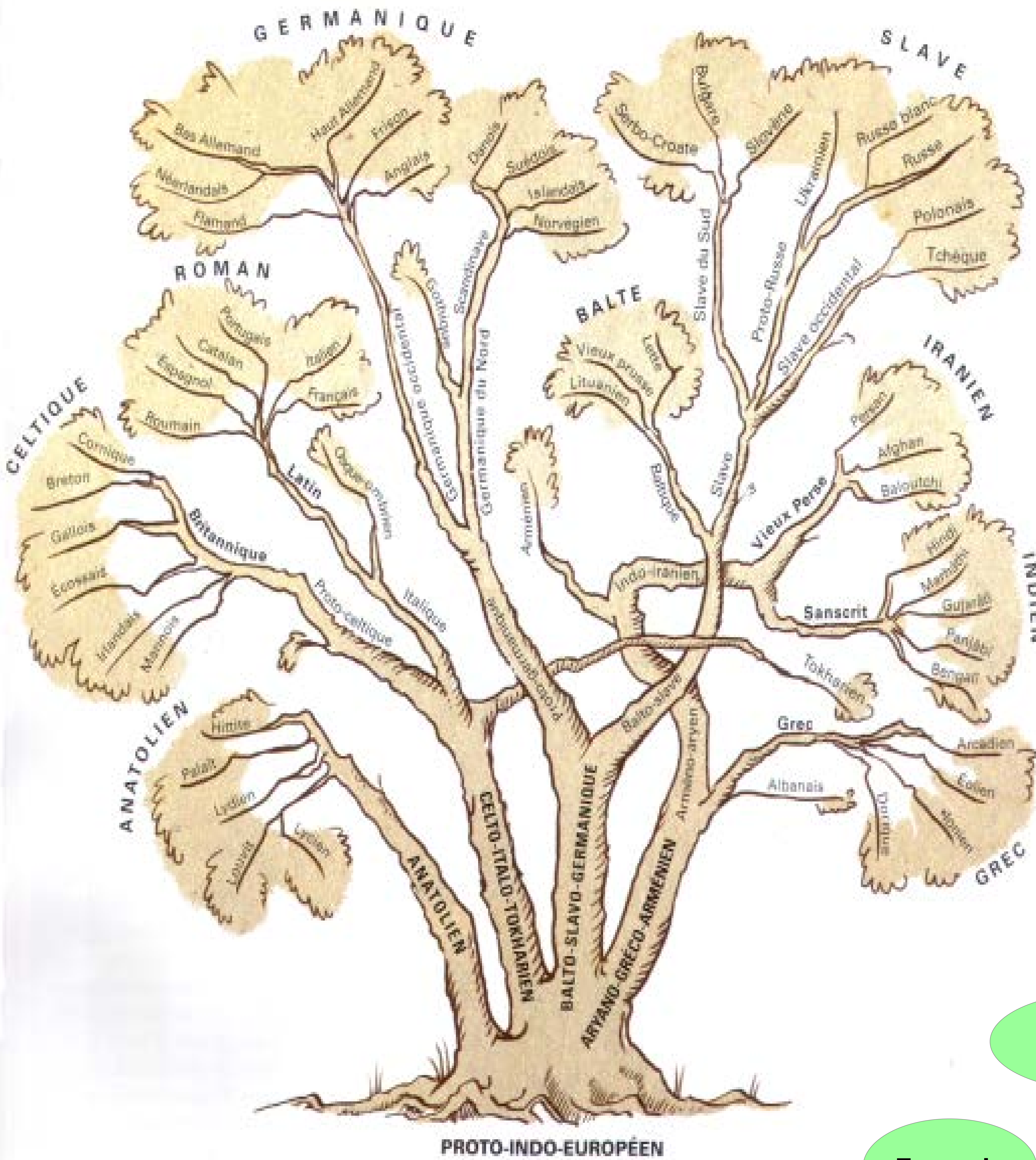


L'évolution des langages indo-européen

Département d'informatique

Antoine CAYPHAS, James TRIPODI, Kevin DEBRY et François TROESSAERT



Nos Buts :

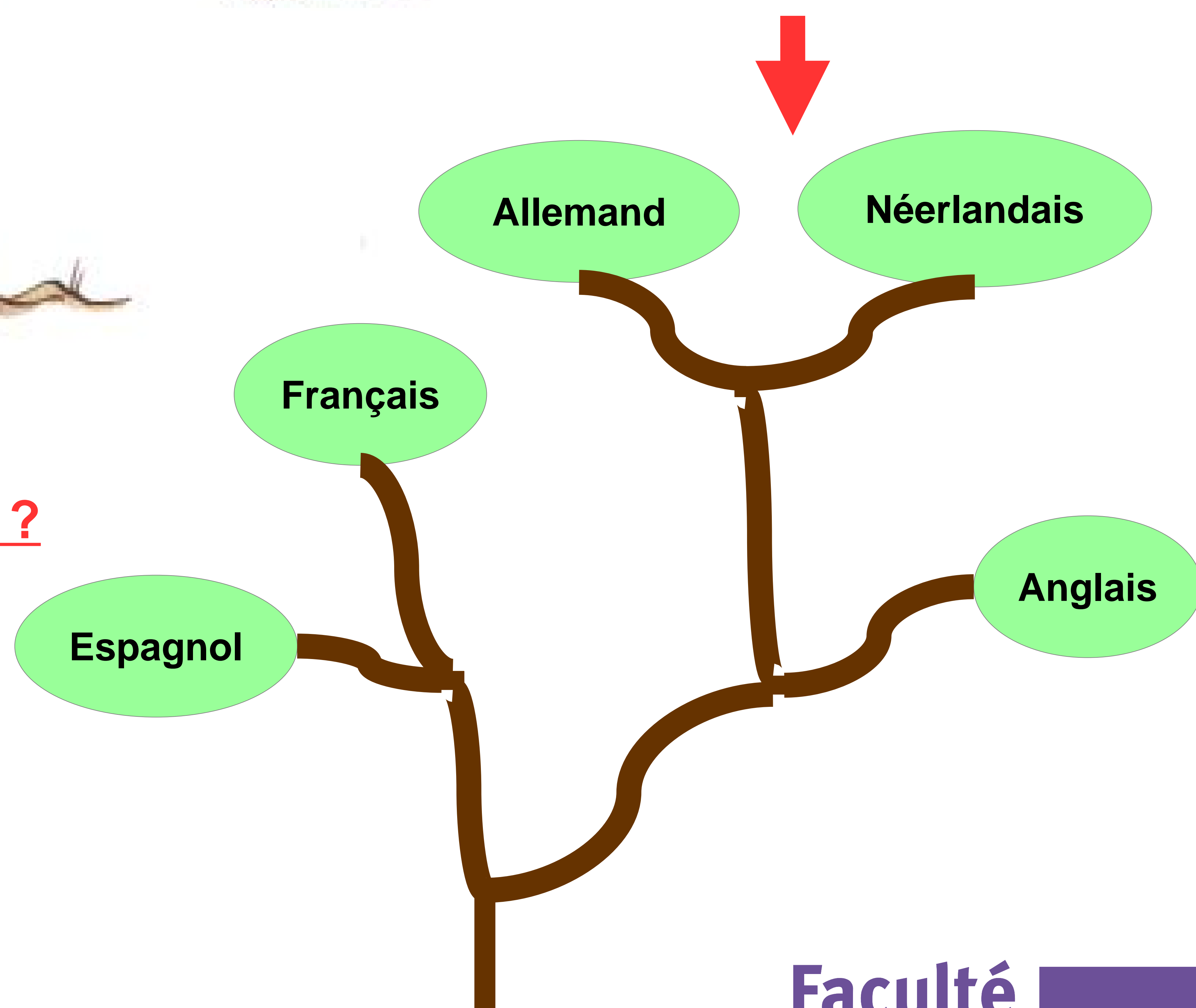
- Déterminer
- Observer

L'évolution des langages

Comment représenter l'évolution de ces langages ?

- Un ARBRE

Comment construire un arbre de ce type ?



Pourquoi l'informatique peut-il être utile ?

Quelle donnée ?

- Lexical

Quelle méthode ?

