

Méthodologie

En savoir plus sur le baromètre...

Les versions successives des Baromètre des langues ont été réalisées par Alain Calvet, docteur ès sciences et Louis-Jean Calvet, docteur ès lettres et sciences humaines, professeur de linguistique. Ont collaboré à leurs réalisations Daniel Prado, Annelore Lemoulinier, François Noctulle, Florica Razumieff et le docteur Alain Couillault. Il a reçu le soutien technique et financier de l'Union latine et de la DGLFLF.

Lorsque l'on s'interroge sur l'importance relative des langues, le critère du nombre de locuteurs est toujours le premier à apparaître : quelle est la langue la plus parlée au monde ? Combien de gens parlent cette langue ? etc. Mais cette approche pose un double problème. D'une part, le décompte des locuteurs n'est pas une science exacte et les différentes sources disponibles donnent des chiffres différents et ne parviennent pas toujours aux mêmes classements. Par ailleurs, d'autres facteurs jouent un rôle dans la détermination du « poids » des langues. Ce baromètre repose pour le moment sur douze facteurs choisis parce qu'ils apparaissent comme pertinents, mais aussi parce que les données étaient disponibles pour chacun d'entre eux. Il est bien sûr possible d'envisager d'autres facteurs que ceux qui sont ici pris en compte. La genèse de ce travail remonte au colloque tenu à Aix-en-Provence en septembre 2007. Le projet a évolué depuis lors et évoluera encore comme nous le verrons plus bas.

Nous avons pris en compte les langues qui selon [Ethnologue](#) ont plus de 500 000 locuteurs, elles sont au nombre de 634. Le choix du site Ethnologue comme point de départ est à nos yeux justifié par le fait que malgré ses imperfections il est le plus complet et le plus cohérent pour ce type d'étude. Les sources des différents facteurs et la manière dont ils sont traités sont décrites plus bas.

Le Baromètre des langues 2010 prenait en compte les 137 langues de plus de 5 millions de locuteurs (selon Ethnologue). Celui de 2012 répertoriait les 563 langues ayant plus d'un million de locuteurs et cette nouvelle version (2017) prend donc en compte celles qui ont plus de 500.000 locuteurs.

Par défaut, nous affectons la même importance à chacun des facteurs, mais en modifiant la position des curseurs de pondération, chacun pourra modifier le classement en fonction de ses besoins.

swahi

Facteurs et sources

Nombre de locuteurs

Il s'agit des locuteurs de langue première, tels que répertoriés par le site [Ethnologue](#). Ce choix implique bien sûr que nous acceptons la liste des langues proposée par ce site, ce qui pose quelques problèmes. Ethnologue a en effet tendance à diviser à l'extrême certaines langues en variantes différentes. Nous n'en prendrons qu'un exemple, celui du malais. Le site donne plus de vingt malais différents, à quoi il faut ajouter l'« indonésien » (ou bahasa indonesia), donné comme partageant 80 % de mots apparentés avec le malais standard qui lui-même est donné comme appartenant à une « macro langue » malais. Mais l'indonésien n'appartient pas, toujours selon le site, à cette macro langue. Et nulle part n'est posée la question de l'éventuelle intercompréhension entre ces différents malais. En outre, Ethnologue ne prend en compte que le nombre des locuteurs L1, ce qui a le désavantage de donner peu de visibilité au phénomène véhiculaire : ainsi, l'indonésien a pour Ethnologue environ 23 millions de locuteurs L1, mais 172 millions de locuteurs L2. Et cette absence de visibilité est encore plus manifeste avec une langue comme le swahili, donné avec un peu plus de 15 millions de locuteurs L1 en Tanzanie et 82 millions de locuteurs L2 en Afrique orientale alors qu'une édition précédente proposait les chiffres de 300000 et 30 millions. Il existe en outre un swahili du Congo (RDC) faisant partie d'une macro-langue swahili et comportant 1000 locuteurs L1 et 9 millions de locuteurs L2, comment s'y retrouver ? Mais, ici comme pour les autres facteurs, nous nous sommes interdits de corriger nos sources, pour d'évidentes raisons de cohérence scientifique. En revanche, nous ne suivons pas toujours Ethnologue pour ce qui concerne les noms des langues, que nous avons bien sûr traduits de l'anglais mais aussi parfois modifiés. Ainsi, le site confond parfois le nom d'un peuple (Mossi) et celui de sa langue (mooré), ou encore distingue entre farsi Est et farsi Ouest là où nous distinguons entre dari et farsi. Mais cela ne change rien à la liste des langues, qui sont définies de manière unique grâce à un code à trois caractères, le code ISO 639-3.

