Jean Rousseau

COMPARAISON DES LANGUES ET INTERCOMPREHENSION

BELC - Bureau d'Etudes pour les Langues et les Cultures

CENTRE INTERNATIONAL D'ETUDES PEDAGOGIQUES

Jean Rousseau

COMPARAISON DES LANGUES ET INTERCOMPREHENSION

BELC - Bureau d'Etudes pour les Langues et les Cultures

CENTRE INTERNATIONAL
D'ETUDES PEDAGOGIQUES

© CIEP, Sèvres. Juillet 1993.

PROLOGUE: LE RUSSE EN HUIT SEMAINES AU DIX-HUITIEME SIECLE?

En 1761, un jeune allemand de vingt-six ans débarquait à Saint-Pétersbourg chez un autre allemand émigré, G.F. Müller, historien officiel de l'Empire Russe et académicien, afin de lui servir d'assistant dans ses travaux d'érudition et de tuteur pour ses enfants. Aussitôt installé dans cette maison où l'on parlait couramment cinq langues[1], cet étudiant, August Ludwig Schlözer, qui s'était formé à la théologie et aux langues orientales à l'Université de Göttingen, avant de passer quelques années en Suède, se mit à apprendre seul la langue russe dont jusque-là il ignorait tout, et, huit semaines après son arrivée, fut en mesure de présenter à son employeur stupéfait la traduction d'un premier oukase. Or un document permet d'éclairer cette prouesse, puisque, dans un court récit autobiographique[2], Schlözer expose lui-même la méthode adoptée et les principes qui l'ont inspirée. Il est déjà d'un intérêt certain d'entrer ainsi dans les secrets de la "boîte noire" d'un brillant intellectuel du siècle des Lumières, mais plus encore quand il s'agit, pour méconnu qu'il soit aujourd'hui, d'un linguiste de premier plan[3], d'un comparatiste hors de pair, qui finira professeur d'Histoire et de Statistique à Göttingen, célèbre dans l'Europe entière.

En l'absence de toute grammaire ou manuel destiné à l'apprentissage de la langue russe par les étrangers, Schlözer raconte qu'il s'est muni d'un dictionnaire de 781 pp. qui faisait figurer tous les mots russes sous le radical avec leur sens en latin, les racines étant en gras et les dérivés ou composés en petits caractères. Son objectif, précise-t-il, était seulement de comprendre la langue écrite, non de l'écrire, et moins encore de la parler. Mais il souligne qu'il avait auparavant abordé "de manière grammaticale" près de 15 langues[4]. D'après son récit, ces connaissances préalables semblent avoir conditionné son apprentissage de trois manières.

---

1. A savoir allemand, russe, suédois, finnois et français.

2. *August Ludwig Schlözer's öffentliches und privat-Leben, von ihm selbst geschrieben. Erstes fragment Aufenthalt und Dienste in Russland, vom J. 1761 bis 1765. LiterarNachrichten von Russland in jenen Jaren.* Göttingen 1802.

3. Toutes les citations de Schlözer seront tirées des pages 39 à 48 de l'ouvrage mentionné à la note précédente.

4. Il ne précise pas lesquelles, mais dit, par plaisanterie, qu'il faut pour arriver à ce nombre inclure le Plattdeutsch et le Chaldéen. Selon toute vraisemblance, d'après ses études, ses travaux, ses intérêts, ses voyages, et ses exemples, il faut ajouter à allemand, grec, latin, hébreu, et arabe, celles-ci: français, anglais, hollandais, espagnol, italien, suédois et finnois.

D'abord, elles ont créé une sorte d'horizon d'attente: "plus on comprend de langues, plus on en acquiert facilement de nouvelles: leurs singularités n'étonnent plus, quand on les a déjà trouvées ailleurs". Elles l'ont ensuite rendu attentif à la présence en toute langue d'une organisation, d'un système, dont la maîtrise est nécessaire. Schlözer, qui dit avoir appris dès l'âge de 11 ans l'hébreu qui lui avait "empli la tête de racines"[5], déclare avoir eu, par la suite, l'idée "que les mêmes généalogies de mots pouvaient très bien se produire aussi en d'autres langues, et que les significations les plus diverses en apparence devaient se rencontrer en un seul foyer". Dès lors, il dit avoir toujours été à l'affût des "significations fondamentales" et s'être livré "en toutes les langues, à la chasse aux racines". Les langues sont donc pour lui traversées de rapports unissant forme et significations[6].

Mais, à cette prise de conscience d'une structuration du lexique par combinaison d'éléments récurrents, parcourue en outre par une organisation sémantique à découvrir, Schlözer ajoute encore une dimension parfaitement inédite pour son temps. En effet, et c'est le troisième bénéfice tiré de sa familiarité avec de nombreuses langues, il utilise, au second degré, les rapports du même ordre qui les unissent, exploitant toutes les possibilités qu'offre leur comparaison: "je compris très vite que sur dix mots russes, il y en avait neuf qui se trouvaient aussi ou en allemand, ou en latin, ou en grec, ou bien encore à la fois dans ces trois langues, et leur identité originelle pouvait être prouvée par des règles certaines et non des étymologies fantaisistes". Schlözer ne fait pas là allusion à des transparences évidentes pour tout le monde (même pour Müller qu'il n'avait pas en grande estime intellectuelle), comme le fait, rappelle-t-il, que "*Nacht, natt, nox, notte, nuit, nux, night, notsch* soient un seul et même mot". D'autres exemples montrent en effet qu'il vise plutôt des parentés bien plus lointaines, provenant de relations génétiques impliquant une profondeur temporelle qui les rend difficiles à découvrir, sinon par la médiation de relais que le nombre de langues impliquées autorise néanmoins: il a ainsi appris à Müller "que *viel* [all. "beaucoup"] et *plurimus* [lat. "la plupart"], *finster* [all. "sombre"] et *tenebrae* [lat. "obscurité"], avaient surgi d'une seule racine (les premiers, par les mots intermédiaires

---

5. Cf. infra l'encadré "La racine en Sémitique".

6. Il donne quelques exemples de tels rapports de sens surprenants entre mots de même origine. Pour l'allemand: *fünf* ("cinq") / *Finger* ("doigt"); *alt* ("âgé") / *Eltern* ("parents"); pour le latin: *pecus* ("troupeau") / *pecunia* ("argent"); *pes* ("pied") / *impedio* ("empêcher").

polus [gr. "beaucoup] et le suédois *flere* ["plus"], les autres, par l'ancien allemand *thimster* ["obscur"] et le slavonique *tma* ["obscurité"])", les écarts de forme étant à imputer aux effets dévastateurs du temps. Sa *SprachMethod* est donc, comme il le déclare, une *WurzelMethod*, tournant autour de l'existence de racines, autant dans l'approche structurale que dans celle d'une comparaison des langues historiquement reliées.

Pour aborder le russe, ces deux perspectives furent exploitées systématiquement. Non seulement Schlözer s'efforçait toujours de ramener les "racines" découvertes en russe à leur signification de base, mais il cherchait aussi à trouver la valeur d'emploi commune, le "concept général" de chaque élément de dérivation, soucieux ensuite de démultiplier au maximum ses acquis: la connaissance de 100 racines, dit-il, pouvait lui fournir sans effort 400 dérivés et plus encore de composés. Une forme complexe comme *vseMILostiveiszemu*, "au très clément", était décomposée, puis chacun de ses éléments composants était fixé, soit quant à son mode de formation, soit pour sa forme même, par un rapprochement avec les langues parentes: *MIL* correspondant au *mild* des dialectes allemands, au grec *meil-* (*meilia*, "offrande, ce qui apaise"); la marque de superlatif *-eisz-* à celle du latin, de l'allemand et du grec; celle de datif *-eMu* à l'allemand *deM*, à l'ancien suédois; la préfixation de *use/vse* est analogue à celle du grec *pas*, "tout". Une telle stratégie autant soucieuse des structures que de comparaison trouverait d'ailleurs un prolongement moderne dans l'approche pédagogique du russe que défendait L. Tesnière[7]. Mais l'intérêt que présente pour nous Schlözer ne se borne pas à ces précieuses confidences expliquant la rapidité de ses progrès par un recours massif à la connaissance d'autres langues de même famille, et par la mobilisation de tous les rapprochements fournis par leur comparaison pour faciliter l'apprentissage.

Schlözer est en effet aussi l'auteur de la première grammaire du russe qui tire parti de la parenté de la langue avec allemand, latin et grec[8]. Il a classé en familles plusieurs dizaines de langues ouraliennes et sibériennes de l'empire russe, en s'appuyant sur l'examen de listes de

---

7. Dont témoigne sa *Table étymologique: Les mots russes classés d'après leur racine* (Documents de Linguistique quantitative N° 8) Paris 1970, qui incorpore timidement des rapprochements avec le français. On peut aussi invoquer la perspective, non comparative cette fois, du *Manuel de Russe à l'usage des scientifiques* Paris 1964, de Y. Gentilhomme.

8. Malheureusement restée manuscrite pour d'obscures raisons d'idéologie nationaliste de la part de son grand rival Lomonossov.

vocabulaire, et ce, dans un ouvrage où sont aussi posés, pour comparer les langues, des principes méthodologiques qui seront au fondement de tout travail de cet ordre pour deux générations[9]. De même, il fut à l'initiative du rapprochement entre finnois et hongrois - premier grand succès, avant l'ère du sanscrit, d'une comparaison génétique appuyée sur des preuves grammaticales[10]. C'est encore lui qui créa le terme de "sémitique" pour désigner le groupe des langues génétiquement apparentées que forment hébreu, araméen, arabe etc. Bref, par tout un pan de ses recherches historiques, dans lesquelles il se plaçait sous la double invocation de Linné et de Leibniz, Schlözer fut un des linguistes les plus doués et les plus novateurs de l'époque des Lumières.

C'est précisément pour avoir été à la fois un praticien conscient des bienfaits de la comparaison afin d'accélérer le processus conduisant à la compréhension passive d'une langue, et l'un des premiers acteurs de la recherche comparative, qu'il méritait de figurer, à titre de père fondateur encore peu connu, au seuil d'une esquisse visant à présenter principes et acquis de la comparaison historique et à évoquer quelques-unes de ses applications possibles pour faciliter la compréhension textuelle entre langues génétiquement apparentées.

---

9. C'est l'*Allgemeine Nordische Geschichte* Halle 1771, dont Humboldt fera encore un rare éloge cinquante ans après.

10. Par S. Gyarmathi *Affinitas linguae Hungaricae cum linguae fennicae originis grammatice demonstrata* Göttingen 1799.

PREMIERE PARTIE: LA COMPARAISON GENETIQUE.

