

**Le “poids” des langues:
une présentation
de la situation linguistique du monde à
l’heure de la mondialisation**

Louis-Jean Calvet

louis-jean.calvet@wanadoo.fr

1. Présentation de la situation linguistique du monde en termes gravitationnels

1. Présentation de la situation linguistique du monde en termes gravitationnels
2. Rapide présentation d'un projet d'un « index » des langues du monde en termes de « poids »

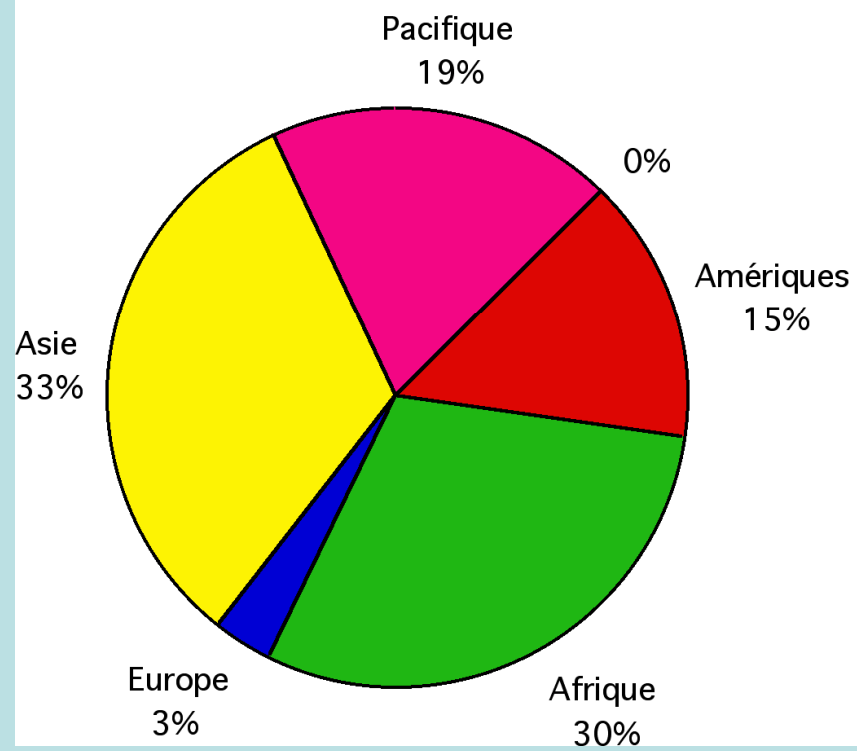
Quelques sites...

<http://perso.orange.fr/Louis-Jean.Calvet/index.htm>

<http://www.ethnologue.com/>

www.tlfq.ulaval.ca/AXL/Langues/acces_Inguesmonde.htm

Les langues parlées dans le monde



Nous avons donc des continents
« pauvres » en langues...

...et d'autres « riches » en langues



Le modèle gravitationnel

- Au niveau 1 une langue hypercentrale
- Au niveau 2 des langues supercentrales
- Au niveau 3 des langues centrales
- Au niveau 4 des langues périphériques

- Niveau 1 une langue hypercentrale
Dont les locuteurs montrent une forte
tendance au monolinguisme
- Niveau 2 des langues supercentrales
- Niveau 3 des langues centrales
- Niveau 4 des langues périphériques

- Niveau 1 une langue hypercentrale

Dont les locuteurs montrent une forte tendance au monolinguisme

- Niveau 2 des langues supercentrales

Dont les locuteurs montrent une tendance vers un bilinguisme horizontal ou vertical

- Niveau 3 des langues centrales

- Niveau 4 des langues périphériques

- Niveau 1 une langue hypercentrale

Dont les locuteurs montrent une forte tendance au monolinguisme

- Niveau 2 des langues supercentrales

Dont les locuteurs montrent une tendance vers un bilinguisme horizontal ou vertical

- Niveau 3 des langues centrales

Dont les locuteurs montrent une tendance vers un bilinguisme vertical

- Niveau 4 des langues périphériques

- Niveau 1 une langue hypercentrale

Dont les locuteurs montrent une forte tendance au monolinguisme

- Niveau 2 des langues supercentrales

Dont les locuteurs montrent une tendance vers un bilinguisme horizontal ou vertical

