

Poids des langues et « prospective » : essai d'application au français, à l'espagnol et au portugais

Louis-Jean Calvet
Université de Provence



Synergies Brésil n° spécial 1 - 2010 pp. 41-58

Résumé : *Il existe différents facteurs pour évaluer l'importance d'une langue aujourd'hui ou dans l'avenir. Un regard scientifique peut aider à ne pas recourir à des points de vue passionnels, ou à ne pas considérer un unique facteur pour déterminer le poids d'une langue en toute circonstance. Ainsi, le recours à des méthodes statistiques et mathématiques, ou le recours à des clusters, permet de pondérer différents facteurs concernant l'importance relative des langues, comme le poids démographique ou la présence sur Internet, par exemple. Cette démarche indique des atouts, des faiblesses, des ouvertures et des menaces possibles pour le français. Quant aux trois langues en question, espagnol, français et portugais, il ne faut sans doute pas s'attendre à des bouleversements dans les quarante années à venir, et ceci par rapport aux trois composants suivants : expansion démographique, technico-économique et prestige culturel.*

Mots-clés : *importance des langues, poids relatif des langues, situation du français*

Resumo: *Existem vários fatores permitindo avaliar a importância de uma dada língua hoje ou no futuro. Um olhar científico contribui a evitar um ponto de vista emotivo, ou a não considerar apenas um único fator na determinação do peso de uma língua em quaisquer circunstâncias. Desse modo, recorrer a métodos estatísticos e matemáticos, ou recorrer a clusters, permite ponderar diferentes fatores relativos à importância relativa das línguas, como o peso demográfico ou a presença na Internet, por exemplo. Esse procedimento indica vantagens, fraquezas, aberturas e ameaças possíveis para o francês. No que concerne às três línguas em questão, espanhol, francês e português, não se deve esperar mudanças extremas nos quarenta anos futuros, e isso no tocante aos três componentes a seguir: expansão demográfica, técnico econômico e prestígio cultural.*

Palavras-chave: *importância das línguas, peso relativo das línguas, situação do francês*

Abstract: *There are many elements that allow us to evaluate the importance of a given language today or in the future. A scientific look contributes to avoid an emotional point of view or not to considerate only a single element in the determination of the weight of a language in any circumstances. Therefore, the use of statistic and mathematic methods, or the use of clusters, allow us to ponderate different elements related to*

the relative importance of languages, as the demographic weight or the presence on the Internet, for instance. This procedure indicates advantages, weaknesses, openness and possible threatens to French. Concerning the three languages Spanish, French and Portuguese, we shouldn't wait extreme changes in the next forty years, specially if we relate this to the three components that follow: demographic expansion, economical-technical and cultural prestige.

Keywords: *language importance, relative weight of language, French situation*

Il est possible que certains d'entre vous connaissent déjà mes travaux actuels, que d'autres non, et je prie donc les premiers de bien vouloir m'excuser pour certaines redites, mais je vais devoir commencer par une rapide description des effets de la mondialisation sur la situation linguistique du monde avant de présenter des instruments nous permettant d'élaborer quelques hypothèses sur l'évolution de cette situation.

Ma démarche répond à la volonté d'éviter deux obstacles, lorsque l'on parle de la situation linguistique du monde, celui de la passion d'une part, que nous trouvons chez quelques ayatollahs de la francophonie, et d'autre part celui du discours anti-anglais, courant comme vous savez, et parfois aux frontières de la paranoïa. Disons tout simplement que je voudrais tenter de porter sur la question de la situation du français un regard scientifique et quantifié, en partant du fait qu'on ne peut bien entendu pas évaluer la situation d'une langue en elle-même, mais uniquement par rapport aux autres langues.

Tout d'abord, un rappel de quelques données chiffrées que vous connaissez sans doute. Vous savez que, d'un point de vue géographique, les langues du monde sont assez inégalement réparties (Afrique : 30% ; Amériques : 15% ; Asie : 33% ; Europe : 3% ; Pacifique : 19%). Nous avons un espace inégalement occupé avec des continents pauvres en langues, comme l'Europe, et d'autres riches, comme l'Asie et l'Afrique.

Pour illustrer les choses d'une autre façon, une représentation graphique montre qu'un petit nombre de langues (0,2% soit une vingtaine de langues) sont parlées par un grand nombre de locuteurs (plus de 100 millions chacune) et représentent 44,3% des locuteurs, puis que 1,2% des langues (72 langues) sont parlées par 38% des locuteurs, c'est-à-dire qu'1,2% des langues du monde représentent 82,3% des locuteurs. En revanche, le plus grand nombre des langues, plus de 80%, sont parlées par très peu de locuteurs. Et les trois langues dont je vais vous parler, l'espagnol, le français et le portugais, sont dans les 0,2% des langues les plus parlées dans le monde.

Cela, c'est une réalité incontournable, par rapport à laquelle il nous faut réfléchir. Bien sûr les langues qui nous retiennent ici, ne figurent pas dans les langues menacées de disparition, mais leur devenir est inséparable de celui des autres langues, et je vous propose d'analyser sa situation à partir de mes recherches actuelles sur ce que j'appelle le « poids » des langues.

Commençons par une rapide évocation historique. En 1532, dans *Pantagruel*, Rabelais fait écrire par Gargantua une lettre à son fils, parti étudier à Paris, dans laquelle il dresse son programme d'éducation. On y trouve, pour ce qui concerne les langues, les propositions suivantes :

« *J'entens et veulx que tu aprenes les langues parfaitement. Premierement la grecque, comme le veult Quintilian, secondement la latine ; et puis l'hebraïcque pour les saintes lettres, et la chaldaïcque et arabicque pareillement ; et que tu formes ton style quant a la grecque, a l'imitation de Platon ; quant a la latine, de Ciceron...* »

Le grec, le latin, l'hébreu, le chaldéen, l'arabe... Ce programme, qui peut faire aujourd'hui sourire ou paraître suranné, témoigne pourtant d'une vision des langues qui comptaient à cette époque, des langues qui semblaient utiles, en un mot qui avaient du poids. Elles ne pèseraient guère aujourd'hui aux yeux d'un parent d'élève qui aurait à choisir les langues étrangères que ses enfants devraient étudier. Ainsi, dans une enquête menée en 2007 auprès d'un public d'étudiants brésiliens de six universités de Rio de Janeiro¹, à la question « *quelles langues aimeriez-vous qu'apprennent vos enfants ?* » il est répondu : l'anglais (82%) l'espagnol (30%) le français (25%).

