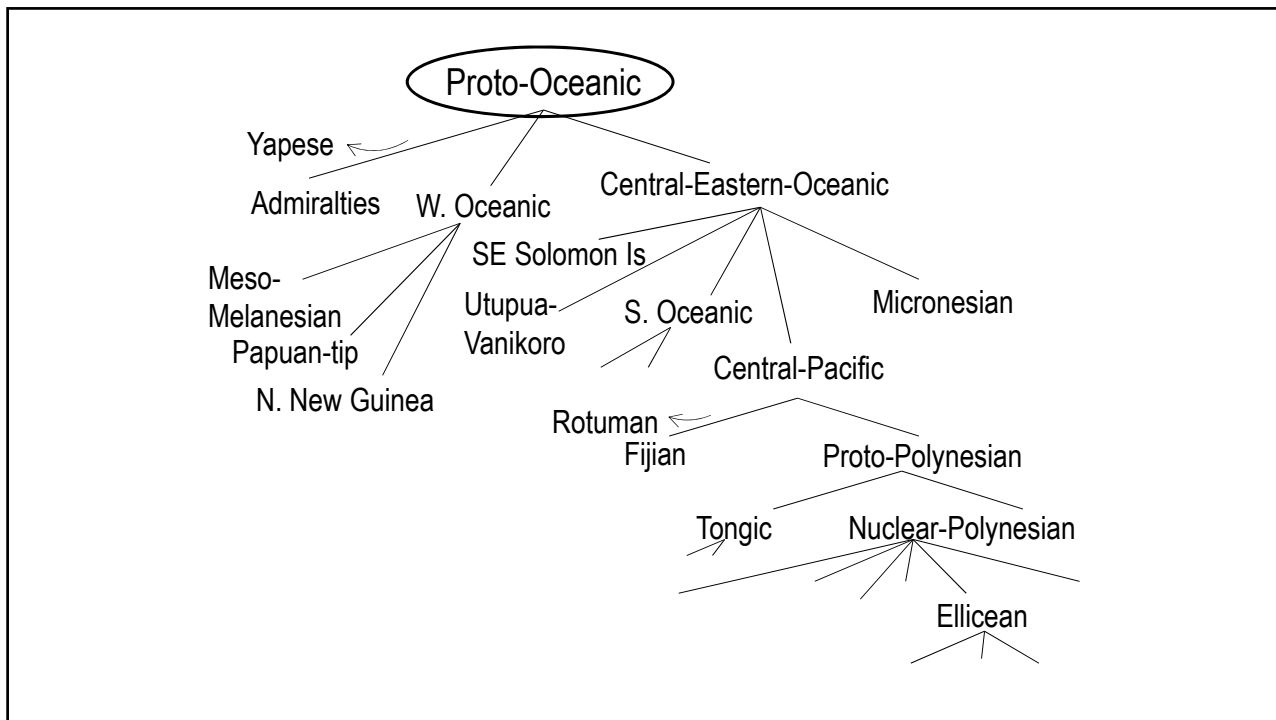
背景

• 本報告では、比較方法に基づく言語の比較再建と系統分類の概要を手法を概観しつつ、統計処理やコンピューター技術を用いて自動化したい側面を、歴史言語学者側からの視点で描写する。



オーストロネシア語族





太平洋の人の移動

語族とは

規則的な音対応により、共通祖先から発達したことを示すことができる言語の集合体

オーストロネシアとは

言語の親縁関係のみで結ばれるグループ

- 1 はじめに
目的／オーストロネシア語族とは
- 2 比較（歴史）言語学の方法
比較方法と数理処理
- 3 統計処理の可能性
系統分類と数理処理
- 4 応用編
統語構造の比較再建と数理処理
- 5 まとめにかえて（予告編）
垂直伝搬と水平伝搬

言語の発達史と統計処理

言語の垂直伝播

- 分岐と言語の発達
- 直接継承 (vs. 間接継承)

