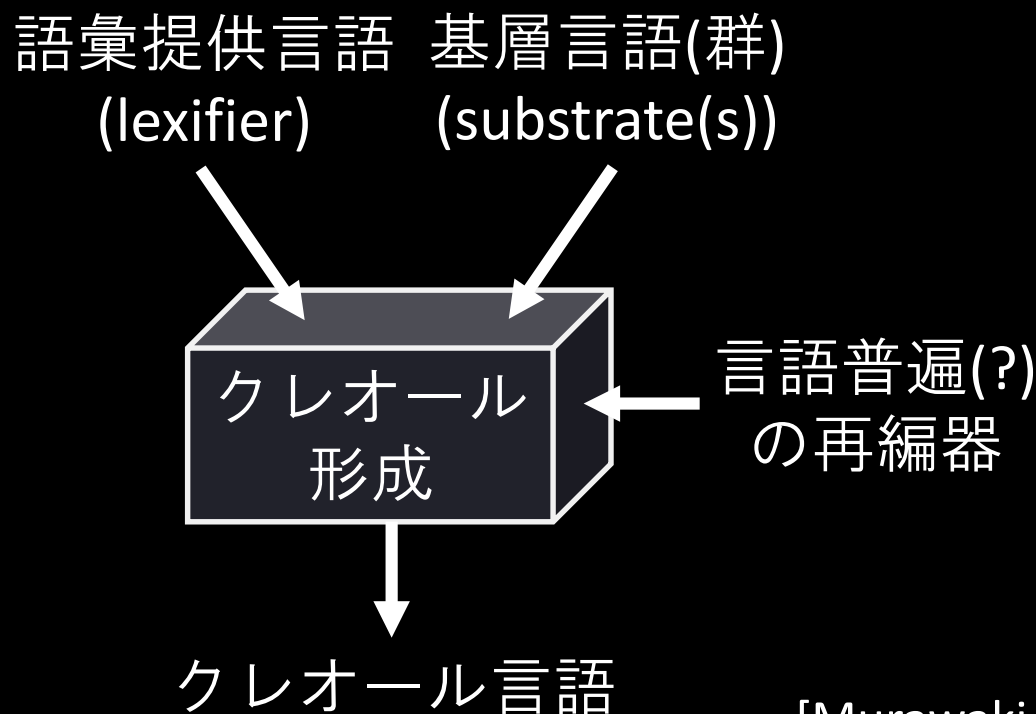


# クレオール形成に対する 混合モデル

京都大学  
村脇 有吾

# Take-home Message

クレオール形成の定量的分析には  
混合モデルがふさわしい



# 言語類型論に基づく系統推定

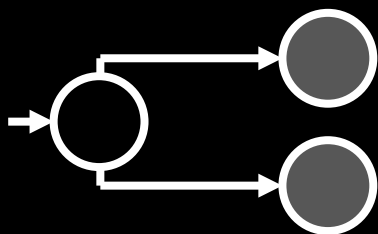
..を試みているが結果に説得力がない

日本語:	1	1	2	...	0	4
朝鮮語:	2	2	1	...	0	4
アイヌ語:	0	1	1	...	0	4

特徴81A  
Order of SOV

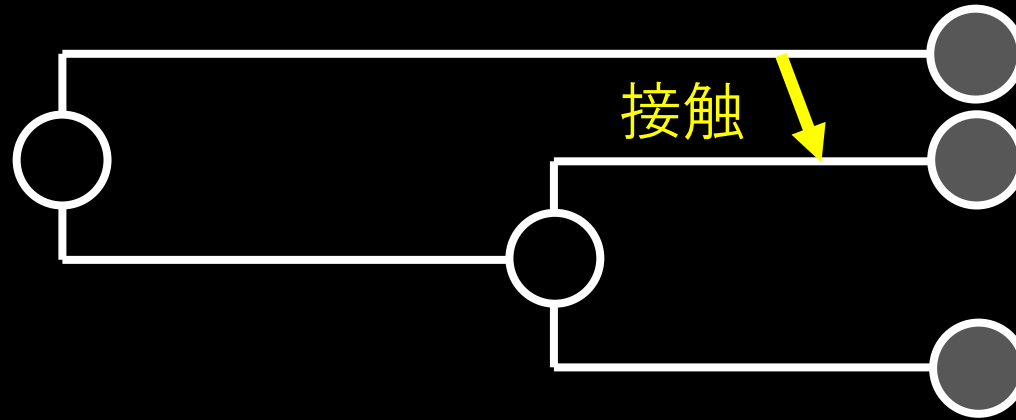
- 0: SOV
- 1: SVO
- 2: VSO

⋮



系統上近い  ベクトル同士が似ている

# 系統推定の仮定に反する現象の性質を理解したい



## 接触到り起因する類型論上の現象例

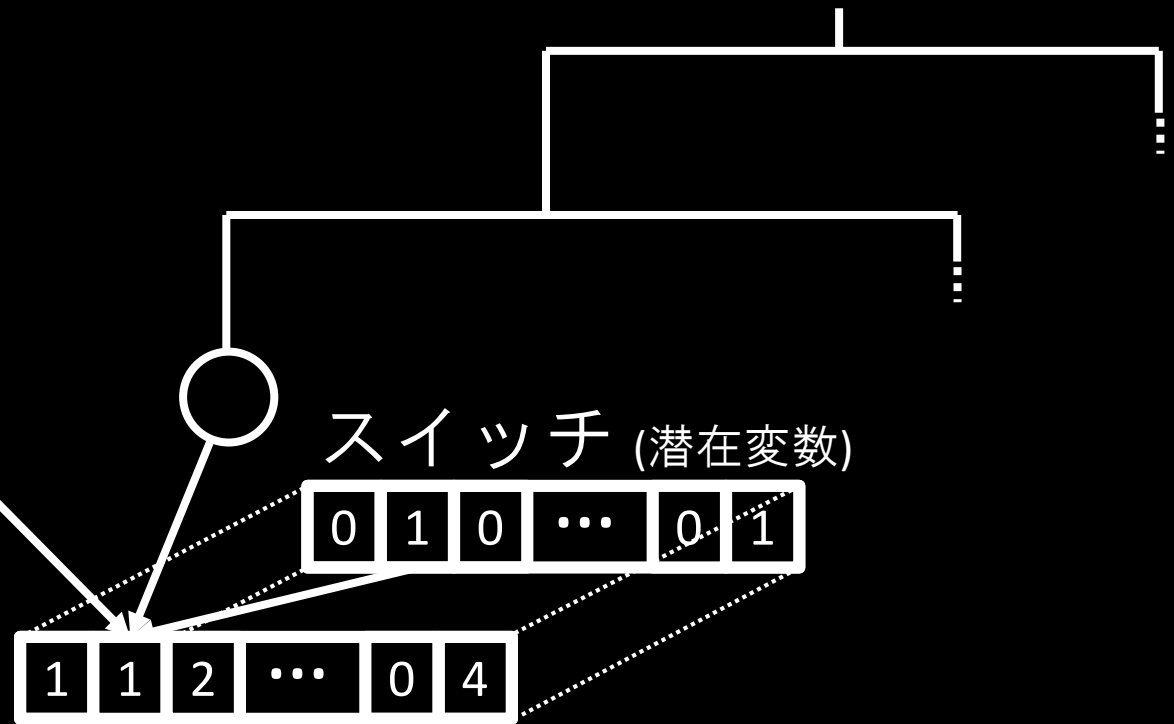
1. 言語連合 [Daumé III, 2009]
2. ピジン・クレオール形成 ← 今回はコレ

# 系統推定における 言語連合のモデル化 1/2

言語連合



系統樹



[Daumé III, NAACL2009]

# 系統推定における 言語連合のモデル化 2/2

p(gen)	#f	Feature Category
.00	1	Tea
.73	19	Phonology
.73	9	Lexicon
.74	4	Nominal Categories / Numerals
.79	5	Simple Clauses / Predication
.80	5	Verbal Categories / Tense and Aspect
.87	8	Nominal Syntax
.87	8	Simple Clauses / Simple Clauses
.91	12	Nominal Categories / Articles and Pronouns
.94	17	Word Order
.99	10	Morphology
.99	6	Simple Clauses / Valence and Voice
.99	7	Complex Sentences
.99	7	Nominal Categories / Gender and Number
.99	5	Simple Clauses / Negation and Questions
1.0	1	Other / Clicks
1.0	2	Verbal Categories / Suppletion
1.0	9	Verbal Categories / Modality
1.0	4	Nominal Categories / Case

ほとんどの特徴  
が系統樹に由来  
するという推定

[Daumé III, NAACL2009]

# クレオール言語の例 (宜蘭クレオール)

[真田+, 2012]