Nombre de pays dans lesquels la langue a un statut officiel

Nous ne prenons en compte que les langues officielles de jure, dans la liste établie par le site de [Université Laval](#). Nous nous en séparons cependant sur un point. Ce site comptabilise parfois la même langue pour un pays et une ou deux régions du pays en question. Par exemple, l'allemand est considéré comme officiel en Suisse, dans le canton de Berne, de Fribourg, etc., ou encore le Français est considéré comme officiel en Belgique, à Bruxelles capitale, dans la communauté francophone de Belgique. Dans ces cas, nous ne comptons le pays qu'une fois.

Il faut alors noter que la taille d'un pays n'est pas ici prise en compte, seule la souveraineté importe ce qui conduit à considérer que Monaco, le Liechtenstein ou Andorre sont plus importants que la Bavière ou la Californie. Nous avons en outre corrigé quelques oublis, comme le quechua.

Nombre d'articles dans Wikipédia

Nous utilisons ici les données que l'on trouve sur le site de [Wikipédia](#).

Le nombre que nous retenons est la somme de tous les articles publiés dans Wikipédia depuis l'origine de l'encyclopédie jusqu'à la date de collecte des données. En cas d'ambiguïté, par exemple lorsque Wikipédia affecte un nombre d'articles à une langue alors qu'il en existe des variantes, nous affectons à chaque variante un nombre d'articles calculé au prorata du nombre de ses locuteurs. C'est le cas des azéris (du nord et du sud), des farsis (de l'est et de l'ouest), des panjabis (de l'Est et de l'Ouest) et des pashtouns (central et du nord). En revanche dans le cas des ouzbeks, la totalité des articles a été affectée à l'ouzbek du nord parlé en Ouzbékistan, les 1.5 millions de locuteurs de l'ouzbek du sud vivant pour la plupart en Afghanistan ont été ignorés.

Prix littéraires

Nous utilisons ici les données que l'on trouve sur les sites des prix littéraires sélectionnés, par exemple <http://nobelprize.org/> pour le prix Nobel. Les valeurs ont été mises à jour en octobre 2016. Nous affectons le prix à la langue dans laquelle l'auteur a écrit l'essentiel de l'œuvre qui l'a distingué. Ainsi, Rabindranath Tagore écrit en Bengali, Gao XingJian en Mandarin, bien que titulaire d'un passeport français et Herta Müller en Allemand, même si elle est roumaine.

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. Nous l'utilisons ici pour différencier une langue parlée dans un seul pays d'une langue parlée dans plusieurs pays, p_i sera donc la proportion des locuteurs d'une langue donnée vivant dans les pays concernés. Classiquement l'expression mathématique de l'entropie est la suivante :

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

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, le symbole S indique que l'on fait la somme de tous les états p_i possibles. Dans notre cas nous utilisons évidemment p_i ainsi que défini plus haut. La valeur minimale de cette fonction est zéro et il n'existe pas de valeur maximale définie.

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

$$(0,98 \times \text{Log}0,98 + 0,02 \times \text{Log}0,02) = 0,098$$

Une langue dont les locuteurs sont répartis de manière égale sur trois pays aura une entropie de $(0,33 \times \text{Log}0,33 + 0,33 \times \text{Log}0,33 + 0,34 \times \text{Log}0,34) = 1,099$

Voyons à présent quelques exemples réels, ceux du marathi, de l'amharique, et de l'espagnol et du chinois mandarin:

Langue	Amharique	Marathi	Mandarin	Espagnol
Entropie	0.116	0.010	0.123	2.536
Locuteurs	30 M	62 M	900 M	430 M

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. Elle est calculée à partir des données de population décrites plus haut.