言語の水平伝播

- 言語接触と借用
- 地理的分布

⇒ 樹形図と波状モデル

機械処理の強みと弱み

- 大量の情報処理能力
- パターン認識
- パターン同士の横のつながりは×

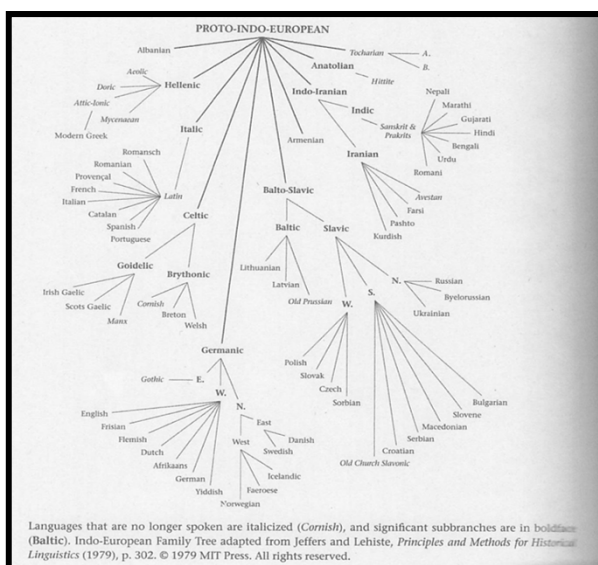


Figure 1: An Indo-European family tree (cited from Mihalceck and Wilson 2011: 528)

樹形図と波状モデル

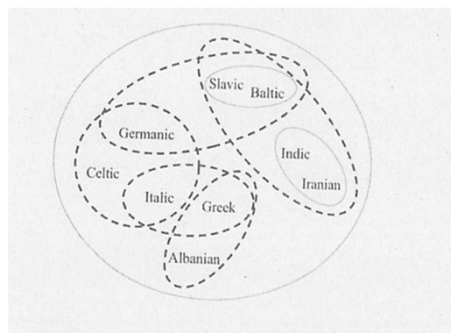


Figure 2: An Indo-European wave diagram (cited from Mihalceck and Wilson 2011: 531)



比較方法と系統分類

1. 語彙データの収集
2. 同源語の特定
3. 音対応の確立
4. 語彙の再建
5. 系統分類（共有変異に基づく）



プロセスの概要

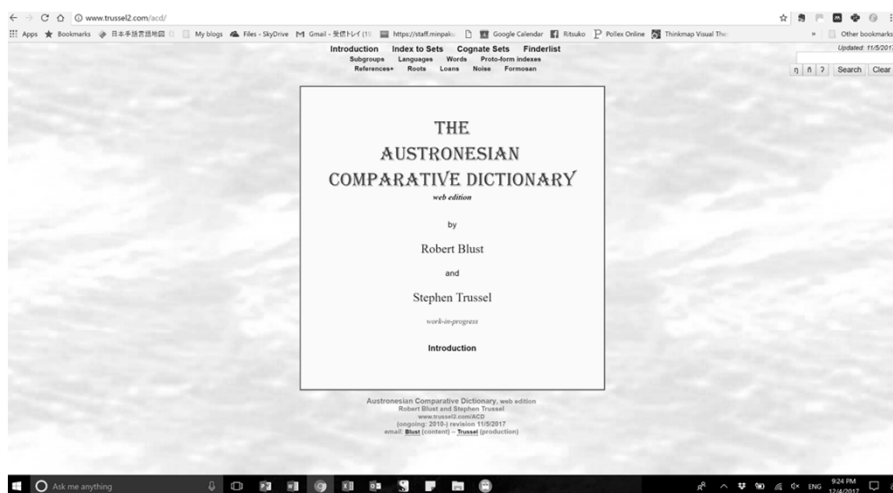
なぜ音対応の確立に機械処理がほしいのか



比較方法と系統分類

1. 語彙データの収集
2. 同源語の特定
3. 音対応の確立
4. 語彙の再建
5. 系統分類（共有変異に基づく）

1. 語彙データの収集



1. 語彙データの収集

- 手作業に依存
- 辞書
- 言語学および文化人類学等における記述
- 個人のフィールドノート
- 将来的にはウェブからロボット収集？アノテーションがついてなくてはならないので、難しいかも。

