Analyzing Correlated Evolution of Multiple Features Using Latent Representations

MURAWAKI Yūgo
Kyoto University

EMNLP2018 @ Brussels, Belgium 4 Nov 2018

Two Types of Data

1. Modern languages represented by typological features

152 discrete features

Tetum	1	2	…	1

Two Types of Data

1. Modern languages represented by typological features

152 discrete features

Tetum	1	2	…	1

Feature 26A: Prefixing vs. Suffixing in Inflectional Morphology

1. Little affixation
2. Strongly suffixing
3. Weakly suffixing
4. Equal prefixing and suffixing
5. Weakly prefixing
6. Strong prefixing

Two Types of Data

1. Modern languages represented by typological features

152 discrete features

Tetum	1	2	…	1

Feature 2 … fixing vs. Suffixing in Inflectional Morphology

1.
2.
3.
4.
5.
6.

Feature 81A: Order of Subject, Object and Verb

1. SOV
2. SVO
3. VSO
4. VOS
5. OVS
6. OSV:
7. No dominant Order

Two Types of Data

1. Modern languages represented by typological features

152 discrete features

2,557 languages

1	2	⋯	1
2	4	⋯	0
1	1	⋯	1
⋮	⋮	⋮	⋮
3	1	⋯	1

Two Types of Data

1. Modern languages represented by typological features

152 discrete features

2,557 languages

1	2	⋯	1
2	4	⋯	?
1	?	⋯	1
⋮	⋮	⋮	⋮
?	1	⋯	?

Two Types of Data

1. Modern languages represented by typological features

152 discrete features

2,557 languages

1	2	⋯	1
2	4	⋯	0
1	1	⋯	1
⋮	⋮	⋮	⋮
3	1	⋯	1

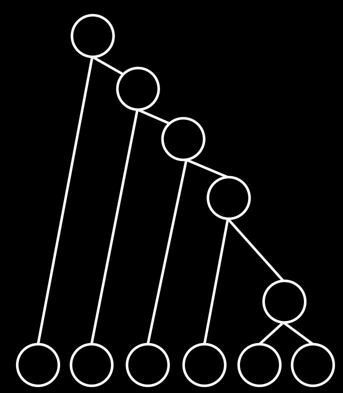
Two Types of Data

1. Modern languages represented by typological features

152 discrete features

Tetum	1	2	⋯	1
	2	4	⋯	0
	1	1	⋯	1
	⋮	⋮	⋮	⋮
	3	1	⋯	1

2. Phylogenetic trees relating these modern languages



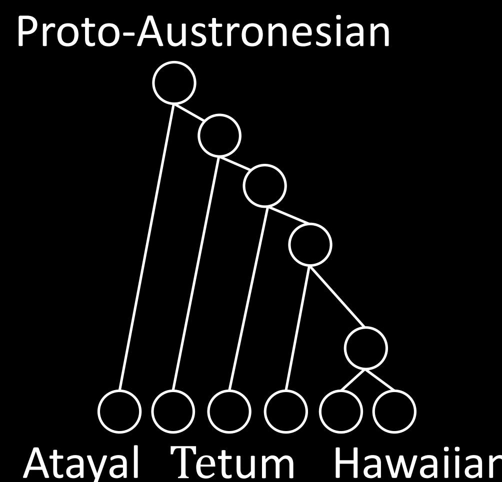
Two Types of Data

1. Modern languages represented by typological features

152 discrete features

Tetum	1	2	⋯	1
	2	4	⋯	0
	1	1	⋯	1
	⋮	⋮	⋮	⋮
	3	1	⋯	1

2. Phylogenetic trees relating these modern languages



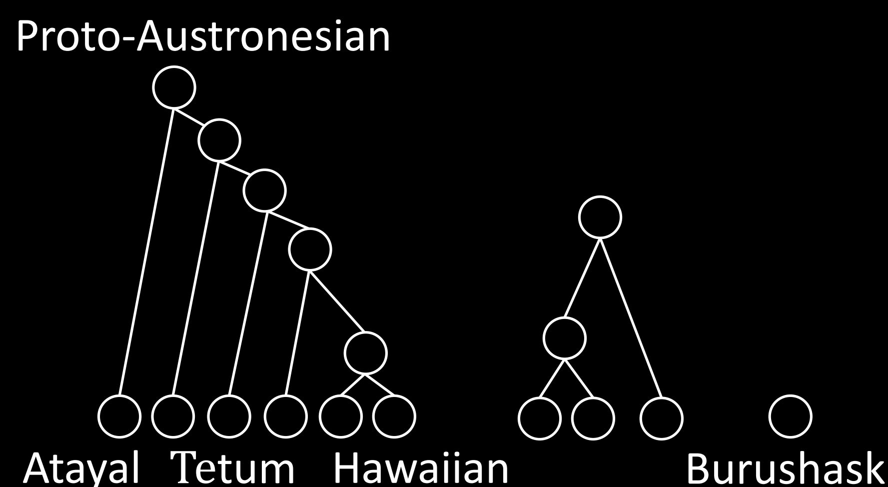
Two Types of Data

1. Modern languages represented by typological features

152 discrete features

Tetum	1	2	⋯	1
	2	4	⋯	0
	1	1	⋯	1
	⋮	⋮	⋮	⋮
	3	1	⋯	1

2. Phylogenetic trees relating these modern languages



Two Types of Data

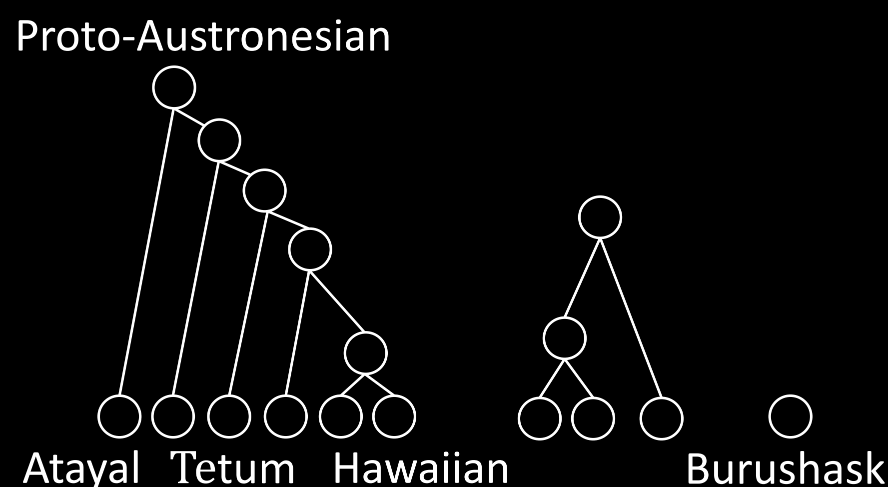
1. Modern languages represented by typological features

152 discrete features

Tetum	1	2	⋯	1
	2	4	⋯	0
	1	1	⋯	1
	⋮	⋮	⋮	⋮
	3	1	⋯	1

2. Phylogenetic trees relating these modern languages

309 language families
including 154 language isolates

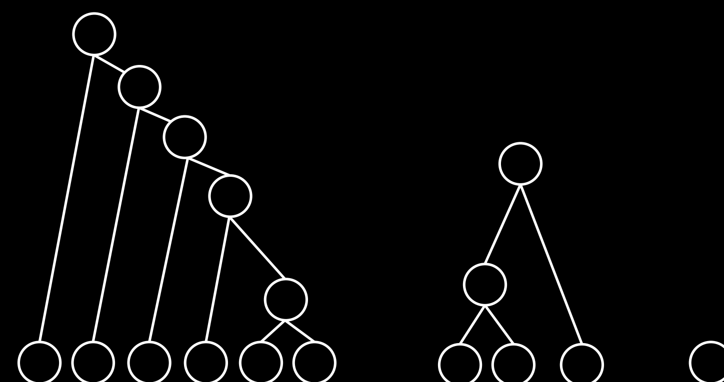


How have languages changed in the past?
How are they likely to change in the future?

