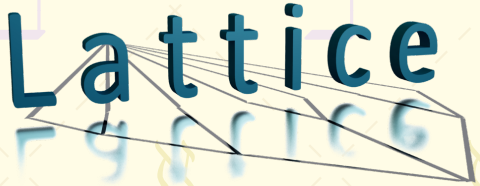
Language Evolution and Diachrony Generation

Research Project Report

Matthieu Boyer



Laboratoire Lattice

CNRS — ENS-PSL — Université Sorbonne Nouvelle

Under supervision of Mathieu Dehouck

Introduction

Few database available in diachrony :

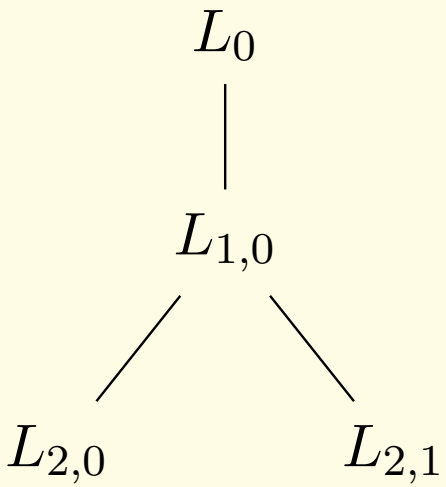
- The Index Diachronica [ind]
- The $\mathcal{E}$vosem [FKD$^+$25]

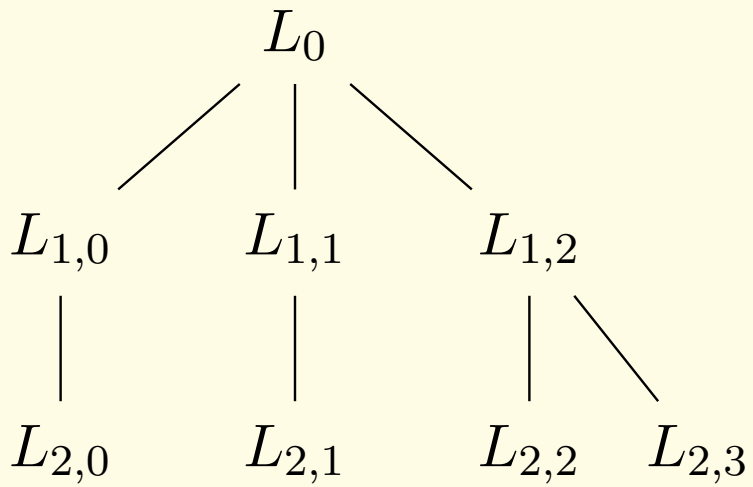
There is a need for less localized data.

Plan

Evolution as Random Trees I

Consider a language L_0, which we will call our *base language*.