- Les méthodes de distance

L'évolution des langages indo-européen

Département d'informatique

Antoine CAYPHAS, James TRIPODI, Kevin DEBRY et François TROESSAERT

Distance de Levenshtein

Langue	Mots
Français / Fr	Champ
Néerlandais / NI	Veld
Anglais / En	Field
Allemand / De	Feld
Espagnol / Es	Campo

Calcul du nombre minimum de :
- Insertion / Délétion
- Substitution

C h a m p - 2 Insertions / Délétions
C - a m p o 0 Substitution
Distance = 2 + 0 = 2

V - e l d 1 Insertion / Délétion
F i e l d 1 Substitution
Distance = 1 + 1 = 2

	Champ	Veld	Field	Feld	Campo
Champ	0	5	5	5	2
Veld	5	0	2	1	5
Field	5	2	0	1	5
Feld	5	1	1	0	5
Campo	2	5	5	5	0

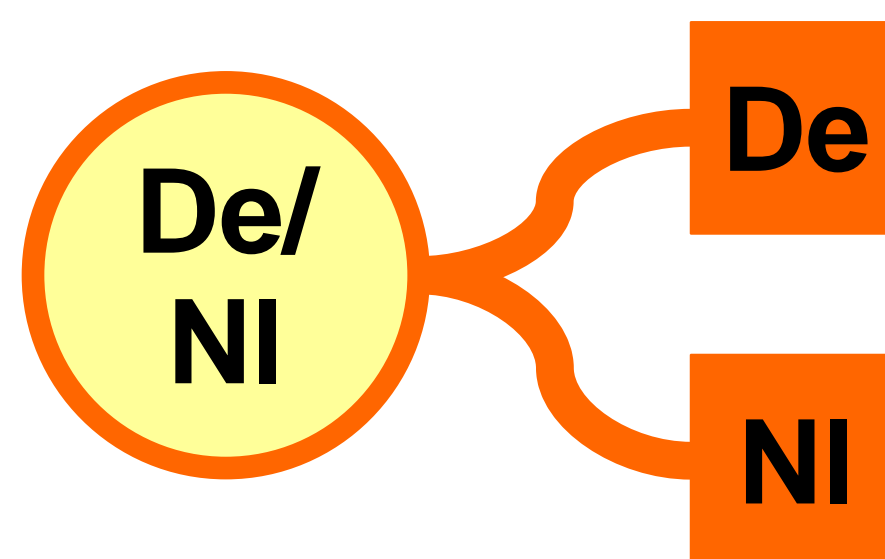
Algorithme UPGMA

Matrice de Distance

Cycle 1

	Fr	NI	En	De	Es
Fr	-	82	82	83	58
NI	82	-	70	45	84
En	82	70	-	72	84
De	83	45	72	-	84
Es	58	84	84	84	-

Arbre



Calcul de la nouvelle distance

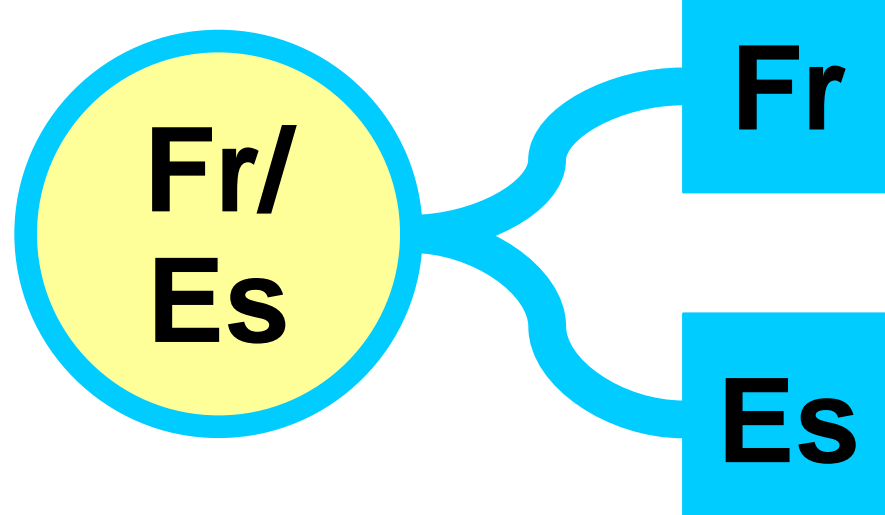
$$\text{Dist}(\text{De/NI}, \text{En}) = (70+72) / 2 = 71$$