To answer these questions,
I develop statistical models that make use of

1	2	⋯	1
2	4	⋯	0
1	1	⋯	1
⋮	⋮	⋮	⋮
3	1	⋯	1

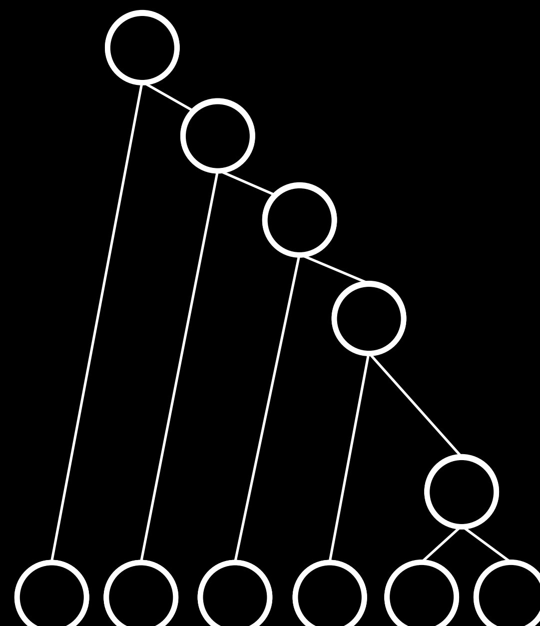
and



Phylogenetic Comparative Method

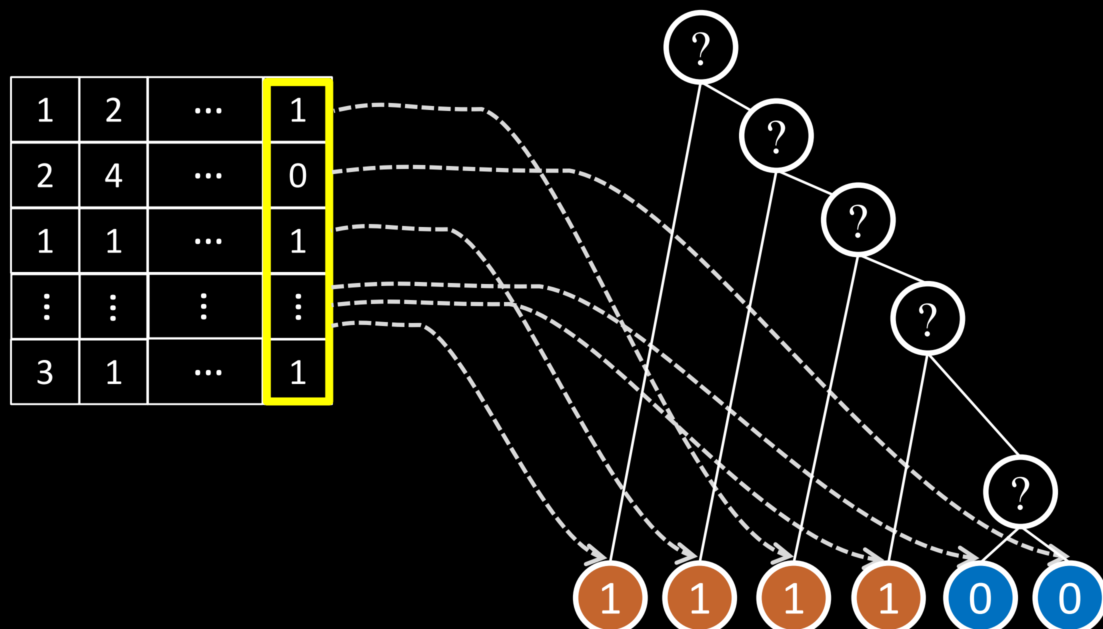
Trees allow us to infer ancestral states, with varying degrees of confidence