Taux de fécondité

Nous utilisons ici les données que l'on trouve sur le site des Nations Unies ayant trait au programme de développement humain. L'indice de fécondité est le nombre d'enfants par femme. Pour affecter une valeur à chaque langue, nous faisons une moyenne pondérée de l'indice dans chacun des pays dans lesquels la langue est parlée, sous réserve que le nombre de locuteurs dans le pays considéré soit au moins égal à 1 % du nombre de locuteurs de la langue dans le monde. Par exemple, 80 % des locuteurs de l'ourdou vivent en Inde et 18 % au Pakistan, l'indice de fécondité affecté à l'ourdou sera donc :

$$\text{Fécondité}_{\text{ourdou}} = (0,80 \times \text{Fécondité}_{\text{Inde}} + 0,18 \times \text{Fécondité}_{\text{Pakistan}})/(0,80 + 0,18)$$

En revanche, 0,7% des locuteurs de l'ourdou vivent au Royaume-Uni et 04 % au Bangladesh. Nous négligeons la contribution de ces locuteurs, leur proportion étant trop faible pour que les valeurs des indices de fécondité du Royaume-Uni et du Bangladesh aient un impact significatif sur la valeur relative à l'ourdou. Le site de l'UNDP cité ci-dessus ne fournit de données que pour les pays affiliés à l'ONU et pour lesquels un indice a été effectivement calculé, ce qui exclut notamment les pays non membres de l'ONU et les pays en guerre. Dans les cas où le pays n'est pas référencé sur le site nous utilisons les données fournies par : *L'état du monde*, (édition disponible plus récente), Paris, La Découverte 2008.

Indice de développement humain (IDH)

Nous utilisons ici les données que l'on trouve sur le site des [Nations Unies](#) ayant trait au programme de développement humain.

L'indice de développement humain est un indice composite prenant en compte le produit national brut par individu, l'espérance de vie à la naissance et le niveau d'éducation. Il quantifie le niveau de développement d'un pays. Pour affecter une valeur à chaque langue, nous faisons une moyenne pondérée de l'indice dans chacun des pays dans lesquels la langue est parlée sous réserve que le nombre de locuteurs dans le pays considéré soit au moins égal à 1 % du nombre de locuteurs de la langue dans le monde. Par exemple, 80 % des locuteurs de l'ourdou vivent en Inde et 18 % au Pakistan, l'IDH affecté à l'ourdou sera donc :

$$IDH_{\text{ourdou}} = (0,80 \times IDH_{\text{Inde}} + 0,18 \times IDH_{\text{Pakistan}}) / (0,80 + 0,18)$$

En revanche, 0,7 % des locuteurs de l'ourdou vivent au Royaume-Uni et 0,4 % au Bangladesh. Nous négligeons la contribution de ces locuteurs, leur proportion étant trop faible pour que les niveaux de développement du Royaume-Uni et du Bangladesh aient un impact significatif sur la valeur relative à l'ourdou. Le site de l'UNDP cité ci-dessus ne fournit de données que pour les pays affiliés à l'ONU et pour lesquels un indice a été effectivement calculé, ce qui exclut notamment les pays non membres de l'ONU et les pays en guerre pour lesquels les données ne sont pas disponibles. Dans ce cas nous affectons au pays non documenté un indice supposé que nous décidons par analogie avec les pays voisins et/ou comparables.

Taux de pénétration d'internet

Nous utilisons ici les données que l'on trouve sur le site [Internet World Stats](#) qui maintient à jour le nombre de liaisons internet dans chacun des pays du monde.

L'indice de pénétration d'Internet est le nombre d'utilisateurs d'Internet divisé par la population totale du pays. Pour affecter une valeur à chaque langue, nous faisons une moyenne pondérée de l'indice dans chacun des pays dans lesquels la langue est parlée sous réserve que le nombre de locuteurs dans le pays considéré soit au moins égal à 1 % du nombre de locuteurs de la langue dans le monde. Par exemple : 80 % des locuteurs de l'ourdou vivent en Inde et 18 % au Pakistan, l'IDH affecté à l'ourdou sera donc :

$$Internet_{\text{ourdou}} = (0,80 \times Internet_{\text{Inde}} + 0,18 \times Internet_{\text{Pakistan}}) / (0,80 + 0,18)$$

En revanche, 0,7 % des locuteurs de l'ourdou vivent au Royaume-Uni et 0,4 % au Bangladesh. Nous négligeons la contribution de ces locuteurs, leur proportion étant trop faible pour que les valeurs des indices de pénétration d'Internet du Royaume-Uni et du Bangladesh aient un impact significatif sur la valeur relative à l'ourdou.

Nombre de traductions, langue cible et langue source

Nous utilisons ici les données de l'[Index translationum](#) que l'on trouve sur le site de l'UNESCO.