$$\text{Dist}(\text{De/NI}, \text{Fr}) = (82+83) / 2 = 83$$

$$\text{Dist}(\text{De/NI}, \text{Es}) = (84+84) / 2 = 84$$

Cycle 2

	Fr	En	De/NI	Es
Fr	-	82	83	58
En	82	-	71	84
De/NI	83	71	-	84
Es	58	84	84	-

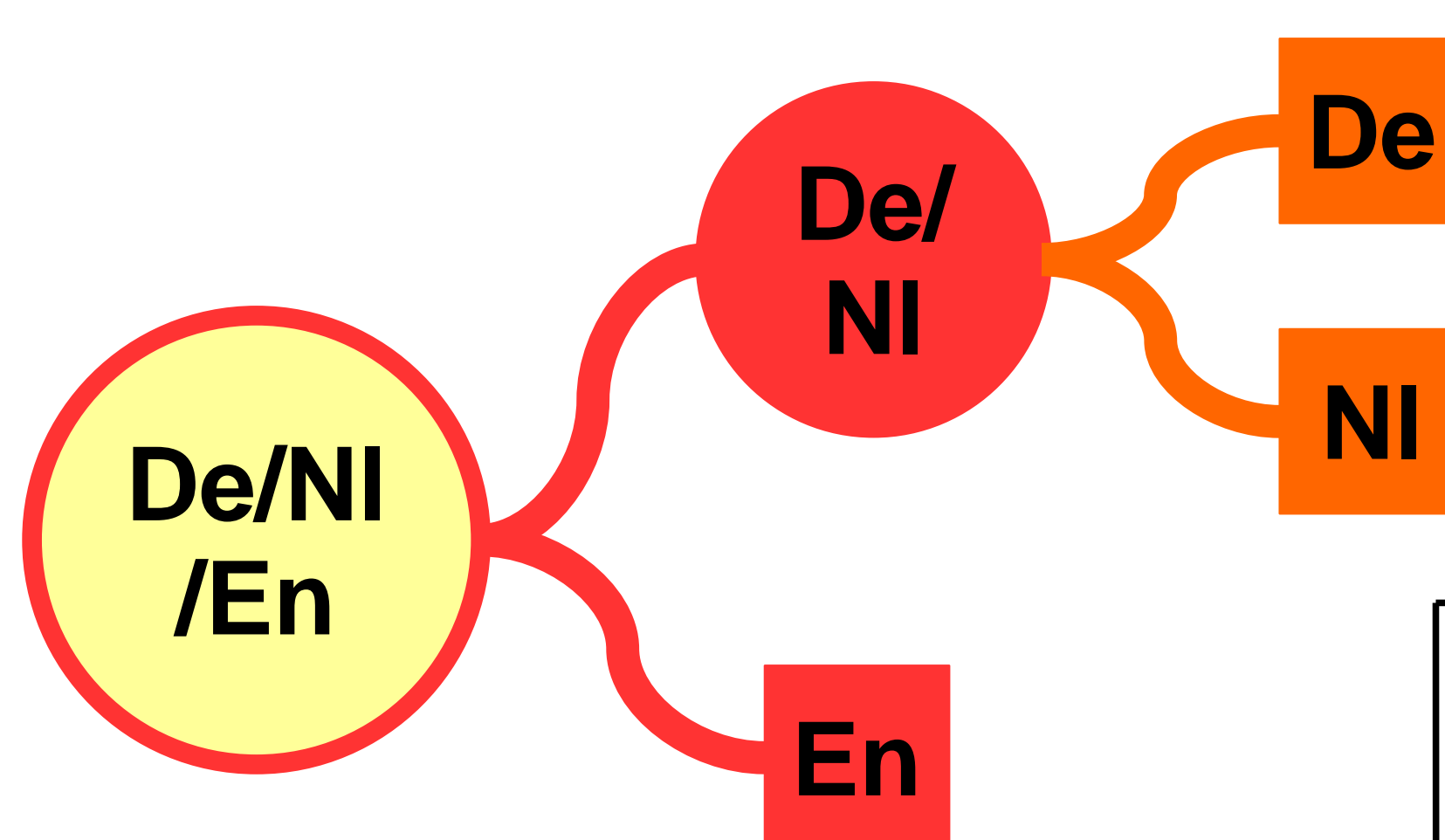


$$\text{Dist}(\text{Fr/Es}, \text{En}) = (82+84) / 2 = 83$$

$$\text{Dist}(\text{Fr/Es}, \text{De/NI}) = (83+84) / 2 = 84$$

Cycle 3

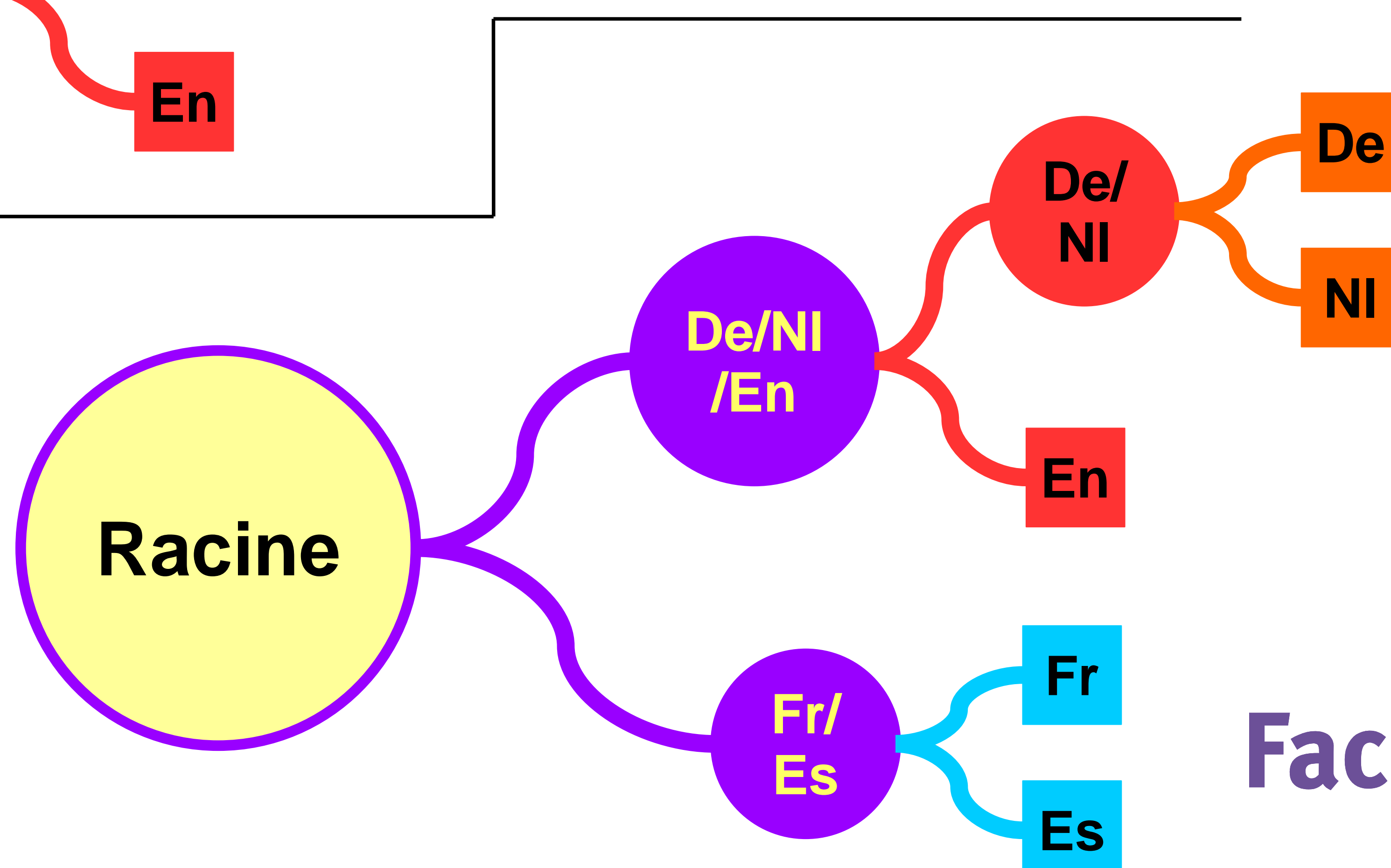
	Fr/Es	En	De/NI
Fr/Es	-	82	83
En	82	-	71
De/NI	83	71	-



$$\text{Dist}(\text{De/NI/En}, \text{Fr/Es}) = (82+83) / 2 = 83$$

Cycle 4

	Fr/Es	De/NI/En
Fr/Es	-	83
De/NI/En	83	-



Évolution des langues indo-européennes

Département d'informatique

Antoine CALEGARO, Yehudi HOLLEVOET, Antoine LIGOT, Jean LOBET et Michaël STEVENS

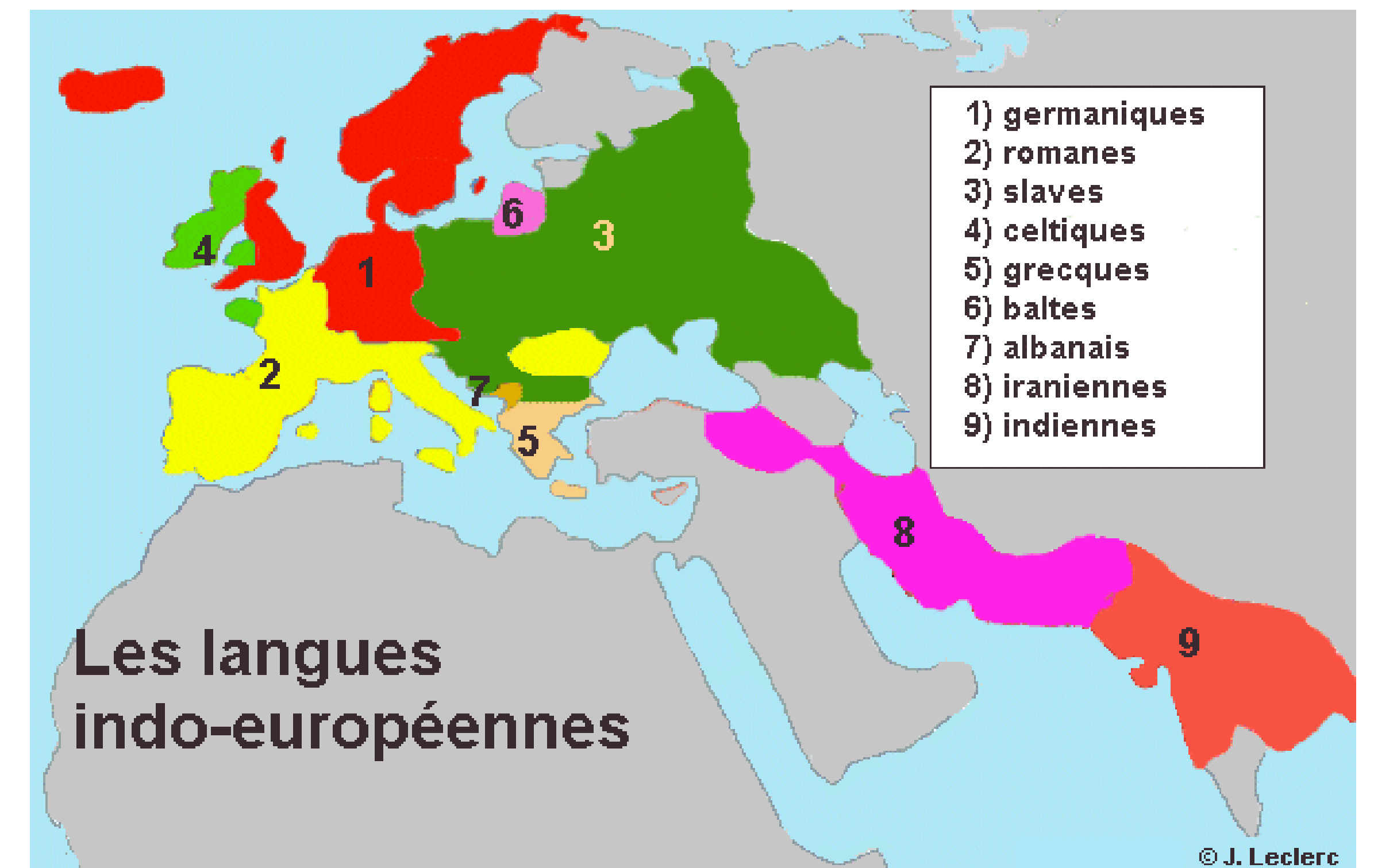
L'indo-européen

Langue préhistorique, sans trace écrite, supposée être à l'origine de toutes les langues dites indo-européennes.