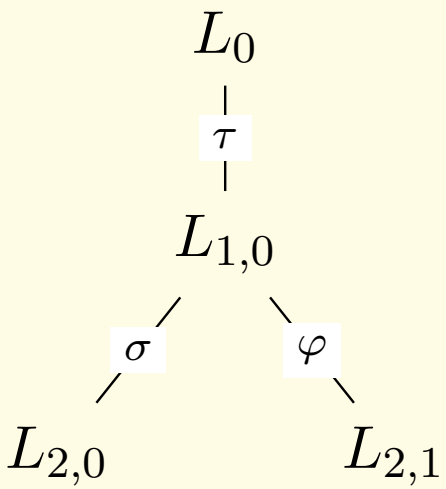
Evolution as Random Trees II

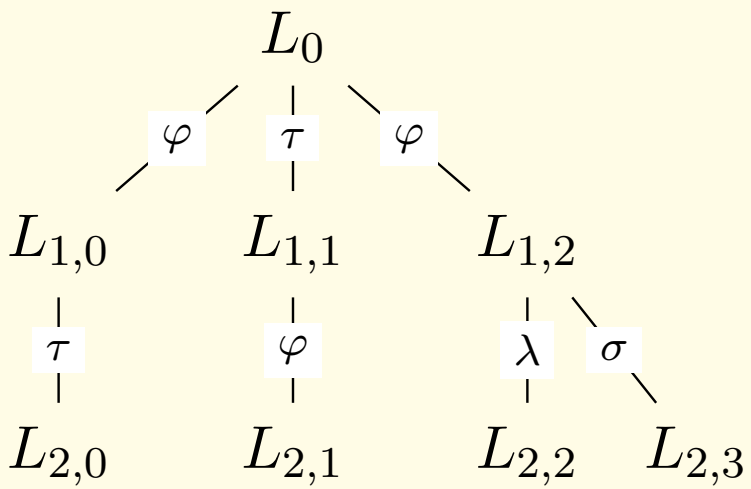
Algorithm One Language Evolution

$leaves \leftarrow \{L_0\}$
$\mathcal{T} \leftarrow \mathrm{Tree}\,(L_0, \varnothing)$
for $n \leq$ `Epochs` **do**
  for $l \in \mathrm{Leaves}\,\mathcal{T}$ **do**
    $S \leftarrow \mathrm{Evolve}\,l$
    $l \leftarrow \mathrm{Tree}\,(l, S)$ ▷ Here $\mathcal{T}$ is modified in place.
return $\mathcal{T}$

Specification of Evolve I

We want to choose between *evolution types* $(\varphi, \sigma, \tau, \lambda)$ at computation :





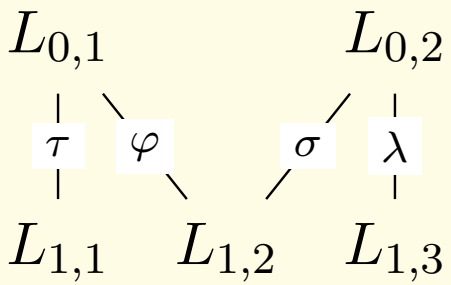
Specification of Evolve II

We want Evolve to be *easily revertible* :

$$\mathbb{P}\left(\text{Evolve}\left(l_1\right) = l_2\right) \neq 0 \Leftrightarrow \mathbb{P}\left(\text{Evolve}\left(l_2\right) = l_1\right) \neq 0$$

Collision Hypothesis I

We assume language interacting create evolutions :



Collision Hypothesis II

Algorithm Two Language Evolution

$\mathcal{T} \leftarrow \text{Tree}\,(L_0, \varnothing)$
for $n \leq$ `Epochs` **do**
  for $l \in \text{Leaves}\,\mathcal{T}$ **do**
    $S \leftarrow \text{Evolve}\,l$
    $\text{Push}\,(\text{Stack}, l \leftarrow l \cup \text{Tree}\,(l, S))$
  for $l \in \text{Leaves}\,\mathcal{T}$ **do**
    $l^\dagger \leftarrow \mathcal{P}_l\,(\text{Leaves}\,\mathcal{T})$
    $S \leftarrow \text{Collision}\,(l, l^\dagger)$
    $\text{Push}\,(\text{Stack}, l \leftarrow l \cup \text{Tree}\,(l, S))$
  $\text{Apply}(\text{Stack})$
return $\mathcal{T}$

Specification of Collision

Collision should :

- be *easily revertible*.

Specification of Collision

Collision should :

- be *easily revertible*.
- provide a way to choose collision type, for each parent.

Specification of Collision

Collision should :

- be *easily revertible*.
- provide a way to choose collision type, for each parent.
- take *linguistic* proximity of the parents into account.

Specification of Collision

Collision should :

- be *easily revertible*.
- provide a way to choose collision type, for each parent.
- take *linguistic* proximity of the parents into account.
- take the probability distribution $\mathcal{P}$ as the strength of the collisions.

References

Alexandre François, Siva Kalyan, Mathieu Dehouck, Martial Pastor, and David Kletz.

Evosem : A database of dialexification across language families.

Online database., 2025.

Index diachronica.

`https://chridd.nfshost.com/diachronica/`.