

La notion de googol et le problème de l'origine des basques

RESUME

L'examen du trésor lexical de l'indo-européen, pour ne citer que ce groupe linguistique, met en relief le fait qu'un certain nombre de concepts sont rebelles à l'innovation lexicale et corrélativement à l'exogénisme: ainsi ceux afférents à la nuit, le nez, les oreilles, la guêpe, la laine, etc.

Les noms de nombre participent de la même tendance et témoignent de la même résistance à la mutation lexicale, mais non, au-delà du googol, à l'exogénisme.

Pas plus qu'une autre langue, le basque n'échappe à cette double règle. L'observation de l'application de la première de ces deux règles confirme l'apparement originel du basque avec les langues caucasiennes (ou japhétiques) tandis que la seconde règle explique les influences chamitiques subséquentes. Ainsi se trouve levée l'apparente contradiction qui oppose tenants des affinités chamitiques et tenants des affinités caucasiennes.

Loin d'être contradictoires, les deux systèmes s'avèrent ainsi solidaires et complémentaires. Historiquement il est probable que lors de leur scission Basques et Caucasiens ne disposaient que d'un googol extrêmement bas; il est par conséquent normal que leurs numérations respectives diffèrent et aient pu faire, pour les premiers, l'objet d'emprunts aux langues chamitiques. Emprunts qui rendent probable un périple méditerranéen vers l'époque protohistorique.

* * *

Certains concepts s'avèrent particulièrement rebelles à l'innovation lexicale. Les mots qui les désignent présentent les caractères suivants:

1.° les appellations sont stables et ne sont assujetties qu'aux seules différenciations attribuables à l'érosion phonétique - laquelle nous apparaît dès lors comme l'unique facteur érosif qui atteint sans exception tous les vocables;

- 2.° la situation est très proche de celle d'un état idéal de la langue où il n'y aurait qu'un signifiant par signifié;
- 3.° d'où la rareté des faits de synonymie ou de polyonymie;
- 4.° la polysémie elle-même, sauf emplois métaphoriques, est rare;
- 5.° la résistance à l'exogénisme est remarquable, même à l'intérieur des langues ou des dialectes non protégés par l'enseignement, l'emploi officiel et obligatoire, le régime politique, etc.

Tel est le cas, par exemple, des appellations européennes de la neige ou de la nuit.

Les appellations it. *notte*, port. *noite*, esp. *noche*, cat. *nit*, roum. *noapte*, angl. *night*, all. *Nacht*, norv. *natt*, néerl. *nacht*. pol. *noc*, serbo-cr. *notch*, russ. *nótchy*, gr. mod. *nux*, finn. *yö*, hongr. *éjjel*, turc. *gece*, basq. *gau*, etc., n'exigent pas un gros effort d'attention pour que l'on observe que:

- 1.° tous les termes procédant de la même souche indo-européenne ne sont différenciés que par la phonétique;
- 2.° dans chacune des langues considérées la nuit ne porte guère qu'un nom et s'il y a polysémie, les sens secondaires ne sont jamais que métaphoriques (obscurité, ténèbres, etc.);
- 3.° lat. *nocte* est prolongé dans toute la Romania, roumain y compris; les langues slaves et les langues germaniques ne sont pas plus différenciées (autant d'indices de stabilité);
- 4.° à la prononciation près, le grec moderne calque le grec homérique (même remarque);
- 5.° quant aux langues agglutinantes, même les moins protégées statutairement (comme le basque) ont résisté aux influences étrangères.

Nous appellerons fixisme primaire celui dont on est assuré qu'il remonte aux temps où les Indo-européens parlaient une même langue (par exemple appellations du nombre 100). Fixisme secondaire celui dont les origines se situent vers l'époque protohistorique ou au début de l'histoire: c'est à ce niveau que se sont fixées des appellations telles que celle du chat domestique ou celle du nombre 1.000 (dans cette dernière option, l'on observe bien une fixation du matériel lexical, mais sans la belle unité observable quant aux appellations, bien plus anciennes, de la nuit). Nous appellerons enfin fixisme tertiaire celui qui n'est pas antérieur aux temps modernes: par exemple les appellations, en Europe, du café, du garage ou du million.