DE LA COMPARAISON EN LINGUISTIQUE A LA LINGUISTIQUE COMPAREE.

Une remarque liminaire s'impose. Toute analyse du langage, toute étude d'une langue est d'essence comparative. Au sens où elle implique l'exercice d'une faculté intellectuelle complexe qui se traduit par une série de décisions du type: X est identique à Y; A est différent de B; C est à D comme E est à F, autant d'assertions consistant toujours à trancher du même et de l'autre, selon les divers points de vue sous lesquels les objets examinés sont considérés. Aucune description de quelque composante d'une langue que ce soit - phonétique, morphologique, syntaxique, sémantique - naïve ou consciente, platement empirique ou lourdement théorisée, ne saurait faire l'économie de la procédure comparative. Quel que soit le modèle théorique de référence, si, sur des énoncés, nous ne pouvions effectuer des séries de commutations, substitutions, permutations et transformations, portant sur des éléments de nature variable, en un mot sans pratiquer de comparaisons, nous serions incapables de produire la liste des phonèmes d'une langue, d'inventorier les combinaisons possibles des morphèmes, de poser des règles de syntaxe, de lister des sèmes etc. Identifier des unités et les relations qui les unissent, leurs niveaux de fonctionnement, poser des classes de phénomènes et définir des catégories, en fin de compte, tout travail de réflexion linguistique se fonde sur notre aptitude cognitive à comparer des objets.

On voit que le propre de cette activité comparante est de pouvoir être conduite sans jamais franchir les bornes de la langue ainsi étudiée. Même s'il est vrai que, pour accéder au statut de linguistique générale et voir assise définitivement leur validité, les conclusions acquises par l'examen du fonctionnement d'un système donné doivent être généralisables, et, pour prendre place dans un cadre unifié, être aussi confrontées à celles qu'autorise l'examen dans les mêmes termes d'autres langues. Aussi, pour y parvenir, devra-t-on pratiquer, comme au second degré, une comparaison des résultats des premières comparaisons portant sur les langues individuelles.

Pourtant, malgré cette omniprésence dans la recherche linguistique d'une heuristique fondée sur la comparaison, ce n'est pas la démarche obligée qu'on vient d'esquisser qui est en général désignée quand on fait de nos jours de la "linguistique comparée", ou de "la linguistique

comparative", voire quand on évoque, sans plus de précisions, "la comparaison".

Pour expliquer cet emploi restreint, et le fait que, dans son acception la plus communément reçue, on ne parle aujourd'hui de "comparaison" que si un linguiste prend pour objet d'étude deux langues au moins, avec l'objectif bien précis d'éclairer leur histoire, il convient de repartir des conditions dans lesquelles la linguistique a accédé au statut de véritable science, c'est-à-dire précisément de l'époque des Lumières, quand s'est développée une approche systématique et historique des langues du monde.

UN PEU D'HISTOIRE

Aux alentours de la seconde moitié du dix-huitième siècle, il s'est produit en Europe occidentale une révolution mentale, un changement de paradigme comme on dirait aujourd'hui, dont on commence tout juste à mesurer l'exacte portée.

Jusqu'alors les langues étaient essentiellement abordées soit dans une perspective normative et pédagogique, guidée par un purisme souvent dépendant d'une idéologie de la Nation, soit universaliste et philosophique, qui les traitait comme autant de reflets du fonctionnement immuable de la pensée. En simplifiant à l'extrême, étudier les langues revenait à pencher soit vers Vaugelas, soit vers la Logique de Port-Royal. Mais pour des raisons qui tiennent au développement des conquêtes coloniales ainsi qu'au souci d'apporter aux peuples rencontrés les secours de la religion, un grand nombre de langues furent, après l'ère des Grandes Découvertes, l'objet de descriptions plus ou moins achevées. Des grammaires furent écrites, des lexiques rassemblés, des textes traduits, à l'initiative presque toujours des missionnaires qui, catholiques, souhaitaient confesser, et, protestants, offrir aux indigènes un accès direct à la parole divine.

Longtemps ces données, quand elles parvenaient en Europe, restèrent d'une étrangeté irréductible, faisant seulement figure de curiosité: elles subissaient un sort analogue à celui que les objets de provenance ethnologique connurent jusqu'au milieu du dix-neuvième siècle, qui voisinaient dans des cabinets de curiosités avec divers monstres ou des merveilles de la nature. Il n'y avait, à la lettre, pas de place pour ces

langues, dans la mesure où la profondeur temporelle allouée à l'espèce humaine en fonction d'une création du monde de fraîche date (-4004 av. JC selon certains calculs) et l'obligation de postuler, d'après la Bible, une monogenèse, avec l'hébreu comme langue-mère, s'appliquaient déjà fort mal aux langues d'Europe, sinon au prix d'étymologies forcées. Pourtant, l'accumulation croissante des descriptions linguistiques en provenance du monde entier rendait d'une part intenable, vu leur diversité, de traiter toutes ces langues "exotiques" comme ayant une identité originelle, et, de l'autre, indispensable d'introduire un ordre dans cette profusion chaotique. Certes on pouvait toujours se contenter d'une simple énumération, et répartir les langues selon des critères géographiques, en les illustrant de la traduction du Notre Père, car cette approche fut encore, au début du dix-neuvième siècle, celle d'un recueil de près de 500 langues, paru en Allemagne et intitulé *Mithridate*[11].

Mais, par ailleurs, ceux qui s'intéressaient aux langues subissaient la forte pression qu'exerçait le modèle théorique illustré par les sciences naturelles. Dans ce domaine, zoologistes et botanistes, en particulier le suédois Carl Linné - sans doute le savant le plus célébré du siècle après Newton - avaient réussi à produire une classification exhaustive des plantes et des animaux selon une procédure rationnelle oeuvrant à l'aide de critères simples. C'est d'abord ce modèle scientifique que les professionnels des langues cherchèrent plus ou moins consciemment à rejoindre. Tant par les circonstances de son apparition que par les questions qu'elle se propose de résoudre, la comparaison linguistique a toujours relevé de la pensée classificatoire, a été une entreprise d'essence taxinomique, son horizon de recherche étant initialement de répartir selon des principes de classement opératoires les centaines de langues alors accessibles.

Or, parallèlement, sous l'impulsion notamment de Leibniz, dont on oublie toujours qu'il fut, de profession, historien, on a commencé à traiter les langues les plus diverses comme des outils de la connaissance historique. L'idée s'impose peu à peu au cours du dix-huitième siècle que les problèmes posés par l'origine et les migrations des peuples connus depuis l'Antiquité peuvent, en l'absence de renseignements comme ceux que César avait donnés sur les Gaulois ou Tacite sur les Germains, être

---

11. Du nom du roi du Pont dont le royaume incluait plus de vingt communautés linguistiques et qui pouvait, dit-on, s'adresser dans sa langue à chacun de ses sujets.

partiellement résolus par l'observation de leurs noms propres, une fois confrontés à des témoignages linguistiques modernes. Dans le cas de peuples plus récemment découverts et sur lesquels des matériaux deviennent disponibles, Leibniz pensait qu'en comparant des listes de mots désignant des notions élémentaires, des actions simples, il était possible de repérer entre deux ou plusieurs langues des analogies prouvant, de la part des peuples qui les parlent, une origine commune.

Selon cette perspective, plus immédiatement décisive que la précédente, la comparaison des langues s'est donc développée, surtout dans les milieux universitaires allemands, chez ceux qui travaillaient à Göttingen ou à Halle, comme une science auxiliaire de l'histoire. Et sa fonction classificatoire n'est presque qu'un effet secondaire, l'une des retombées annexes des succès remportés, des méthodes élaborées pour répondre à des questions qui touchaient d'abord à l'histoire des langues. La méthode comparative est ainsi le produit d'une interrogation sur le passé, sur les rapports entretenus par les langues du point de vue de leur origine, ayant progressivement débouché sur une construction théorique assez solide pour que soit du même coup satisfaite une visée taxinomique.

La possibilité ainsi offerte de résoudre des questions touchant à l'histoire des peuples par une méthode rigoureuse explique que cette direction de travail ait rapidement acquis un statut social éminent, au premier chef en Allemagne, puis en Grande-Bretagne, et finalement en France, au long du dix-neuvième siècle. La constitution de la linguistique comparée en une discipline institutionnelle dans le champ universitaire, la mobilisation de crédits, la création de nombreuses chaires n'ont été possibles que parce que le corps social tout entier se trouvait impliqué dans les questions soulevées, avec l'espoir de trouver des réponses à une quête désespérée de l'origine. Alors même que la Nation allemande était encore purement idéale, attendant encore son unité politique, elle pouvait être au moins fondée, du point de vue de la langue, dans un passé mythique, dans une figure glorieuse de ses origines. Fut ainsi défini autour de la comparaison, et pour un siècle environ, un programme de recherches cohérent qui fonctionna comme un paradigme remarquablement opératoire et productif quant à ses découvertes. Limité aussi dans son projet, car, jusqu'à la première guerre mondiale, la science de la langue s'est à peu près bornée à faire l'histoire des langues au moyen de leur comparaison.

Dès les tous débuts du travail comparatif, un certain nombre de principes ont été posés, implicitement ou non, qui rendaient valide l'entreprise. Moyennant quelques aménagements, ils continuent jusqu'à aujourd'hui à fonder les recherches.

Pour en savoir plus: S. Auroux et al. *Histoire des idées linguistiques. T. 2: Le développement de la grammaire occidentale* Liège 1992; R.H. Robins *A short history of Linguistics* Londres 1987; B. Malmberg *Histoire de la Linguistique de Sumer à Saussure* Paris 1991.

I. Principes et conditions de possibilite de la comparaison.

Si l'examen comparé de deux ou plusieurs langues permet de résoudre des questions portant sur leur passé et leur origine, c'est parce qu'une langue est un fait social d'une nature assez singulière pour permettre une telle investigation.

En premier lieu, il importe de souligner qu'un signe linguistique quelconque n'a de valeur qu'en vertu d'une tradition en acte dans un groupe social donné, dans une communauté particulière. Si les groupes phoniques transcrits graphiquement par *Hund*, *cane*, *perro*, *sobaka*, et *kalb* désignent, pour les locuteurs respectivement de langue allemande, italienne, espagnole, russe, arabe etc. l'animal que les français nomment *chien*, ce n'est pas évidemment pas parce que chacune de ces émissions vocales aurait, en elle-même, de par sa nature phonique, un rapport quelconque avec le concept de "chien", mais uniquement en vertu du fait que tel est l'usage de tous ceux qui appartiennent à chacun des groupes concernés, usage auquel se conformeront les enfants qui apprendront d'eux la langue qui y est parlée et qui sera ainsi perpétuée par transmission orale continue.