La linguistique

Discipline s'intéressant à l'étude des langues qui sont modélisées à l'aide de trois classes de caractères :

- ➔ Les caractères morphologiques : les règles grammaticales
- ➔ Les caractères phonologiques : la structure des sons
- ➔ Les caractères lexicaux : la signification des mots



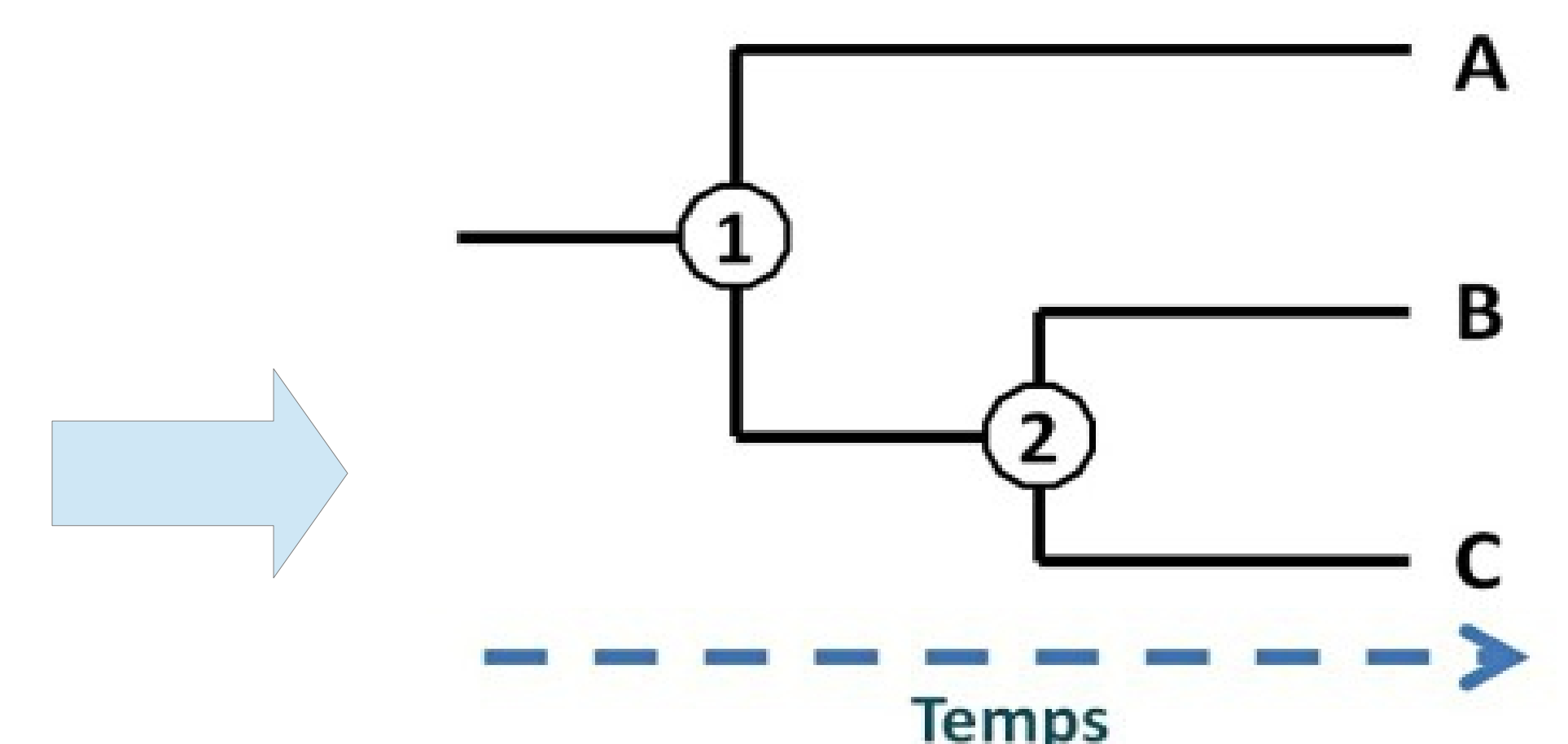
Objectif

Déterminer l'histoire évolutive d'un ensemble de langues et exprimer cette connaissance au travers d'arbres phylogénétiques

Arbre phylogénétique

- ➔ Montre les relations de parenté entre des groupes de langues.
- ➔ Chaque noeud interne représente l'ancêtre commun de ses deux descendants.

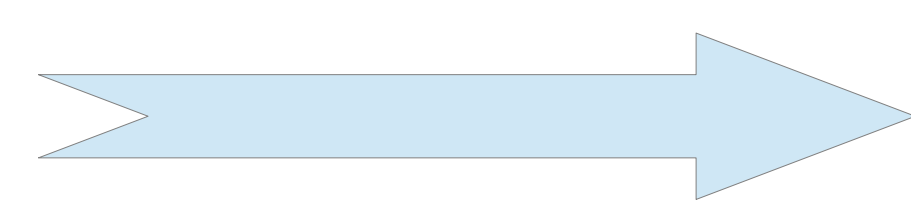
	Présence du subjonctif	Le son /f/ peut s'orthographier [ph]	Présence d'un pronom de politesse
Langue A	✗	✗	✓
Langue B	✓	✓	✗
Langue C	✓	✓	✓



- ⊗ : noeud = ancêtre hypothétique.
- : lien = branche reliant le taxon à un ancêtre.
- Y : feuille = taxon (ou gène).

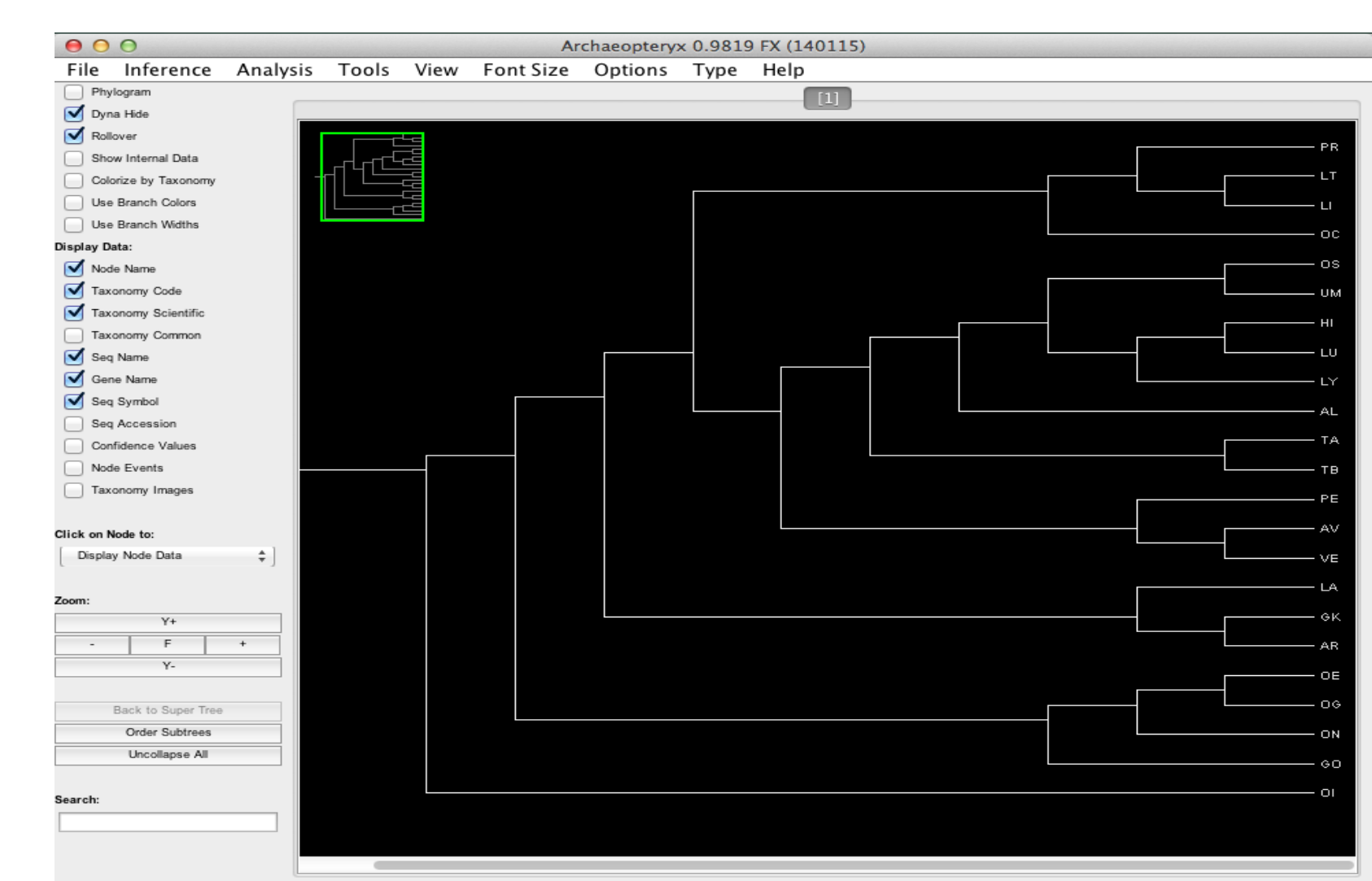
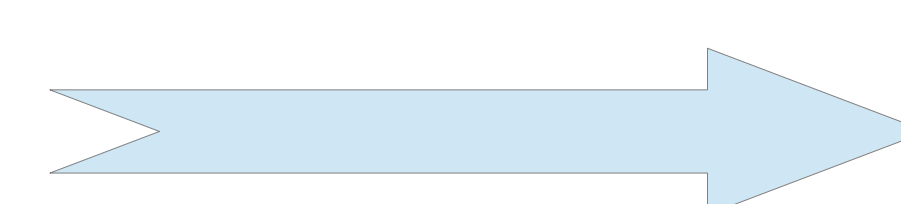


Récupération des données



Traitement des données

Affichage du résultat



Exploitation des résultats

- ➔ Dans le domaine de l'Histoire, l'étymologie, ...
- ➔ Dédire un modèle sur l'évolution des langues. Comment une langue évolue-t-elle ?
- ➔ L'étrusque ? Le méroïtique ?

