Laboratoire Lattice

CNRS — ENS-PSL — Université Sorbonne Nouvelle

Under supervision of Mathieu Dehouck

Introduction

Few database available in diachrony :

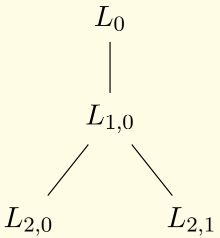
- The Index Diachronica [?]
- The $\mathcal{E}$vosem [?]

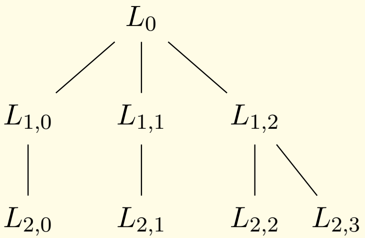
There is a need for less localized data.

Plan

Evolution as Random Trees I

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





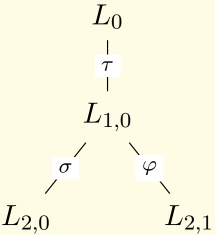
Evolution as Random Trees II

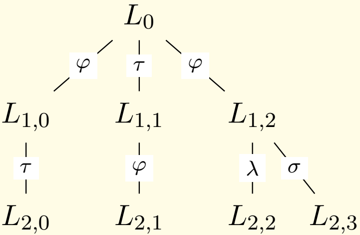
Algorithm One Language Evolution

$leaves \leftarrow \{L_0\}$
$\mathcal{T} \leftarrow \mathrm{Tree}\,(L_0, \varnothing)$
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* :

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

Plan

Collision Hypothesis I

We assume language interacting create evolutions :

$$\begin{tikzcd}
L_{0,1} \arrow[d, dash, "\tau" description] \arrow[dr, dash, "\varphi" description] & & L_{0,2} \arrow[dl, dash, "\sigma" description] \arrow[d, dash, "\lambda" description] \\
L_{1,1} & L_{1,2} & L_{1,3}
\end{tikzcd}$$

Collision Hypothesis II

Algorithm Two Language Evolution

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

A Manifold of Languages I

$\mathcal{P}_l$ models the probability of interaction with l. Defining a *geographical* embedding gives :

$$\mathcal{P}_l \propto \frac{1}{d_l\left(x\right)}$$

A Manifold of Languages II

- The simplex, that is, $d_l(x) = 1$ for all l, x.
- $\mathbb{R}^3$, with the ℓ^2 distance.
- The 2-sphere $\S^1$ where each language is a pairλ, φ :

$$d_{(\theta_1,\lambda_1)}(\theta_2, \lambda_2) = \arccos\left(\sin\left(\varphi_1\right)\sin\left(\varphi_2\right) + \cos\left(\varphi_1\right)\cos\left(\varphi_2\right)\cos\left(\lambda_2\right.\right.$$

A Manifold of Languages III

We suppose a language only interacts with languages from the same epoch, for now. We could add a new dimension to the manifold to modelize time.
Moreover, we are not required to use a metric but simply a positive separated function.