Arbitraire et contingence du signe linguistique.

La connexion, le lien entre son et sens est, en principe, parfaitement arbitraire, dans la mesure où, pour toute langue, n'importe quelle signification pourra être représentée par n'importe quelle combinaison de sons, et, corollairement, les milliers de formes signifiantes disponibles en chaque idiome sont globalement indépendantes les unes des autres: appeler *chien* tel quadrupède n'implique en rien la nécessité de désigner tel autre par *chat*. Ce premier point est fondamental, car si le lien était naturel et nécessaire, si le signe, de par sa valeur

sonore, était apte à évoquer d'une manière ou d'une autre une notion, un concept, aucune comparaison ne serait possible, ou plutôt elle s'exercerait d'une autre manière, nous éclairant par exemple sur le fonctionnement de l'esprit humain en général, puisque de tels phénomènes passeraient pour la traduction immédiate de son unité.

Ainsi, il est possible que diverses religions éparses dans le monde entier conçoivent le ciel comme "paternel" et la terre comme "maternelle", dans la mesure où la fertilisation des fruits et des plantes qu'apporte la pluie est analogue à une semence reçue par un réceptacle fécond. Mais, parce qu'elle peut résulter d'une vision de la nature universelle, une telle conception ne saurait déboucher sur une conclusion historique qui y verrait un héritage commun[12]. De même, la thématique d'une boisson d'immortalité[13], si elle se restreint à ce seul motif, est parfaitement insuffisante pour trancher d'une origine commune des cultures où on la rencontre, mais, prise dans un ensemble de motifs concomitants (une cuve gigantesque, une fausse fiancée, une lutte entre dieux et êtres démoniaques, etc.) cette série pourra être déclarée non fortuite, même si on la trouve dans des aires géographiques très éloignées et constituer alors un argument fort en faveur de l'héritage d'une telle conception religieuse.

La langue s'écarte par là fondamentalement des autres artefacts sociaux, en ce que des formes culturelles comme la religion, les institutions, les coutumes etc. sont toujours susceptibles de recevoir une justification naturaliste, tandis que les traits linguistiques relevant du même type d'explication rationnelle sont en très petit nombre. Il peut s'agir de certaines catégories, comme l'existence d'une distinction de genre qui s'enracine évidemment dans la division sexuelle, mais s'applique aussi à des objets dont le genre grammatical n'obéit plus à aucune motivation. Ou encore de la structuration du lexique, comme l'illustrent le recours à des noms de parties du corps pour désigner des mesures, le comput digital etc.

L'expression de certains concepts au moyen de formes stylisées d'impressions acoustiques est également un phénomène général en toute

---

12. L'exemple est donné par Paul Thieme, "The Comparative Method for Reconstruction in Linguistics" in D. Hymes ed. *Language in Culture and Society. A Reader in Linguistics and Anthropology*. New York 1964, 585-598, p. 586.
13. A. Meillet *La Méthode comparative en linguistique historique*. Oslo 1925, 1-2.

langue, susceptible de déboucher sur des convergences. On songe non seulement aux onomatopées et aux divers mots désignant des phénomènes sonores[14], mais aussi aux mots qui, non-onomatopéiques, présentent, dans la plupart des langues du monde, une analogie dans la relation son / sens: ainsi *mama*, "mère" et *papa* (ou *tata*, *dada*), "père", ou encore, quoique moins nette, au recours fréquent à la voyelle d'avant [*i*] pour "ici", "ceci", et aux voyelles d'arrière [*a*], [*u*] pour "là", "cela".

De tels exemples d'iconicité phonétique restent marginaux et, pour l'essentiel, étant arbitraire, la relation qui s'établit dans les signes linguistiques entre le plan de l'expression et le plan du contenu est une relation d'ordre factuel, particulier. Elle n'a d'existence que dans l'histoire de chacun des groupes humains que réunit l'usage d'une langue donnée. L'ensemble des rapports entre forme et signification qui caractérise en propre une langue apparaît, se maintient et s'efface à l'intérieur des limites de communautés précises, dans la mesure où une langue n'existe qu'en fonction d'une convention sociale. D'où cette conséquence que les faits de langue sont toujours par essence des faits singuliers; envisagés du point de vue de l'histoire de l'humanité, ils ne viennent à l'existence qu'une seule fois et ne sauraient renaître après leur disparition.

Il s'ensuit de cette double caractéristique de l'arbitraire et de la contingence des formes linguistiques que les ressemblances qu'elles manifestent d'une langue à une autre ne sauraient trouver d'autre explication qu'historique.

LA COMPARAISON, ENTRE HASARD ET PERTINENCE.

Les seules ressemblances pertinentes pour l'examen comparatif sont celles qui portent à la fois sur la forme et le sens. Une analogie de contenu à elle seule - le fait qu'une langue ait parallèlement à une autre telle ou telle catégorie grammaticale, tel ou tel mécanisme (un système de tons ou l'harmonie vocalique) - ne prouve rien quant à un lien historique unissant ces deux langues. Pour avoir une validité pour la comparaison les ressemblances doivent donc impérativement concerner son et sens. Mais cette

---

14. Sans qu'il faille exagérer la portée de telles désignations expressives. Meillet rappelle que beaucoup d'entre elles ne sont pas prévisibles *a priori*, le français *siffler* différant nettement de l'allemand *pfeifen* ou du russe *svistet'*(*Introduction à l'étude comparative des langues indo-européennes* [reprint] Alabama 1964, 15).

condition nécessaire n'est pas suffisante, puisque certaines ressemblances découlant de l'expressivité phonosymbolique, on l'a vu, ne peuvent être invoquées à ce titre. En outre, il convient d'éliminer les convergences dues au hasard. Le nombre des sons distincts utilisés dans les langues étant fonction de l'organisation de l'appareil phonatoire, est déjà inscrit dans des limites; de plus certains sons [*t, k, m, n, s, i, a, u*] sont communs à presque toutes les langues.

Il en résulte qu'entre toutes se manifesteront des analogies fortuites. Ces rencontres ont évidemment plus de chance de survenir pour des mots d'une syllabe: on trouve *nass*, en allemand et *nas* en Zuni, langue du Nouveau Mexique, au sens de "humide"[15]; en persan *bad*, et en didinga, langue du Soudan, *badh*, font écho à l'anglais *bad* avec le même sens de "mauvais"; *man* signifie "homme" en coréen comme en anglais[16]; *pneu-* est une racine grecque signifiant "respirer, souffler", comme *pniw-* en klamath, langue indienne de l'Oregon[17]; *dog* désigne un chien en anglais et en mbabaram, langue aborigène d'Australie[18]. De telles coïncidences sont déjà infiniment plus rares pour des mots de deux syllabes, mais on peut avoir *dori*, "souhaiter, désirer" en roumain comme en lau, langue d'Océanie[19]. Il s'agit cependant toujours d'un simple jeu de la nature, où le hasard seul est en cause, compte tenu de l'impossibilité de postuler des contacts historiques entre les deux peuples et de la distance les séparant, surtout pour des concepts qui ne sont pas suspects de fournir des mots voyageurs, à la différence de *tomate* (du nahuatl), *kangourou* (d'une langue aborigène d'Australie), *véranda* (de l'hindi), *sofa* (de l'arabe), *robot* (du tchèque), *gong* (du javanais) ou *thé* (du malais).

La chance ne saurait pourtant se répéter indéfiniment et, pour l'éliminer, on peut user de considérations de probabilité. La possibilité de trouver dans trois langues une ressemblance de forme et de sens est le carré de sa probabilité pour deux langues[20]. Si cinq langues offrent chacune 8% de ressemblances avec une autre, la probabilité qu'une telle

---

15. C.F. Hockett *A Course in Modern Linguistics* New York 1958, 486.
16. A. Meillet *Introduction à l'étude comparative des langues indo-européennes* [reprint] Alabama 1964, 16; J.H. Greenberg "Genetic Relationship among Languages" in *Essays in Linguistics* Chicago 1957, 35-45, p. 36.
17. C. Watkins *The American Heritage Dictionary of Indo-European Roots* Boston 1985 XI.
18. Où il n'est pas un emprunt à l'anglais car il est normalement issu de **gudaga* (B. Comrie *Language Universals and Linguistic Typology* Chicago 1981, 29.
19. M. Ruhlen *A Guide to the World's Languages Vol. I: Classification* Stanford 1987, 11.
20. La probabilité pour une seule langue doit, pour *n* langues, être élevé à la puissance *(n - 1)*.

concordance soit due au hasard n'est que de $(0, 08)^4$ ou 0,00004096, soit une chance sur 25.000 approximativement, et même pour trois langues la probabilité serait inférieure à 1%[21]. Un nombre significatif de correspondances de forme et de sens dans deux ou plusieurs langues est donc une indication certaine d'une connexion historique. Mais il est deux facteurs de ressemblance entre langues qui relèvent également de processus historiques: d'une part, les phénomènes d'emprunt, de l'autre, une relation génétique entre plusieurs langues.

LES EMPRUNTS.

Quant aux emprunts, l'expérience montre qu'il n'est aucun élément signifiant d'une langue qui ne puisse être emprunté, et il est possible de produire des exemples de toute sorte pour illustrer. Néanmoins, sur le plan du lexique, l'emprunt se limite généralement à certains domaines sémantiques, et affecte de préférence les concepts culturels ou techniques venus d'une autre aire géographique, du type *téléphone*, *tabac*, ou *igloo*. En revanche, l'emprunt n'affecte pour ainsi dire jamais les catégories grammaticales[22]. Le français a ainsi pu emprunter tel ou tel mot étranger au pluriel mais jamais la formation de pluriel du russe ou de l'arabe. Les emprunts de mots peuvent d'ailleurs se produire en nombre illimité, mais, pour autant, la langue dans ses structures n'en serait pas affectée: même si à chacun des mots français se substituait un mot anglais, cette langue resterait du français par sa morphologie et sa syntaxe, les affixes de dérivation et flexionnels. On peut donc considérer les formes grammaticales particulières comme étant à l'épreuve de l'emprunt et à ce titre probantes pour la comparaison qui vise à reconnaître les parentés génétiques. Et s'il est vrai que tout mot peut être emprunté, il reste que le vocabulaire fondamental (*oeil*, *nez*, *tête*, *jambe*) ou les pronoms manifestent une relative résistance au remplacement.