- Niveau 3 des langues centrales

Dont les locuteurs montrent une tendance vers un bilinguisme vertical

- Niveau 4 des langues périphériques

Dont les locuteurs montrent une tendance vers un bilinguisme horizontal et vertical

Selon le nombre de langues à chaque
niveau

Langue hypercentrale

Langues supercentrales

Langues centrales

Langues périphériques

Selon le « poids » des langues

Langue hypercentrale

Langues supercentrales

Langues centrales

Langues périphériques

Ce modèle a une valeur
prédictive

Par exemple un bilingue français/alsacien
a, dans près de 100% des cas,
l'alsacien pour langue première

Par exemple un bilingue français/alsacien
a, dans près de 100% des cas,
l'alsacien pour langue première

Et un bilingue cantonnais/mandarin a,
dans près de 100% des cas, le
cantonnais pour langue première

La modèle gravitationnel nous donne à
voir le versant linguistique de la
mondialiation: il nous donne une photo
fixe

Mais la situation est en évolution

Le « poids » des langues

De façon générale:

Comment mesurer l'importance relative des langues ?

Comment les classer ?

On pense en général au nombre de leurs locuteurs

Mais le décompte du nombre de locuteurs n'est pas une science exacte :

Voici des données provenant de trois sources différentes

(2003)

	Quid	Linguasphere	S.I.L. (ethnologue)
1. Chinois	1 milliard	1 milliard	885 millions
2. Anglais	500 millions	1 milliard	322 millions
3. Hindi	497 millions	900 millions	182 millions
4. Espagnol	392 millions	450 millions	332 millions
5. Russe	277 millions	320 millions	170 millions
6. Arabe	246 millions	250 millions	---
7. Bengali	211 millions	250 millions	189 millions
8. Portugais	191 millions	200 millions	177 millions
9. Malais	159 millions	160 millions	23 millions
10. Français	129 millions	125 millions	72 millions
11. Allemand	128 millions	125 millions	98 millions
12. Japonais	126 millions	130 millions	125 millions

Nous essayons de mettre de
l'ordre dans ce désordre...

En travaillant à aide de
facteurs objectifs et
différenciés

Dix facteurs ...

- Nombre de locuteurs
- Nombre de pays dans lesquels la langue a un statut officiel
- Nombre d'articles dans in Wikipedia
- Nombre de prix Nobel de littérature
- Entropie
- Taux de natalité
- Indice de développement humain (IDH)
- Présence des langues sur Internet
- Nombre de traductions comme langue source
- Nombre de traductions comme langue cible

... et nos références

- Nombre de locuteurs et statut officiel :
 - <http://www.ethnologue.com/web.asp>
- Wikipedia
 - http://en.wikipedia.org/wiki/Wikipedia:Multilingual_statistics
- Prix Nobel
 - <http://nobelprize.org/>
- Entropie
 - calculée à partir des données de population
- Taux de fécondité
 - <http://www.prb.org/FrenchContent.aspx>
- IDH
 - <http://www.undp.org/french/>
- Présence sur internet
 - <http://www.internetworldstats.com/stats.htm>
- Index translationum :
 - <http://databases.unesco.org/xtrans/stat/xTransStat.html>

Classification selon le nombre de locuteurs

- 1. Mandarin
- 2. Hindi
- 3. Anglais
- 4. Espagnol
- 5. Arabe
- 6. Portugais
- 7. Bengali
- 8. Russe
- 9. Japonais
- 10. Allemand
- 11. Pendjabi
- 12. Javanais
- 13. Wu
- 14. Vietnamien
- 15. Tagalog
- 16. Tamoul
- 17. Min
- 18. Coréen
- 19. Français
- 20. Marathi

Notez que 0,17% des langues ont plus de 100 millions de locuteurs... et que 81% des langues ont moins de 100.000 locuteurs

Nombre de locuteurs	nombre de langues	pourcentage
Plus de 100 millions	12	0,17%
Entre 100 et 10 m.	72	1,2%
Entre 10 et 1 m.	239	3 %
Entre 1 million et 100.000	795	13,19%
Entre 100.000 et 10.000	1605	26,5%
Entre 10.000 et 1.000	1782	29,4%
Entre 1.000 et 100	1075	17,7%
Entre 100 et 10	302	5%
Entre 10 et 1	181	3 %