Dans certains cas l'*Index translationum* indique une langue qui aurait « disparu ». C'est le cas du serbo-croate, code iso [hbs] qui indique qu'avant 1992 ce que parlaient les Croates, les Bosniaques et les Serbes était considéré comme une seule et même langue. Après 1992, *translationum* compile les traductions à partir du croate [hrv], du bosniaque [bos] et du serbe [srp]. Pour tenir compte des traductions antérieures à 1992, nous les répartissons entre les trois langues « nouvelles » au prorata du nombre de traductions à partir de ces trois langues après 1992. Cette méthode n'est bien sûr pas parfaite, mais nous avons pensé que l'affectation des traductions antérieures à 1992 au seul serbe favoriserait celui-ci par rapport au croate et au bosniaque. L'affectation du nombre total de traductions antérieures à 1992 aux trois langues favoriserait celles-ci par rapport aux autres langues. Un cas analogue de langue hypothétique se présente avec le « chinois ». *Translationum* reporte des données pour le chinois [zho], le chinois ancien [och] et le chinois yué [yue]. Dans ce cas, nous affectons les données relatives au « chinois » [zho] au seul mandarin et les données relatives au yué à celui-ci. Les données relatives au chinois ancien [och] ne sont pas prises en compte, tout comme celles relatives au grec ancien et au latin.

Une troisième situation est celle de l'arabe, il existe un arabe « standard » [ara] et plus de 25 arabes dialectaux. Tous les arabes dialectaux sont évidemment des langues vivantes et l'arabe standard ou littéraire est une langue « commune » qui ne possède aucun locuteur en langue maternelle. Cependant, l'*Index translationum* reporte des données pour l'arabe standard [ara] l'arabe du Tchad [shu] et l'arabe marocain [ary] mais rien pour les arabes égyptien et algérien qui sont les plus parlés avec le marocain ni pour aucune des autres variétés dialectales. Dans ce cas contrairement au serbo-croate et au chinois les données relatives à l'arabe standard sont affectées à lui et à lui seul. La raison de ce choix réside pour une part dans notre volonté de respecter les décisions des peuples. Les Croates, Bosniaques et Serbes souhaitent qu'il existe trois langues différentes correspondant à leurs trois pays. Dans le cas des arabes, la communauté arabo-musulmane souhaite qu'il existe une langue commune supposée comprise de tous au travers du Coran. Le cas du chinois est plus difficile à rationaliser. Il existe évidemment plusieurs chinois parlés, non inter-compréhensibles, mais « un seul » chinois écrit commun à

toutes les variétés dialectales. L'information compilée sous le code [zho] chinois ne permet pas d'affecter la traduction à l'une ou l'autre des variétés. Nous aurions pu choisir d'affecter les valeurs au prorata du nombre de locuteurs de celles-ci. Nous avons fait l'hypothèse que la grande majorité des traductions venaient du chinois mandarin et donc affecté toutes les données du chinois [zho] au chinois mandarin [cmn]. Certains lecteurs mieux informés pourront être en désaccord avec notre choix, il fallait en faire un et nous assumons celui-ci.

Le cas du tchèque [ces] et du slovaque [slk] est encore différent. L'*Index translationum* ne reporte pas de langue commune antérieure à la séparation des deux pays, il n'existait pas de langue tchécoslovaque, mais bien un tchèque et un slovaque.

Ces considérations s'appliquent au cas des langues sources que des langues cibles.

Universités

L'idée de ce facteur, introduit pour la première fois dans le baromètre des langues est de quantifier l'importance d'une langue par son enseignement au niveau universitaire. Il s'agit d'examiner les sites internet d'un échantillon des universités dans tous les pays du monde pour en extraire l'information sur les langues enseignées aux premiers niveaux de l'enseignement supérieur, le niveau doctorat (ou "post-graduate") étant exclu. Sont également pris en compte :

- Les organismes universitaires ou para-universitaires consacrés à l'enseignement de "langues rares" ou plutôt rarement enseignées. Il s'agit là de l'INALCO (Institut National des Langues et Civilisations Orientales, en France), du SOAS (School of Oriental and African Studies, au Royaume-Uni), du NARLC (National African Languages Resource Center, aux États-Unis), du CIIL (Central Institute of Indian Languages, en Inde), etc.,
- Des "foreign/modern languages centers" qui permettent aux étudiants de quelque niveau que ce soit de se familiariser avec une langue sans que celle-ci fasse formellement partie de leur cursus universitaire.