Les emprunts sont d'ailleurs faciles à identifier: si l'on repère

---

21. J.H. Greenberg "Genetic Relationship among Languages" in *Essays in Linguistics* Chicago 1957, 35-45, p. 39.

22. C'était l'opinion de Meillet ("Le problème de la parenté des langues" in *Linguistique historique et Linguistique générale I* [reprint] Paris 1958, 76-109, p. 82) ou d'E. Sapir *Language: An Introduction to the Study of Speech* New York 1921, 217), que combattait H. Schuchardt (*Brevier. Ein Vademekum der allgemeinen Sprachwissenschaft* Halle 1928, 195), pour qui même les désinences flexionnelles ne sont pas à l'abri d'une invasion par un matériau étranger. Voir le chapitre consacré à la question par U. Weinreich *Languages in contact: findings and problems* La Haye 1963, 29-46, où est mentionné l'emprunt par le roumain méglénite de désinences personnelles au bulgare et de l'instrumental géorgien *-iw* à l'arménien.

des similarités entre une langue A et une langue B, alors même que l'on sait que la langue B appartient à une famille de langues (CDEF) et si les similarités ne se retrouvent pas entre A et C, A et D, A et E etc. on peut conclure que les similarités ainsi limitées aux deux langues A et B sont des emprunts. Les ressemblances unissant anglais et français sont plus étroites et plus nombreuses que celles qui tiennent à ce que les deux langues se rattachent à l'ensemble indo-européen, et de plus, si l'anglais relevait des langues romanes, il manifesterait une connexion aussi systématique avec italien ou espagnol, ce qui n'est pas le cas.

L'HYPOTHESE GENETIQUE ET LE CHANGEMENT PHONETIQUE.

On bute donc finalement sur ce fait massif que certaines langues présentent entre elles, dans leur vocabulaire de base et leurs structures grammaticales, des similarités si nombreuses et si précises qu'il est exclu de les attribuer à des caractéristiques universelles ou au hasard, telles, en outre qu'elles ne peuvent pas être non plus le produit d'emprunts. L'hypothèse d'ordre historique à laquelle on est alors conduit consiste à voir dans ces langues le résultat d'une évolution à partir d'un original commun. Ce qui autorise à postuler une relation génétique entre les langues ainsi rapprochées est une nouvelle conséquence de cette donnée essentielle: la langue est avant tout une réalité sociale. Le système complexe que constitue la langue est certes propre à chaque individu et lui est immanent, mais comme il s'impose aussi à tous les membres d'un groupe donné, il n'existe qu'en tant que chacun d'eux en possède un sensiblement identique. La marge de variation qu'autorise l'usage idiosyncrasique est toujours étroitement circonscrite dans les limites imposées par l'exigence de continuer à être compris.

Or, quand on observe l'usage qui est fait de l'un de ces systèmes, on constate d'abord des innovations qui sont des accidents purement individuels (prononcer telle occlusive avec une mouillure, articuler telle voyelle un peu plus ouverte ou fermée que la moyenne) et prennent fin avec la mort de la personne qui les exhibe. Mais d'autres sont de nature collective, car elles tendent à apparaître chez tous les enfants apprenant à parler dans un espace géographique donné, à un moment donné. Nous pouvons en observer de semblables à Paris même où l'opposition entre un *a* d'avant, celui de *patte*, et un *a* d'arrière, dans *pâte*, disparaît progressivement au détriment de ce dernier, tout comme celle des deux voyelles nasales de *brun*

et *brin*, qui ne laisse plus subsister que celle-ci[23]. De tels changements, communs à tous les enfants, et transmis aux générations suivantes, sont d'ordre articulatoire, ils sont spontanés et inconscients. Il est même rare qu'ils soient perçus, car ils ne sont pas assez marqués pour que les locuteurs cessent d'avoir l'impression de parler la même langue que leurs aînés.

Les causes de telles altérations sont sans doute en dernière instance à chercher du côté de facteurs sociologiques [VOIR EXCURSUS: LES CAUSES DU CHANGEMENT PHONETIQUE], mais il convient surtout de souligner ceci, qui est décisif pour l'entreprise comparative: de tels processus se produisent avec une régularité absolue. Pour ce qui touche à la phonétique, on constate en effet que si, en une langue donnée, une articulation est conservée dans un mot donné, elle se maintiendra aussi dans tous les mots de la même langue où elle se présente dans les mêmes conditions. De même, pour une innovation articulatoire. Quand survient un changement phonétique, il apparaît certes dans quelques mots seulement, mais se généralise rapidement: la substitution d'un son à un autre se produira bientôt partout où ce son figurait auparavant dans des conditions identiques.

LES CAUSES DU CHANGEMENT PHONETIQUE.

Les linguistes ne sont jamais tombés d'accord pour fournir une explication du changement phonétique et il se pourrait bien que comme l'affirmait sèchement en 1933 le linguiste américain L. Bloomfield: "The causes of sound change are unknown" (*Language* New York 1933, 385). Pourtant, les tentatives n'ont pas manqué. Certaines d'entre elles sont aujourd'hui condamnées sans appel. Il s'agit d'abord des raisons d'ordre anatomique ou physiologique: les mutations des organes vocaux ou des modifications des centres cérébraux étaient censées refléter les dispositions particulières de tels ou tels peuples et finalement imputées à leurs différences intellectuelles. Le racisme naïf de ces explications n'est évidemment plus tenable. Il est impossible de corréler des traits de structure des langues, qu'ils soient phonétiques ou grammaticaux, avec des différences culturelles, lesquelles ne peuvent d'ailleurs jamais se lire en termes d'inégalités. La preuve en est que tout enfant, quel que soit son patrimoine génétique, pourra, dans des conditions normales d'exposition, apprendre n'importe quelle langue à la perfection.

On a pu postuler également une connexion entre les conditions physiques ou climatiques et l'évolution d'un système phonétique. Mais la "dureté" de ce dernier n'est pas corrélée avec la sévérité du climat. Si les langues du Caucase accumulent les sons à

---

23. Voir A. Martinet *La prononciation du français contemporain. Témoignages recueillis en 1941 dans un camp d'officiers prisonniers* Paris/Genève 1945 (rééd. 1971), et H. Walter *Enquête phonologique et variétés régionales du français* Paris 1982.

articulation glottale et uvulaire, les Esquimaux en revanche, encore moins privilégiés par leur environnement, possèdent un système parfois décrit, avec d'ailleurs un impressionnisme non moins coupable, comme "agréable", et celui d'Aborigènes australiens vivant dans un désert impitoyable a pu être déclaré "euphonique". Pas davantage l'altitude et les difficultés respiratoires qu'elle implique n'influent sur le changement consonantique. On a aussi avancé, entre autres le psychologue allemand W. Wundt, qu'il y avait une relation entre certaines conventions sociales et des caractéristiques phonétiques: les Iroquois tiennent pour déplacé de fermer le bouche en parlant, d'où l'absence de consonnes labiales. Mais on peut aussi bien tirer argument des particularités phonétiques pour inférer des traits de comportement.

D'autres théories paraissent simplement plus plausibles en ce qu'elles ne sont pas carrément intenables a priori. C'est le cas de l'hypothèse du substrat: l'arrivée d'immigrants et leur absorption par une population indigène qui acquiert leur langue occasionnerait pour celle-ci certains changements imputables à la langue des autochtones. Ainsi, en France, les différences entre langue d'oc et langue d'oïl correspondraient à une division et à une répartition dialectale des tribus celtiques présentes antérieurement à l'occupation romaine. Ou bien encore, les changements connus par le latin pour aboutir au roumain s'expliqueraient par le substrat dace. Mais, comme on ignore à peu près tout des langues impliquées, toute preuve fait défaut. De même, les traits particulier du Black English aux USA ont été parfois expliqués par les caractéristiques des langues vernaculaires d'Afrique parlées par les premiers esclaves, alors qu'ils peuvent aussi avoir d'autres causes.

La théorie de la facilité ou de l'économie des efforts articulatoires, mise en avant par O. Jespersen, entend rendre compte des multiples changements interprétables comme une simplification (assimilation, fusion etc.). En outre, on constate que certains changements semblent irréversibles: [*s*] devient [*h*] et non l'inverse. Mais d'un part, les notions de facilité ou de difficulté de l'articulation sont toujours relatives: ce qui est aisé pour le locuteur d'une langue ne l'étant pas pour celui d'une autre langue, ou, dans la même langue, peut s'avérer de compréhension plus difficile pour le destinataire. Ensuite, il resterait à expliquer pourquoi tous les changements appelés à tout moment par l'exigence de simplification ne se produisent pas. Enfin une simplification supposée peut très bien engendrer une complexité nouvelle, comme quand la syncope vocalique engendre des groupes consonantiques inédits, ou quand des réductions phonologiques entraînent des complications dans la morphologie. Tout compte fait, la thèse de la simplification est impuissante à expliquer la diversification en dialectes: pourquoi la complexité devrait-elle subsister concurremment à l'évolution positive que connaîtrait l'un d'entre eux seulement?

Plus séduisante apparaît au premier abord l'hypothèse selon laquelle, si un système phonologique offre une dissymétrie en présentant une place vide, une pression structurale interne est susceptible d'occasionner un changement afin d'instaurer un système plus équilibré. Mais ces dissymétries semblent plutôt inhérentes au fonctionnement linguistique et on ne voit pas pourquoi seules certaines d'entre elles susciteraient le besoin d'un changement. Bien des langues ont des "trous" dans leur système, qui ne montrent aucune évolution tendant à les combler par création d'un nouveau phonème.

On a pu également avancer que dans la mesure où chaque nouvelle génération doit maîtriser sa langue, c'est dans le processus imparfait d'apprentissage que le changement phonétique est le plus susceptible de se manifester. Pourtant, les difficultés, incontestables, de prononciation rencontrées par les jeunes enfants sont finalement

surmontées; surtout elles demeurent idiosyncrasiques et non systématiques. La cause du changement n'est donc pas dans l'apprentissage, même s'il est vrai qu'une fois qu'un tel changement s'introduit dans la langue il est diffusé par les plus jeunes locuteurs.

Il arrive aussi que l'on puisse faire dépendre les changements phonétiques d'une volonté de différenciation animant plus ou moins consciemment un groupe social déterminé. Dans le Nord de l'Australie, les membres mâles et initiés du peuple Lardil recourent à un style d'élocution particulier dans des situations sociales spécifiques. La grammaire restant identique, des mots spéciaux sont alors utilisés qui comportent des phonèmes totalement absents de la langue commune. Un petit groupe de pêcheurs de Martha's Vineyard étudiés par le sociolinguiste W. Labov se sont mis à exagérer une tendance préexistante pour s'affirmer comme groupe social indépendant avec un statut supérieur aux touristes affluant durant l'été.