La même enquête effectuée aux quatre coins du monde donnerait des résultats comparables, à quelques variantes locales près (le chinois par exemple apparaîtrait sans doute au Japon ou en Australie...), et l'on n'y trouverait certainement pas les langues choisies par Gargantua pour son fils.

Entre 1532 et 2007, l'évaluation de l'importance des langues a ainsi changé radicalement. De tous temps les êtres humains ont eu à la fois une vision utilitaire des langues, une évaluation intuitive de leur importance, et un jugement sur elles. On entend ainsi dire qu'une langue est « *belle* », « *chantante* », « *gutturale* », qu'on aime telle ou telle langue, qu'on n'aime pas telle autre, que celle-ci est utile, celle-là moins... Bref l'espèce humaine n'a pas cessé d'établir des échelles de valeur plus ou moins approximatives ou intuitives, d'émettre des stéréotypes, dont Einstein disait qu'ils étaient plus difficiles à désintégrer qu'un atome... En d'autres termes, dès lors que des populations parlant des langues différentes sont entrées en contact, ces contacts et les rapports de force qui leur étaient associés ont produit une certaine *vision des langues*.

Comment mesurer le « poids » des langues ?

Dès lors se pose une question : Comment mesurer l'importance relative des langues, en évitant les stéréotypes, les idées toutes faites ? Comment les classer ?

Cette approche pose bien sûr un certain nombre de questions. Les *langues* tout d'abord sont-elles des entités *identifiables* et *comptables* de façon univoque ? Le site ethnologue par exemple, auquel on se réfère généralement, fluctue souvent dans ses classifications, considérant l'arabe comme *une* langue, ce qu'il a fait à une époque, ou prenant en compte les différents arabes, ce qu'il fait aujourd'hui. Mais le SIL (dont ethnologue est l'expression) a mis au point une norme ISO des langues, la norme ISO 639-5, dans sa version la plus récente,

celle de 2008, qui utilise un code alpha-3 (à trois lettres) pour désigner les langues du monde. On y trouve par exemple pour l'arabe 35 codes différents, allant d'une forme parlée par plus de 47 millions de personnes, comme l'arabe égyptien, à des formes parlées par 7000 personnes, comme l'arabe chypriote ou l'arabe tadjik, ce qui, d'un point de vue linguistique, n'est pas faux : Il existe beaucoup de dialectes arabes, pas toujours intercompréhensibles. En revanche on peut se demander pourquoi il n'y a qu'un code pour l'anglais, l'espagnol ou le portugais et deux pour le français. On divise d'un côté, on unifie de l'autre, et peut-être y a-t-il un agenda caché, peut-être divise-t-on pour mieux régner. Quoiqu'il en soit cette norme ISO existe, et les langues que nous répertorions ont donc toutes un code alphanumérique.

Revenons à la question de l'importance relative des langues, des rapports qu'elles entretiennent, de l'intérêt qu'elles peuvent représenter. Lorsque l'on s'interroge aujourd'hui sur le problème du classement des langues, ou de l'importance relative des langues, on pense en général à un critère unique, celui du nombre de leurs locuteurs : combien de gens parlent telle ou telle langue ? Soulignons tout de suite que le calcul du nombre de locuteurs d'une langue n'est pas une science exacte, que les évaluations varient considérablement. Trois sources qui datent de la même année (2003) nous donnent des classements très différents. Ainsi pour l'encyclopédie Quid et pour Ethnologue du Summer Institut of Linguistics le français atteint la dixième position, alors que Linguasphere (UNESCO) le situe à la ozième. De même, l'anglais est classé deux fois deuxième et une fois troisième ; l'espagnol, de son côté, deux fois quatrième et une fois deuxième.

Mais, même si nous avions les moyens de savoir en temps réel le nombre exact de locuteurs des différentes langues du monde, nous n'aurions là qu'un facteur d'évaluation de leur importance. Le chinois par exemple est certes la langue la plus parlée, mais elle n'est parlée que dans quelques pays, elle n'est langue officielle que de rares pays, etc...

D'où l'idée de réfléchir sur le « poids » des langues. Nous nous proposons en fait de tenter une mesure et une comparaison de l'importance des langues, c'est-à-dire une classification, à partir du plus grand nombre de facteurs discriminants possible, dont il faudra chaque fois tester la pertinence. Ces facteurs nous permettront d'effectuer une classification que nous analyserons ensuite plus finement à l'aide de méthodes statistiques.

Voici donc une liste non exhaustive des différents facteurs utilisables :

- Nombre de locuteurs « langue maternelle »
- Nombre de locuteurs « langue seconde »
- Traductions «source »
- Traductions «cible »
- Prix Nobel de littérature
- Production cinématographique
- Existe-t-il un traitement de texte?
- Existe-t-il un correcteur orthographique?
- Nombre de pays dans lesquels la langue est officielle/co-officielle

- Nombre de pays dans lesquels on peut étudier ces langues
- Importance sur Internet

Certains de ces facteurs sont très importants, comme la production cinématographique. Voici la moyenne du nombre de films produits par an :

- Inde (709)
- Anglais (Australie, USA, Royaume Uni, Nlle Zélande) (571)
- Japonais (273)
- France (155)
- Espagnol (Argentine, Cuba, Espagne, Mexique, Pérou) (108)
- Italien (96).