Classification selon le statut officiel des langues

- 1. Anglais (45)
- 2. Français (30)
- 3. Arabe (25)
- 4. Espagnol (20)
- 5. Portugais (8)
- 6. Allemand (6)
- 7. Italien
- 8. Russe
- 9. Bahasa
- 10. Néerlandais
- 10. Hongrois
- 10. Mandarin
- 10. Roumain
- 10. Farsi
- 10. Croate
- 10. Slovène
- 10. Albanais
- 10. Tamoul
- 10. Swahili
- 10. Bambara

Présence sur Internet

- 1. Anglais
- 2. Mandarin
- 3. Espagnol
- 4. Japonais
- 5. Allemand
- 6. Français
- 7. Coréen
- 8. Italien
- 9. Portugais
- 10. Malais
- 11. Néerlandais
- 12. Arabe
- 13. Polonais
- 14. Suédois
- 15. Thaï
- 16. Turc
- 17. Russe
- 18. Vietnamien
- 19. Farsi
- 20. Roumain

Prix Nobel de littérature

- 1. Anglais 24
- 2. Français 12
- 3. Allemand 11
- 4. Espagnol 10
- 5. Russe 5
- 6. Italien 5
- 7. Suédois 4
- 8. Polonais 4
- 9. Grec 2
- 10. Japonais 2
- 11. Danois 2
- 12. Ex-aequo : 1
Mandarin, Finnois,
Tchèque, Islandais,
Arabe, Provençal,
Yiddish, Portugais,
Bengali, Hébreu,
Hongrois, Turc, Serbo-
croate
- 24. Les autres 0

traduction langue source

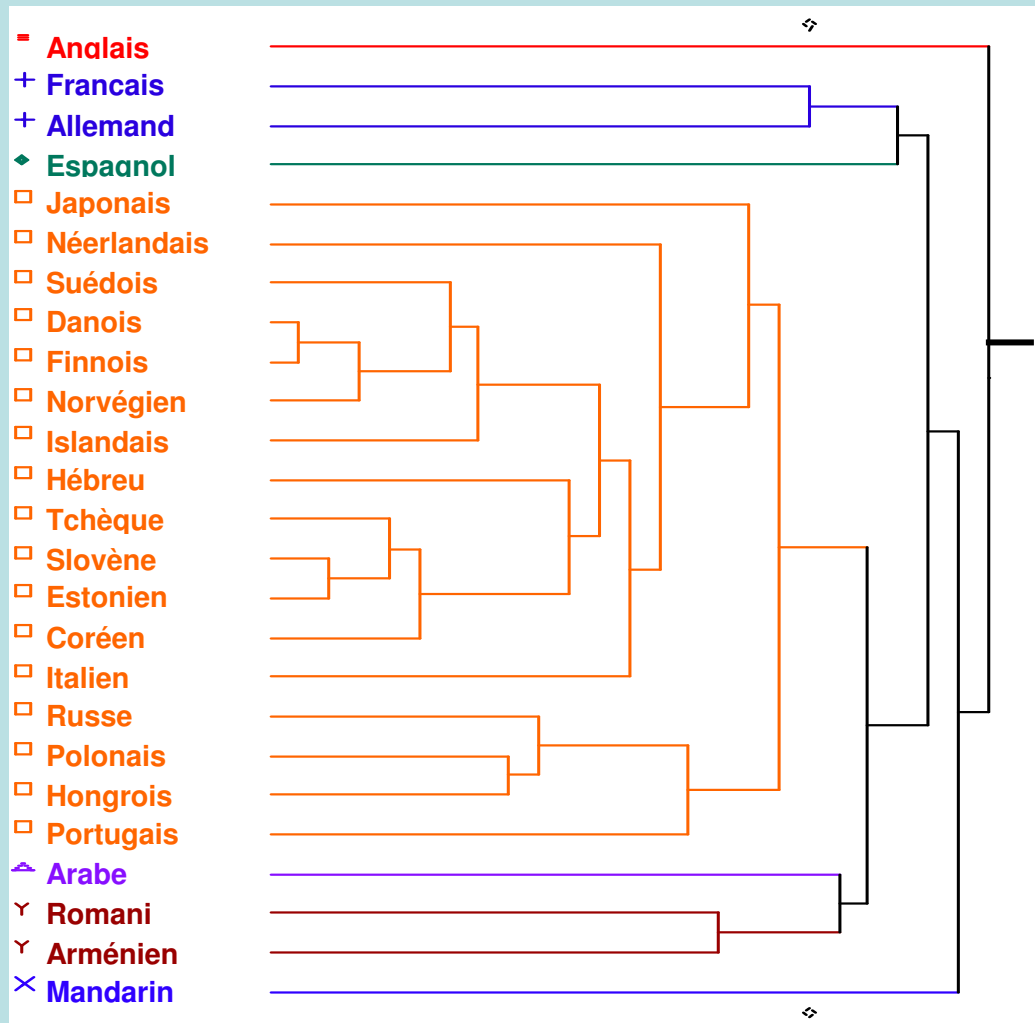
- 1. Anglais
- 2. Francais
- 3. Allemand
- 4. Russe
- 5. Italien
- 6. Espagnol
- 7. Suedois
- 8. Danois
- 9. Tcheque
- 10. Neerlandais
- 11. Polonais
- 12. Japonais
- 13. Hongrois
- 14. Arabe
- 15. Norvegien
- 16. Portugais
- 17. Hebreu
- 18. Mandarin
- 19. Finnois
- 20. Bahasa