LE MYSTÈRE DU FIXISME LEXICAL

Les raisons du fixisme lexical sont loin d'être toujours claires. Pourquoi la stabilité, primaire ou secondaire, des vocables référant à des concepts aussi hétéroclites que le sel, le lait, les oeufs, le nom, les étoiles, le miel, le chanvre, la barbe, le nez, le coeur, la dent ou la veuve? Pourquoi le nom du chat est-il stable alors que celui du coq ne l'est pas? Confessons n'avoir trouvé à ce problème encore aucune explication vraiment satisfaisante¹.

Il y a pourtant un cas au moins où des raisons paraissent plausibles: celui des oronymes, hydronymes, etc., d'une part²; des noms de nombre d'autre part. Termes qui présentent ce caractère commun de pouvoir être considérés comme des signes distinctifs à l'intérieur d'un ensemble homogène susceptible d'être dissocié sans que cette dissociation entraîne automatiquement une hétérogénéisation des composants³. A cet égard le nom de nombre est conceptuellement très proche du nom propre (auquel il sert d'ailleurs de nos jours couramment de substitut sous la forme du numéro-matricule)⁴. En matière de noms de nombre la polysémie est évidemment inconcevable: l'on n'imagine pas que le même vocable puisse signifier à la fois 3 et 7. La polyonymie est presque aussi rare et, en tout état de cause, chose instable: en germanique v. h. a. *zëhanzo*, littéralement *dix fois dix*, n'a pas survécu longtemps à la concurrence de *hundert*; en basque *amarreum* (litt. *dix cents*), synonyme parfait de *mila*, n'est guère utilisé; en France *quatre-vingts* a éliminé *nonante* (c'est l'inverse en Belgique) et *milliard* a pratiquement évincé *billion*, etc.

Les appellations des ordinaux témoignent de la même tendance: en espagnol *ochavo*, synonyme parfait de *octavo*, est devenu rare et *noveno*, synonyme parfait de *nono*, obsolète.

1 Tout au plus peut-on envisager quelques explications de caractère négatif. Il est probable, par exemple, que l'appellation de la nuit n'a fait en indo-européen l'objet d'aucun tabou lexical. Tandis qu'à Tahiti, où le nom des souverains est taboué, le vocable indigène *po* (= nuit) a été remplacé par *mi* à l'avènement de la reine Pomaré.

2 En matière d'onomastique, le confusionnisme est pour ainsi dire toujours le fait d'une circonstance particulière. Voici un cas extrême: lorsque les missionnaires espagnols baptisèrent en masse les Aztèques, il s'avéra que les hagionymes disponibles n'étaient pas en nombre suffisant. Les Indiens furent alors répartis en groupes, dont chacun reçut collectivement le même prénom. Ladislas Jagellon, duc de Lithuanie, ne procéda pas autrement lorsqu'en 1387 il fit baptiser d'un seul coup tous ses sujets.

3 Cf. H. POLGE, *La notion d'ensemble appliquée aux disciplines humaines*, Auch, Société archéologique du Gers, 1971, p. 60.

4 Il est des langues, comme l'espagnol, qui ne distinguent pas le nombre du numéro: c'est l'inverse en allemand (*Zahl*, *Nummer*). Le numéro réfère à l'emploi d'un nombre cardinal là où une logique parfaite exigerait l'emploi d'un ordinal. C'est pourquoi le vocable *numéro* n'est jamais suivi d'un ordinal.

En ces domaines particuliers, la langue a donc atteint cet état idéal, auquel nous faisons allusion un peu plus haut, en vertu duquel il y a un signifiant, et un seul, par signifié. D'où sans doute la résistance qu'elle offre en pareil cas à l'innovation lexicale, surtout lorsque celle-ci peut être source de confusion.