- C: akong, anta raran no...anta raran ga nani no tiyaw sigototeru  
おじいちゃん、あなた、昔の、あなた、昔は何のお仕事をしてたの?
- B: yama ha  
山だよ
- A: raran yama ano  
昔、山、あの
- B: minna onci umah umah ha raran  
みんな、私たち、除草、除草、え、昔
- C: nani uyeteru?  
何を植えてた?
- B: ima ga nai mo  
今はない、もう

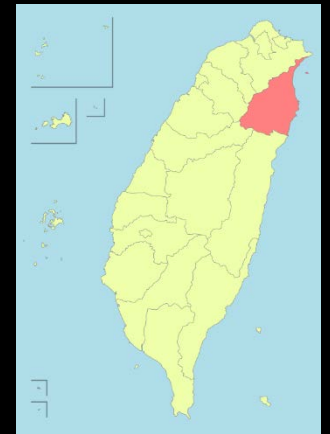


Image source:  
Author: Luuva  
License: CC BY-SA 3.0

# 日本語系統論における 困ったときのクレオール

[川本, 1988]

日本列島でもこれに似たようなことが起こったに違いない。古アジア系の「私・あなた・好き」型の言語が「アイ・ラヴ・ユー」型の南島語を受け入れた。その結果古アジア語化した南島語が成立した。その後弥生時代の直前に再び南島語族が水田稲作と共に渡来する。それまでの焼畑農耕に比べて大きな規模の水田農耕が多数の労働者の結集を必要としたのではなからうか。サトウキビプランテーションがトクピシンの発生をうながしたように、水田稲作がピジンを生んだ。これがやがてクレオール化して日本祖語となった。



# クレオール形成を定量的に検証 するためのデータは整いつつある



THE ATLAS OF PIDGIN AND CREOLE  
LANGUAGE STRUCTURES  
ONLINE



Home

Languages

Features

WALS-APiCS

Examples

Sources

Authors

Credits

Legal

Download

Contact

Help

## Welcome to *APiCS Online*



Fishing boats: Fishermen selling their catch at Abandze, Ghana, the site of the first British trading station on the Gold Coast, Fort Kormantin,

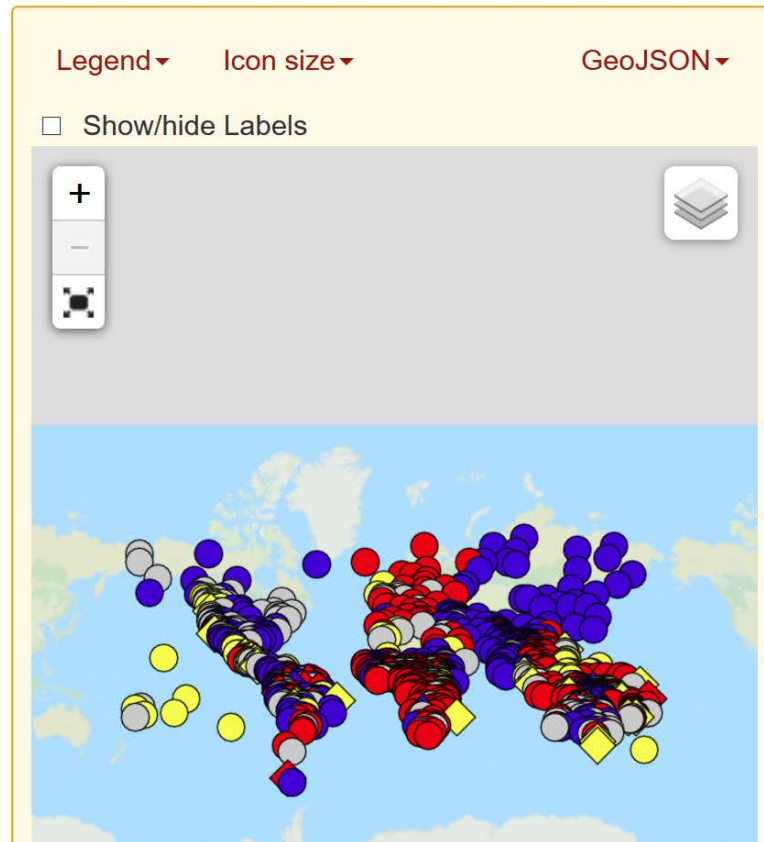
This web site contains supporting electronic material for the *Atlas of Pidgin and Creole Language Structures (APiCS)*, a publication of Oxford University Press. APiCS shows comparable synchronic data on the grammatical and lexical structures of 76 pidgin and creole languages. The language set contains not only the most widely studied Atlantic and Indian Ocean creoles, but also less well known pidgins and creoles from Africa, South Asia, Southeast Asia, Melanesia and Australia, including some extinct varieties, and several mixed languages.