Dans la même perspective il est avéré qu'à certaines des variantes phonologiques toujours existantes viennent s'attacher des connotations sociales impliquant prestige ou au contraire stigmatisées. S'il devient socialement valorisé de fermer ses voyelles, on constatera une tendance de la part des locuteurs se vivant comme d'un statut inférieur à épouser ce mouvement de fermeture par hypercorrection. Le snobisme, au sens d'effort pour singer une élite, est sans doute un puissant agent de diffusion et de généralisation d'innovations ayant connu un marquage sociolinguistique.

Le changement linguistique s'opère quand une règle est variable: deux options étant en concurrence selon le contexte social, l'une d'elles se trouve marquée positivement. Ne sont évidemment pas éclairées dans cette perspective les raisons qui font que c'est dans telle ou telle direction que s'opère le changement, car il semble que ce soit pur hasard si telle ou telle articulation se voit conférer un prestige conduisant à les imiter.

Pour en savoir plus: J. Aitchison *Language change: Progress or Decay?* Londres 1981, 111-169; R. Lord *Teach yourself Comparative Linguistics* Londres 1966, 82-87; R. Burling *Patterns of language* San Diego CA. 1992, 195-200; R. Antilla *An Introduction to Historical and Comparative Linguistics* New York 1972, 179-206; H. H. Hock *Principles of Historical Linguistics* Berlin 1988, 627-662; T. Crowley *An Introduction to Historical Linguistics* Auckland 1992, 191-203; T. Vennemann "Language change as language improvement" in C. Jones ed. *Historical Linguistics. Problems and Perspectives* Londres 1993, 319-344.

Par sa régularité, sa systématicité, le changement qui touche à l'articulation indépendamment du sens est le fait de langue qui se prête le mieux à une description rigoureuse et à des conclusions avérées. De tous les phénomènes linguistiques, c'est celui qui, le premier, a donné l'impression d'être soumis à de véritables lois. Tout l'édifice théorique que constitue la comparaison génétique dépend en fin de compte de l'existence, empiriquement vérifiable, de correspondances phonétiques régulières entre deux formes chronologiquement successives d'une même langue, encore appelées **lois phonétiques**.

Du latin classique au français, en passant par le latin vulgaire, on sait qu'il n'y a aucune solution de continuité, puisque, dans la succession des générations, aucune d'elles n'a eu l'impression de parler une langue différente de celle des ancêtres. L'une des lois auxquelles obéit cette évolution, saisie à travers les deux états de langue observés, est celle-ci: tout *a* latin accentué est devenu en français un *e* (*pa'ter* > *père*, *ama'tum* > *aimé*). Le principe de régularité des changements phonétiques, ou, selon une formulation plus ancienne, la constance des lois phonétiques, est le principe qui fonde toute comparaison génétique.

Le saut qualitatif effectué par la méthode comparative au début du dix-neuvième siècle a consisté à analyser les ressemblances de forme et de sens éventuellement constatées entre diverses langues à l'aide de ce principe, c'est-à-dire à les traiter comme des correspondances phonétiques entre langues différentes, et non plus à l'intérieur d'une continuité strictement verticale, comme dans le cas du *a* latin devenu *é* en français. Mais ces correspondances désormais horizontales, en quelque sorte, qu'on identifie de langue à langue, sont traitées en vertu du même principe explicatif que le rapport latin / français. Elles sont en effet considérées comme le reflet, la trace, seule à subsister, d'une évolution différenciée qu'aurait connue chaque langue au cours de son histoire à partir d'une identité originelle.

La régularité phonétique constatée de manière transversale, entre plusieurs langues, reçoit, de par sa nature identique, une explication analogue à la systématicité observée pour l'évolution unilinéaire et s'analyse dans les mêmes termes. L'hypothèse fondatrice de la comparaison des langues tient dans ce postulat que des différences ou plutôt les variantes repérées dans le cadre d'une ressemblance globale sont le produit de développements particuliers connus par chaque langue selon une évolution dont nous n'avons pas d'autres témoignages que le résultat final. Les transformations subies par le latin vulgaire et qui ont abouti aux langues romanes selon un développement en ce cas observable - celui qui mène du latin *pira* (avec *i* bref) à l'italien *pera*, à l'espagnol *pera*, au sicilien *pira*, à l'ancien français *peire* - sont déclarées analogues à celles opérées en d'autres langues qui, elles, offrent seulement des correspondances analogues à celles que les langues romanes présentent entre elles, et également imputables à une source unique, mais, dans leur cas, aujourd'hui disparue. Toute comparaison génétique est ainsi dans la situation où serait

la grammaire comparée des langues romanes si l'on ignorait tout du latin.

En conséquence de ces deux principes que sont l'arbitraire du signe et la régularité du changement phonétique, si deux ou plusieurs langues offrent aussi bien dans leurs signes ayant une fonction appellative, les racines, que désignative, particules, suffixes, désinences, un nombre significatif de concordances de détail, et que celles-ci peuvent être présentées sous forme de correspondances phonétiques systématiques, c'est que les langues réunies par ces rapports résultent d'évolutions différentes d'une même langue parlée auparavant. Leur relation dépend d'une genèse commune, puisque toutes ont été, à un moment du passé, une seule et même langue, leur divergence actuelle ne tenant qu'à la forme différente prise au cours de son développement par cette langue aujourd'hui perdue et inconnaissable, sauf au travers de toutes celles qui continuent à la refléter.

Il faut souligner que l'hypothèse unitaire ici présentée pour rendre compte des correspondances régulières repérées entre des langues différentes est la plus commode, mais non la seule possible. On pourrait aussi expliquer les systèmes de correspondances découverts en recourant "à un ensemble de dialectes qui n'ont jamais connu l'unification"[24], ou aborder les familles de langues comme résultant de processus de convergence tels ceux dont témoignent certaines aires géographiques comme les Balkans [VOIR EXCURSUS: LA LINGUISTIQUE DES AIRES].

LA LINGUISTIQUE DES AIRES.

Plusieurs aires géographiques où coexistent des langues d'origine différente présentent cette caractéristique que les langues ainsi mises en contact offrent non seulement beaucoup d'emprunts lexicaux, mais surtout manifestent, dans leur grammaire, nombre de traits communs qui ne paraissent pas avoir été hérités, alors que ce domaine est pourtant réputé plus résistant à la diffusion. Ce sont donc des langues qui, en dépit de leur hétérogénéité génétique, ont connu un développement tendant à la convergence typologique: les similitudes partagées par ces langues parlées dans une même région sont telles que chacune d'elles finit par ressembler davantage aux autres qu'à celles qui lui sont génétiquement apparentées.

On désigne ce type de regroupement où les langues, du fait de leurs contacts mutuels, partagent des traits de structure à l'aide du terme allemand de *Sprachbund* (union, ou association de langues).

---

24. M. Cohen *Le Langage (structure et évolution)* Paris 1950, 60.

Dans la région des Balkans se rencontrent ainsi grec moderne, bulgare, macédonien, albanais et roumain qui, ensemble, présentent des traits qu'ils ne partagent pas avec les langues qui leur sont les plus proches sur le plan génétique - traits qui ne peuvent donc être considérés comme hérités. Toutes ces langues sont certes indo-européennes mais appartiennent les unes, à des rameaux indépendants (grec, albanais), les autres, soit au groupe des langues slaves (bulgare, macédonien), soit à celui des langues romanes (roumain), il est donc facile de cerner sur la base de données comparatives générales quelles similarités sont imputables au seul contact géographique et ne peuvent être mis au compte d'un héritage commun.

Parmi ces traits, on peut compter le syncrétisme du génitif et du datif (le bulgare étant par exemple la seule langue slave à avoir remplacé son système casuel en recourant, de manière strictement analytique, à des prépositions); l'article postposé (sauf pour le grec moderne); la perte générale de l'infinitif, remplacé par une subordonnée introduite par une conjonction, ainsi en grec *dos mu na pjo*, m. à m. "donne moi que je boive". L'origine de ce dernier phénomène doit sans doute être attribuée à la prononciation en usage dans la période byzantine: la majeure partie des verbes présentant pour la troisième personne du singulier du présent une désinence en *-i* et en *-in* à l'infinitif, la chute du *-n* final a amené l'identité des terminaisons, et comme le grec était, durant et après cette époque, utilisé comme *lingua franca* dans tous les Balkans, le bilinguisme a assuré la diffusion de cette substitution dans les autres langues du domaine.

D'autres domaines géographiques offrent des exemples analogues. Dans le sud-est asiatique, la présence de tons en chinois, thaï et vietnamien a été de même attribuée au contact. En Inde, les langues indo-européennes, possèdent, comme leurs voisines dravidiennes et Munda, non parentes, une série de rétroflexes inconnue des autres langues indo-européennes et recourent aux classificateurs numéraux. Entre la mer Noire et la Caspienne, l'arménien oriental et l'ossète, indo-européens, ont des consonnes glottalisées, par quoi leur système phonologique se rapproche de celui des langues caucasiques, toutes proches. Certaines langues bantoues ont des clicks, comme leurs voisines les langues Bush-Hottentotes. Pour l'Europe occidentale, le linguiste américain B. L. Whorf avait également isolé un ensemble de traits communs, tels que les articles et les formes verbales périphrastiques qu'il tenait pour si caractéristiques qu'il avait proposé le terme de SAE (Standard Average European, ou européen moyen standard) en le considérant comme une seule langue par opposition à des langues massivement différentes comme l'Amérindien, son objet d'étude primordial.

Pour en savoir plus: T. Bynon *Historical Linguistics* Cambridge 197, 244-256; B. Comrie *Language Universals and Linguistic Typology* Chicago 1981, 197-203; W. P. Lehmann *Historical Linguistics: an Introduction* Londres 1992, 137-139;

D'autres encore voient dans l'*Ursprache*, ou langue originelle, ainsi reflétée, une *lingua franca*[25]. Néanmoins l'hypothèse d'une divergence à partir d'une identité initiale demeure la plus tentante, sans être rien de

25. N.E. Collinge "Language as it evolves: tracing its forms and families" in N.E. Collinge et al. *An Encyclopaedia of Language* London/New York 1990 876-916, p. 888.

plus qu'une hypothèse[26].