1	2	⋯	1
2	4	⋯	0
1	1	⋯	1
⋮	⋮	⋮	⋮
3	1	⋯	1



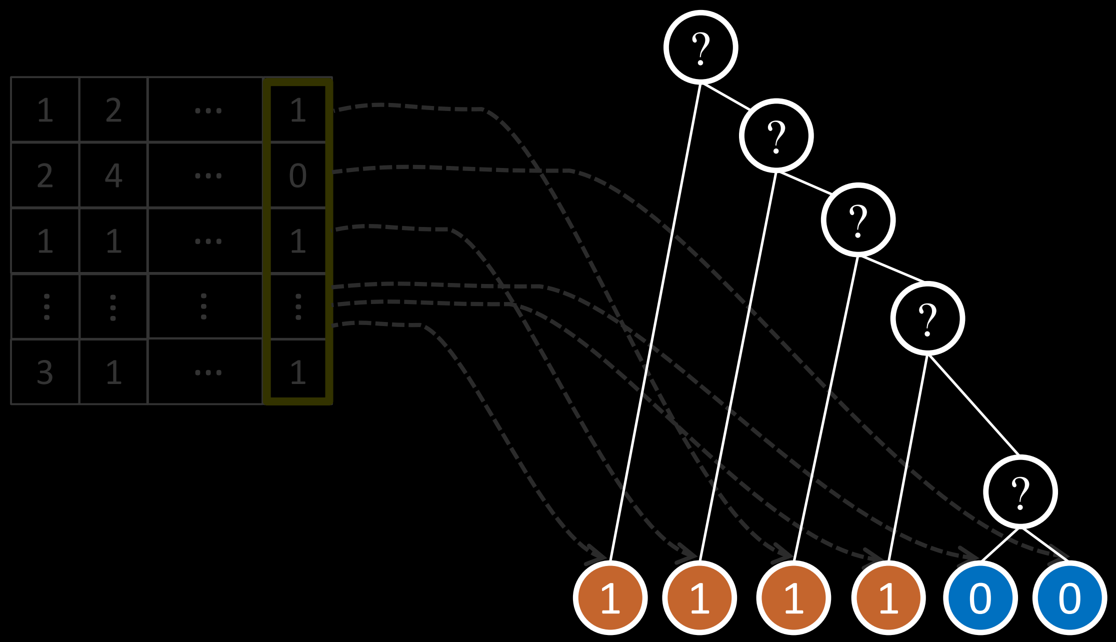
Phylogenetic Comparative Method

Trees allow us to infer ancestral states, with varying degrees of confidence



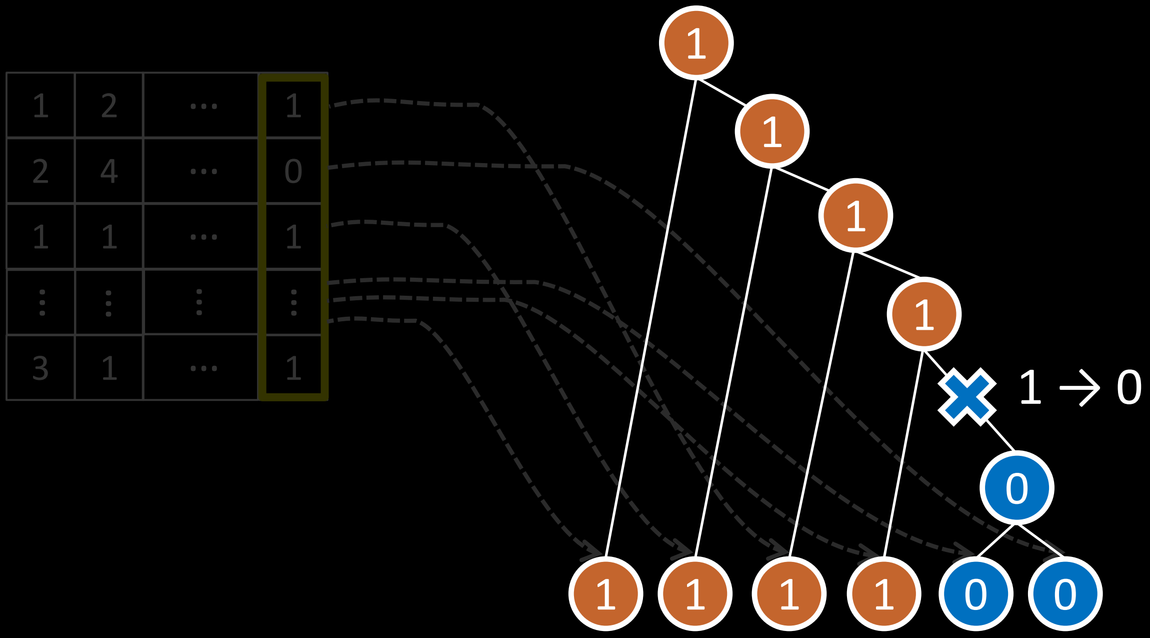
Phylogenetic Comparative Method

Trees allow us to infer ancestral states, with varying degrees of confidence



Phylogenetic Comparative Method

Trees allow us to infer ancestral states, with varying degrees of confidence



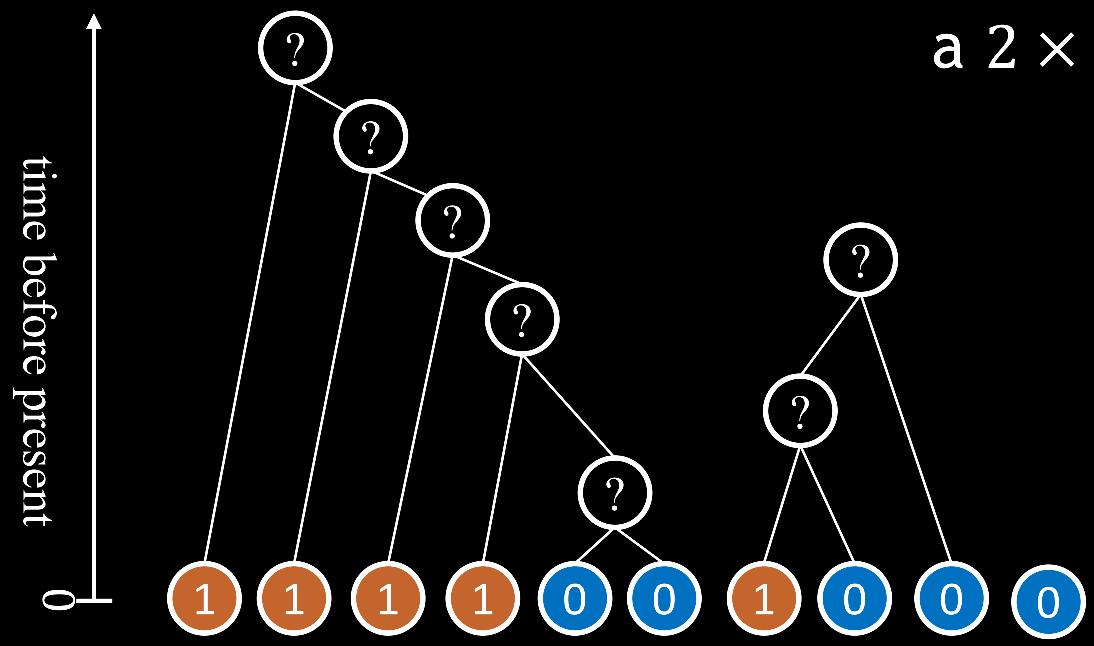
Continuous-time Markov Chains (CTMCs) for Statistical Analysis

Transition probability as a function of *continuous* time

$$P(x = b \mid \mathrm{parent}(x) = a, t) = \exp(tQ)_{a,b}$$

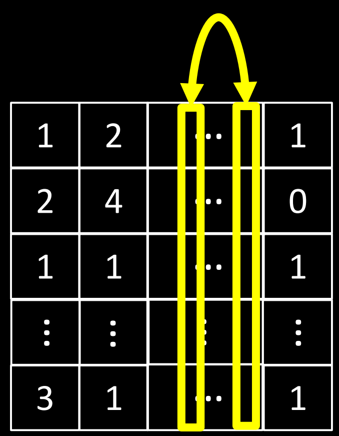
A binary feature has a 2×2 *transition rate matrix*

$$Q = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}$$



[Greenhill+, 2010]
[Maurits+, PNAS, 2014]

Wait, Features are not Independent
Need to Model *Correlated* Evolution



Implicational universals [Greenberg, 1963]

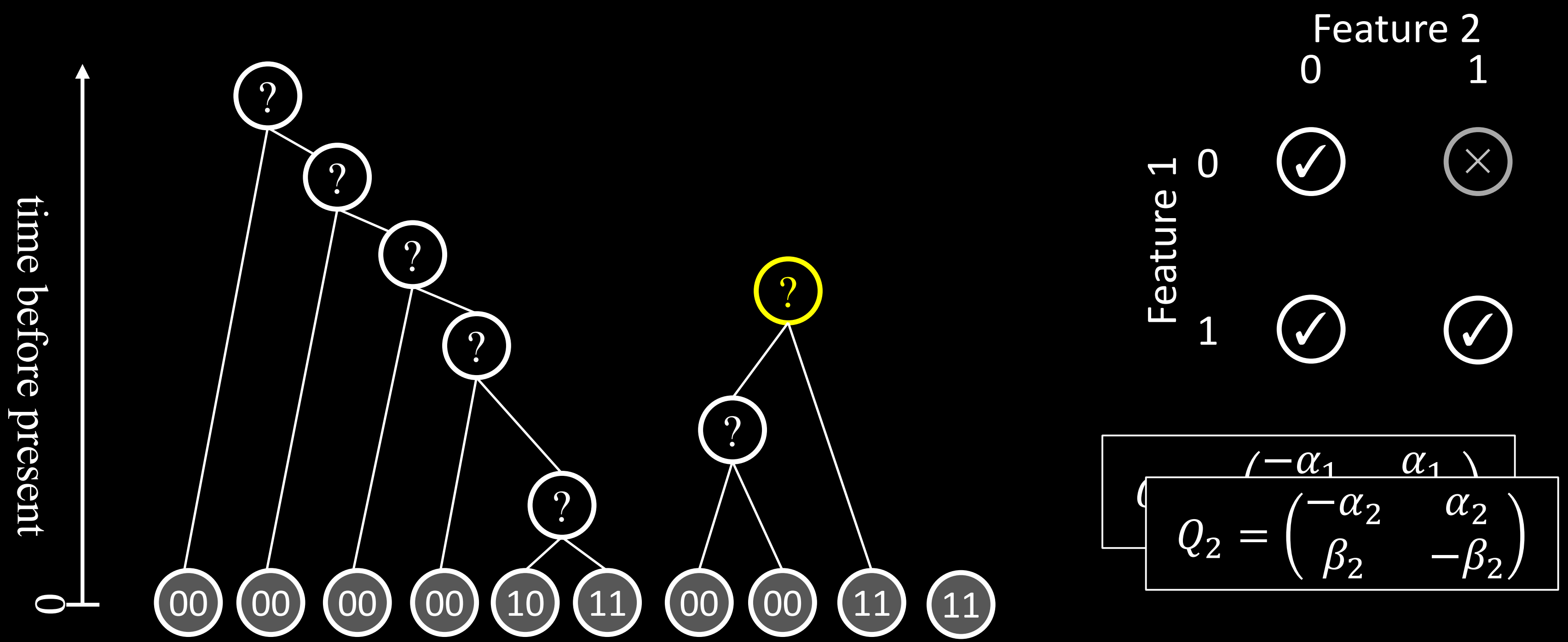
Order of noun and relative clause

Order of object and verb		NRel	RelN
	VO	✓	×
	OV	✓	✓

[Dryer, 2011]