Comment croiser ces données?

Voici deux classifications sur dix facteurs

Somme des données standardisées		Somme des rangs	
Anglais	Islandais	Anglais	Japonais (-6)
Français	Finnois	Français	Polonais (+ 3)
Espagnol	Romani	Espagnol	Danois (-3)
Allemand	Russe	Allemand	Hongrois (+8)
Japonais	Polonais	Néerlandais (+1)	Coréen (+8)
Néerlandais	Portugais	Russe (+8)	Turc (+10)
Arabe	Norvégien	Portugais (+9)	Finnois (-5)
Suédois	Mandarin	Italien (+1)	Tchèque (+2)
Italien	Hébreu	Arabe (-2)	Hébreu
Danois	Tchèque	Suédois (-2)	Serbe (+22)

Analyse en clusters : un exemple



- Certains clusters sont des singletons, d'autres des doubletons et un grand nombre de langues se trouvent dans le même cluster
- Un singleton montre que la langue concernée a un comportement singulier face à un ou plusieurs facteurs

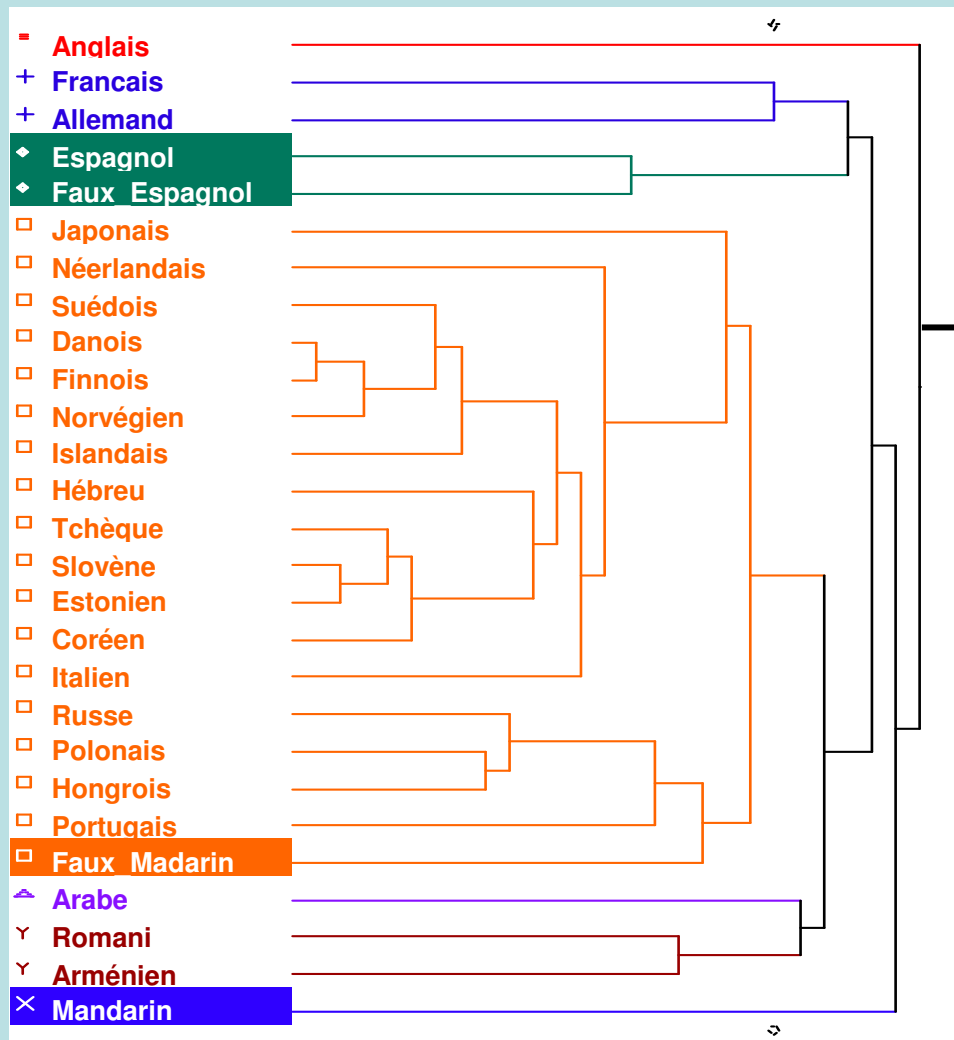
Deux « fausses » langues

Un faux espagnol, langue officielle dans un pays
au lieu de 21

Un faux mandarins avec seulement 50 millions de
locuteurs

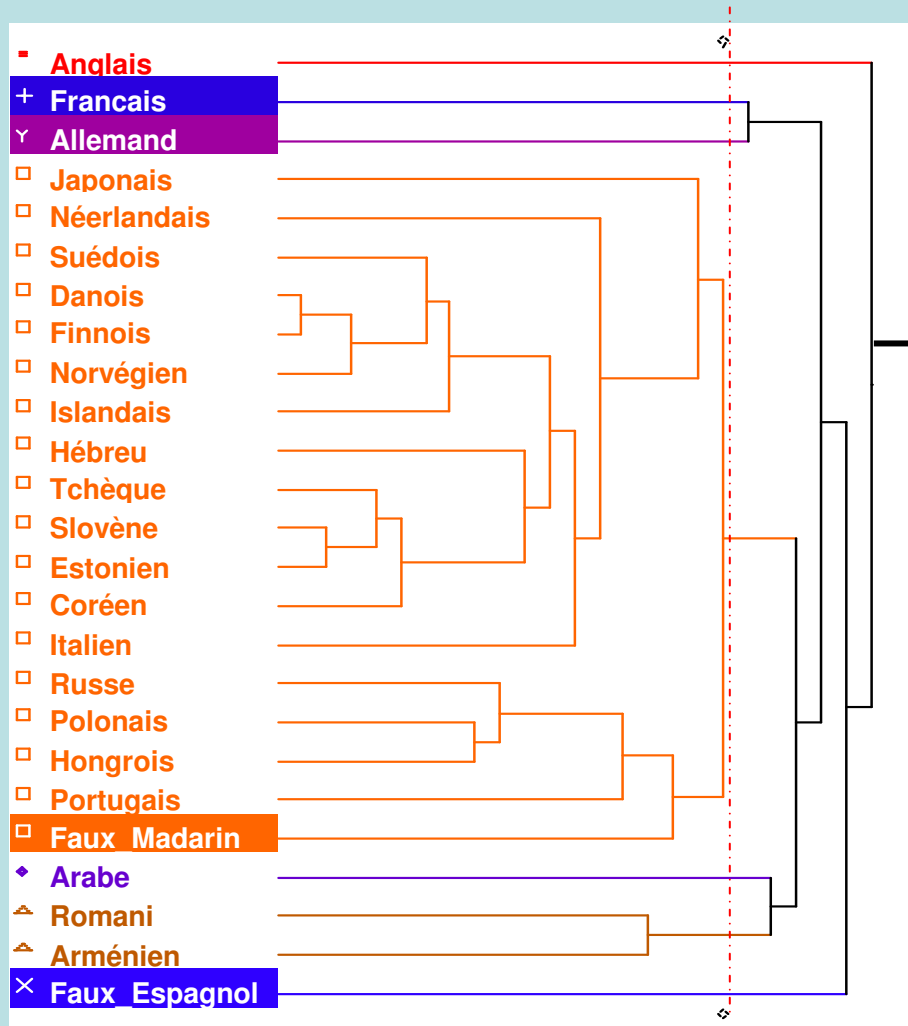
Les autres facteurs restent inchangés

Ajoutons ces deux fausses langues



- Le faux espagnol a le même comportement que le vrai
- Car le poids de l'espagnol repose sur de nombreux facteurs (Nobel, Internet, nombre de locuteurs, etc ...)
- Le mandarin reste un singleton