Pour les cinémas anglais, japonais, français, espagnol, italien, les choses sont claires : nous savons en quelles langues sont tournés les films. Mais la première place est occupée par le cinéma indien, avec 700 films par an, sans que nous sachions encore quelle est la répartition par langues (hindi, ourdou, tamoul, pendjabi, telougou, marathi, etc.) et nous ne pouvons donc pas utiliser ce facteur pour l'instant.

Voici donc la liste des dix facteurs que je vais utiliser aujourd'hui et, entre parenthèse, les sources des données utilisées :

- Nombre de locuteurs (<http://www.ethnologue.com/web.asp>)
- Nombre de pays dans lesquels la langue a un statut officiel (<http://www.ethnologue.com/web.asp>)
- Nombre d'articles dans Wikipedia (http://en.wikipedia.org/wiki/Wikipedia:Multilingual_statistics)
- Nombre de prix Nobel de littérature (<http://nobelprize.org/>)
- Entropie (calculée à partir des données de population)
- Taux de fécondité (<http://www.prb.org/FrenchContent.aspx>)
- Indice de développement humain, IDH (<http://www.undp.org/french/>)
- Taux de pénétration d'internet (<http://www.internetworldstats.com/stats.htm>)
- Nombre de traductions, langue cible (<http://databases.unesco.org/xtrans/stat/xTransStat.html>)
- Nombre de traductions, langue source (<http://databases.unesco.org/xtrans/stat/xTransStat.html>)

La grande majorité de ces facteurs n'appelle ni commentaires ni explications. Certains en revanche peuvent faire problème ou demandent des précisions.

Il en va ainsi de l'entropie. L'entropie est une fonction qui permet de quantifier le "désordre". Elle a été utilisée à l'origine en thermodynamique, puis a trouvé des applications en théorie de l'information et plus récemment en linguistique. Son expression mathématique est la suivante :

$$\text{Entropie} = -\sum(p_i \cdot \text{Log}(p_i))$$

Dans laquelle p_i est la probabilité pour un système de se trouver dans un état donné et $\text{Log}(p_i)$ le logarithme naturel de cette probabilité. La valeur minimale de cette fonction est zéro et il n'existe pas de valeur maximale définie. Nous

l'utilisons ici pour différencier une langue parlée dans un seul pays d'une langue parlée dans plusieurs pays, p_i sera donc pour nous la proportion des locuteurs d'une langue donnée vivant dans un pays donné et tous les pays du monde dans lesquels cette langue est parlée seront pris en compte.

Considérons une langue parlée très majoritairement (98%) dans un pays un dont quelques locuteurs vivent dans un second, l'entropie sera :

$$- (0.98 * \text{Log}0.98 + 0.02 * \text{Log}0.02) = 0.098$$

Une langue parlée dans trois pays de démographie comparable aura une entropie de

$$- (0.33 * \text{Log}0.33 + 0.33 * \text{Log}0.33 + 0.34 * \text{Log}0.34) = 1.099$$

Voyons à présent quelques exemples réels. Le marathi, l'amharique, l'espagnol et l'arabe. L'entropie n'a donc rien à voir avec le nombre global de locuteurs d'une langue mais bien avec la façon dont ces locuteurs sont répartis dans l'aire ou les aires dans lesquelles cette langue est parlée.

Un problème : ces différents facteurs nous donnent une classification des langues en deux groupes (valeurs catégorielles : [oui/non] ou en une hiérarchie de 1 à 7000 [valeurs continues]).

Nous procédons donc à une transformation linéaire en ramenant, pour chaque facteur, la valeur maximale à 1 et la valeur minimale à 0, la valeur des langues intermédiaires étant produite par une simple règle de 3.

Voici, à titre de comparaison, ces deux types de classements pour le nombre de locuteurs.

Langue	Rang	Valeur absolue (nombre de locuteurs)	Valeur normée
Mandarin	1	725.5	1
Portugais	2	174.5	0.240283
Haoussa	3	45	0.061729
Xiang	4	36	0.04932
Malayalam	5	35.8	0.049044
Bahasa	6	30.3	0.04146
Visayan/Cebuano	7	20	0.027259
Népal	8	17.2	0.023398
Hongrois	9	13.1	0.017745
Slovaque	10	8	0.010713
Norvégien	11	4.5	0.005887
Islandais	12	0.23	0

Le second (valeur normée) nous paraît plus précis et c'est celui-ci que nous utiliserons.

Pour illustrer le fait que l'analyse des rapports entre les langues uniquement en termes de nombre de locuteurs ne suffit pas, je vous propose de considérer ces classements des premières langues selon quelques-uns des facteurs retenus.

- Nombre de locuteurs :

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

- Nombre de pays dans lesquels les langues sont officielles :

1. Anglais	5. Portugais	9. Bahasa
2. Français	6. Allemand	10. Néerlandais ²
3. Arabe	7. Italien	10 Hongrois
4. Espagnol	8. Russe	10 Mandarin.

- Présence des langues sur Internet : 1. Anglais ; 2. Mandarin ; 3. Espagnol ; 4. Japonais ; 5. Allemand ; 6. Français ; 7. Coréen ; 8. Italien ; 9. Portugais ; 10. Malais.