Independent CTMCs

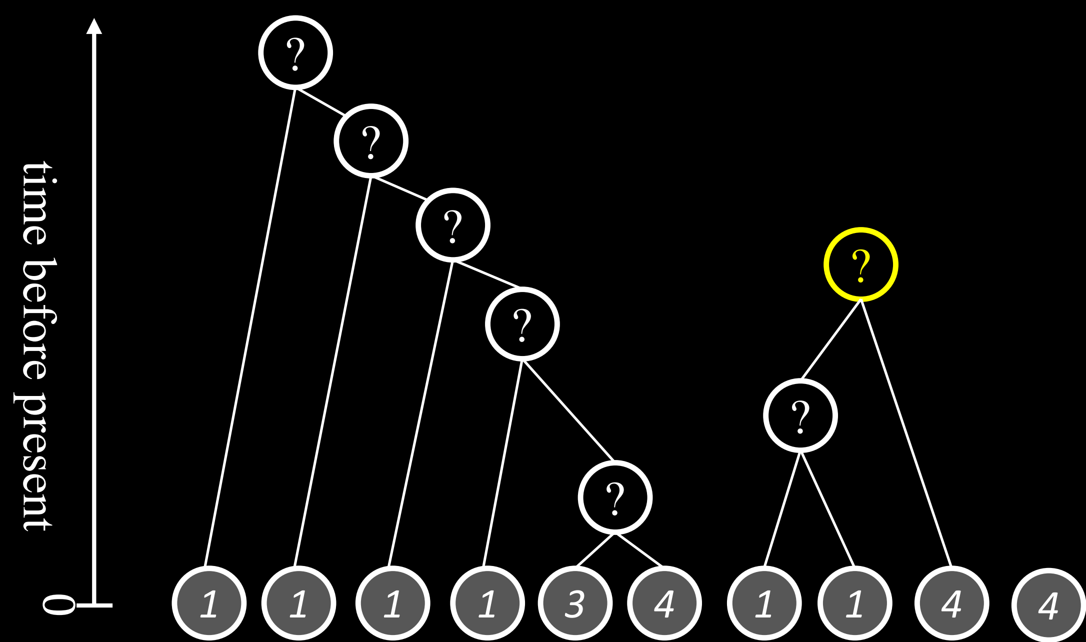
Unable to take into account the observation that the feature combination 01 is unnatural

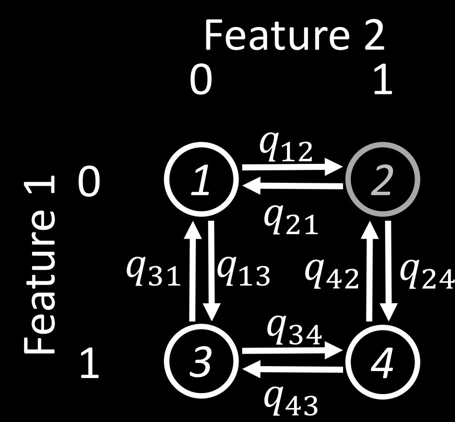


Expanding Feature Combinations Does not Scale

[Dunn+, Nature, 2011]

Combinatorial explosion prevents us from modeling interactions involving *multiple* features

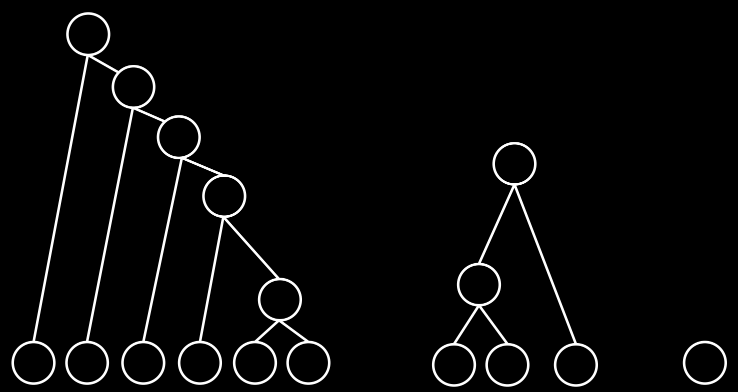




$$Q = \begin{array}{c} \\ 00 \\ 01 \\ 10 \\ 11 \end{array} \begin{array}{c} \begin{array}{cccc} 00 & 01 & 10 & 11 \end{array} \\ \begin{pmatrix} * & q_{12} & q_{13} & - \\ q_{21} & * & - & q_{24} \\ q_{31} & - & * & q_{34} \\ - & q_{42} & q_{43} & * \end{pmatrix} \end{array}$$

My Goal: Model Correlated Evolution Covering All Possible Dependencies

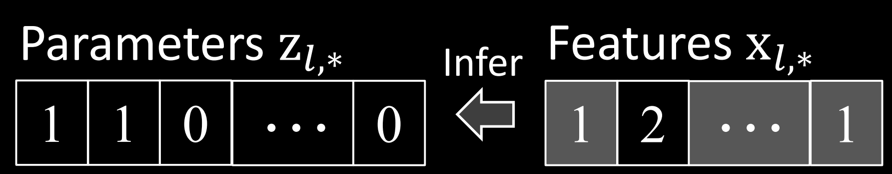
1	2	⋯	1
2	4	⋯	0
1	1	⋯	1
⋮	⋮	⋮	⋮
3	1	⋯	1



My Solution: Latent Representations

Idea originally presented in [Murawaki, NAACL2015]

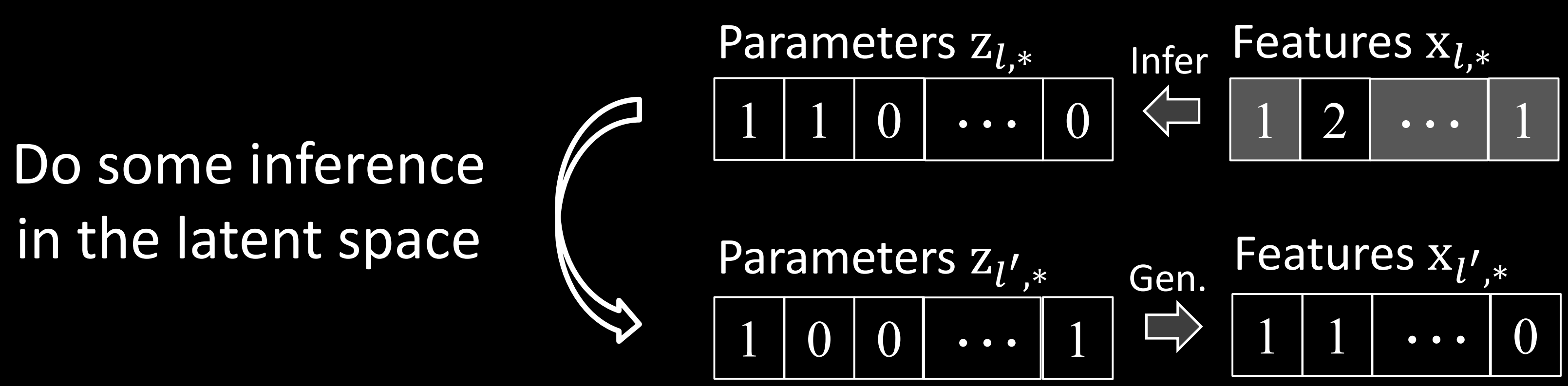
- Reorganize 152 discrete surface features into 100 binary latent *parameters*
  - Parameters are independent by assumption



My Solution: Latent Representations

Idea originally presented in [Murawaki, NAACL2015]

- Reorganize 152 discrete surface features into 100 binary latent *parameters*
  - Parameters are independent by assumption
- Inference in the latent space implicitly captures correlated evolution