*APiCS Online* is a separate publication, edited by Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath, and Magnus Huber. It was made possible by support from the Deutsche

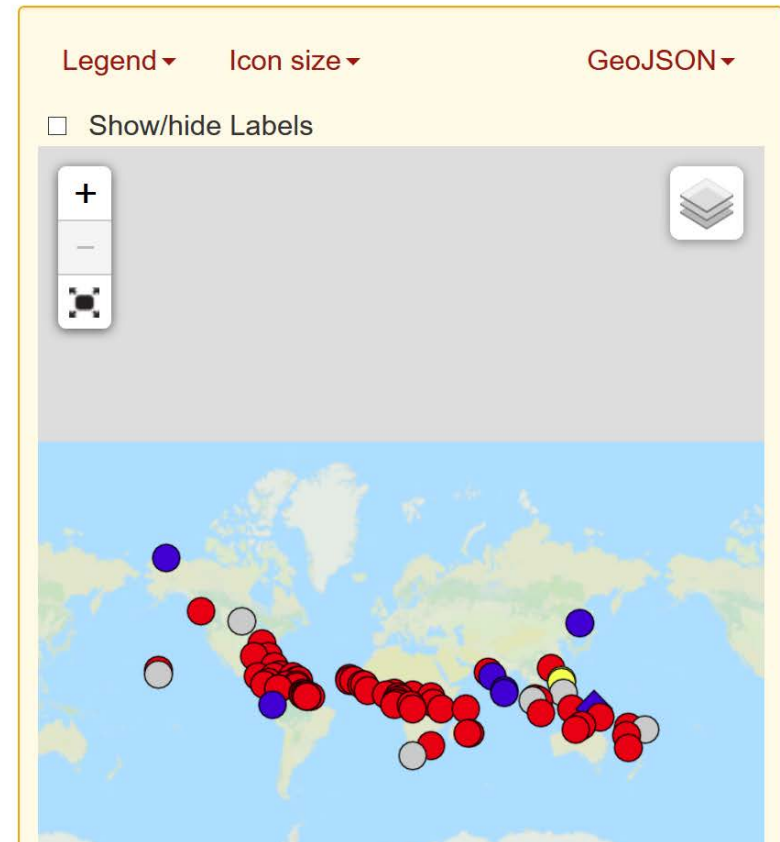
# クレオール形成を定量的に検証するためのデータは整いつつある

WALS map:  **81A Order of Subject, Object and Verb**

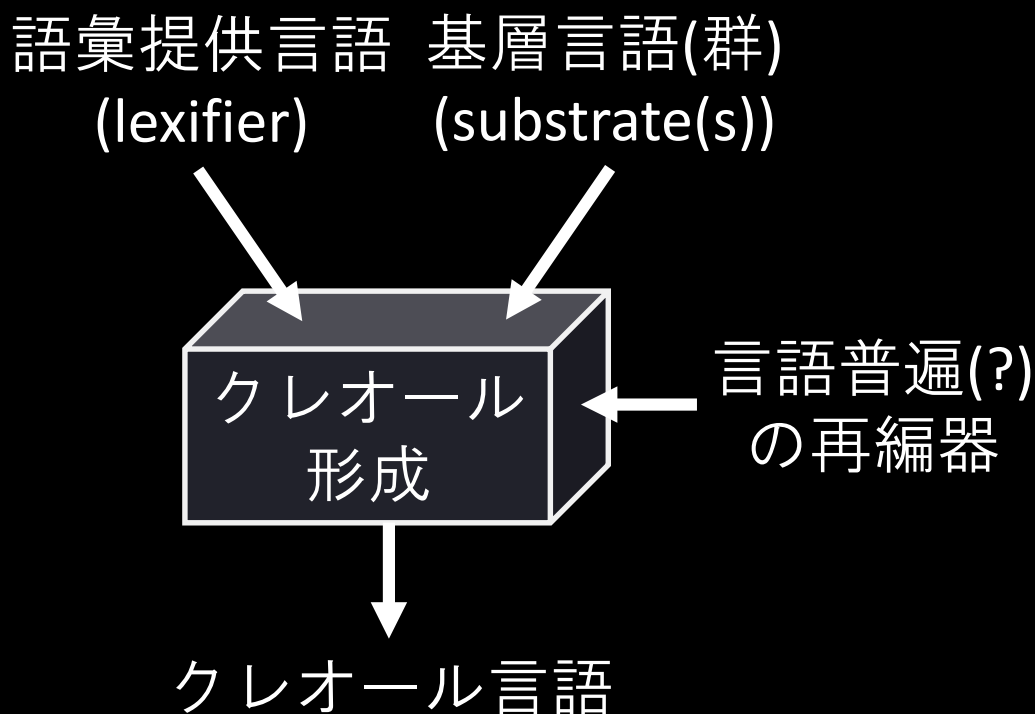
by Matthew S. Dryer



WALS-like APiCS map: **1 Order of subject, object, and verb**