Pour en savoir plus: A. Meillet *La Méthode Comparative en Linguistique historique* Oslo 1925 [reprint Paris 1970]; A. Meillet *Introduction à l'étude comparative des langues indo-européennes* Paris 1937 [reprint Alabama 1964], 13-50; A. Meillet "Sur la Méthode de la grammaire comparée" in *Linguistique historique et Linguistique générale I* Paris 1938, 19-43; "Le problème de la parenté des langues", ib. 76-109; J. Manessy-Guitton "La Parenté généalogique" in A. Martinet et al. *Le Langage* Paris 1968, 814-864; H.H. Hock *Principles of Historical Linguistics* Berlin/New York/Amsterdam 1986, 34-51; T. Bynon The Neo-Grammarian model in *Historical Linguistics* Cambridge 1977, 17-75; W.P. Lehmann *Historical Linguistics: an Introduction* Londres 1992, 1-22; R. Anttila *An Introduction to Historical and Comparative Linguistics* New York 1972. P. Thieme "The Comparative Method for Reconstruction in Linguistics in D. Hymes ed. *Language in Culture and Society. A Reader in Linguistics and Anthropology* New York 1964, 585-598; J.H. Greenberg "Genetic Relationship among Languages" in *Essays in Linguistics* Chicago 1957; M. Ruhlen *A Guide to the World's languages Vol. I Classification* Stanford 1987, 1-23

II. Methodes et procedures de la comparaison.

Quoi qu'il en soit, le matériau essentiel de la comparaison génétique, celle qui vise à la classification des langues du point de vue de leur origine et de leur histoire, saisie au travers de leurs relations, est constitué par les correspondances phonétiques entre les langues attestées. Ces correspondances sont "la seule réalité à laquelle [la grammaire comparée] ait affaire"[27]. Il convient de souligner l'importance d'une telle approche sur plusieurs plans.

Des correspondances a la restitution.

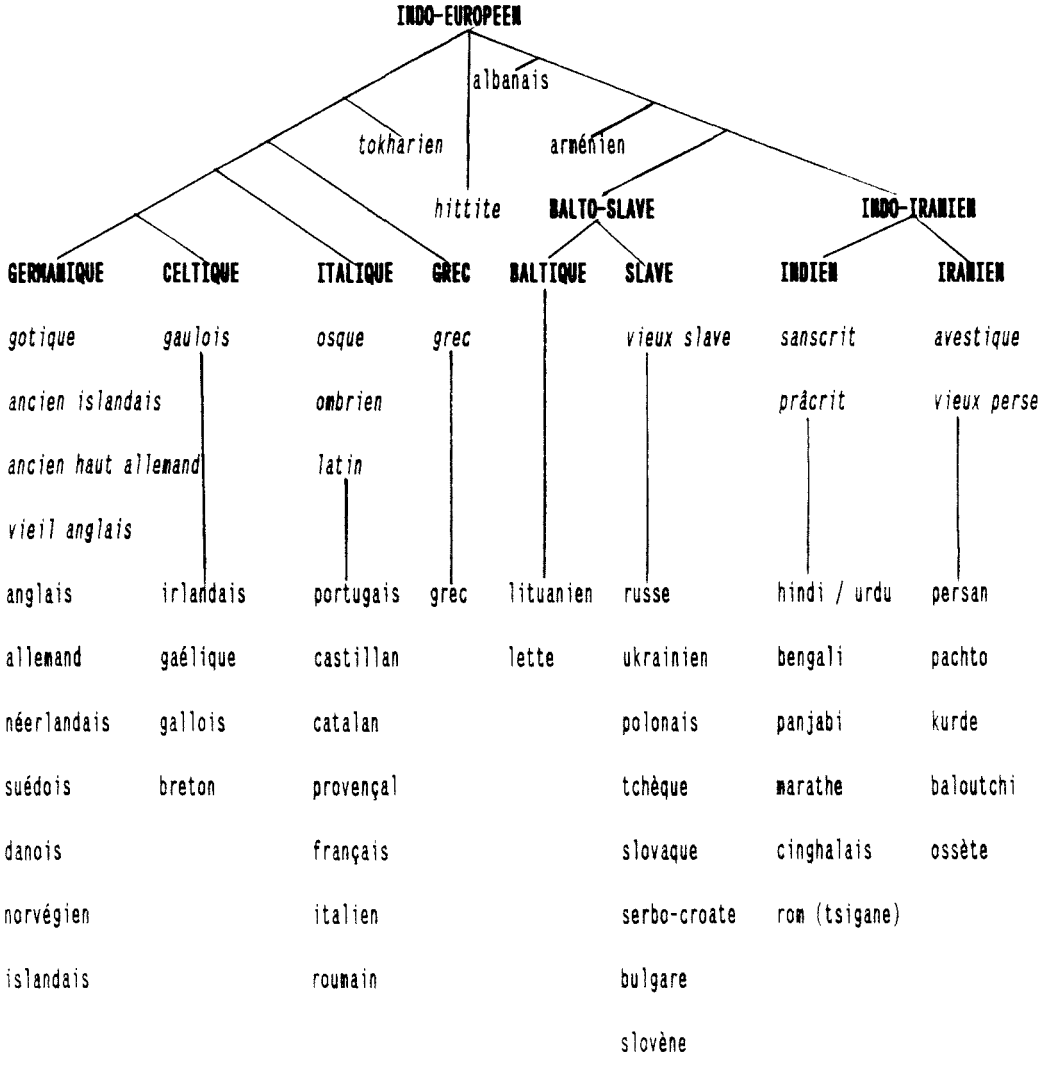
D'abord, même si le point de départ d'une investigation de ce type est toujours une intuition fondée sur des ressemblances, il est néanmoins permis de s'en affranchir: "l'histoire phonétique ne se fait pas avec des ressemblances, mais avec des systèmes de correspondances [..] le linguiste n'opère pas avec des faits concrets plus ou moins homologues, mais avec des correspondances pouvant porter sur des faits hétérogènes"[28]. Ce qui

26. A. Meillet en souligne souvent les limites: "les correspondances supposent une réalité commune; mais de cette réalité on ne peut se faire une idée que par des hypothèses, et ces hypothèses sont invérifiables" *Introduction à l'étude comparative des langues indo-européennes* [reprint] Alabama 1964, 41.

27. A. Meillet *Introduction à l'étude comparative des langues indo-européennes* [reprint] Alabama 1964, 41.

28. A. Meillet *Introduction à l'étude comparative des langues indo-européennes* 1964, 470-471; Cf. *Linguistique historique et Linguistique générale I* [reprint] Paris 1958, 82.

signifie en pratique qu'une correspondance peut être parfaitement probante alors même que la ressemblance est inexistante. Ainsi, au mépris apparent de toute vraisemblance, il est permis d'effectuer un rapprochement tout à fait valide, pour le nom de "deux", non seulement entre le sanscrit *d(u)vâ*, le grec *dyo*, et le latin *duo*, mais aussi avec l'arménien *erku*, parce que deux autres correspondances montrent que *erk-* de l'arménien peut répondre à *dw-* en une autre langue[29]. De telles correspondances se laissent reconnaître entre de nombreuses langues attestées, parmi lesquelles on peut citer l'ancien indien, ou sanscrit, le lituanien, l'islandais ancien, l'arménien, le grec, le latin, le gotique, et d'autres encore, selon le tableau généalogique suivant:



29. Voir les précisions que fournit Meillet: "de même que le grec a pour l'idée de "craindre" une racine *dwi-*, l'arménien a *erki-* (*erkiwil*, "crainte"), et de même que le grec a pour dire "longtemps" un vieil adjectif *dwâron*, l'arménien a *erkar*, "long" [..] la concordance se laisse ramener à une règle générale de correspondance: un ancien *dw-* aboutit à arm. *erk-*." *La Méthode comparative en linguistique historique* Oslo 1925, 6, Cf. p. 31.

L'autre conséquence est que la restitution constitue toujours une extrapolation hasardeuse. Assurément, si l'on découvre pour le nom "mère", la correspondance suivante: ancien nordique *modir* - irlandais *mathir* - latin *mâter* - grec *mâter* - lituanien *mote*, "femme"; *motina*, "mère" - ancien slave *mati* - arménien *mayr* - ancien indien *mâta*; puis pour "souris", celle-ci: a. n. *mus* - lat. *mus*, *muris* - gr. *mus* - arm. *mukn* - a. ind. *mûḥ*; pour "lune" ou "mois", cette autre: gotique *mena*, "lune", *menoþs*, "mois" - irl. *mí*, "mois" - lat. *mênsis*, "mois" - gr. *mên*, "mois", *mênê*, "lune" - lit. *menuo*, "lune", *menesis*, "mois" - a. sl. *meseci*, "lune, mois" - arm. *amis*, "mois"; enfin pour "mort": got. *maurþ*, "meurtre" - lat. *mors*, *mortis*, "mort" - gr. *mortos*, "mortel" - lit. *mirtis*, "la mort" - a. sl. *su-mrti*, "la mort" - arm. *mard*, "un être humain (un mortel) - a. ind. *mrtiḥ*, nous serons légitimés à poser, en position initiale, cette équation: (got.) *m-* = (celt.) *m-* = (lat.) *m-* = (gr.) *m-* = (lit.) *m-* = (a. sl.) *m-* = (arm.) *m-* = (a. ind.) *m-*. Et même, plutôt que d'avoir à répéter une série aussi encombrante, nous pouvons choisir de l'identifier par un signe conventionnel. Ce dernier pourrait être arbitraire (un chiffre par exemple), mais il est évidemment préférable que le signe "rappelle quelque chose de ce qu'[il] doit désigner"[30], et l'on choisira de symboliser la correspondance par *m-* en l'affectant d'un astérisque: **m-*. Dès lors, au lieu de répéter qu'il y a une concordance à l'initiale entre l'*m* gotique, l'*m* celtique, l'*m* latin, et ainsi de suite, nous parlerons de l'*m* indo-européen [VOIR EXCURSUS: QU'EST-CE QUE L'INDO-EUROPEEN?], noté **m* i.e, comme étant l'une des correspondances que l'on constate entre certaines langues historiquement attestées.

QU'EST-CE QUE L'INDO-EUROPEEN?

On donne ce nom, par convention, depuis 1813 (Thomas Young, en anglais, indo-european) à la langue originellement unique dont on doit postuler l'existence pour expliquer les ressemblances que manifestent dans leur vocabulaire et leur formes grammaticales un certain nombre de langues actuellement parlées en Asie et en Europe occidentale. Le premier terme utilisé (par un danois écrivant en français, en 1810) était celui d'indo-germanique, qui malgré les apparences n'est pas inspiré par une volonté nationaliste de la part des Allemands, car il se justifiait par le désir de décrire cette famille de langues du point de vue des langues parlées aux deux pôles de son extension géographique, à un moment où les langues celtiques n'y étaient pas incluses, cet usage est d'ailleurs persistant en langue allemande (*indo-germanische*). En revanche, le terme *aryen*, fréquent au siècle dernier, et justifié tant qu'on pensait que le sanscrit était le plus fidèle reflet de la langue première, a disparu

30. L. Hjelmslev *Le Langage* (trad.) Paris 1966, 37.

sauf comme synonyme, à vrai dire peu usité, d'indo-iranien

Au sens large du terme l'indo-européen est aussi l'ensemble des langues existantes ou ayant existé qui dérivent de ce tronc commun: on appelle ainsi langue indo-européenne, selon la définition de Meillet, "toute langue qui, à un moment quelconque, en un lieu quelconque, à un degré d'altération quelconque, est une forme prise par [l'indo-européen], et qui continue ainsi, par une tradition ininterrompue l'usage de l'indo-européen", ce qui permet de dire que l'albanais, le grec, le français, le tokharien ou l'hindi sont des langues indo-européennes.