- « prix Nobel de littérature » :

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

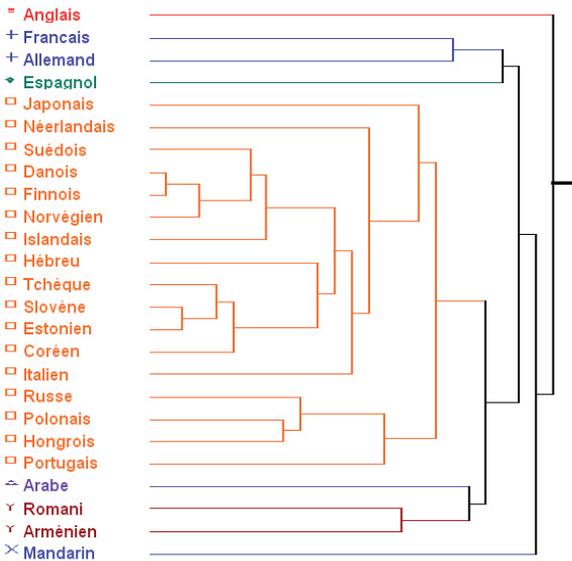
Ce dernier facteur nous pose problème. Il peut s'agir en effet qu'un regard occidental sur la littérature mondiale, comme le montre la portion congrue réservée aux langues arabe ou chinoise. Mais en même temps il témoigne d'une réalité : les littératures arabe ou chinoise ne sont pour l'instant pas très accessibles aux lecteurs occidentaux (le facteur « traduction langue source » en témoigne également) et nous avons donc là un indice à suivre car il pourra dans l'avenir refléter des évolutions.

- Traductions langue source :

1. Anglais	8. Danois	15. Norvégien
2. Français	9. Tchèque	16. Portugais
3. Allemand	10. Néerlandais	17. Hébreu
4. Russe	11. Polonais	18. Mandarin
5. Italien	12. Japonais	19. Finnois
6. Espagnol	13. Hongrois	20. Bahasa.
7. Suédois	14. Arabe	

Il ne s'agit là que de quelques exemples, et nous n'insisterons pas plus sur les difficultés de recueil de ces différentes données : elles posent des problèmes techniques et financiers mais peu de problèmes scientifiques. En d'autres termes nous savons exactement ce que nous voulons faire, et nous savons le faire, il nous manque simplement les moyens humains et donc financiers de saisir toutes les données pour toutes les langues...

Mais il y a une autre façon de tester les facteurs: l'analyse en "clusters". L'ensemble des critères définit un espace à 10 dimensions dans lequel les langues se répartissent. La question est alors de savoir si elle se répartissent de manière régulière ou si elle se regroupent en nuages plus ou moins homogène, les "clusters", selon leur comportement face aux différents facteurs. Deux langues appartenant à un même cluster sont plus proches l'une de l'autre, dans l'espace considéré, que des autres langues. Voici le résultat d'une telle analyse pour les 25 premières langues réparties en 7 clusters.



Ce type de représentation appelée dendrogramme permet de visualiser en deux dimensions ce qui se passe dans un espace à dix dimensions.

Pour tester les différents facteurs, nous pouvons introduire dans le tableau des "fausses" langues en modifiant légèrement leurs caractéristiques. Par exemple, en "inventant" un espagnol qui ne serait officiel que dans un pays, et un mandarin qui n'aurait que 50 millions de locuteurs. Dans ce cas nous constatons que le "faux" espagnol se comporte comme le vrai, les deux langues se regroupant dans le même cluster, ceci parce que de nombreux autres facteurs (prix Nobel, nombre de locuteurs, place sur Internet...) assurent son classement, tandis que le "faux" mandarin se sépare du "vrai" parce que le facteur qui assure le poids de cette langue est essentiellement son nombre de locuteurs, le faux mandarin se trouve dans le cluster le plus peuplé, en quelque sorte dans le "lieu commun".

Et ceci est confirmé si nous supprimons les deux “vraies” langues pour ne conserver que les “fausses” langues : le “faux” espagnol reste seul, en singleton, tandis que le “faux” mandarin se regroupe avec le portugais, le bahasa, etc. Cette technique des fausses langues nous permet ainsi de juger de la pertinence des facteurs: s’il en est qui ne changent en rien le classement, ils peuvent être considérés comme redondants et éliminés.

Cette analyse est multifactorielle. Parmi les dix facteurs retenus certains sont partiellement redondants, par exemple une langue bien classée sur le facteur prix nobel de littérature le sera également sur le facteur langue source de traduction. On dit que ces facteurs sont corrélés. Il existe en mathématique une méthode pour réduire le nombre de facteurs en perdant le minimum d’information : l’analyse en composantes principales. Cette analyse consiste à réorganiser les facteurs originaux en un nombre égal de facteurs “latents” et à les classer par ordre décroissant de “quantité d’information” qu’ils apportent. Cela donnerait par exemple un facteur intermédiaire entre prix Nobel et langue source. Nous n’en sommes pas là, mais c’est à ce type d’analyse que nous voudrions aboutir.

Pour l’instant nous utilisons nos facteurs normés et voici la classification des vingt premières (entre parenthèses, la somme des valeurs normées) selon cette technique :

- | | |
|-------------------------|------------------------|
| 1. Anglais (7.1696) | 11. Islandais (2.3055) |
| 2. Français (4.4703) | 12. Finnois (2.1844) |
| 3. Espagnol (4.3887) | 13. Romani (2.1790) |
| 4. Allemand (4.0885) | 14. Russe (2.1784) |
| 5. Japonais (2.7013) | 15. Polonais (2.1450) |
| 6. Néerlandais (2.6956) | 16. Portugais (2.1222) |
| 7. Arabe (2.6580) | 17. Norvégien (2.1175) |
| 8. Suédois (2.5820) | 18. Mandarin (2.0652) |
| 9. Italien (2.4290) | 19. Hébreu (1.9912) |
| 10. Danois (2.3633) | 20. Tchèque (1.9435) |

Il s’agit là de la préfiguration d’un projet en cours dont la version finale, un “index des langues du monde”, constituera un observatoire du « poids » des langues, et sera remis à jour régulièrement si le comportement des langues face aux différents facteurs change.

Reste bien sûr une question : Donner un classement des langues du monde, A QUOI CELA PEUT-IL SERVIR ?