Évolution des langues indo-européennes

Département d'informatique

Antoine CALEGARO, Yehudi HOLLEVOET, Antoine LIGOT, Jean LOBET et Michaël STEVENS

Les différentes méthodes de reconstruction d'arbres phylogénétiques

Les méthodes exactes

Recherche exhaustive

→ Énumère toutes les constructions possibles pour ensuite déterminer *la solution optimale*.

Les méthodes par rapprochement

Méthodes agglomératives

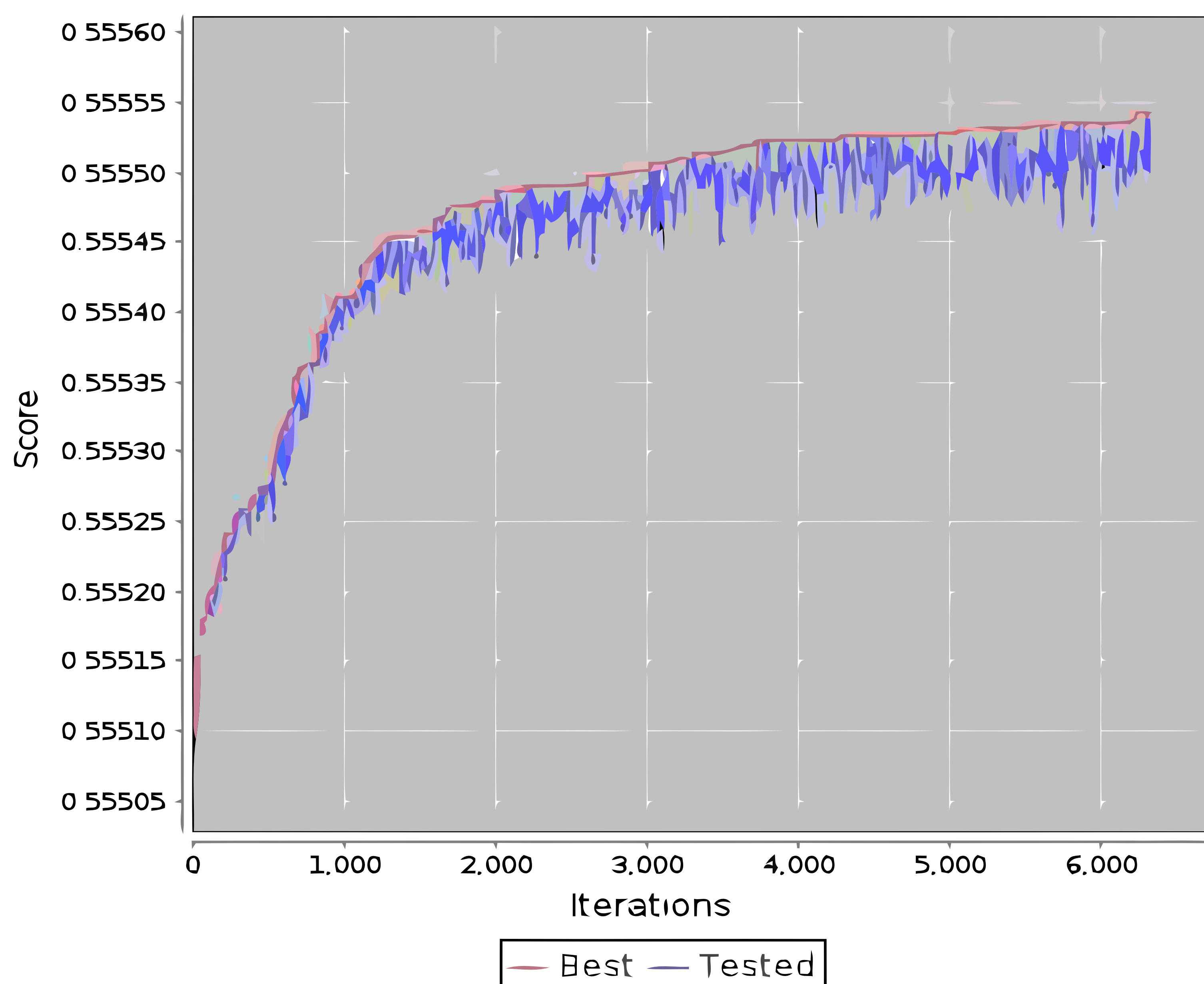
→ Fonctionnent *par itérations successives*. À chaque itération, les deux langues les plus proches sont assemblées. Au final, nous obtenons un arbre phylogénétique.

Méthodes heuristiques

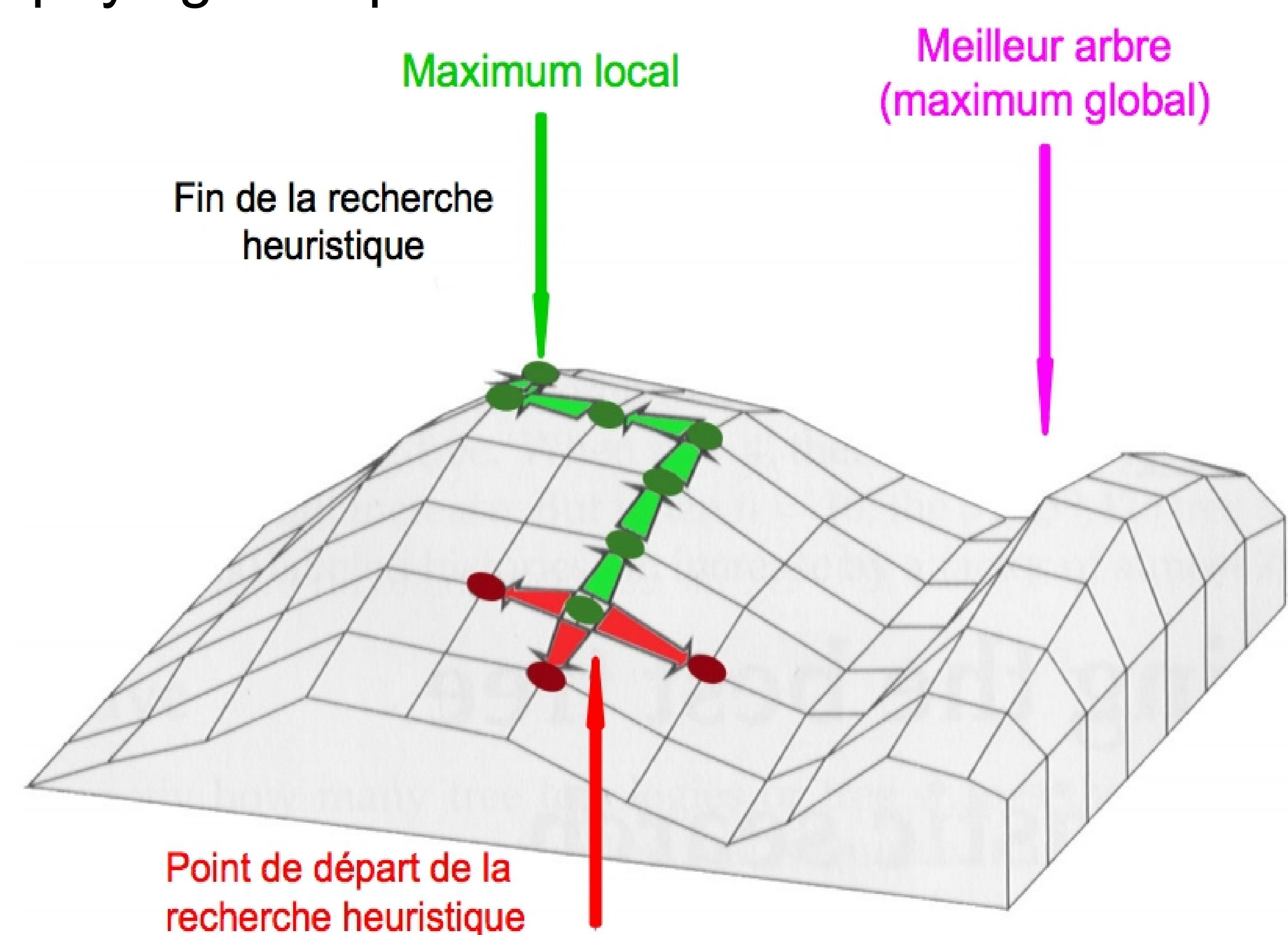
→ Permettent de trouver *une solution approchant la solution exacte* à l'aide de processus aléatoires.