手作業OK

	'eye'	'sky'	'road'	'hand'	'two'	'three'
Paiwan (Taiwan)	matsa	kalevlevan	djalan	lima	ɟusa	tjelu
Bontok (Philippines)	matá	dáya	dálan	líma	duwá	tulú
Malagasy (Madagascar)	másu	lánitra	lálana	tánana	rúa	télu
Manam (Papua New Guinea)	mata	lang	jala	debu	rua	toil
Tuvaluan (Tuvalu)	mata	lagi	ala	lima	-	tolu
Rapanui (Easter I.)	mata	rangi	ara	rima	rua	toru

2. 同源語の特定

- 研究者の目に依存
- 仮作業
- 作業上、大きな（心理的）負担ではない
- 人の目が入ることの是非？
- 次項3.のプロセスが機械処理できるようになれば、必要なくなる（自動的にはじけるようになる）かも？

手作業OK？ cf. The BWB (Beck-Wichmann-Brown) system

	'eye'	'sky'	'road'	'hand'	'two'	'three'
Paiwan (Taiwan)	matsa	kalevlevan	djalan	lima	ɟusa	tjelu
Bontok (Philippines)	matá	dáya	dálan	líma	duwá	tulú
Malagasy (Madagascar)	másu	lánitra	lálana	tánana	rúa	télu
Manam (Papua New Guinea)	mata	lang	jala	debu	rua	toil
Tuvaluan (Tuvalu)	mata	lagi	ala	lima	-	tolu
Rapanui (Easter I.)	mata	rangi	ara	rima	rua	toru

	'eye'	'sky'	'road'	'hand'	'two'	'three'
Paiwan (Taiwan)	matsa		djalan	lima	ɖusa	tjelu
Bontok (Philippines)	matá		dálan	líma	duwá	tulú
Malagasy (Madagascar)		lánitra	lálana		rúa	télu
Manam (Papua New Guinea)	mata	lang	jala		rua	toil
Tuvaluan (Tuvalu)	mata	lagi	ala	lima	-	tolu
Rapanui (Easter I.)	mata	rangi	ara	rima	rua	toru

3. 音対応の確立

- 扱うデータ量が多い
- 分析対象が多い
- 客観的な状況が知りたい
思い込みの影響？
見えていない現象？

機械にやってほしい！

3. 音対応の確立－機械にとっての難しさ？

- 扱うデータ量は多いが、ひとつひとつの組み合わせの出現数は少ない
- ギャップの存在=データの組み合わせがすべての言語で同じでない
- 対応する位置を指定できない
⇒位置を指定してから機械処理にかける？

cf. “WordCorr”



	‘eye’	‘sky’	‘road’	‘hand’	‘two’	‘three’
Paiwan (Taiwan)	matsa		djaLan	Lima	Dusa	tjeLu
Bontok (Philippines)	matá		dáLan	Líma	Duwá	tuLú
Malagasy (Madagascar)		Lánitra	lálana		Rúa	téLu
Manam (Papua New Guinea)	mata	Lang	jaLa		Rua	toiL
Tuvaluan (Tuvalu)	mata	Lagi	aLa	Lima	-	toLu
Rapanui (Easter I.)	mata	Rangi	aRa	Rima	Rua	toRu

	*l	*z	*d			
Paiwan (Taiwan)	L	Dj	D	Lima	Dɤusa	tjeLu
Bontok (Philippines)	L	D	D	Líma	Duwá	tuLú
Malagasy (Madagascar)	L	L	R		Rúa	téLu
Manam (Papua New Guinea)	L	J	R		Rua	toiL
Tuvaluan (Tuvalu)	L	L		Lima	-	toLu
Rapanui (Easter I.)	R	R	R	Rima	Rua	toRu

オーストロネシア系の言語における単語の例 (1)