# クレオール形成に対する諸仮説



1. **L**exifierの影響が強い
2. **S**ubstrateの影響が強い
3. LexifierとSubstrateの両方が強い
4. 言語普遍的な構造再編 (**R**) が強い

# 混合モデル

- クレオールはL, S, Rの確率的混合

$$\theta_{i,j,k} = \frac{\exp(m_{j,k} + n_{i,k})}{\sum_k \exp(m_{j,k} + n_{i,k})}$$

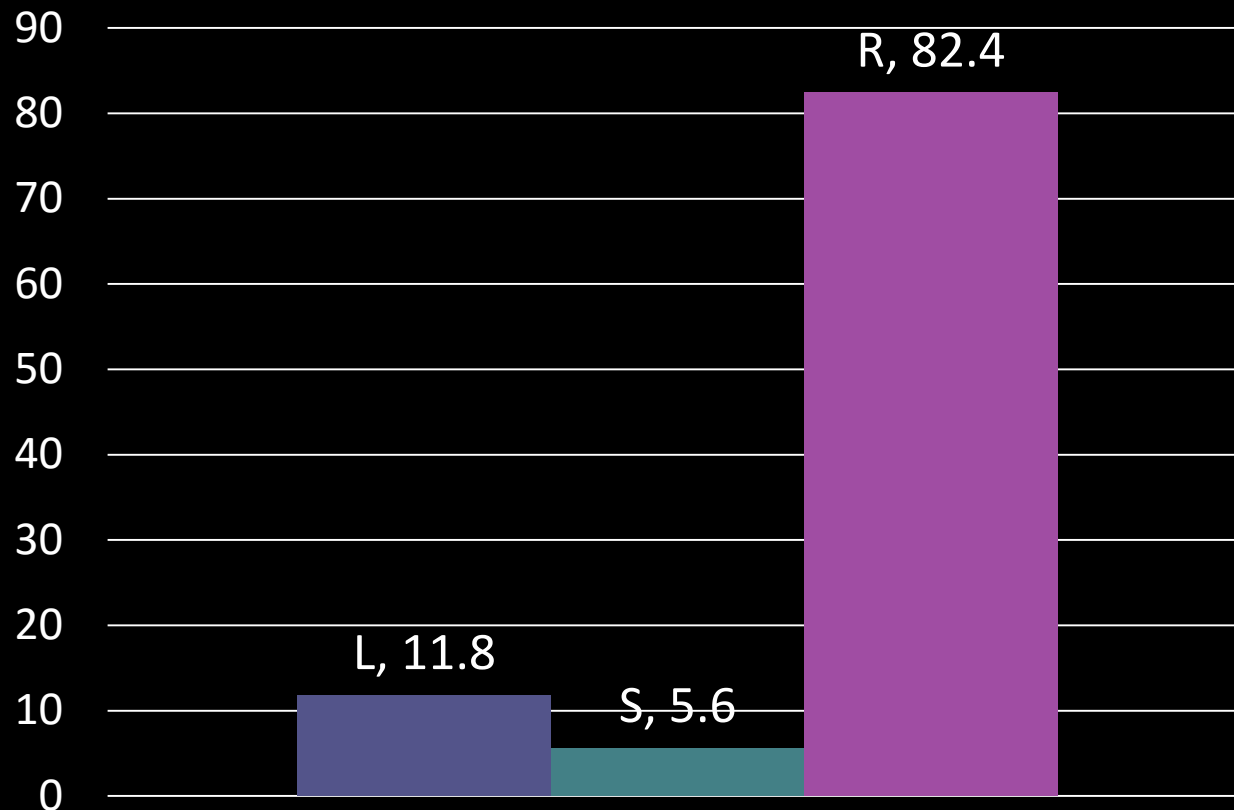
- $\theta_{i,j,k}$ : クレオール*i*の特徴*j*がソース*k*に由来する比率
- $m_{j,k}$ : 特徴*j*がソース*k*に由来する場合のパラメータ
- $n_{i,k}$ : クレオール*i*がソース*k*に由来する場合のパラメータ