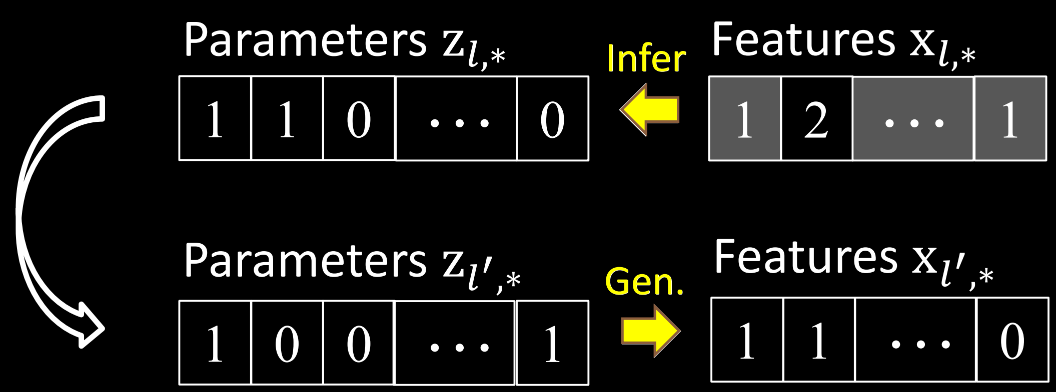


My Solution: Latent Representations

Idea originally presented in [Murawaki, NAACL2015]

- Reorganize 152 discrete surface features into 100 binary latent *parameters*
  - Parameters are independent by assumption
- Inference in the latent space implicitly captures correlated evolution

Do some inference in the latent space



Latent Representations Capture Inter-Feature Dependencies

[Murawaki, IJCNLP2017]

Parameters $z_{l,*}$

1	0	1	⋯	0

Features $x_{l',*}$

1	2	⋯	1

Latent Representations Capture Inter-Feature Dependencies

[Murawaki, IJCNLP2017]

Weight matrix W

2.9	0.4	-0.3	⋯	-0.2
6.3	-4.3	-5.7	⋯	5.9
8.2	-0.2	-2.5	⋯	0.3
⋮	⋮	⋮	⋮	⋮
0.2	0.3	1.2	⋯	-2.4

Parameters $\mathrm{z}_{l,*}$

1	0	1	⋯	0

$\times$ W $=$

Feature score vector

10.2	-9.8	-8.9	⋯	-4.9

Drawn from locally normalized distributions $\Downarrow$

Features $\mathrm{x}_{l',*}$

1	2	⋯	1

Latent Representations Capture Inter-Feature Dependencies

[Murawaki, IJCNLP2017]

Parameters $z_{l,*}$

1	0	1	⋯	0

×

Weight matrix W

2.9	0.4	-0.3	⋯	-0.2
6.3	-4.3	-5.7	⋯	5.9
8.2	-0.2	-2.5	⋯	0.3
⋮	⋮	⋮	⋮	⋮
0.2	0.3	1.2	⋯	-2.4

=

Feature score vector

10.2	-9.8	-8.9	⋯	-4.9

Drawn from locally normalized distributions $\Downarrow$

Features $x_{l',*}$

1	2	⋯	1

For $z_{l,k} = 1$, $w_{k,f(i_1,j_1)} \gg 0$ and $w_{k,f(i_2,j_2)} \ll 0$ indicate that feature i_1 is likely to take value j_1 and that feature i_2 is unlikely to take value j_2

Estimate Transition Rate Matrices for Parameters and Use Them for Simulation

Parameters $\mathrm{z}_{l,*}$

1	1	0	$\cdots$	0

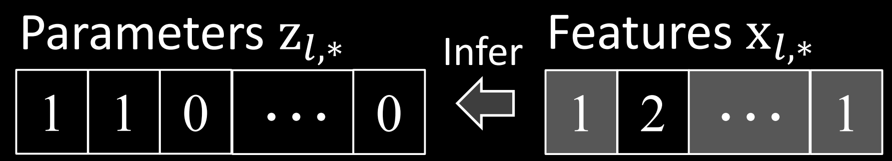
Infer $\Leftarrow$

Features $\mathrm{x}_{l,*}$

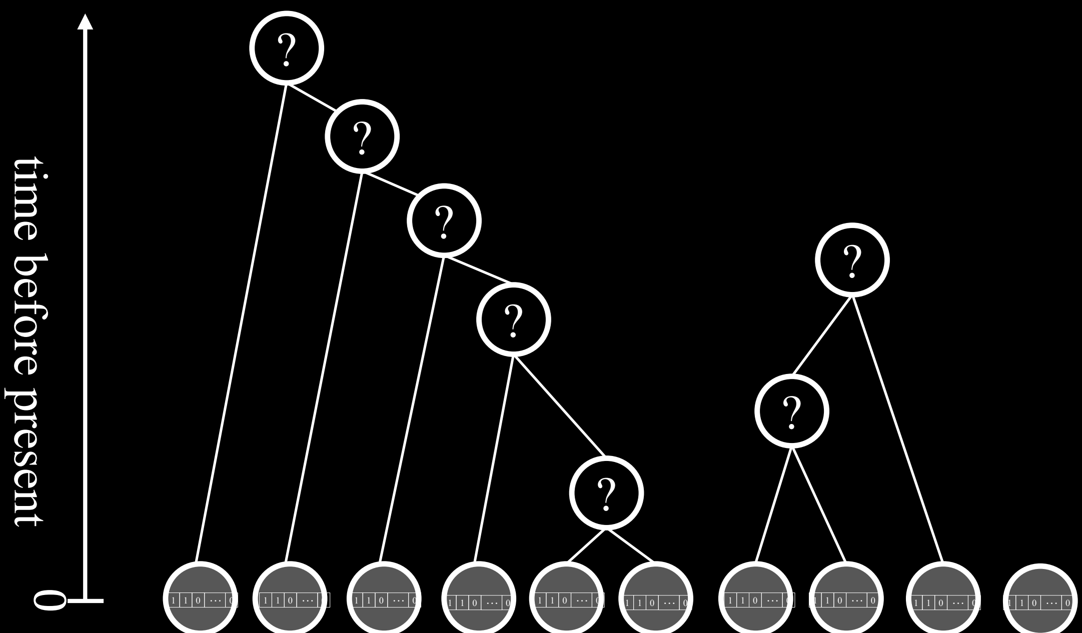
1	2	$\cdots$	1

Do some inference
in the latent space

Estimate Transition Rate Matrices for Parameters and Use Them for Simulation



Do some inference
in the latent space



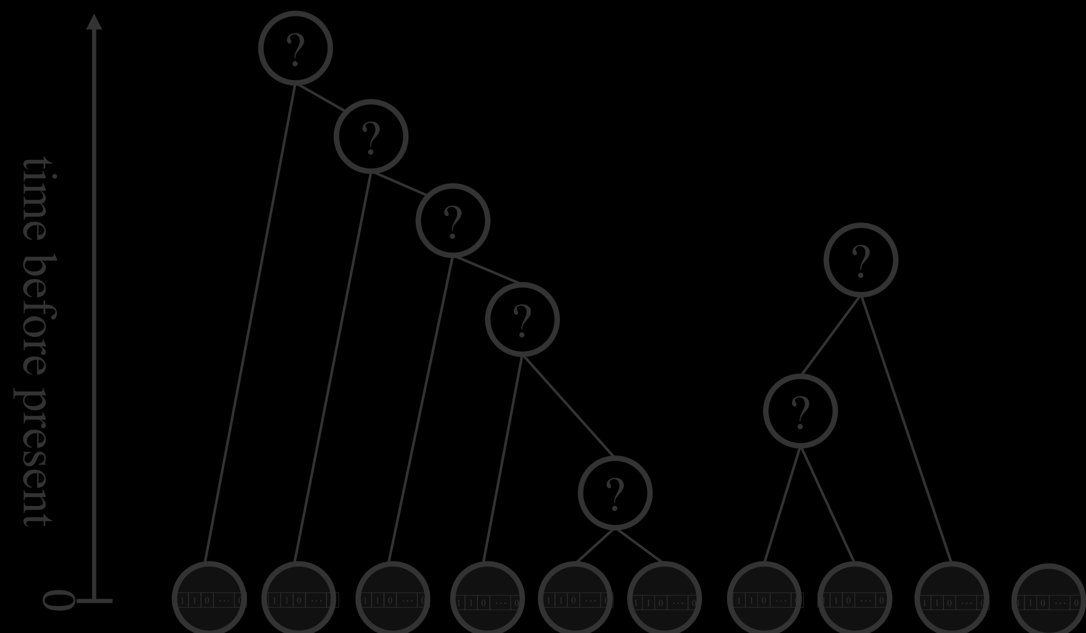
$$Q_{100} = \begin{pmatrix} -\alpha_{100} & \alpha_{100} \\ \beta_{100} & -\beta_{100} \end{pmatrix}$$