地域	言語名	数字の3	数字の5
台湾	パイワン語	təlu	lima
フィリピン	ボントック語	toolø	lima
ミクロネシア	ポーンベイ語	siluh	limau
ミクロネシア	キリバス語	teniua	nimaua
ヴァヌアトゥ	ヴァヌアラヴァ語	nitelø	tafalemø
フィジー	マズアタ方言	ʔolu	lima
サモア	サモア語	tolu	lima
ハワイ	ハワイ語	ekolu	elima
ニュージーランド	マオリ語	toru	rima
ラパヌイ (イースター島)	ラパヌイ語	toru	rima

注：ʔは正門閉鎖音を示す。

オーストロネシア系の言語における単語の例 (1)

地域	言語名	数字の3	数字の5	
台湾	バイワン語	təlu	Lima	L
フィリピン	ポントック語	toLó	Lima	L
ミクロネシア	ポーンペイ語	siluh	Limau	L
ミクロネシア	キリバス語	teNiua	Nimaua	N
ヴァヌアトゥ	ヴァヌアラヴァ語	niteL	tafaLem	L
フィジー	マズアタ方言	ʔoLu	Lima	L
サモア	サモア語	toLu	Lima	L
ハワイ	ハワイ語	ekoLu	eLima	L
ニュージーランド	マオリ語	toRu	Rima	R
ラパヌイ (イースター島)	ラパヌイ語	toRu	Rima	R

注：ʔは正門閉鎖音を示す。

3'. 音対応の確立の評価

- The Beck-Wichmann-Brown (BWB) System

cf.

Brown, Cecil H. 2017. Evaluating proposals of language genealogical relationship. *Language Dynamics and Change* 7: 252-285.

4. 語彙の再建



- 音対応が確立し、祖語形が再建されれば、自動的に変換可。
→規則的な継承形が異なるものをはじきだすような使い方ができるはず。


	*l	*z	*d	*lima	*zuwa	*telu
Paiwan (Taiwan)	L	Dj	D	Lima	Dɤusa	tjeLu
Bontok (Philippines)	L	D	D	Líma	Duwá	tuLú
Malagasy (Madagascar)	L	L	R		Rúa	téLu
Manam (Papua New Guinea)	L	J	R		Rua	toiL
Tuvaluan (Tuvalu)	L	L		Lima	-	toLu
Rapanui (Easter I.)	R	R	R	Rima	Rua	toRu

Proto-X	*mata	*lanjɨC	*zalan	*qalima	*duSa	*telu
	'eye'	'sky'	'road'	'hand'	'two'	'three'
Paiwan (Taiwan)	matsa	kalevlevan	djalan	lima	ɟusa	tjelu
Bontok (Philippines)	matá	dáya	dálan	líma	duwá	tulú
Malagasy (Madagascar)	másu	lánitra	lálana	tánana	rúa	télu
Manam (Papua New Guinea)	mata	lang	jala	debu	rua	toil
Tuvaluan (Tuvalu)	mata	lagi	ala	lima	-	tolu
Rapanui (Easter I.)	mata	rangi	ara	rima	rua	toru

比較方法と系統分類


1. 語彙データの収集
2. 同源語の特定
3. 音対応の確立
4. 語彙の再建
5. 系統分類（共有変異に基づく）

- 
- 1 はじめに
目的／オーストロネシア語族とは
 - 2 比較（歴史）言語学の方法
比較方法と数理処理
 - 3 統計処理の可能性
系統分類と数理処理
 - 4 応用編
統語構造の比較再建と数理処理
 - 5 まとめにかえて（予告編）
垂直伝搬と水平伝搬
- 