Il s'agit d'un compromis entre la rapidité de recherche et la qualité des solutions trouvées.

McMc



Langages	Arbres
5	105
6	945
7	10 395
8	135 135
9	2 027 025
10	34 459 425
20	8.2×10^{21}
30	2.75×10^{76}



Réarrangement des branches

Modification aléatoire qui implique un changement de la structure de l'arbre. Le but étant d'analyser un ensemble d'arbres voisins.

McMc – Monte Carlo et Chaine de Markov

Méthode heuristique permettant de déduire la probabilité que la structure d'un arbre soit optimale, connaissant la probabilité que deux langues soient associées.

Des arbres phylogénétiques linguistiques

Sciences Informatiques

Nicolas CAMBIER, Hadrien GOFFIN, Hassan AMOUHZI, Gabriel SHAKO EKANGA



L'origine des langues

Indo-européens

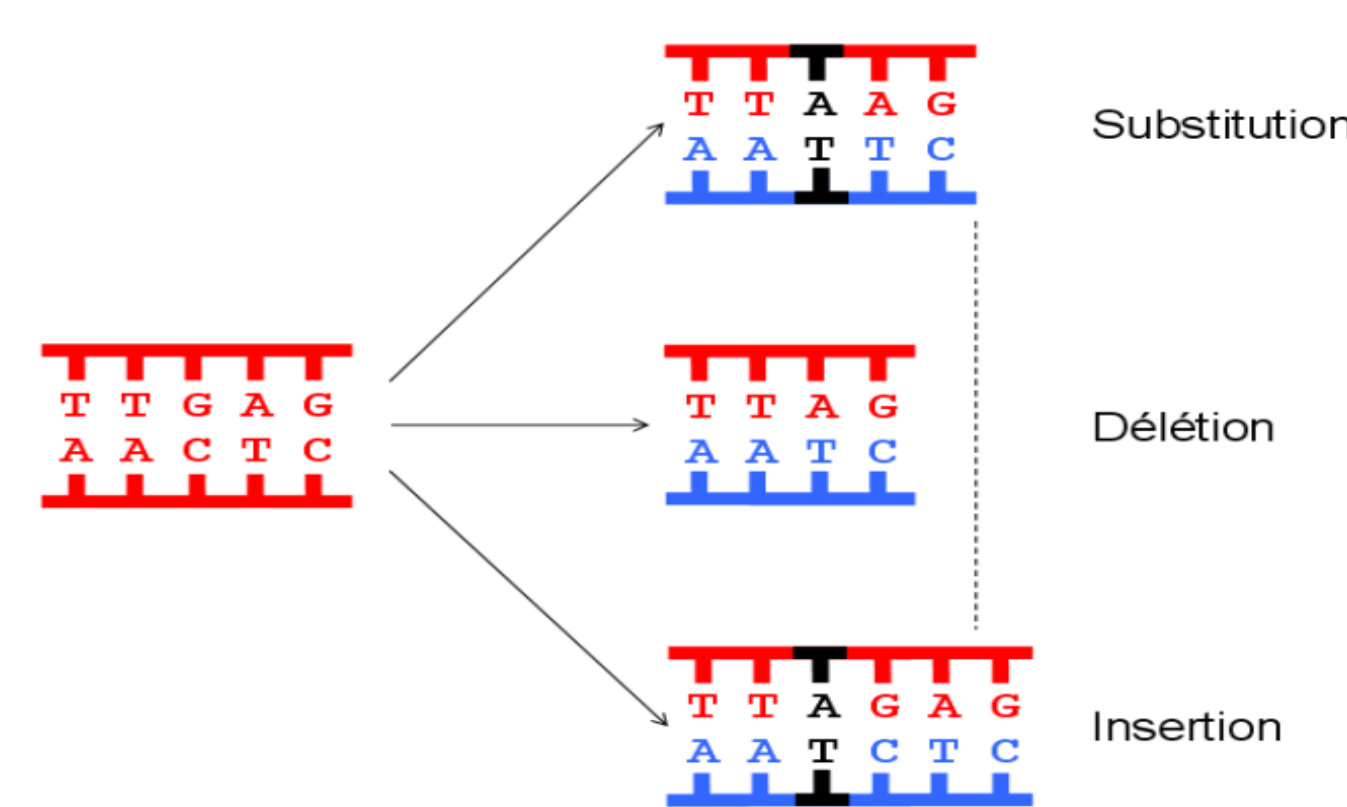
On note entre les langages européens et d'autres très lointains (comme le sanskrit) de nombreuses ressemblances. Un indice vers l'origine de ces peuples ?

Elfique ?!

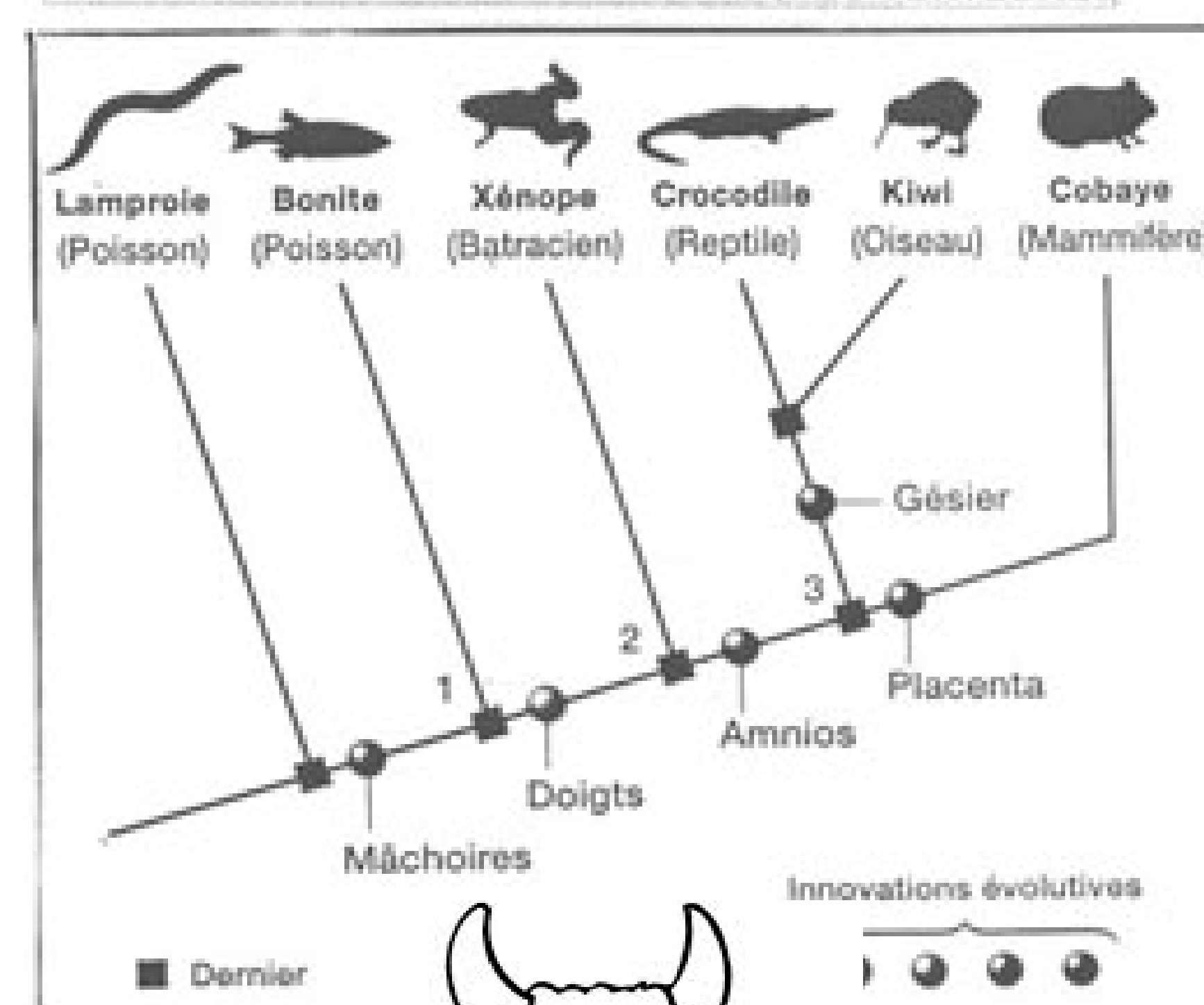
L'étude linguistique s'applique à toutes les langues. Pourquoi pas les langues fictives ? Mettons J.R.R Tolkien à l'épreuve...