Il existe environ 20.000 établissements universitaires dans le monde, il n'est pas question de les examiner de manière exhaustive.

Nous avons sélectionné environ 1000 universités les plus représentatives dans 172 pays et répertorié 291 langues enseignées au niveau du premier cycle universitaire (licence/maitrise).

Traitement des données

Le domaine de variation des valeurs brutes de nos données sont très différents, le nombre de locuteurs d'une langue varie entre 900 millions et quelques individus, l'indice de fécondité d'environ 1.2 à environ 7, alors que l'indice de développement humain, par construction varie entre 0 et 1. Pour donner la même importance à tous les facteurs nous utilisons deux procédures.

Lorsque le rapport entre les valeurs les plus élevées et les valeurs les plus faibles est supérieur à 100 nous utilisons le logarithme de la valeur. C'est le cas du nombre de locuteurs.

D'autre part pour, nous « normons » les valeurs brutes en procédant à une transformation linéaire suivant la formule :

$$\text{Valeur_normée} = (\text{Valeur}_{\text{brute}} - \text{Valeur}_{\text{Bruteminimale}}) / (\text{Valeur}_{\text{Brutemaximale}} - \text{Valeur}_{\text{Bruteminimale}})$$

Cette transformation affecte la valeur normée 1 à la valeur brute maximale du facteur, la valeur normée 0 à la valeur brute minimale et des valeurs intermédiaires réparties de façon linéaire pour les autres valeurs.

Le résultat est que tous les facteurs varient entre 0 et 1 ce qui permet de leur affecter un poids « égal ».

Utilisation d'autres facteurs

Pour établir un classement des langues, on peut évidemment utiliser bien d'autres facteurs et chacun peut les imaginer. A titre d'exemple, nous pouvons citer :

- La langue est-elle officielle dans un organisme international ?
- Nombre de publications scientifiques dans la langue
- Nombre de prix Nobel scientifiques dont les auteurs s'expriment dans la langue
- Taux d'alphabétisation des pays dans lesquels la langue est parlée
- La langue est-elle utilisée dans les moteurs de recherche sur Internet ?
- Existe-t-il une suite bureautique dans la langue ?
- Classement des entreprises des pays dans lesquels la langue est parlée
- Taux de pénétration de la téléphonie fixe et mobile dans les pays dans lesquels la langue est parlée
- Nombre d'étudiants étudiant la langue comme langue étrangère
- Nombre de pays dans lesquels il est possible d'étudier la langue comme langue étrangère
- Taux de croissance du PIB
- Taux de possession d'un ordinateur personnel dans les pays dans lesquels la langue est parlée
- Production et/ou exportation cinématographiques dans la langue considérée
- Et bien d'autres.

Un des problèmes dans le traitement de nos données est que toutes les lignes et colonnes doivent être remplies. En cas de données manquantes, nous avons quatre solutions :

1. Supprimer la ligne (donc une langue)
2. Supprimer la colonne (donc un paramètre)

3. Affecter la valeur zéro à la langue concernée pour ce paramètre
4. Affecter à la valeur manquante la valeur moyenne des valeurs présentes dans la colonne
- La solution 1 est acceptable, mais si elle est employée trop souvent nous nous retrouverons avec un nombre de langues très restreint, c'est-à-dire les grandes langues des grands pays du monde, quel que soit le sens que nous donnons au mot grand.

La solution 2 n'est évidemment pas envisageable, nous souhaitons ici ajouter des paramètres et non en supprimer.

La solution 3 peut être acceptée. Par exemple, affecter la valeur nulle au nombre de publications scientifiques dans une langue mineure de Papouasie ou du Cameroun n'est bien sûr pas une cause d'erreur grave.

Nous pouvons employer la quatrième solution si nous sommes certains que la valeur cherchée n'est pas nulle, mais que nous sommes incapables de la trouver dans les bases de données. Nous pouvons affecter la valeur moyenne des langues « comparables » pour lesquelles le paramètre est disponible. Nous ne sommes pas très loin du choix arbitraire de la valeur. Nous sommes dans un constat d'échec. Il est donc clair que nous souhaitons des données complètes, homogènes dans la source et la méthode de calcul et si possible dans le temps.

Dans certains cas, les données sont parfaitement disponibles et introduire un nouveau paramètre serait facile.

Dans d'autres, nous n'avons pas été capables de localiser une source convenable, ce qui ne signifie pas qu'une telle source n'existe pas.