Une preuve formelle de cette résistance à la confusion nous est fournie par l'aréologie européenne des termes signifiant million et milliard. L'appellation de forme *million* est attestée en italien (*milione*), en espagnol (*millón*), en catalán (*milió*), en roumain (*milion*), en anglais (*million*), en allemand (*million*), en norvégien (*million*), en suédois (*miljon*), en néerlandais (*miljoen*), en polonais, en serbe et en bulgare (*milion*), en russe (*millión*), en finnois (*milliö*), en turc (*milyon*), etc., *mais non en grec*, où le terme se serait rencontré avec *milion* qui signifie mille. De même le thème *milliard* est attesté en italien (*miliardo*), en roumain (*miliard*), en anglais (*milliard*), en allemand (*Milliarde*), en norvégien (*milliard*), en suédois et en néerlandais (*miljard*), en polonais et en bulgare (*miliard*), en serbe (*milijarda*), en russe et en hongrois (*milliárd*), en turc (*milyar*), etc., *mais non en espagnol*, où le terme se serait croisé avec *millar* qui signifie millier. En portugais, où *milhar* signifie aussi millier, le problème a été résolu par le recours à *bilião*, milliard. Ce qui prouve que lorsque le croisement phonétique de deux vocables est véritablement amphibologisant et que l'ambiguïté est effectivement intolérable, la langue se protège préventivement et non *a posteriori*, thérapeutiquement, comme l'ont cru certains tenants de la doctrine de Gilliéron.

Quoi qu'il en soit de ce dernier problème, sur lequel nous avons eu l'occasion de nous pencher déjà⁵, il reste acquis que des noms du type million ou du type milliard peuvent être véhiculés loin de leur point de départ (épicycle) et adoptés par des langues de souches extrêmement variées. Quid de cette catégorie spéciale de mots voyageurs? C'est le problème que nous allons examiner à la lueur de la notion de googol.

LA NOTION DE GOOGOL ET SES IMPLICATIONS LINGUISTIQUES

Le googol est le nom de nombre le plus élevé portant une appellation dans la langue vernaculaire⁶. Il constitue un moyen particulièrement com-

5 Cf. H. POLGE, *Les appellations basques et gasconnes du coq et du chat dans Fontes linguae Vasconum*, 1970, p. 141.

6 Des termes tels que *cent* ou *mille* permettent d'éviter une combinaison [multiplicande-multiplicateur] où le multiplicande est égal ou supérieur au multiplicateur, ainsi dans les expressions imaginaires *dix-dix* ou *onze-dix*.

mode de mesurer le niveau intellectuel d'une civilisation. Nul chez les Veddah, il este de 2 chez les Australiens et les Abipones; de 3 chez les Fuégiens et les Amérindiens de Colombie; de 4 chez certains Malais, Polynésiens et Néocalédoniens; de 5 chez certains Esquimaux, etc. Les peuples qui ne savent pas encore pratiquer la dactylogonomie, usent souvent du duel, du triel ou du quadriiel, et ignorent la notion, abstraite de pluriel. Lorsque les progrès de la numération sont suffisants pour qu'on puisse parler de base, la base n'est pas toujours 5, 10 ou 20, comme on aurait tendance à le croire mais 4 là où le pouce n'est pas assimilé aux autres doigts: ainsi chez quelques Thraces de l'Antiquité ou chez les Chimilas de la Colombie septentrionale (encore aujourd'hui en Grande-Bretagne le majeur porte parfois le nom de *second doigt*).

Chez certains peuples les progrès de la numération sont a la fois spontanés et endogènes. Il arrive un moment où au lieu de montrer la main ou de dire «main», l'on dote d'un nom le chiffre 5. Mais cet effort de création lexicale est inutile lorsqu'un peuple en retard se trouve en contact avec un peuple plus élevé: *alors les appellations des nombres supérieurs au googol sont empruntées à une langue voisine.*

De cette règle nous connaissons plusieurs exemples. Chez les Conibos, dont le googol est de 2 seulement, l'on a recours à la langue quichua pour exprimer les notions de 3, 4, 5, etc. De même chez les Amérindiens de la tribu Cocana, à cette nuance près que leur googol atteint le niveau 4. Au Brésil les indigènes de la province du Para ont recours au portugais lorsqu'ils comptent au-delà de 300. En malais le nombre le plus élevé usité dans les transactions est de 10.000, chiffre qu'on exprime par le mot *laxa*, qui est d'origine hindoue, etc. *Il n'y a aucune raison pour penser qu'il en ait été autrement dans l'Europe préhistorique.*

La numération parlée n'est parfaitement logique que dans des langues spontanées comme le chinois ou dans des langues artificielles comme l'esperanto, qui avec douze mots seulement parviennent à exprimer les mille premiers nombres sous la forme un, deux, trois, quatre, cinq, six, sept, huit, neuf, dix; puis dix-un, dix-deux, etc.; puis deux-dix, trois-dix, etc., jusqu'à cent; puis deux cents, trois-cents, quatre-cents, etc., jusqu'à mille. Si cette logique était prolongeable, il y aurait des mots spéciaux pour dire dix-mille (*wan* en chinois), cent-mille, etc., ce qui est le cas en sanscrit (*ayuta*=10.000; *laksa*=100.000; *prayuta*=1.000.000; *koti*=10.000.000).