Bioinformatique

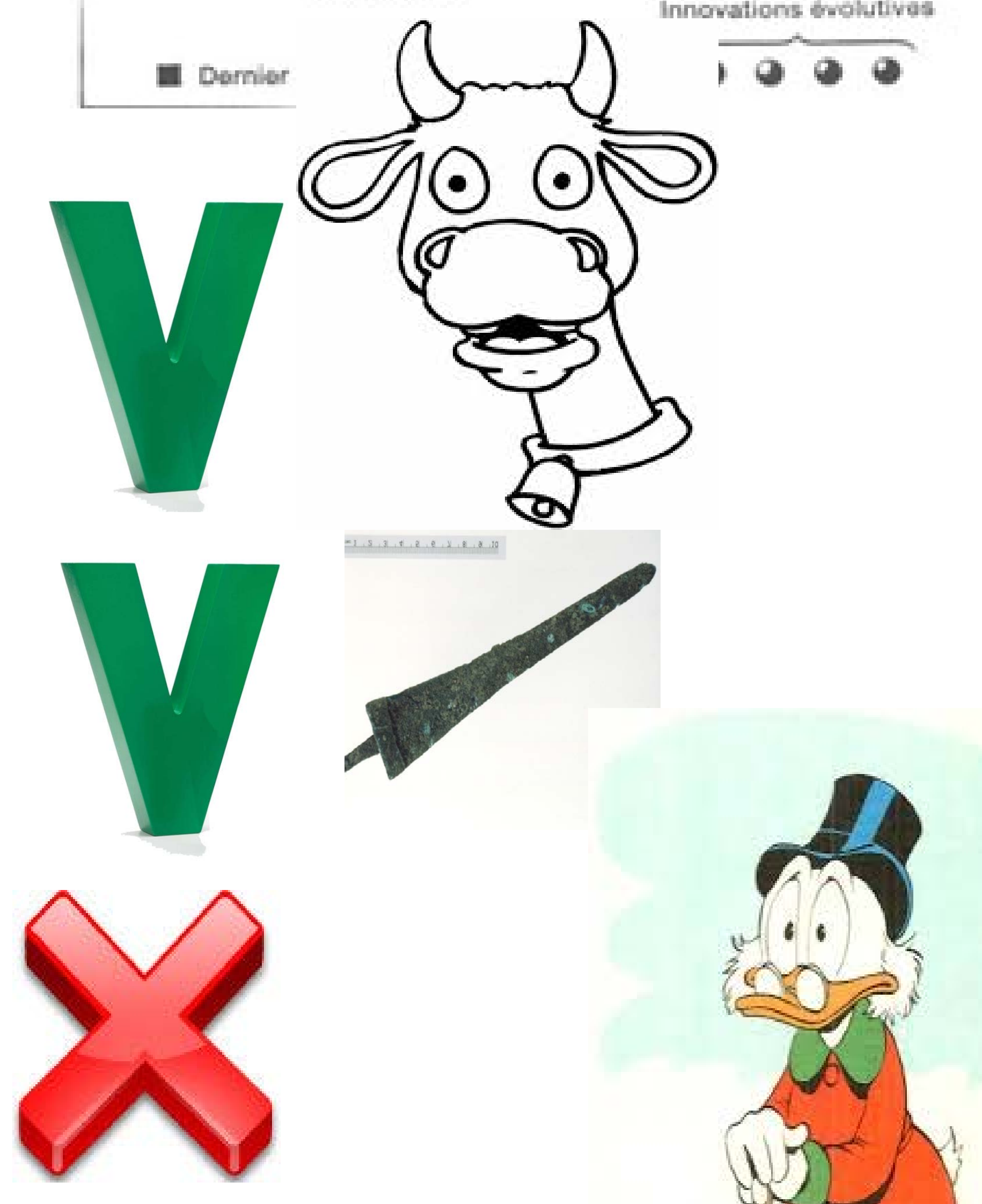
Pour établir le lien entre ces différentes langues, les linguistes se sont tournés vers les techniques utilisées en biologie pour comparer des séquences d'ADN.



ARBRE PHYLOGENETIQUE DE QUELQUES VERTEBRES



Français	Latin	Proto-germanique	Sanskrit	Grec ancien
Vache	Bos	Kowz	Go	Boús
Cuivre	Aes	Aiz	Ayas	Kansya
Or	Aurum	Gulba	SvarNa	Agkhourous

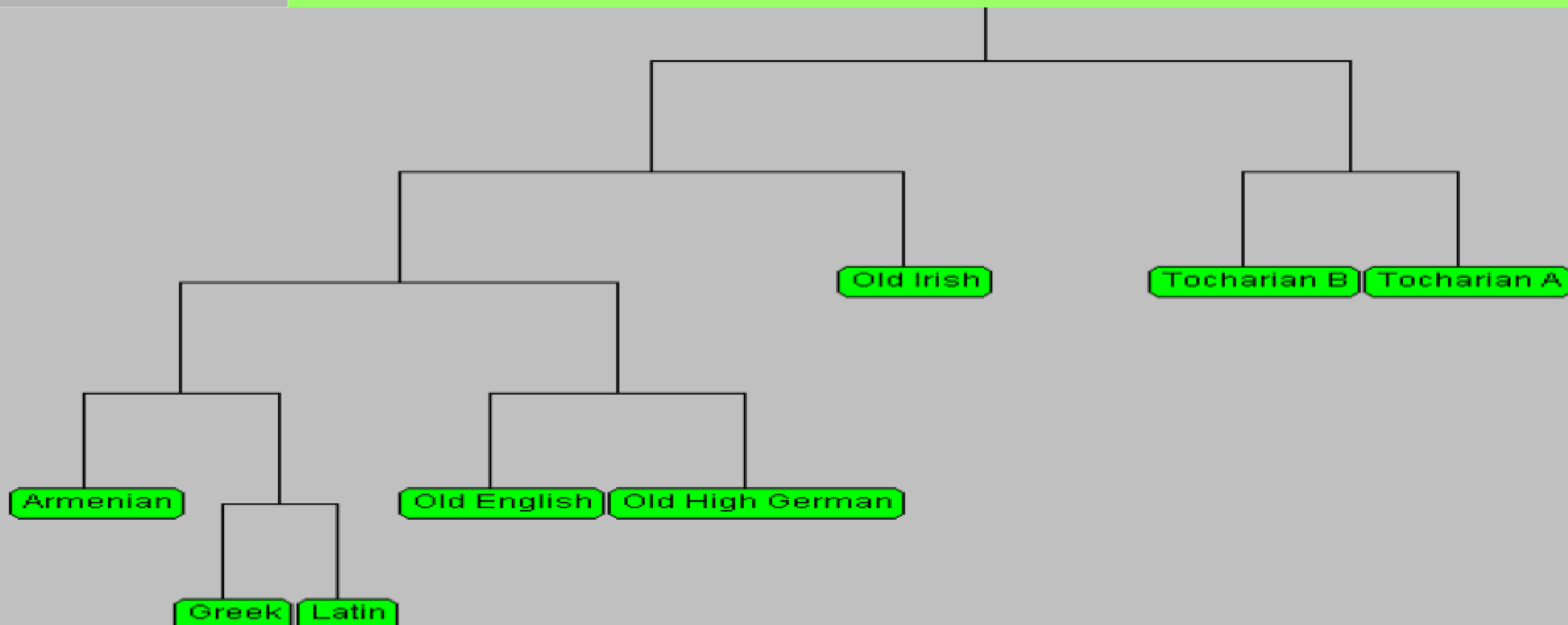


Anglais

Elvish ?!

Un anneau pour les gouverner tous

One ring to rule them all



Des arbres phylogénétiques linguistiques

Sciences Informatiques

Nicolas CAMBIER, Hadrien GOFFIN, Hassan AMOUHZI, Gabriel SHAKO EKANGA

Comment construire un arbre phylogénétique ?