Aussi est-il devenu courant de désigner la langue commune d'où toutes les langues indo-européennes sont supposées être issues (en vertu de l'hypothèse fondamentale de la comparaison génétique) par le terme d'indo-européen commun ou de proto-indo-européen, en particulier lorsqu'on en tente une restitution.

Pour en savoir plus: J. Manessy-Guitton "L'indo-européen" in A. Martinet et al. *Le Langage* Paris 1968, 1240-1287; W.P. Lehmann *Theoretical Bases of Indo-European Linguistics* Londres/New York 1993; A. Martinet *Des Steppes aux Océans. L'indo-européen et les "Indo-Européens"*. Paris 1986; A. Meillet *Introduction à l'étude comparative des langues indo-européennes* Paris 1937 [reprint Alabama 1964] 35-50.

On voit que la notion de restitution ne saurait être maniée sans précautions. Pour Meillet, à proprement parler, on ne reconstruit pas l'indo-européen: "les correspondances sont les seuls faits positifs, et les «restitutions» ne sont que les signes par lesquels on exprime en abrégé les correspondances."[31]. Pour autant, on ne saurait refuser les indications que les langues attestées fournissent sur l'articulation vraisemblable du prototype, et c'est ainsi que procède le linguiste le plus rigoureux, en cherchant à s'appuyer sur l'évolution phonétique la plus plausible[32]. Le choix du symbole vise toujours à l'économie dans l'explication du changement et, du coup, au réalisme.

Ainsi, quand la comparaison du sanscrit *bharami*, du grec *pherô*, de l'arménien *berem* et du gotique *baira*, tous au sens de "je porte", puis d'autres (scr. *nabhah.* / gr. *nephos* / ancien saxon *nebal*) fournit la correspondance suivante (scr.) *bh* = (gr.) *ph* = (arm.) *b* = (germ.) *b*, le comparatiste est conduit à se demander si l'une de ces réalisations phonétiques n'est pas chronologiquement première, et, si oui, laquelle et pourquoi. Ses réponses aimanteront le choix du symbole à poser comme forme

31. A. Meillet *Introduction à l'étude comparative des langues indo-européennes* [reprint Alabama 1964], 42.

32. Par exemple le phonème *s* dental peut devenir un simple souffle, ou devenir fricatif *P*, et même occlusif, *t*, ou se sonoriser en *z* qui passe ensuite à *r* (phénomène dit "rhotacisme"), ainsi que devenir *ch* (A. Meillet *Introduction à l'étude comparative des langues indo-européennes* [reprint Alabama 1964], 34).

hypothétique restituée.

Il procède alors par essais et erreurs: *b* peut effectivement déboucher sur *bh* par une évolution qui requiert seulement l'aspiration et qui est cautionnée par de nombreux exemples, mais, pour devenir *ph*, deux étapes sont nécessaires (aspiration et dévoisement) et l'évolution serait sans exemple dans le domaine indo-européen; la seconde hypothèse *ph* > *bh*, qui réclame le voisement, est possible, mais sans précédent en i.e., tandis que *ph* > *b* est plus problématique, en ce qu'elle exigerait deux étapes (perte de l'aspiration et voisement).

En revanche, une dernière hypothèse est plus satisfaisante, partant de *bh*, car aussi bien *bh* > *b* que *bh* > *ph* ne réclament qu'un seul changement (dévoisement d'un côté et perte de l'aspiration de l'autre), et c'est là l'évolution phonétique la plus simple et la mieux attestée. Mais pour trancher définitivement, resterait encore à étudier dans leur ensemble le comportement des aspirées labiales (*bh*, *ph*) comme des occlusives labiales (*b*, *p*) i.e. et à trouver une cohérence dans leur traitement. Au-delà, c'est le traitement de la totalité de la classe des aspirées et des occlusives sonores qui devra être également abordé et n'offrir aucune contradiction avec l'hypothèse émise dans le cas précis étudié.

LA RECONSTRUCTION INTERNE.

Préalablement à toute comparaison effectuée de langue à langue, il est souvent possible, en partant d'un état de langue donné, de pratiquer une forme de reconstruction interne qui livrera un état antérieur, quoique non attesté de cette langue. On peut montrer sur un corpus hypothétique les bénéfices d'une telle démarche. Soit une langue dont l'examen livrerait les unités significatives suivantes:

pak	*pal*	*pat*	*map*	*lap*
fik	*fil*	*fit*	*maf*	*laf*
wuk	*wul*	*wut*	*maw*	*law*

Si nous examinons la distribution des sons représentés par des lettres dans les trois premières colonnes nous constatons que *k* peut suivre n'importe quelle voyelle *a*, *i*, *u*, et que c'est aussi le cas de *l* et de *t*. Mais les autres consonnes manifestent des restrictions dans leur distribution: *p* n'apparaît que devant *a*; *f* seulement devant *i*; et *w*

seulement devant *u*. Ces trois consonnes sont donc en distribution complémentaire et peuvent être considérées comme les allophones d'un seul phonème que nous symboliserons provisoirement par **P*, afin de désigner l'ensemble *p* / *f* / *w*, dont chacun des membres se trouve dans un environnement spécifique.

Dans les deux dernières colonnes, les membres de l'ensemble *p* / *f* / *w* apparaissent tous en position finale, et, à ne considérer que ces deux colonnes, nous pourrions les croire en contraste plutôt qu'en distribution complémentaire. Les deux conclusions peuvent pourtant être conciliées si nous postulons que les mots des deux dernières colonnes se terminaient jadis par une voyelle aujourd'hui disparue. Auquel cas, le *p* aurait précédé un *a*, le *f* un *i*, le *w* un *u*, selon le modèle de distribution déjà reconnu. En outre, on peut alors poser que tous les mots de la langue obéissaient, antérieurement à l'état où nous la saisissons, au schéma *CVCV*, même si, pour les trois premières colonnes, nous ne pouvons reconstruire la voyelle finale disparue autrement que sous la forme non spécifiée *V*:

**PakV*	**PalV*	**PatV*	**maPa*	**laPa*
**PikV*	**PilV*	**TitV*	**maPi*	**laPi*
**TukV*	**TulV*	**TutV*	**maPu*	**laPu*

Plutôt que le symbole **P*, nous pourrions adopter l'un des membres de l'ensemble qu'il coiffe. De préférence **p*, parce qu'il est plus commun dans les langues que *f* ou *w*, et aussi parce qu'il se développe fréquemment en *f* devant *i* et en *w* devant *u*, alors que l'inverse est plus inhabituel.

UN EXEMPLE DE LA METHODE COMPARATIVE.

La seconde étape consiste à pratiquer la comparaison externe en posant des correspondances entre langues probablement parentes. Esquissons sur un autre corpus inventé (et violemment réducteur) quelques-unes des démarches élémentaires de la comparaison. Supposons que l'on découvre en quelque contrée quatre langues bien distinctes qui, outre de nombreuses analogies dans le fonctionnement morphologique et syntaxique, offrent aussi des ressemblances frappantes dans leur lexique, en particulier pour les mots suivants:

L1	L2	L3	L4	Sens
tuha	*duga*	*tuka*	*tuk*	"arbre"
rifu	*rifu*	*lifu*	*rif*	"mouche"
wani	*wani*	*weni*	*wan*	"ciel"
rana	*rana*	*lena*	*lena*	"pierre"
kaha	*gaga*	*kaka*	*kak*	"abri"
puti	*budi*	*puti*	*put*	"éclair"
paltufo	*niri*	*nili*	*nir*	"lance"
fanu	*fanu*	*fenu*	*fan*	"ruisseau"
napu	*nabu*	*napu*	*nap*	"flèche"
mita	*mida*	*mita*	*mit*	"océan"

Les ressemblances sont, dès le premier coup d'oeil, nombreuses, et l'emprunt ne pourrait suffire à les expliquer. Avant même d'examiner précisément les régularités de ces données, nous butons sur deux formes inattendues. En L1, le mot qui désigne la lance, *paltufo*, n'a rien de commun avec *niri*, *nili*, *nir*, dans L2, L3, L4; en outre, les principes phonologiques qui paraissent en vigueur pour les autres mots de L1, qui n'ont que deux syllabes, chacune de structure *CV*, sont violés par ses trois syllabes; de plus le groupe *-lt-* semble contredire le patron normal des mots. Selon toute vraisemblance, *paltufo* est un emprunt de L1 à une autre langue, inconnue. Un autre mot doit être suspect, *lena*, en L4, car, d'après notre corpus de cette langue, il est le seul à se terminer par une voyelle, à présenter la voyelle *e* et la consonne *l*. En revanche, s'il est emprunté, il doit l'être d'une langue apparentée, compte tenu de similitude des mots de L1, L2, L3 pour "pierre": l'explication la plus plausible est alors qu'il a dû être emprunté à L3, à cause de l'identité des mots dans les deux langues et du fait que les caractéristiques phonologiques de *lena* sont normales pour les mots de L3.

Les sons des quatre langues offrent entre eux des correspondances régulières. Le mot pour "flèche" en L1 commence par un [*n*], comme les mots qui correspondent dans les trois autres langues, le mot pour "ciel" a un [*n*] comme troisième élément dans les quatre langues, et c'est aussi le cas pour le mot "ruisseau". Les sons correspondants dans les deux langues sont donc [*n*]-[*n*]-[*n*]-[*n*]. Deux autres correspondances complètement régulières sont [*f*]-[*f*]-[*f*]-[*f*], ainsi que [*m*]-[*m*]-[*m*]-[*m*]. Une autre correspondance sans exception est [*r*]-[*r*]-[*l*]-[*r*]. En toute position où L1 a [*r*], L2 et L4 sont dans le même cas, en position correspondante L3 a [*l*], même si les données, pour les raisons exposées plus haut, sont incomplètes pour les mots "pierre" et "lance", les formes disponibles se conforment à ce schéma. On sait qu'une correspondance peut être parfaitement régulière, alors même

que les membres de la correspondance sont différents: l'écart qu'offre L3, avec son [*l*], est systématique, et c'est la régularité, bien plus que l'identité phonétique, qui est décisive pour prouver la filiation génétique et tenter une reconstitution du proto-langage. D'autres correspondances sont du même type [*t*]-[*d*]-[*t*]-[*t*], et [*p*]-[*b*]-[*p*]-[*p*].