De ces considérations, il est loisible d'induire l'hypothèse que des termes du type *dix-dix* (ou *cinq-vingt* dans un système vicésimal) ont dû précéder dans le temps l'élaboration de la notion de cent, hypothèse confirmée par v. h. a. *zëhanzo*, postérieurement évincé, comme nous l'avons vu par *hundert*. De même des expressions du type *dix-cents* ont dû précéder l'explicitation

lexicale de la notion de mille. Aujourd'hui encore d'ailleurs, en français comme en allemand (mais non en espagnol), le millésime est souvent exprimé sans qu'intervienne la notion de mille, dans des expressions telles que dix-neuf-cent-trente-trois (all. *neunzehn-hundert-drei-und-dreizig*).

Avant leur dispersion, les Aryens avaient certainement élaboré la notion de cent et l'apparente diversité des appellations européennes actuelles tient surtout au fait que le vieux slave a utilisé des formes fléchies (v. sl. *suto* > russ. mod. *sto*) tandis que lat. *centum*, germ. *hund*, gr. *hekatón*, etc., devenaient invariables. L'on peut donc considérer *cent* comme le googol des Indo-européens.

Par contre la notion de mille n'a dû être exprimée qu'une fois amorcée la dispersion des Aryens. Elle s'exprime en effet sous des formes aussi irréductibles entre elles qu'all. *tausend* et fr. *mille*. Sans doute les Slaves ont-ils emprunté à leurs voisins le nom de nombre qui chez eux exprime mille, car en russe comme en polonais l'apparement avec le germanique est évident-tandis que l'apparement avec le grec est évident chez les Bulgares et chez les Serbo-croates.

Vers la fin du Moyen âge, l'Italie, qui dispose du googol alors le plus élevé d'Europe, servira d'épicentre aux appellations du million, puis du milliard. En d'autres termes ce qui est vrai des Conibos ou des Cocanas l'est aussi, *mutatis mutandis*, du monde européen: les peuples arithmétiquement les plus en avance fournissent à leurs voisins plus attardés les appellations des noms de nombre supérieures à leurs googols.

Quelques anomalies sont cependant perceptibles, dont il n'est sans doute pas impossible de rendre compte: par exemple gr. mod. *milion*, mille en remplacement de gr. class. *chilioi*. Le googol, lorsqu'il est élevé, ne doit pas être une notion très populaire, donc bien stabilisée: combien de français, en 1972, seraient capables de définir le trillion ou le nonillion et de lire correctement et sans hésiter un nombre comportant par exemple 16 chiffres? Dans une telle perspective, le googol n'ayant qu'une existence quasi théorique, peut faire l'objet de substitutions par exogénisme.

GOOGOL ET INNOVATION LEXICALE

Une apparente entorse à la règle en vertu de laquelle la numération est réfractaire au flottement sémantique et à l'instabilité lexicale est observable dans l'Europe moderne. Elle est afférente à la lecture des très grands nombres, lesquels ne sont guère usités qu'au sein du monde savant.

Anciennement en France, *pour exprimer les grands nombres*, l'on séparait les chiffres en tranches de six à partir de la droite: le nombre qui

occupait la deuxième tranche exprimait un nombre de millions; celui qui occupait la troisième tranche un nombre de billions et ainsi de suite jusqu'à nonillion en passant par les trillions, quadrillions, quintillions, sextillions, septillions et octillions. Autrement dit la loi de formation du nom de nombre, à la N e tranche avait pour forme $(N-1)$ illions (formule où N est le nombre de tranches complètes ou incomplètes). Il en fut ainsi jusqu'au XVII^e siècle: à cette époque se répand l'usage de procéder par tranches de trois (et non plus de six) chiffres⁷. La loi de formation du nom de nombre élevé revêt alors la forme $(N-2)$ illions, le nombre qui occupe la 4^e tranche est celui des billions; la 5^e tranche celui des trillions, etc. Ce procédé est aussi celui de l'Italie, de l'Espagne, des Stats-Unis, etc. Mais des pays comme le Danemark, l'Allemagne, l'Angleterre, etc., sont restés fidèles aux tranches de six chiffres. C'est pourquoi la règle ancienne est souvent qualifiée de règle germanique et le règle moderne de règle latine.