- Mais le faux chinois rejoint le cluster le plus « peuplé » : il doit essentiellement son poids au nombre de ses locuteurs

Supprimons les deux « vraies » langues



- Le faux espagnol reste un singleton
- Le faux mandarin rejoint le cluster le plus peuplé, etc...

Pourquoi cet « index des langues du monde » ?

Pourquoi cet « index des langues du monde » ?

Un observatoire des langues du monde,
mis à jour régulièrement en tenant
compte des changements de
comportements des langues face à un
ou plusieurs facteurs

Pourquoi cet « index des langues du monde » ?

-Qui sera en libre accès sur Internet

Pourquoi cet « index des langues du monde » ?

- Qui sera en libre accès sur Internet
- Avec nos sources et la façon dont nous les traitons statistiquement

Pourquoi cet « index des langues du monde » ?

- Qui sera en libre accès sur Internet
- Avec nos sources et la façon dont nous les traitons statistiquement
- Afin de donner la possibilité d'un débat scientifique

Pourquoi cet « index des langues du monde » ?

- Qui sera en libre accès sur Internet
- Avec nos sources et la façon dont nous les traitons statistiquement
- Afin de donner la possibilité d'un débat scientifique
- Et d'offrir une aide à la décision en matière de politique linguistique

Nous travaillons par exemple sur:

- 1) Le poids des langues en Méditerranée, pour un colloque sur les politiques linguistiques en Méditerranée
- 2) Le poids des langues dans une petite niche écolinguistique: la région frontalière entre la Brésil et la Guyane française

Nous pourrions travailler sur:

- 1) Le poids des langues dans un seul pays (Inde, Chine...)
- 2) Le poids des langues dans zone d'intégration régionale (l'Afrique francophone, le Mercosur...)

Le versant linguistique de la mondialisation:

-Environ 7.000 langues

- Environ 7.000 langues
- Inégalement réparties

- Environ 7.000 langues
- Inégalement réparties
- Organisation gravitationnelle

- Environ 7.000 langues
- Inégalement réparties
- Organisation gravitationnelle
- Avec des poids différents

- Environ 7.000 langues
- Inégalement réparties
- Organisation gravitationnelle
- Avec des poids différents
- Déterminés par différents facteurs

- Environ 7.000 langues
- Inégalement réparties
- Organisation gravitationnelle
- Avec des poids différents
- Déterminés par différents facteurs
- Offrant la possibilité d'une classification

- Environ 7.000 langues
- Inégalement réparties
- Organisation gravitationnelle
- Avec des poids différents
- Déterminés par différents facteurs
- Offrant la possibilité d'une classification
- Et la possibilité d'intervention sur cette classification