L’intérêt premier est pour moi d’observer le comportement des langues face à un certain nombre de facteurs et ainsi de réfléchir sur leur devenir. Il sera présenté en accès libre sur Internet, et l’on pourra consulter à la fois nos sources et la façon dont nous les avons traitées, conditions à nos yeux essentielles pour que puisse s’ouvrir un débat scientifique. On pourra aussi y faire son propre classement en cliquant sur les facteurs qui nous intéressent, en supprimant certains d’entre eux ou en leur donnant un coefficient différent. L’utilisateur du site pourrait ainsi réaliser un classement correspondant à ses questions, ses attentes ou ses besoins.

De ce point de vue, il constituera une aide non négligeable à la décision en matière de politiques linguistiques. Un pays qui s'interrogerait sur les langues à enseigner dans le secondaire ou le supérieur, ou encore pour un groupe de locuteurs qui s'interrogerait sur le statut de sa langue et pourrait aller voir pourquoi d'autres sont mieux classées, et ce qu'il convient de faire pour faire « monter » la sienne. Mais le fait de réaliser son propre classement est plus intéressant.

En voici deux exemples, deux cas concrets : Imaginons quelqu'un qui voudrait savoir en quelles langues il faudrait rédiger un nouveau logiciel révolutionnaire qu'il prépare pour dans dix ans. Il pourra ne retenir que trois facteurs :

- nombre de locuteurs
- nombre d'articles dans Wikipedia
- pénétration d'internet

Imaginons maintenant une radio internationale qui voudrait savoir en quelles langues il faudrait émettre pour toucher le plus de gens possibles. Elle pourra choisir les facteurs :

- Nombre de locuteurs
- Pays dans lesquels les langues sont officielles
- Entropie

Les résultants seraient :

Classement «Logiciel » (Nbre de locuteurs, Wikipedia, Internet)	Classement «Radio » (Nbre de locuteurs, langues officielles, entropie)
Anglais	Espagnol
Mandarin	Anglais
Japonais	Arabe
Allemand	Mandarin
Français	Français
Espagnol	Hindi
Italien	Bengali
Portugais	Russe
Hindi	Portugais
Coréen	Bahasa/Malais

Cet « index » des langues, cette classification, peut se décliner régionalement (Méditerranée, Amérique latine, etc.). Nous avons ainsi établi, pour un colloque sur les politiques linguistiques en Méditerranée, le « poids des langues en Méditerranée », où l'on constate que les langues ne se comportent pas de la même façon régionalement et mondialement. Voici par exemple deux classements du 'poids des langues' en Méditerranée, sans l'anglais et avec l'anglais :

Sans l'anglais

- français
- arabe
- italien
- espagnol
- hébreu
- grec
- catalan
- turc
- maltais

Avec l'anglais

- arabe
- français
- anglais
- italien
- espagnol
- hébreu
- grec
- catalan
- turc
- maltais

Après cette rapide présentation du modèle, je voudrais en venir à ce qui nous intéresse, une analyse prospective qui, en partant de la situation actuelle tentera d'imaginer ce qu'elle pourrait être en 2025. Nous allons partir du cas du français.

Voici pour commencer les 20 premières langues :

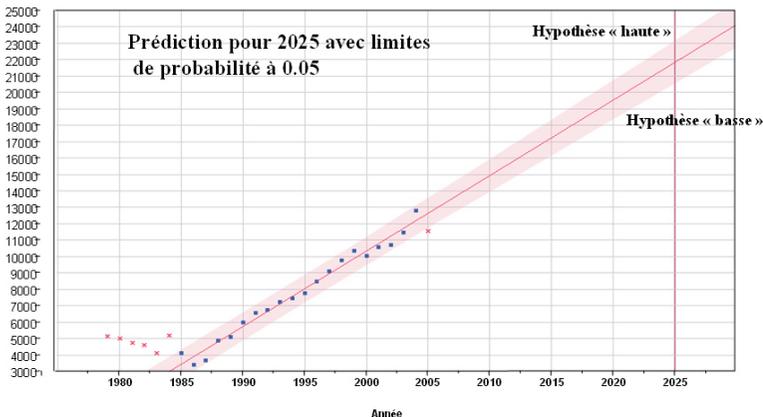
Anglais	7.1696	Romani	2.1790
Français	4.4703	Russe	2.1784
Espagnol	4.3887	Polonais	2.1450
Allemand	4.0885	Portugais	2.1222
Japonais	2.7013	Norvégien	2.1175
Néerlandais	2.6956	Mandarin	2.0652
Arabe	2.6580	Hébreu	1.9912
Suédois	2.5820	Tchèque	1.9435
Italien	2.4290	...	
Danois	2.3633	...	
Islandais	2.3055	Bahasa/Malais (32ème)	
Finois	2.1844		1.5511

auxquelles j'ajoute le malais, actuellement 32^{ème} mais dont nous verrons qu'il va connaître une progression importante.

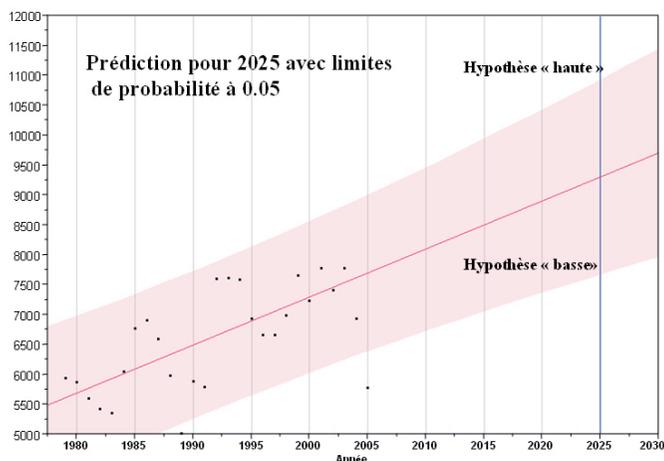
En considérant ces données, nous remarquons:

- Le statut nettement séparé de l'anglais (7.1696), dont la première place n'est nullement en danger.
- Le fait que français (4.4703) et espagnol (4.3887) sont très proches, et que leurs places pourraient s'inverser.
- Le fait que d'autres langues sont très proches les unes des autres et que leurs places pourraient également s'inverser : arabe, japonais et néerlandais.