Le terme de milliard n'a théoriquement droit de cité ni dans l'une ni dans l'autre des deux règles. Pourtant dans de nombreux pays il est utilisé pour désigner mille millions et se trouve par conséquent synonymisé, *dans la règle latine*, avec billion.

Autre conséquence de la divergence des règles; angl. *quadrillion* correspond à un nombre un milliard de fois plus grand que son homologue fr. *quadrillion*. En Allemagne un *billion* est un million de millions; un *trillion*, un million de billions, etc.

APPLICATION AUX DIALECTES

Il est piquant de constater que les seuls flottements lexicaux que l'on observe en matière de numération sont le fait non des milieux populaires, mais du monde savant. Telle est l'une des raisons pour lesquelles ils n'affectent pas le monde des dialectes.

Le *Dictionnaire béarnais* de S. Palay donne bien *trilhoù* au sens de trillion, mais il est douteux que ce mot soit communément employé. La numération gasconne étant entièrement de souche latine, il est difficile d'évoquer à son sujet les incidences lexicales de la notion de googol. Par contre, grâce au basque qui use du latin à partir de mille (*mila*), il est permis de penser que les Aquitains, antérieurement à leurs premiers contacts avec les Romains, ne devaient pas pouvoir exprimer verbalement un nombre supérieur à 999 autrement qu'en disant 11×100 , 12×100 , etc.

⁷ Sans doute en application des phénomènes de vision qu'a mis en valeur la psychologie de la forme (*Gestalttheorie*).

Pour des raisons probablement identiques, lat. *mille*, est passé à l'irlandais (*mile*) et au brittonique (*mil*).

Autre observation: les dialectes, même ceux qui lexicalement sont fort diversifiés, reproduisent en petit les phénomènes observables à grandes échelle, par exemple à l'échelle du continent européen. A quelques nuances phonétiques près, les vocables basques désignant la neige, la nuit ou les nombres relèvent de la *koinè*.

CONCLUSION

En résumé:

1.° Les noms de nombre s'inscrivent au rang des vocables particulièrement rebelles à l'innovation lexicale (non à l'emprunt), vocables qui, mis à part le domaine apparenté de l'oronymie, de l'hydronymie, de la macrotoponymie, etc, sont relativement peu nombreux (surtout au niveau du fixisme primaire) et ne sont différenciés que par la seule érosion phonétique;

2.° Ils présentent sensiblement les mêmes caractéristiques que ces vocables, à savoir extrême rareté de la synonymie et très grande instabilité, le cas échéant, de la polyonymie: d'une manière générale l'on observe qu'il y a un signifiant, et un seul, par signifié;

3.° Lorsque d'aventure, du fait d'un télescopage phonétique, il y a risque de confusion, la langue y remédie *ab ovo* et non thérapeutiquement autrement dit *a posteriori*⁸;

4.° Lorsque deux civilisations d'inégal niveau intellectuel se trouvent en contact, les noms de nombre supérieurs au googol de la civilisation la plus attardée font l'objet d'emprunts à la nomenclature numérique de la civilisation voisine plus évoluée. Ce qui explique la facilité avec laquelle les noms afférents à des nombres élevés se muent en mots voyageurs⁹. Ce qui explique aussi qu'à l'intérieur d'une langue les noms de nombre peuvent très bien procéder de souches diverses (ainsi les appellations russes de 1000, d'origine germanique, et du million, d'origine romane);

8 Les journaux espagnols contemporains, lorsqu'ils évoquent les mouvements français dits de contestation, n'usent point de l'équivalent phonétique castillan *contestación*, qui signifie généralement réponse, mais de la forme française telle quelle, mise entre guillemets, "contestation". Ainsi l'ambiguïté est palliée *ab ovo*. Mais il n'est pas exclu que la polysémie française passe un jour ou l'autre à l'Espagne.