- $z_{i,j} \sim \text{Categorical}(\theta_{i,j,L}, \theta_{i,j,S}, \theta_{i,j,R})$

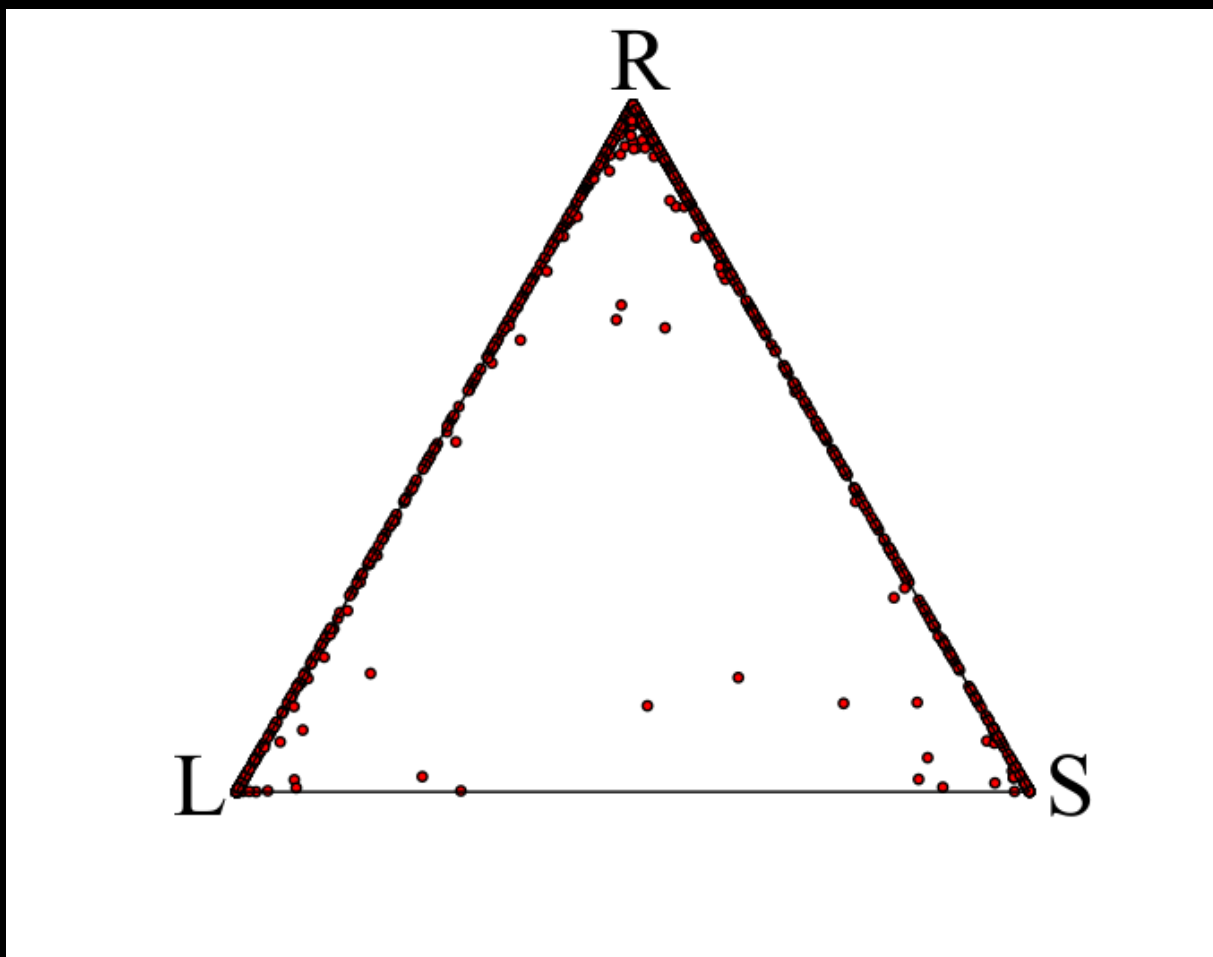
$$x_{i,j} \sim \begin{cases} \delta(y_{i,j,L}) & z_{i,j} = L \text{なら} L \text{の値をコピー} \\ \delta(y_{i,j,S}) & z_{i,j} = S \text{なら} S \text{の値をコピー} \\ \text{Categorical}(\phi_j) & z_{i,j} = R \text{なら} R \text{の確率分布から生成} \end{cases}$$