Tester la cohérence de Tolkien !

Comme le quenya et le telerin d'Aman, le sindarin descend de l'eldarin commun, la langue des Eldar pendant leur grande marche de Cuiviénen vers le Valinor, lui-même dérivé du quendien.

« Et ils se nommèrent les Quendi, signifiant ceux qui parlent avec des voix, car ils n'avaient pas encore rencontré d'autres êtres vivants qui parlassent ou chantassent. »

Morgoth's Ring — Partie II « Les Annales d'Aman »

Français	Sindarin	Quenya	Quendien
Poisson	LOK	KHAL	TIR
Homme	DER	NDER	NÎ



Thranduil, Le Hobbit : La désolation de Smaug, Peter Jackson

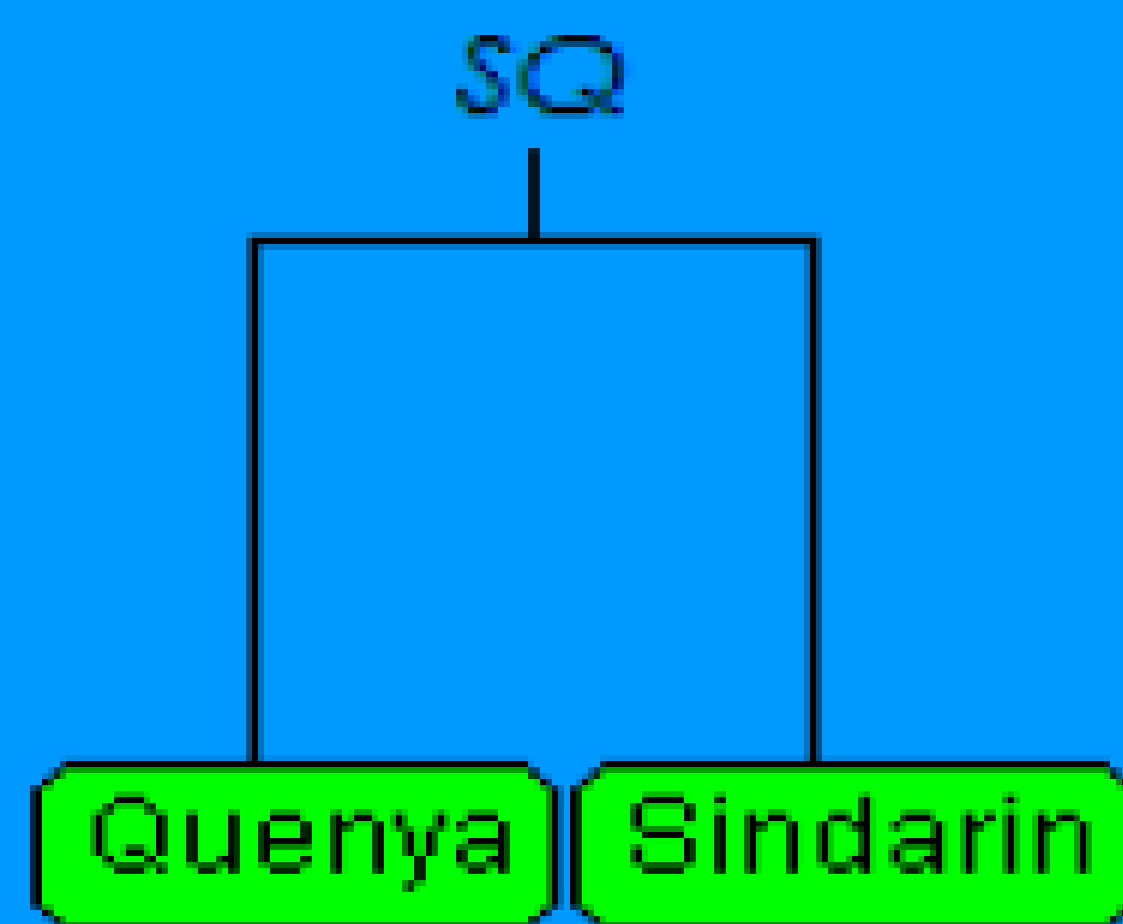
Unweighted Pair Group Method with Arithmetic Mean

DER (S) et NDER (Qy) ont même origine. Sindarin et Quenya ne sont espacés que d'un mot. Le Quendien a une distance de 2 avec eux.

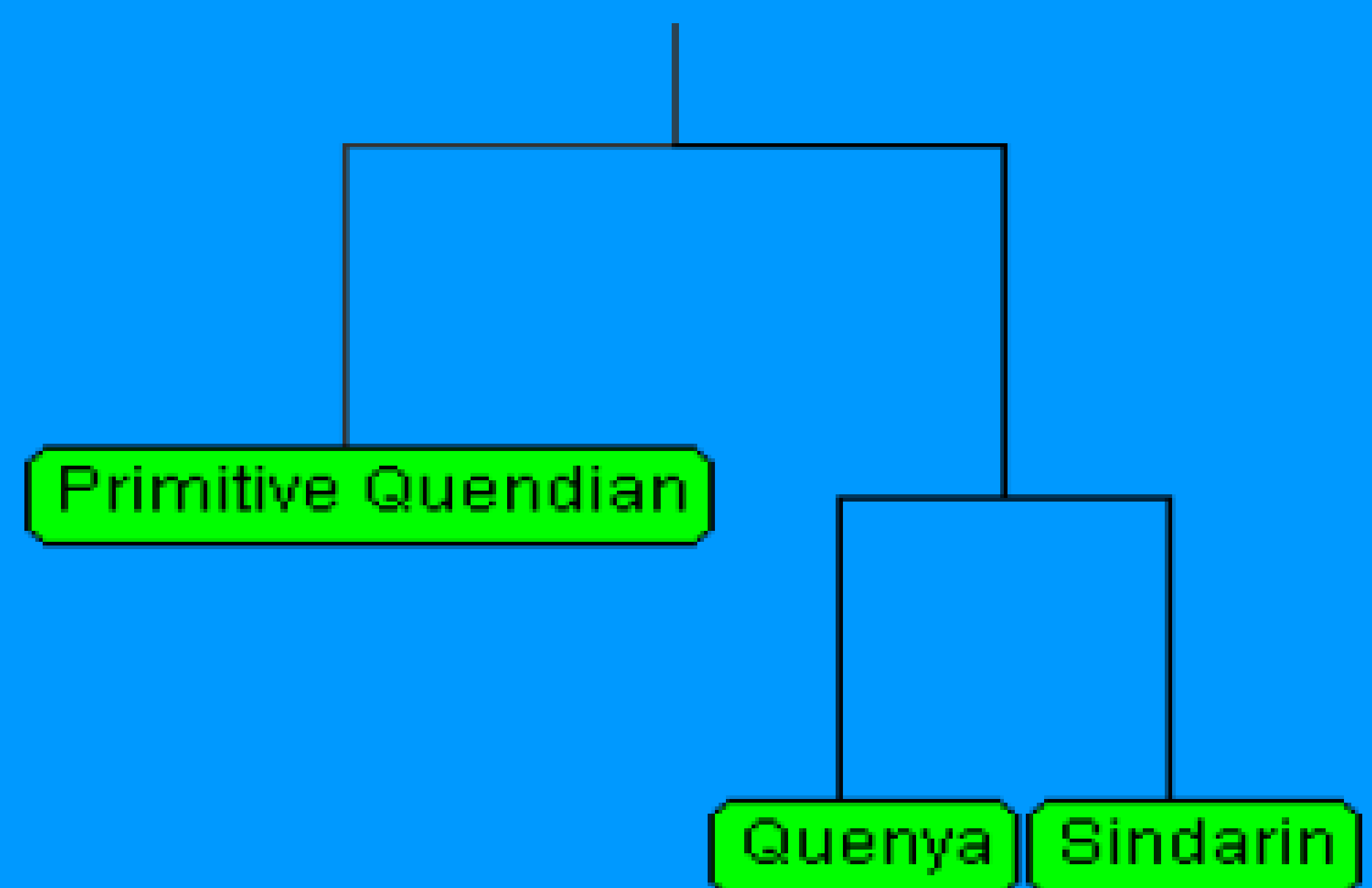
Distance	Qd	Qy
S	2	1
Qy	2	

On prend la plus petite distance et on rejoint les deux langues.

$$\text{Distance } Qd/SQ = (2+2)/2 = 2$$



On recommence jusqu'à vider le tableau (ça va vite, ici) !



Maximum Parsimony

On crée tous les arbres possibles et on prend celui qui a le coût le plus bas :