Estimate Transition Rate Matrices for Parameters and Use Them for Simulation

Parameters $\mathrm{z}_{l,*}$ Infer Features $\mathrm{x}_{l,*}$

$$\begin{array}{|c|c|c|c|c|}\hline 1 & 1 & 0 & \cdots & 0 \\ \hline \end{array} \xleftarrow{\text{Infer}} \begin{array}{|c|c|c|c|}\hline 1 & 2 & \cdots & 1 \\ \hline \end{array}$$

Do some inference
in the latent space



$$Q_{100} = \begin{pmatrix} -\alpha_{100} & \alpha_{100} \\ \beta_{100} & -\beta_{100} \end{pmatrix}$$

Estimate Transition Rate Matrices for Parameters and Use Them for Simulation

Do some inference in the latent space

Parameters $\mathrm{z}_{l,*}$ $\xleftarrow{\text{Infer}}$ Features $\mathrm{x}_{l,*}$

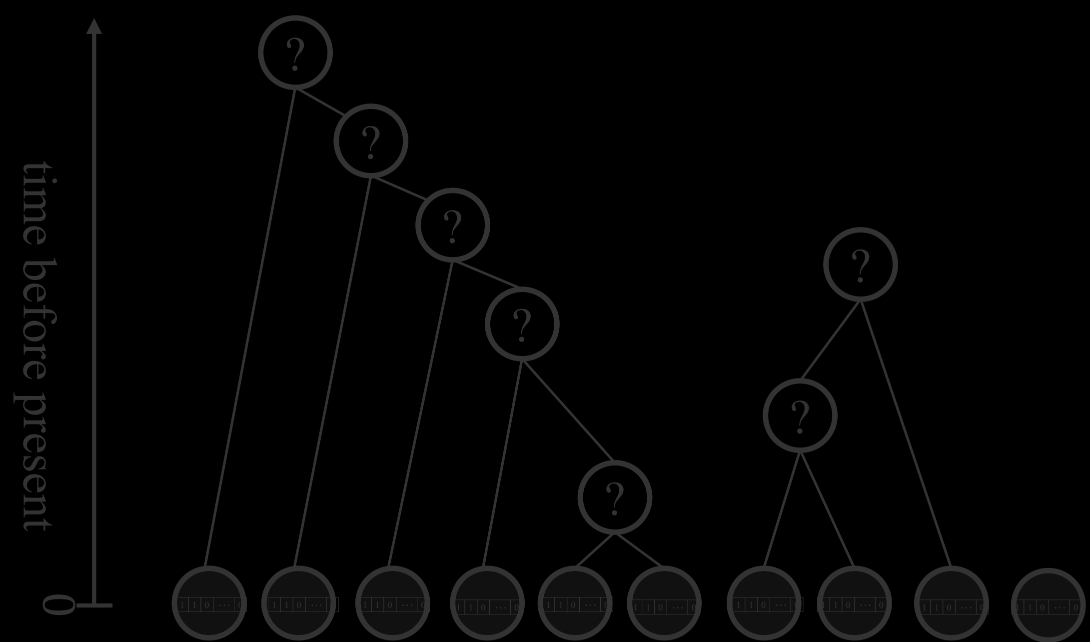
1	1	0	$\cdots$	0

1	2	$\cdots$	1

Parameters $\mathrm{z}_{l',*}$ $\xrightarrow{\text{Gen.}}$ Features $\mathrm{x}_{l',*}$

1	0	0	$\cdots$	1

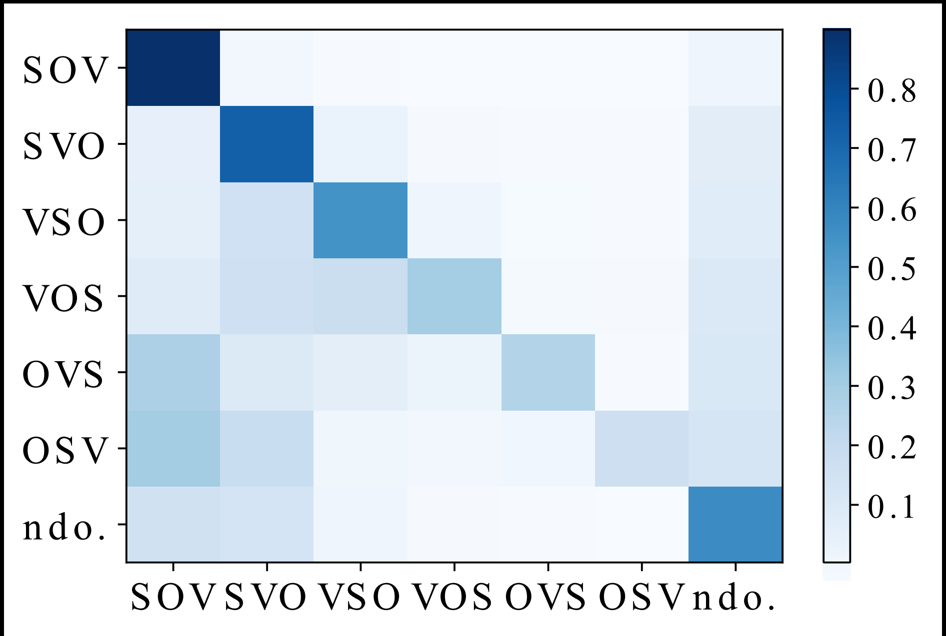
1	1	$\cdots$	0



$$Q_{100} = \begin{pmatrix} -\alpha_{100} & \alpha_{100} \\ \beta_{100} & -\beta_{100} \end{pmatrix}$$

Results: Order of Subject, Object and Verb

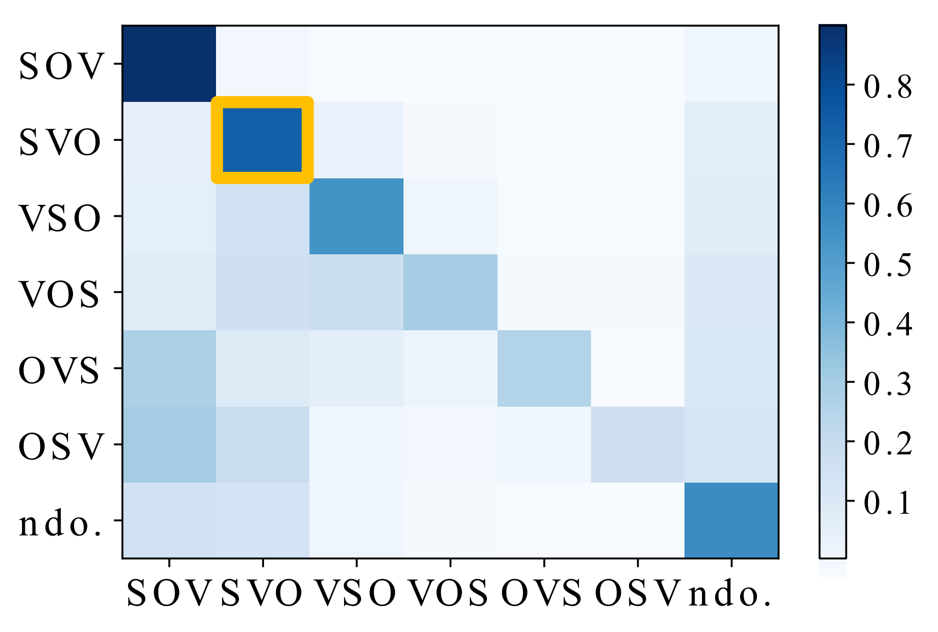
Transition prob. with $t = 2{,}000$



Largely agree with the results of [Maurits+ PNAS, 2014]