Pour avoir une idée de l'évolution possible de cette situation, la méthode la plus simple serait bien sûr de partir des données concernant les années précédentes et de prolonger les courbes pour chacun des facteurs : c'est ce qu'on appelle la méthode de **régression linéaire**, et c'est celle que recommande d'utiliser le CAS. Voici à titre d'exemple ce que cela donnerait pour le facteur « langue cible» pour le français :



et pour le facteur « langue source ».



En utilisant la méthode de régression linéaire pour chacun des dix facteurs retenus nous pouvons élaborer différents scénarii. Nous allons donc essayer de présenter ce que serait la situation du français en 2025, puisque telle était la commande du CAS. Selon une hypothèse « moyenne », le score (et le rang) des quatre premières langues ne changerait pas (1. anglais ; 2. français ; 3. espagnol ; 4. allemand). Il faut cependant noter la montée de l'arabe (du 7^{ème} rang au 5^{ème} rang) et la baisse du japonais (du 5^{ème} au 6^{ème}) et du russe (du 9^{ème} au 10^{ème}), ainsi que la montée remarquable du bahasa (du 32^{ème} au 11^{ème}).

Nous pouvons imaginer un scénario favorable « extrême », dans lequel le français serait toujours dans l'hypothèse haute et les autres langues dans l'hypothèse basse, qui ne changerait rien aux quatre premières langues mais serait défavorable à l'arabe. A l'inverse, un scénario défavorable « extrême », dans lequel le français serait toujours dans l'hypothèse basse et les autres langues dans l'hypothèse haut, renverrait le français après l'espagnol et ferait passer l'arabe devant l'allemand.

Mais la date de 2025, choisie par le CAS est très proche et nous pouvons tenter de voir plus loin. Pour cette rencontre nous avons appliqué la même technique jusqu'en 2050. Voici ce que cela donnerait :

2025			2050		
1.	Anglais	6,7968	1.	Anglais	6,8396 (=)
	Français	4,5775		Français	4,6033 (=)
	Espagnol	4,2815		Espagnol	4,5935 (=)
	Allemand	3,5737		Arabe	3,7461 (+1)
5.	Arabe	3,0545	5.	Allemand	3,5934 (-1)
	Italien	2,8093		Mandarin	2,8272 (+3)
	Portugais	2,3068		Italien	2,7963 (-1)
	Japonais	2,2557		Portugais	2,7574 (-1)
	Mandarin	2,2533		Russe	2,5992 (+1)
10.	Russe	2,0517	10.	Hindi	2,3875 (+5)
	Coréen	2,0150		Japonais	2,3406 (-3)
	Turc	1,6663		Coréen	2,3314 (-1)
	Bahasa	1,6420		Bengali	2,2896 (+3)
	Javanais	1,5981		Bahasa	2,2019 (-1)
15.	Hindi	1,5415	15.	Javanais	2,1206 (-1)
	Bengali	1,4509		Turc	2,0852 (-4)
	Vietnamien	1,3663		Vietnamien	1,9296 (=)
	Cantonais	1,2630		Cantonais	1,8546 (=)
	Tamoul	1,1451		Tamoul	1,8496 (=)
20.	Wu	1,1084	20.	Telugu	1,8112 (+1)
	Telugu	1,0768		Marathi	1,7559 (+1)
	Marathi	1,0135		Wu	1,6888 (-2)

Vous voyez que si les trois premières langues ne bougent pas (anglais, français, espagnol) il y a ensuite des langues qui « montent » (arabe, mandarin) et d'autres qui descendent (allemand, japonais...)

Cette méthode est simple mais insuffisante car pour une partie des facteurs retenus se présentent différentes possibilités d'évolution. Surtout elle s'apparente à une météo qui se contenterait d'annoncer chaque soir que le temps de demain sera à peu près le même qu'aujourd'hui. On aurait ainsi de bonnes chances de ne pas se tromper mais on ne comprendrait rien aux phénomènes atmosphériques, à la mécanique des fluides appliquée aux mouvements de l'air, etc.

Or ce qui est intéressant, c'est précisément ces phénomènes atmosphériques, ces mouvements de l'air, c'est-à-dire, pour ce qui concerne les langues, les mouvements sous-jacents à leurs rapports, que la simple prolongation des courbes, même en jouant sur les hypothèses hautes ou basses, ne permettent pas de percevoir. Rien par exemple ne permettait il y a trois siècles de prévoir l'émergence brusque des créoles. Ou encore rien ne permettait il y a un siècle de prévoir l'éclatement de l'hindoustani en hindi et ourdou. C'est pourquoi il nous faut considérer ce qui peut contredire les prédictions obtenues par la méthode de la régression linéaire, trop mécaniste.

Un exemple : il y a actuellement 12% d'obèses en France et la technique de régression linéaire nous dit qu'en 2020 ce taux sera de 30%, le même qu'aux USA. Mais les services de santé feront tout pour que cela ne s'avère pas, et la régression linéaire n'est qu'un indicateur, encore une fois une météo approximative.

Ce qui nous mène à ce que nous appelons des « facteurs de rupture » :

- Langues sources de traduction
- Statut international, UE, OTAN, ONU, OSCE, OCDE, OMC, BM, FMI, et diverses ONG
- Langues officielles
- Diffusion internationale radio et télévision
- Le FLE(S).