9 Un exemple caractéristique de ce phénomène se situe chez les Patagons, lesquels parlent trois dialectes (tehuel-che, araucan et pampéen) tellement différenciés que l'intercompréhension est nulle. Nulle sauf en ce qui concerne la numération, qui est commune aux trois langues et empruntée au quichua.

5.° Dans une langue quelconque, l'intrusion de noms de nombre exogènes est chose banale au-delà du googol, mais, une fois définie, la nomenclature numérique offre à l'innovation une résistance considérables, au seul phonétisme près;

6.° Plus un nom de nombre se situe bas, plus il a de chances d'appartenir à la langue vernaculaire, plus il est proche de sa souche populaire et plus il est exposé, par conséquent, à l'érosion phonétique. Inversement plus il est élevé, plus il a de chances d'affecter la forme d'un cultisme¹⁰;

7.° Une septième conclusion s'impose, la plus importante de toutes du point de vue propre au problème posé par notre titre: entre les concepts réfractaires à l'innovation lexicale d'une manière générale et les noms de nombre en particulier, il existe une différence fondamentale: les premiers (comme la nuit) sont rebelles à l'exogénisme, tandis que les seconds ne le sont pas (basq. *mila*) et ne se fixent, le cas échéant, que postérieurement à leur intégration. Aux uns ne correspondent que rarement des mots voyageurs tandis que le phénomène est banal, pour les autres, au-delà du googol.

* * *

Les linguistes qui se sont penchés sur le problème des apparentements du basque avec d'autres langues n'ont pas été sans être embarrassés par le fait que des affinités lexicales de caractère général sont constatables entre le basque et les langues caucasiques, mais sans apparentement quant à la numération, tandis que les affinités sont plus probantes avec les langues chamitiques en ce qui concerne plus spécialement la numération.

Or les concepts généraux rebelles à l'innovation lexicale qui associent basque et langues caucasiques sont sensiblement les mêmes en basco-caucasique qu'en indo-européen: ils concernent par exemple la nuit, la dent, la barbe, le nez, etc: convergence difficilement attribuable au hasard et qui tient probablement à un parallélisme des démarches de l'esprit.

Par ailleurs si l'on admet que lors de leur dissociation, Basques et Caucasiens ne disposaient que d'un googol extrêmement bas, l'on ne s'étonnera pas que les uns comme les autres aient, au cours de leurs migrations subséquentes, emprunté à des voisins arithmétiquement plus évolués leurs numérations respectives. Remarque qui explique les différences de numération en-

10 Ceci étant vrai aussi des ordinaux. Ainsi en espagnol où les seuls ordinaux vraiment usités ne rendent que les notions de premier à dixième, de vingtième, de centième et de millième. Hormis ces cas, l'on a recours au numéro, dont nous avons dit (cf. ci-dessus note 4) qu'il est un cardinal illogiquement substitué à un ordinal.

Les formes populaires ou semi-populaires du type *cuaresma*, *diezmo* ou *siesta* ont revêtu des sens particuliers.

H. POLGE

tre basque et langues caucasiques et surtout *qui lève la contradiction entre tenants des affinités japhétiques et tenants des affinités chamitiques*. Dès lors les deux thèses ne sont plus opposées, mais complémentaires. Surtout si l'on admet que les Basques ont atteint la péninsule ibérique en migrant par la Méditerranée.

Cette dernière hypothèse présente au surplus l'avantage de cadrer avec les nombreux apparentements que nous avons eu l'occasion de souligner déjà, en matière de techniques, entre Basques et peuples méditerranéens¹¹. Car la numération n'est au fond qu'une technique qu'on est toujours tenté de copier ou d'adopter, comme toutes les autres techniques, lorsqu'on se trouve en contact avec un monde plus avancé.

Enfin notre reconstitution rendrait parallèlement compte des curieuses ressemblances plusieurs fois signalées déjà entre groupes sanguins basques et groupes sanguins de certaines ethnies insulaires du Bassin méditerranéen (Sardaigne), voire peut-être de certaines convergences phonétiques.

H. POLGE

¹¹ V. p. ex. H. POLGE, *Les appellations gasconnes et basques des céréales les plus usuelles* dans *Fontes linguae Vasconum*, numéro 3, pp. 303-317.