Results: Order of Subject, Object and Verb

Transition prob. with $t = 2{,}000$

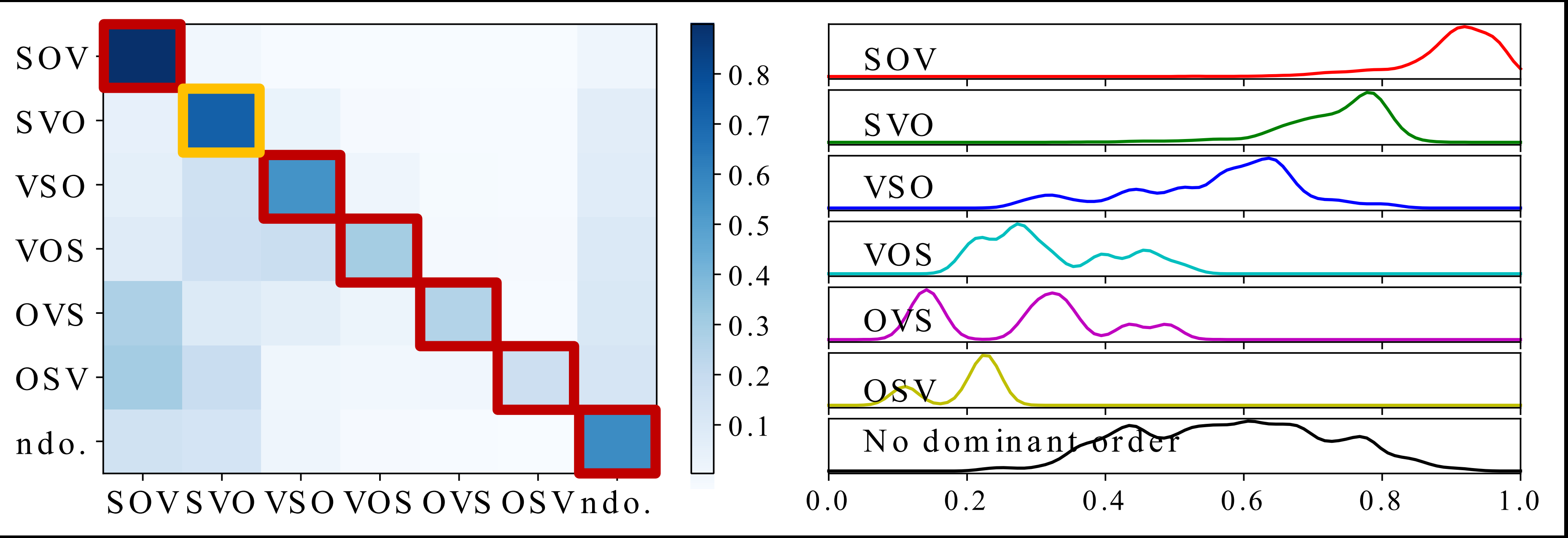


Largely agree with the results of [Maurits+ PNAS, 2014]

Results: Order of Subject, Object and Verb

Transition prob. with $t = 2{,}000$

Variability in prob. of keeping the same order

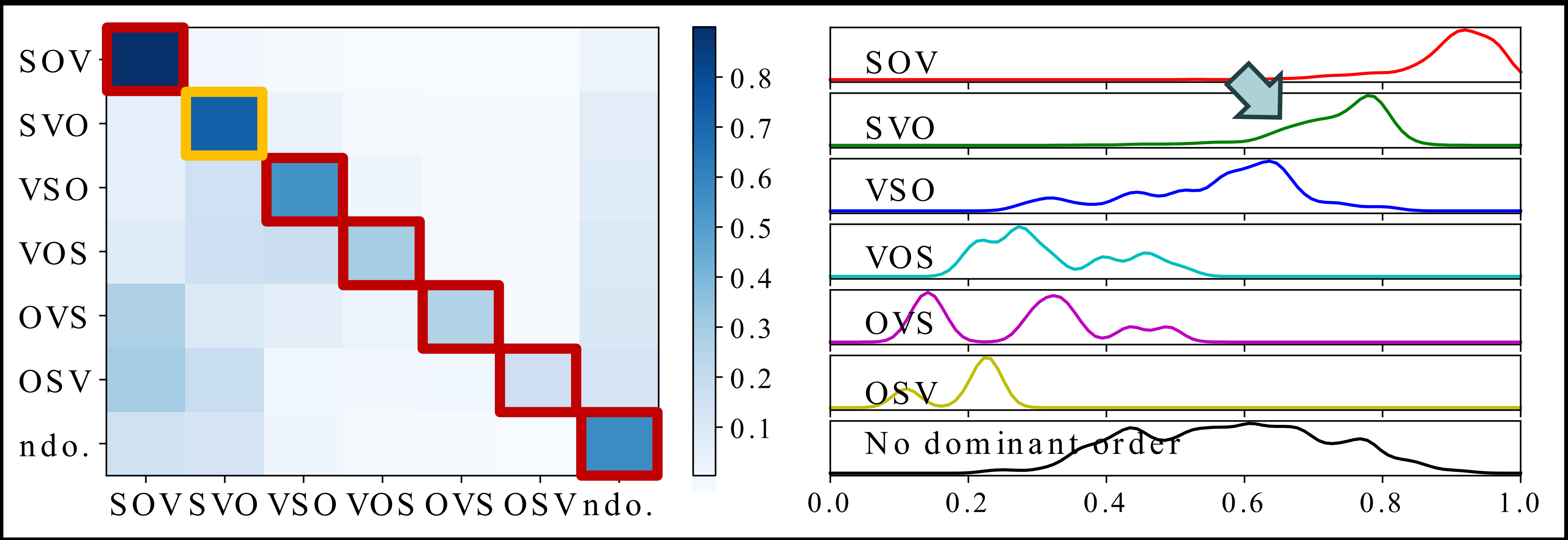


Largely agree with the results of [Maurits+ PNAS, 2014]