Sur ce dernier point

Sur ce dernier point

Quelques précisions éthiques
et théoriques

Aux yeux du linguiste toutes les langues
sont égales

Aux yeux du linguiste toutes les langues
sont égales

Dans la réalité il y en a qui sont plus
égales que les autres

Aux yeux du linguiste toutes les langues
sont égales

Dans la réalité il y en a qui sont plus
égales que les autres

Et de nos jours beaucoup sont menacées
de disparition

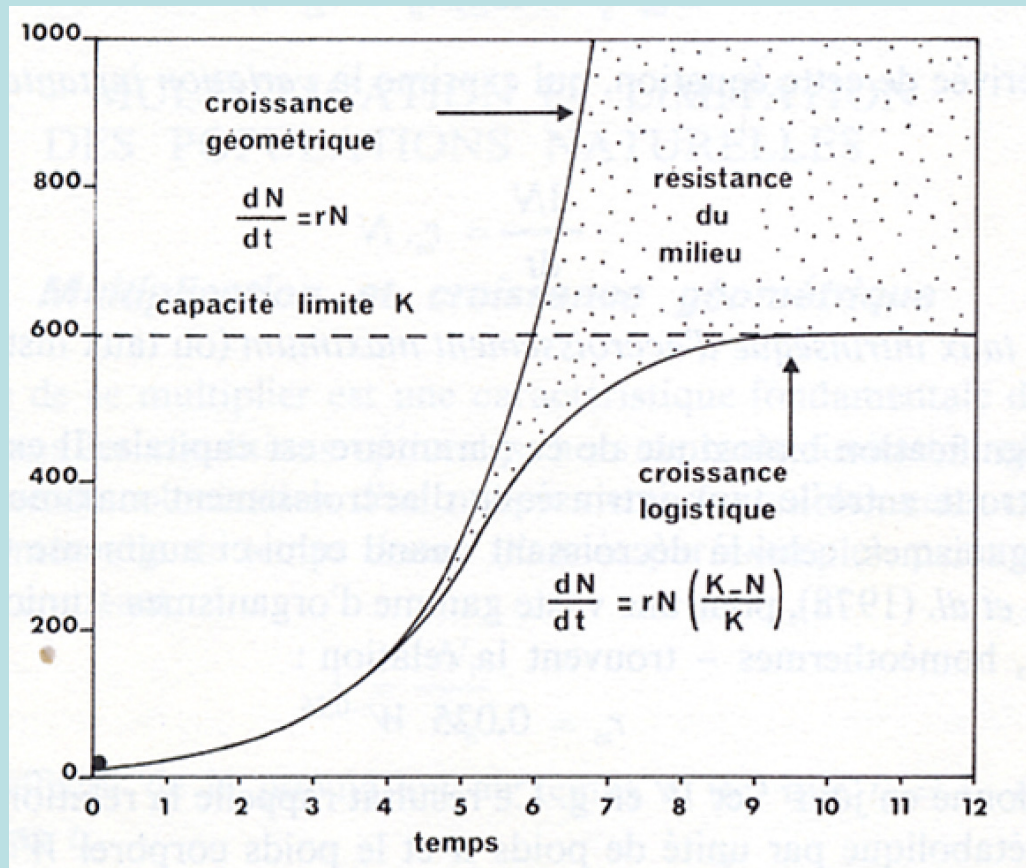
Il n'y a pas de langues sans locuteurs

Deux populations:

Celle des langues

Celle des êtres humains

êtres humains/langues
hôte/parasite



langues/langues
proie/prédateur

Sélection naturelle (cf Darwin)

Il s'agit bien sûr d'une
métaphore

Il s'agit bien sûr d'une
métaphore

Mais d'une métaphore qui
s'appuie sur une réalité
objective:

-Des langues en voie de
disparition

- Des langues en voie de disparition
- Des langues de poids inégal

L'approche en termes de poids

Outre qu'elle constitue un projet scientifique

L'approche en termes de poids

Outre qu'elle constitue un projet scientifique

Offre des possibilités d'interventions militantes, *in vivo*: agir sur les facteurs pour agir sur les langues

Par exemple

Par exemple

-Wikipedia

Par exemple

-Wikipedia

-Internet

ASCII 1

(American Standard Code for Information Interchange)

7 bits (128 caractères)

ASCII 2 (ou étendu)

8 bits (256 caractères)

UNICODE

Tous les caractères de tous les systèmes d'écriture

Présence des langues sur Internet

1. Anglais 35, 20%
2. Chinois 13,70%
3. Espagnol 9%
4. Japonais 8,40%
5. Allemand 6,90%
6. Français 4,20%
7. Coréen 3, 90 %
8. Italien 3,30%
9. Portugais 3,10%
10. Bahasa indonesia 1,80%
11. Néerlandais 1,70%
12. Arabe 1,70%

That's all folks