5. 系統分類

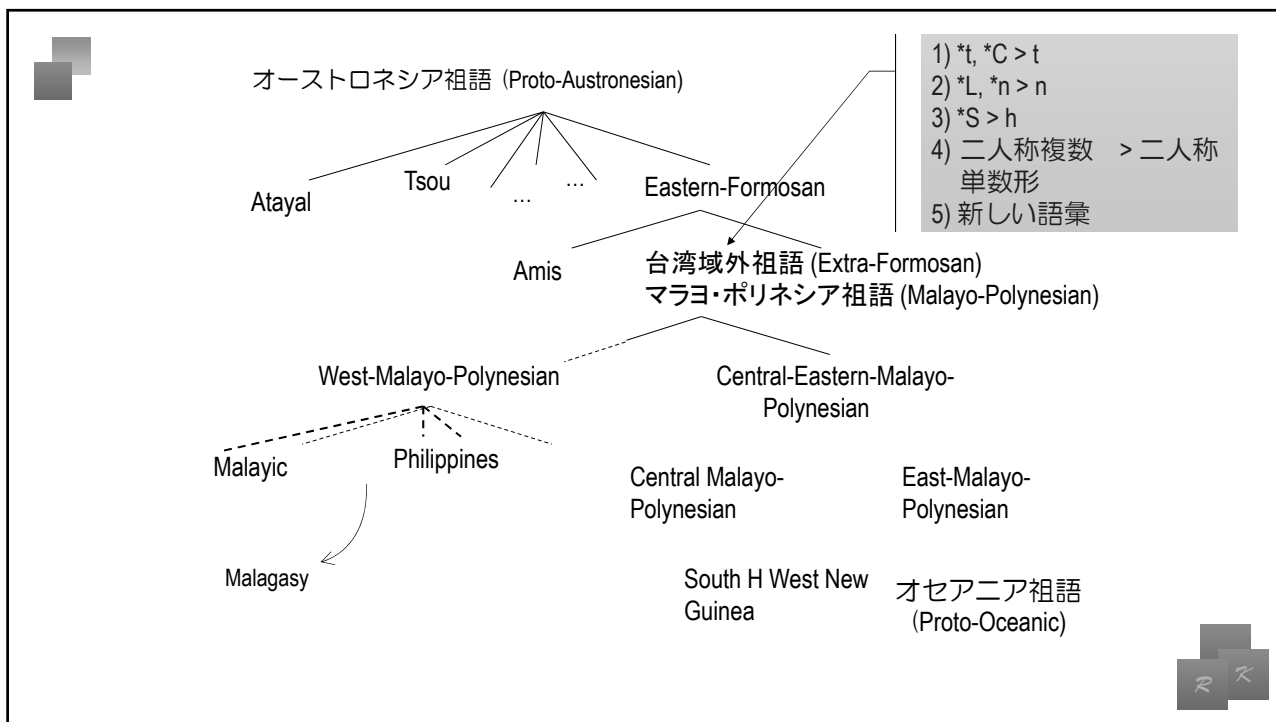
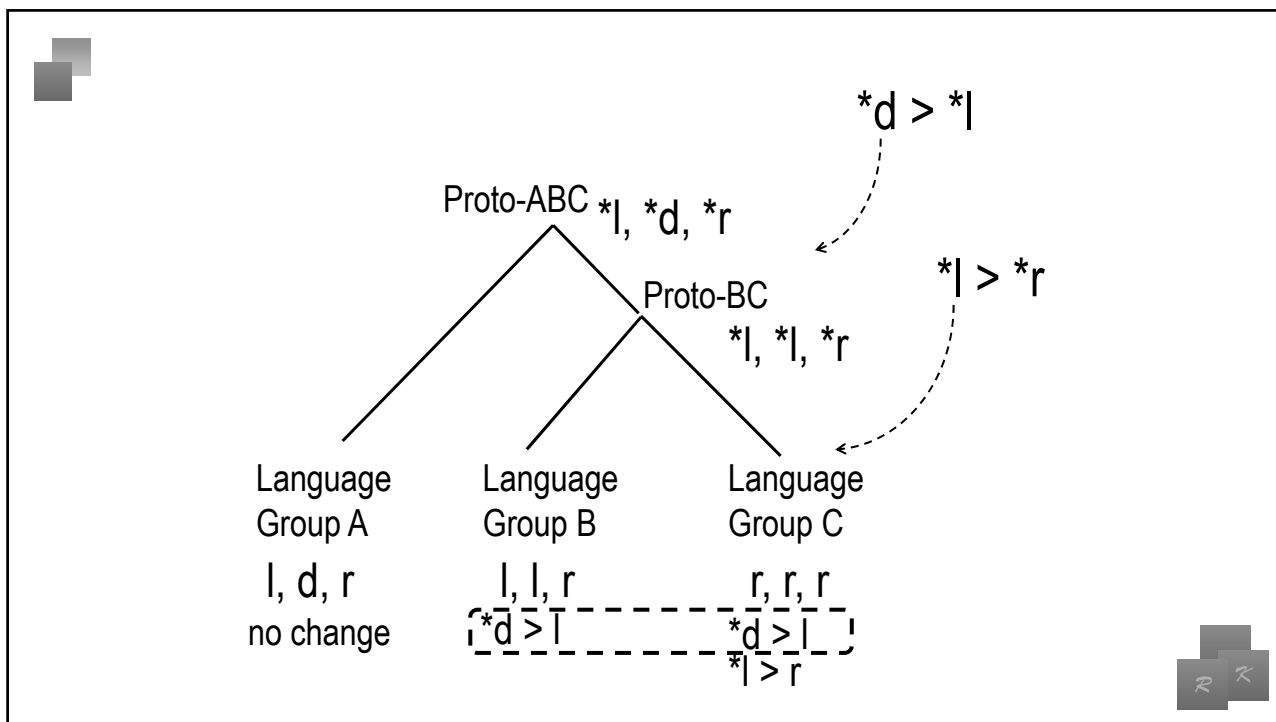
- 伝統的な手法