Results: Order of Subject, Object and Verb

Transition prob. with $t = 2{,}000$

Variability in prob. of keeping the same order

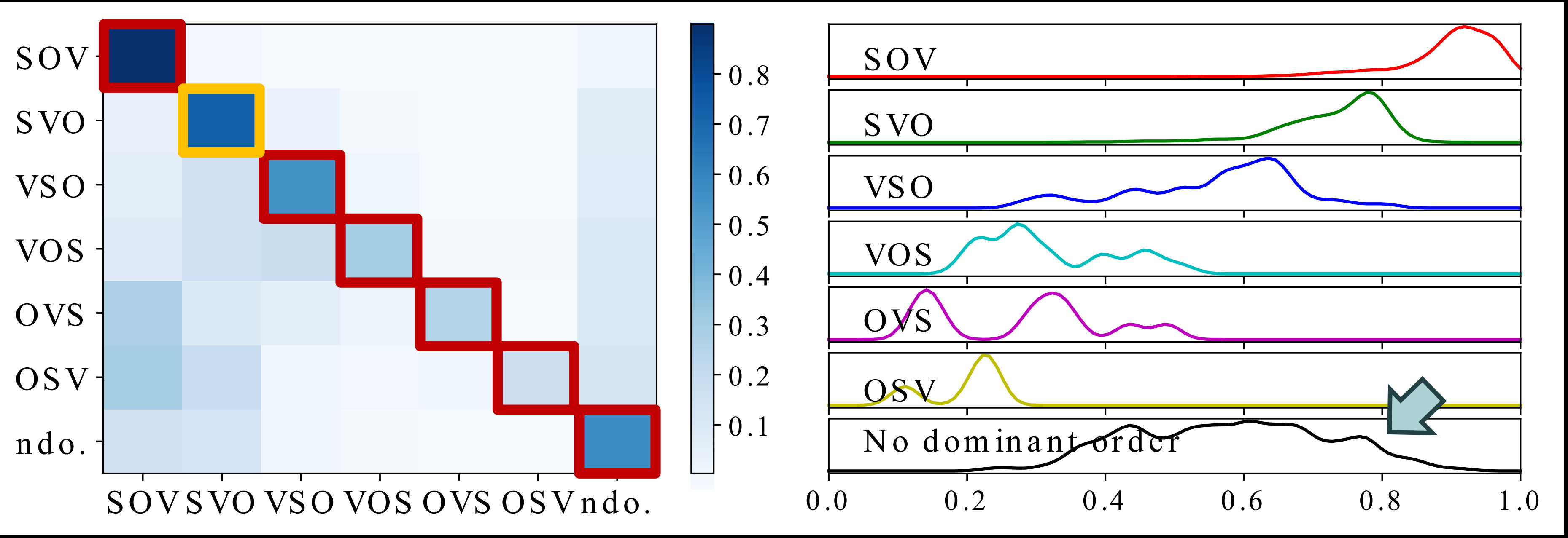


Largely agree with the results of [Maurits+ PNAS, 2014]

Results: Order of Subject, Object and Verb

Transition prob. with $t = 2{,}000$

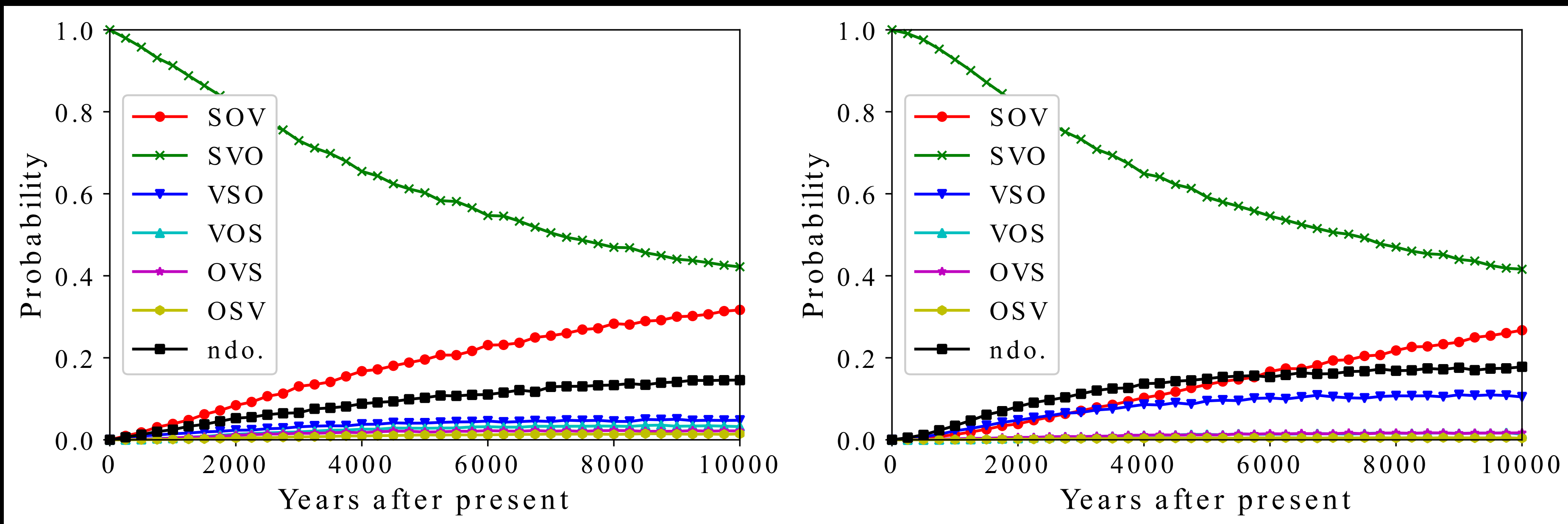
Variability in prob. of keeping the same order



Largely agree with the results of [Maurits+ PNAS, 2014]

Results: Order of Subject, Object and Verb

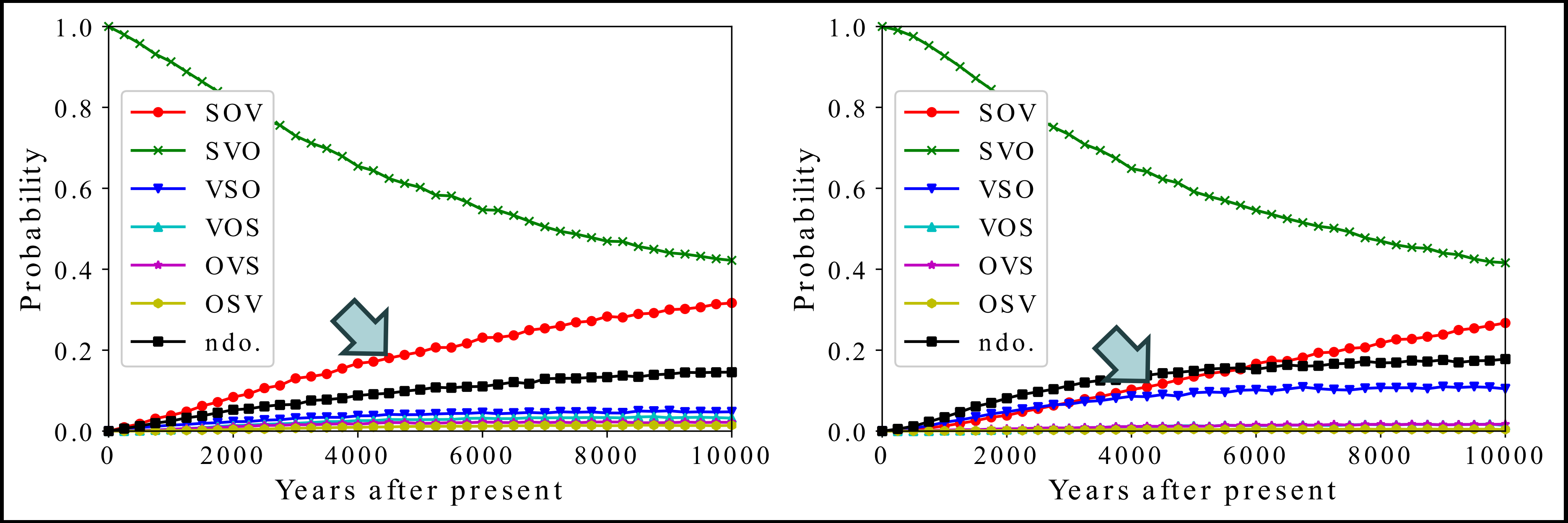
Tetum (Austronesian) South-Central Kikongo (Atlantic-Congo)



The strongly prefixing language, South-Central Kikongo, is more resistant to the change to SVO than Tetum

Results: Order of Subject, Object and Verb

Tetum (Austronesian) South-Central Kikongo (Atlantic-Congo)



The strongly prefixing language, South-Central Kikongo, is more resistant to the change to SVO than Tetum

Conclusions

- Proposed a new framework of latent representation-based analysis of diachronic typology
  - Investigate correlated evolution in an exploratory manner
- Future work
  - Analyze features other than the order of subject, object and verb
  - Inspect inferred ancestral states
  - Modeling contacts