# 推定結果

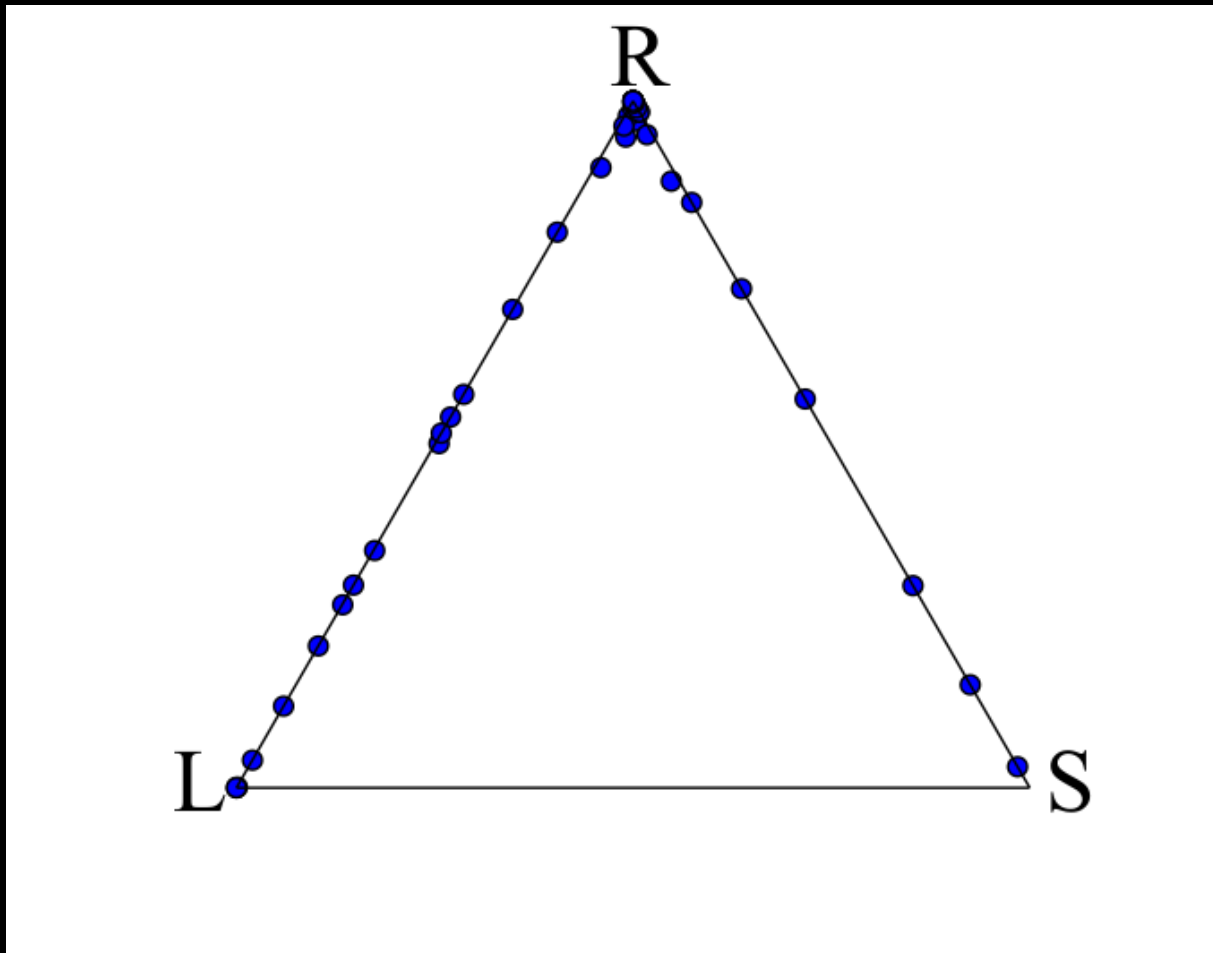
- 混合比の集約結果



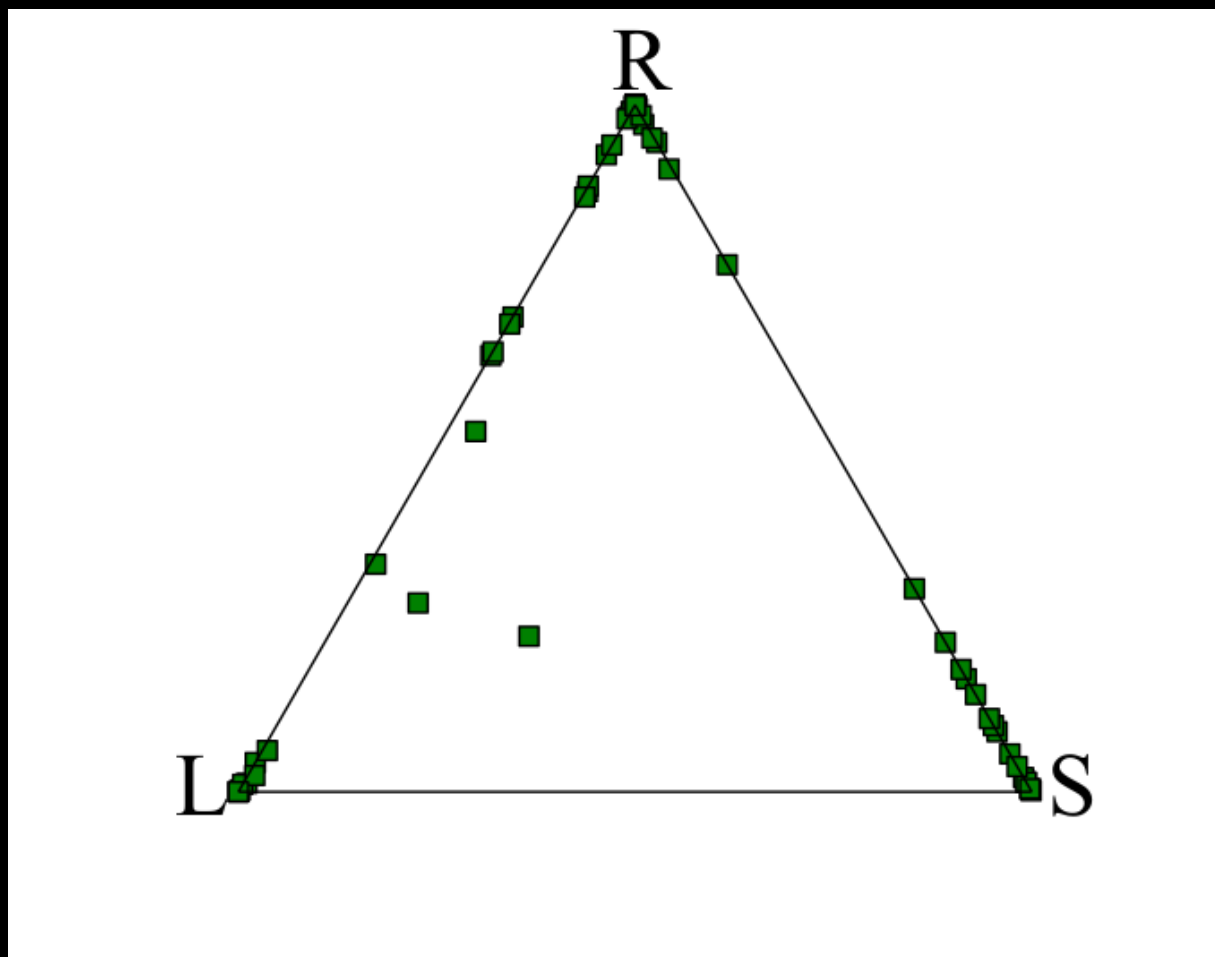
# $\theta_{i,j}$ の分布



# $m_{j,k}$ (特徴ごと) の分布



# $n_{i,k}$ (クレオールごと) の分布





# 構造再編の言語普遍性?

割合	特徴	日本語
90.3%	Numeral Classifiers: Absent	×
85.9%	Order of Numeral and Noun: Numeral-Noun	○
84.7%	Negative Morphemes: Negative particle	×
84.5%	Inclusive/Exclusive Distinction in Independent Pronouns: No inclusive/exclusive	○
81.9%	Predicative Possession: 'Have'	×
81.8%	Applicative Constructions: No applicative construction	○
74.3%	Negative Indefinite Pronouns and Predicate Negation: Predicate negation also present	○
73.8%	Gender Distinctions in Independent Personal Pronouns: No gender distinctions	×
73.0%	Alignment of Case Marking of Full Noun Phrases: Neutral	×
70.5%	Occurrence of Nominal Plurality: All nouns, always optional	×

# まとめ

- クレオール形成の定量的分析に混合モデルを適用
- 言語再編の影響が圧倒的という推定結果
  - ただしデータ処理に怪しい部分がある
- 日本語はあまりクレオールのでない
- クレオール形成に言語普遍性があるのなら、系統推定に組み込むことは可能そう