Ce sont des facteurs pertinents sur lesquels il serait possible d'intervenir de façon positive ou négative, qui constituent des moteurs de changement et, parfois, pour ce qui concerne le français qui est notre sujet, des menaces potentielles. Vous voyez que nous en avons répertorié cinq, je vais en commenter trois, laissant de côté le problème des langues officielles, car je suis incapable de savoir si les pays africains par exemple vont conserver le français ou passer à une langue locale, et nous savons tous quels sont les problèmes de l'enseignement du FLE.

Commençons par les flux de traduction, que nous pouvons analyser à partir de la base de données de l'UNESCO (*index translationum*) ou des chiffres du syndicat national des éditeurs, et qui sont un indicateur fondamental pour analyser la place de la France et du français sur le marché mondial de la traduction. La situation actuelle repose en partie sur l'aide à la traduction : la diffusion de la pensée et de la production culturelle françaises à l'étranger ne tombe pas du ciel, elle est le sous-produit d'une politique (aide à la traduction, mais aussi tournées de conférences d'auteurs, financement de colloques à l'étranger, lycées français, alliance française etc.).

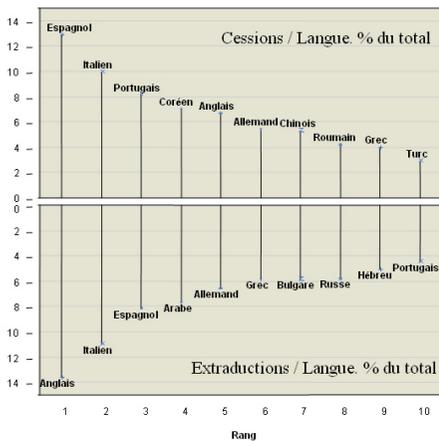


Tableau 1

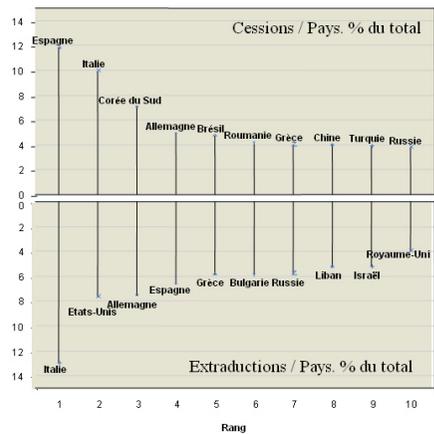


Tableau 2

Le tableau 1 nous montre que les pays de langues romanes sont les premiers acheteurs de titres français (partie haute du tableau) : espagnol, italien, portugais... La partie basse du tableau (extraduction) concerne les aides à la traduction et nous montre que la France subventionne d'abord les traductions vers l'anglais, l'italien et l'espagnol et, subsidiairement, vers l'arabe et l'allemand.

Le tableau 2 concerne les pays et non plus les langues. Nous y voyons que, concernant l'anglais, l'aide va surtout vers les Etats-Unis, la Grande-Bretagne finançant elle-même les traductions à partir du français.

Un ouvrage récent (sous la direction de Gisèle Sapiro, *Translatio, le marché mondial de la traduction en France à l'heure de la mondialisation*, CNRS éditions, Paris 2008) souligne que « la politique d'aide à l'extraduction a permis de conserver ou de reconquérir la position du français comme deuxième ou troisième langue centrale dans nombre de pays ». Ici l'aide à la traduction dispensée par le CNL est donc déterminante. Si cette aide diminuait, la diffusion de la production littéraire et intellectuelle française en pâtirait gravement.

Pour ce qui concerne le statut de langue internationale, le français est, avec l'anglais, l'une des deux langues de travail du secrétariat général de l'ONU, ce qui implique par exemple que le personnel recruté soit bilingue, et l'on sait que cela n'est pas toujours le cas. Par ailleurs il est théoriquement, au sein des institutions de l'UE, à égalité avec toutes les langues officielles des pays membres. Mais cette situation pose problème et l'UE n'arrive pas à s'entendre sur une politique linguistique efficace. Le refus de choisir favorise pour l'instant l'anglais, le français et dans une moindre mesure l'allemand. Il pourrait bien à terme ne favoriser que l'anglais. Ce dossier est extrêmement délicat, et les différentes présidences de l'Europe se le repassent comme une patate chaude, mais il conviendrait d'entamer une réflexion apaisée sur ce point.

Pour ce qui concerne les langues comme vecteurs d'information sur les radios internationales, nous avons identifié 115 radios nationales émettant vers l'étranger en différentes langues et interrogé leurs sites pour déterminer en quelles langues elles émettent. Le tableau suivant :

Plus de 50 fois	21 à 49 fois	11 à 20 fois	7 à 10 fois	5 et 6 fois
Anglais (52)	Français (37) Russe (29) Arabe (26) Espagnol (25) Allemand (22) Chinois (22)	Portugais (11 + 3 + 5) Farsi (15) Indonésien (15) Albanais (14) Hindi (12) Ourdou (11) Roumain (11) Serbe (11) Swahili (11) Vietnamiens (11)	Bengali Japonais (10) Pashtoun (10) Turc (10) Hausa (9) Thai (9) Dari (8) Grec (8) Italien (8) Birman (7) Bosniaque (7) Croate Géorgien (7) Kluner (7)	Azéri (6) Hongrois (6) Ukrainien (6) Bulgare (5) Coréen (5) Lao (5) Macédonien (5) Népalais (5) Polonais (5) Tamoul (5)

nous montre par combien de radios différentes les langues sont utilisées, et l'on voit qu'ici encore le français occupe la deuxième place, l'espagnol la cinquième et le portugais la huitième. Mais là aussi, les choses peuvent bien sûr changer. La réduction du nombre de langues utilisées par RFI ne menace pas directement le français mais :

- Elle menace la diffusion de la pensée française
- La même chose pourrait se produire sur d'autres radios qui supprimeraient des langues, dont le français