Il se pose alors un nouveau problème, celui de l'indépendance des données. En effet, on observe par exemple que la fécondité est corrélée négativement avec l'Indice de développement humain, le coefficient de corrélation étant égal à -0,87. De même, la valeur enseignement universitaire est très corrélée avec les valeurs statut (0,846) et prix littéraires (0,837). Il existe bien d'autres exemples de corrélations entre les divers paramètres, il est ici inutile de les citer tous. Pour parler clair, l'information que nous traitons est redondante, cela n'apparaît pas de manière évidente, mais une analyse statistique des données met le phénomène en évidence. L'ajout de nouveaux facteurs augmentera cet aspect redondance de l'information. Il existe une méthode mathématique simple pour traiter ce problème : l'analyse en composantes principales (ACP). Elle consiste à calculer de nouveaux facteurs virtuels dont aucun n'est corrélé avec les autres. Le résultat est que le pourcentage d'information fourni par les composantes principales est concentré sur un nombre restreint de facteurs. Dans notre cas, les douze paramètres peuvent être réduits à six paramètres totalement indépendants les uns des autres et qui retiennent 97 % de l'information comme on peut le voir dans le tableau ci-dessous :

Composante principale	Valeur propre	Pourcentage	Pourcentage cumulé
1	6,0806	50,672	50,672
2	2,0553	17,127	60,799
3	1,0091	8,409	76,208
4	0,8826	7,355	83,563
5	0,6448	5,374	88,936
6	0,5332	4,443	93,380
7	0,2428	2,023	95,403
8	0,2050	1,708	97,111
9	0,1477	1,231	98,342
10	0,0887	0,739	99,081
11	0,0554	0,462	99,543
12	0,0549	0,457	10,000

L'avantage est une simplification du problème, l'inconvénient est que l'analyse de la signification des composantes principales qui sont des combinaisons virtuelles des paramètres originaux réclame un effort de réflexion supplémentaire. Ceci constitue une voie de recherche pour l'avenir.

Faites votre propre classement

Nos onze facteurs sont considérés dans notre traitement comme équivalents, c'est-à-dire qu'ils sont affectés du même coefficient. Mais le visiteur de ce site pourra utiliser notre travail pour faire son propre classement. Il dispose pour cela de deux possibilités : soit changer le coefficient de certains facteurs, leur donner plus ou moins d'importance, à l'aide des curseurs, soit supprimer certains facteurs, en mettant le curseur à zéro. Ce baromètre des langues constitue en effet une aide à la décision en matière de politique linguistique, et l'on peut dans ce cadre avoir besoin d'un classement fondé sur une partie seulement des facteurs que nous utilisons. Imaginons par exemple que les concepteurs d'un nouveau logiciel doivent choisir un nombre limité de langues dans lesquelles le rédiger. Ils peuvent sélectionner trois facteurs (nombre de locuteurs, nombre d'articles dans Wikipédia, pénétration d'Internet) particulièrement pertinents pour leur propos et mettre les autres à zéro. Imaginons maintenant que, pour lancer une radio ou une chaîne d'information télévisée internationale des entrepreneurs se demandent en quelles langues émettre pour avoir la diffusion la plus large. Ils peuvent sélectionner trois facteurs (nombre de locuteurs, statut officiel, entropie) et mettre les autres à zéro. Un tel choix assure un grand nombre de locuteurs potentiels, des locuteurs ayant un niveau de vie suffisant pour posséder un récepteur de radio, et la probabilité que la langue d'émission soit reconnue et donc enseignée à l'étranger. Le résultat de ces opérations donnerait pour les dix premières langues les classements suivants :

Classement « logiciel »			Classement « Radio/télé »		
Rang	Langue	Score	Rang	Langue	Score
1	anglais	2.801	1	espagnol	2.652
2	japonais	2.587	2	anglais	2.344
3	allemand	2.511	3	français	1.952
4	espagnol	2.474	4	russe	1.547
5	français	2.470	5	portugais	1.542
6	mandarin	2.420	6	allemand	1.455
7	néerlandais	2.406	7	arabe levantin du Nord	1.435
80	portugais	2.399	80	mandarin	1.433
9	suédois	2.383	9	serbe	1.423
10	russe	2.375	10	coréen	1.395

Bien entendu, le choix des facteurs reste de la responsabilité de l'utilisateur et notre propos n'est ici que de donner deux exemples pour faire comprendre la flexibilité et les possibles utilisations du baromètre.