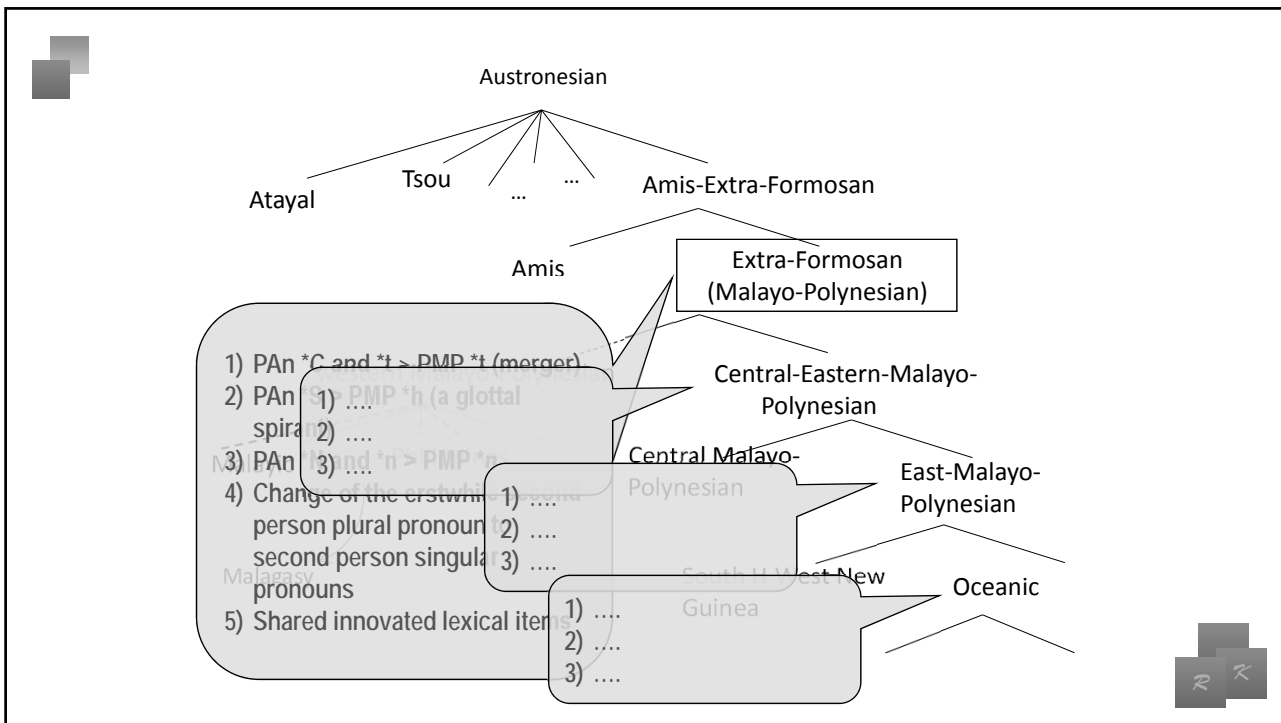
 - 計量的な手法
- 

言語系統特定の手順

- 同系の言語かどうかを調べる
- ↓
- 音対応に基づき、祖語の音と語彙を再建する
- ↓
- 共有変化にもとづき、変化の過程を推論する

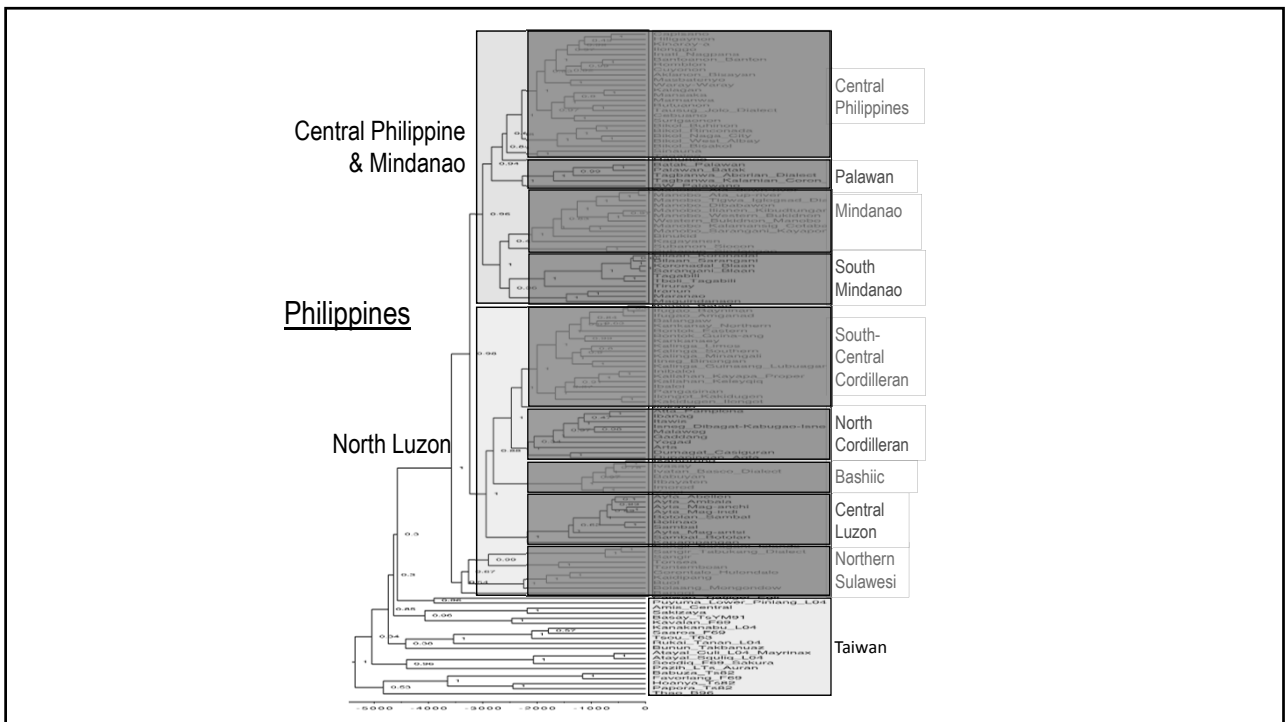
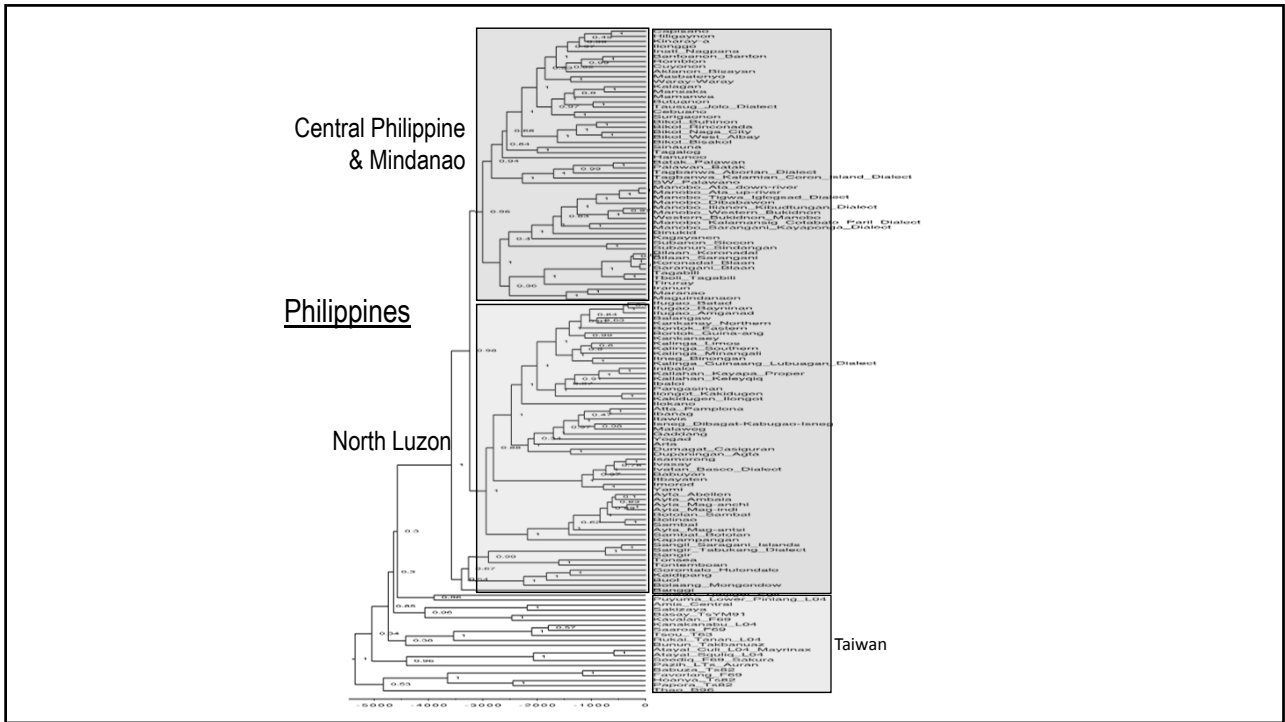
	*l	*z	*d			
Paiwan (Taiwan)	L	Dj	D	Lima	Dɤusa	tjeLu
Bontok (Philippines)	L	D	D	Líma	Duwá	tuLú
Malagasy (Madagascar)	L	L	R		Rúa	téLu
Manam (Papua New Guinea)	L	J	R		Rua	toiL
Tuvaluan (Tuvalu)	L	L		Lima	-	toLu
Rapanui (Easter I.)	R	R	R	Rima	Rua	toRu










5. 系統分類

- 伝統的な手法
- 計量的な手法



- 
-
- <http://www.shh.mpg.de/438157/cobldatabase>
 - <http://calc.digling.org/>
- 

- 
- はじめに
1 目的／オーストロネシア語族とは
 - 比較（歴史）言語学の方法
2 比較方法と数理処理
 - 統計処理の可能性
3 系統分類と数理処理
 - 応用編
4 統語構造の比較再建と数理処理
 - まとめにかえて（予告編）
5 垂直伝搬と水平伝搬
- 

- 
- 1 はじめに
目的／オーストロネシア語族とは
 - 2 比較（歴史）言語学の方法
比較方法と数理処理
 - 3 統計処理の可能性
系統分類と数理処理
 - 4 応用編
統語構造の比較再建と数理処理
 - 5 まとめにかえて（予告編）
垂直伝搬と水平伝搬