Pour nous résumer, la langue française est présente en tête du classement pour un certain nombre de facteurs : nombre de pays dans lesquels elle est officielle (deuxième place), utilisation par les radios internationales (deuxième place), prix Nobel de littérature (deuxième place), langue source de traduction (deuxième place) et nombre d'articles sur Wikipédia (troisième place). Elle est moins bien placée sur Internet (sixième place) et mal placée pour ce qui concerne le nombre de locuteurs (dix-neuvième place). Elle est bien placée mais potentiellement en danger comme langue internationale (secrétariat général de l'ONU, problème des langues de l'UE) et comme langue étrangère dans les systèmes scolaires. Nous avons ainsi des **atouts**, des **faiblesses**, des **ouvertures** possibles et des menaces, et l'analyse « AFOM » nous montre ce qu'il serait possible de faire (« ouvertures ») et ce qu'il faut surveiller de près (« menaces »). Pour les « atouts » et les « faiblesses » le chiffre entre parenthèse indique le classement actuel du français :

Atouts	Faiblesses
Langue officielle (2) Radios internationales (2) Traduction langue source (2) Littérature (Nobel : 2) Wikipedia (3)	Equipement informatique (11) Démographie (19) Fécondité (52)
Action culturelle vers l'étranger Plan informatique Traductions vers l'anglais	ONU, UE Baisse des budgets culturels vers l'étranger Enseignement du FLE(S)
Ouvertures	Menaces

J'ai indiqué que mon approche en termes de poids des langues pouvait constituer, entre autres choses, une aide à la décision en matière de politique linguistique. Ce tableau suggère qu'il conviendrait de développer une politique qui sauvegarde les points forts, tente de renforcer certains points faibles là où des ouvertures sont possibles et soit vigilante face aux menaces

Nous voyons que le statut international de la langue française est aujourd'hui essentiellement assuré par des **facteurs institutionnels** (langue officielle, langue de travail) et **culturels** (littérature, enseignement, traduction, radios internationales). Disons que pour ce qui concerne le premiers points (facteurs institutionnels) la langue française a de beaux restes, peut-être immérités, et sûrement fragiles, et que pour ce qui concerne le second point (facteurs culturels) elle est peut-être menacée. Car on peut avoir l'impression que la situation du français est fragile à l'ONU et à l'UE, et que la France est en

ce moment en train de brûler ses cartes de visite culturelles à l'étranger : réduction des budgets et des postes dans la politique de diffusion de la langue, etc. Ici encore la France dispose de possibilités de choix et d'action : le statut de la langue française est aussi lié aux moyens mis à la disposition des Alliances Françaises et des centres culturels par exemple.

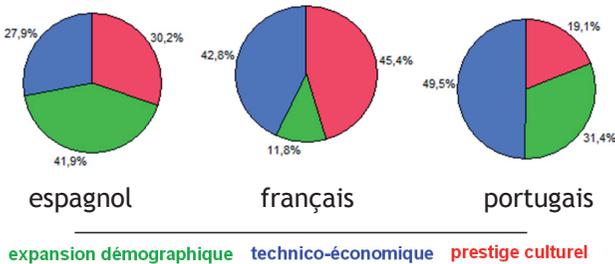
Venons-en aux deux autres langues annoncées, l'espagnol et le portugais. A des fins de comparaison, nous avons décomposé les dix facteurs déterminant le poids des langues en trois composantes :

Une composante démographique (Nombre de locuteurs, entropie, fécondité)

Une composante technico-économique (IDH, Wikipedia, internet)

Une composante culturelle (langue officielle, Nobel, langue cible, langue source).

Les rapports entre ces trois composantes varient bien sûr selon les langues, et voici ce que cela donne, en 2010, pour nos trois langues :

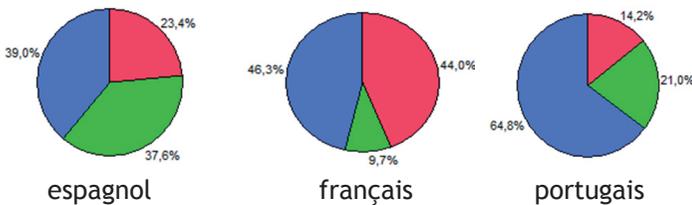


Nous voyons qu'elles se répartissent de façon équilibrée pour l'espagnol, que la composante démographique est faible pour le français, et que c'est la composante technico-économique qui pèse le plus lourd pour le portugais.

Quelle peut être l'évolution de ces composantes ? Entre 2010 et 2050 :

- La structure globale ne change pas, l'espagnol demeure « équilibré » sur la période
- Le français reste une combinaison de poids économique et culturel
- Le portugais combine poids démographique et économique
- Sur la période le poids économique relatif augmente et le poids culturel relatif diminue

La situation en 2050 serait donc la suivante :



encore une fois en prolongeant les courbes. Et le classement serait :

Anglais	6,8396	Russe	2,5992	Vietnamien	1,9296
Français	4,6033	Hindi	2,3875	Cantonais	1,8546
Espagnol	4,5935	Japonais	2,3406	Tamoul	1,8496
Arabe	3,7461	Coréen	2,3314	Telugu	1,8112
Allemand	3,5934	Bengali	2,2896	Marathi	1,7559
Mandarin	2,8272	Bahasa	2,2019	Wu	1,6888
Italien	2,7963	Javanais	2,1206		
Portugais	2,7574	Turc	2,0852		

J'ai dit « serait » car c'est par rapport à ces prévisions que les politiques doivent se déterminer, faire des choix. Des choix individuels, langue par langue (politique de la lusophonie, de la francophonie, etc.) ou des choix collectifs. Nous voyons par exemple sur cette liste qu'il y a quatre langues romanes dans les huit premières, et cela peut donner des idées de politique linguistique (rappel : les trois espaces linguistiques...).

Notes

¹ Menée sous la direction scientifique de Jürgen Heye, Telma Pereira et Louis-Jean Calvet

² A partir du néerlandais toutes ces langues sont officielles dans 3 pays.