Nous avons posé six correspondances complètement régulières: [*n*]-[*n*]-[*n*]-[*n*]; [*f*]-[*f*]-[*f*]-[*f*]; [*m*]-[*m*]-[*m*]-[*m*]; [*r*]-[*r*]-[*l*]-[*r*]; [*t*]-[*d*]-[*t*]-[*t*]; [*p*]-[*b*]-[*p*]-[*p*]. Mais si nous tentons maintenant d'en isoler d'autres, nous butons sur diverses exceptions. Sur la base du mot "arbre", le son [*k*], par exemple en L3, semble correspondre à [*h*], [*g*], et [*k*] en L1, L2, L4, respectivement, ce qui nous conduirait à la correspondance [*h*]-[*g*]-[*k*]-[*k*]. Dans cette hypothèse, en partant du mot de L3 pour "abri", [*kaka*], nous devrions nous attendre à ce que les mots parallèles soient, pour L1, [*haha*], pour L2, [*gaga*], et pour L4, [*kak*]. Or nos espérances sont déçues, puisque le mot de L1 n'est pas [*haha*], mais [*kaha*], la consonne initiale démentant notre belle construction.

Certaines correspondances doivent donc être posées en tenant compte d'un environnement phonologique spécifique. Par exemple la correspondance [*h*]-[*g*]-[*k*]-[*k*] est valide à condition d'exclure la position initiale, pour laquelle la correspondance [*k*]-[*g*]-[*k*]-[*k*] est requise. Pour confirmer cette correspondance conditionnée par l'environnement phonologique, il suffirait de rencontrer à propos d'un autre mot la série [*kiri*], [*giri*], [*kili*], [*kil*]. Quant aux voyelles [*i*] et [*u*], les correspondances dépendent également de l'environnement phonologique. En se fondant sur "arbre" et "océan", les correspondances seraient à poser comme [*u*]-[*u*]-[*u*]-[*u*], et [*i*]-[*i*]-[*i*]-[*i*], où ces voyelles ne sont pas en position finale. Mais les mots de L4 montrent une absence générale de voyelles finales: aussi bien les mots pour "ruisseau" que pour "ciel" témoignent que la série est, à la finale, respectivement [*u*]-[*u*]-[*u*]-[ø], et [*i*]-[*i*]-[*i*]-[ø]. En position finale, la série [*a*] offre, d'après le mot pour "océan", un comportement analogue: [*a*]-[*a*]-[*a*]-[ø]. Mais, ailleurs, [*a*] en L1, L2 et L4 correspond pour certains mots de L3, comme "abri" ou "flèche", à [*a*], tandis que pour d'autres ("ciel", "pierre", "ruisseau") à [*e*]. Si nous nous efforçons de découvrir un principe général qui déterminerait, en L3, l'apparition de [*a*] ou de [*e*], nous découvrons que [*e*] apparaît seulement devant la nasale [*n*], tandis que [*a*] ne se rencontre jamais devant cette nasale. L'alternance entre [*a*] et [*e*] en L3, comme contre-partie de [*a*] dans les autres langues

est donc parfaitement régulière. Nous pouvons rassembler les correspondances jusqu'ici obtenues:

(1)	[*n*]-[*n*]-[*n*]-[*n*]	
(2)	[*f*]-[*f*]-[*f*]-[*f*]	
(3)	[*m*]-[*m*]-[*m*]-[*m*]	
(4)	[*r*]-[*r*]-[*l*]-[*r*]	
(5)	[*t*]-[*d*]-[*t*]-[*t*]	
(6)	[*p*]-[*b*]-[*p*]-[*p*]	
(7a)	[*k*]-[*g*]-[*k*]-[*k*]	en position initiale
(7b)	[*h*]-[*g*]-[*k*]-[*k*]	en toute autre position
(8a)	[*u*]-[*u*]-[*u*]-[*ø*]	en position finale
(8b)	[*u*]-[*u*]-[*u*]-[*u*]	en toute autre position
(9a)	[*i*]-[*i*]-[*i*]-[*ø*]	en position finale
(9b)	[*i*]-[*i*]-[*i*]-[*i*]	en toute autre position
(10a)	[*a*]-[*a*]-[*a*]-[*ø*]	en position finale
(10b)	[*a*]-[*a*]-[*e*]-[*a*]	devant nasale
(10c)	[*a*]-[*a*]-[*a*]-[*a*]	en toute autre position

Ce système de formules régulières définit la forme phonétique des mots des quatre langues apparentées et doit nous permettre de prédire la prononciation d'un mot d'une langue dont nous ne connaîtrions que les correspondants dans les trois autres. Ainsi du mot pour "lance" en L1, qui, s'il n'avait été l'objet d'un emprunt, aurait dû être, d'après ses pendants et les formules posées, de forme [*niri*]; de même le mot originel pour "pierre" en L4 serait [*lan*], n'était son emprunt à L3.

L'ETAPE DE LA RECONSTRUCTION.

L'existence, pour des séries entières de mots, d'une ressemblance phonétique réductible à un réseau de correspondances régulières, appuyée en outre sur cette constatation empirique générale que le changement phonétique est, en toute langue, globalement régulier, permet de poser l'hypothèse que cette identité fondamentale à l'intérieur d'un spectre de variations doit s'expliquer par une origine commune de chacun des mots des différentes langues que l'on regroupe ainsi. Pour chaque série il doit y avoir une forme originelle dont tout membre est le prolongement, la continuation plus ou moins divergente. Le double mouvement qui consiste à reconstruire le prototype et à reconstituer les changements phonétiques qu'il a pu subir est l'étape qui suit l'établissement des correspondances.

Pour chaque correspondance phonétique, nous poserons une unité phonétique de la proto-langue. Chaque fois par exemple que [*n*] apparaît dans un mot des langues apparentées, les autres mots posés en parallèle contiennent aussi [*n*] et dans la même position. Il n'est pas trop aventuré

de supposer que cet accord tient à ce que le mot de la proto-langue qui est au point de départ des quatre formes existantes contenait aussi [*n*] dans cette position. Le postulat est donc que [*napu*], [*nabu*], [*napu*], [*nap*] commencent tous par [*n*] parce que le mot pour "flèche" dans la proto-langue commençait également avec un [*n*], que le troisième segment de [*fanu*], [*fanu*], [*fenu*], [*fan*] est, à chaque fois, [*n*], parce que le troisième segment du mot de la proto-langue pour "ruisseau" était aussi [*n*], etc. Nous reconstruirons donc pour la proto-langue le segment *[*n*], qui n'a été affecté d'aucun changement phonétique, puisque dans les langues attestées il est reflété de manière uniforme. La même reconstruction peut, pour des raisons similaires, être pratiquée pour *[*f*] et *[*m*]. Dans le cas de la correspondance [*r*]-[*r*]-[*l*]-[*r*], où se rencontre une disparité, le segment à reconstruire doit être tel qu'il permette un développement historique vers [*l*] comme vers [*r*]: ce requisit peut aussi bien conduire à reconstruire [*l*] que [*r*] comme proto-segment, dans la mesure où les deux sons sont très proches et interchangeables. Nous reconstruirons *[*r*] pour des raisons d'économie dans la mesure où nous n'avons alors à postuler un changement de [*r*] en [*l*] que pour une seule langue (L3), alors que reconstruire *[*l*] exigerait qu'une altération de [*l*] en [*r*] se soit produite indépendamment en trois langues L1, L2, L4, et un maintien seulement en une seule, L3. Cette éventualité, quoique parfaitement possible, est cependant moins probable. L'exemple montre que la reconstruction n'offre pas toujours un caractère de certitude et que l'on doit parfois se contenter d'une spéculation sur la vraie nature du segment de la proto-langue. Le même type de raisonnement vaut pour reconstruire *[*t*] et *[*p*] à partir respectivement de [*t*]-[*d*]-[*t*]-[*t*] et [*p*]-[*b*]-[*p*]-[*p*]. Dans le cas de la correspondance [*k*]-[*g*]-[*k*]-[*k*] qui n'est vraie qu'en position initiale, nous pouvons reconstruire *[*k*], en postulant que [*k*] est devenu [*g*] en L2. Si nous avons eu raison de reconstruire *[*p*] et *[*t*], la troisième occlusive à reconstruire doit être *[*k*] et non *[*g*], dans la mesure où une certaine symétrie préside toujours à l'organisation d'un système phonologique, en particulier pour les occlusives et les fricatives: il est extrêmement courant qu'une langue ait trois occlusives sourdes, [*p*], [*t*], [*k*], il ne l'est pas qu'elle possède deux occlusives sourdes [*p*], [*t*], et une sonore [*g*]. En reconstruisant *[*k*], aux côtés de *[*p*] et *[*t*], nous prenons en compte cette tendance universelle. Si la proto-langue possédait l'occlusive *[*k*] en position initiale, la même occlusive devait également apparaître en d'autres positions. En conséquence, c'est le même segment qui doit être au

point de départ de la correspondance [*h*]-[*g*]-[*k*]-[*k*], qu'illustre la série des mots [*kaha*], [*gaga*], [*kaka*], [*kak*], pour "abri": le mot originel devait contenir deux occurrences de *[*k*] qui se sont maintenues en L3 et L4. Le changement phonétique opéré en L2 n'étant pas conditionné par l'environnement, en toute position, [*k*] est devenu [*g*]. Mais, en L1, un changement phonétique s'est produit, lié à un environnement spécifique: [*k*] est devenu [*h*] entre deux voyelles, alors qu'il s'est maintenu à l'initiale du mot. Le changement phonétique ainsi postulé est plausible puisqu'il est courant qu'une occlusive devienne une fricative, surtout en position intervocalique, alors que le développement inverse est moins fréquent.

Quant aux voyelles, nous trouvons [*u*]-[*u*]-[*u*]-[*ø*] en position finale et [*u*]-[*u*]-[*u*]-[*u*] ailleurs, de même [*i*]-[*i*]-[*i*]-[*ø*] et [*i*]-[*i*]-[*i*]-[*i*]. Il est naturel de reconstruire *[*u*] et *[*i*], en toute position pour la proto-langue, avec un changement phonétique en L4 seulement où [*u*] et [*i*] tombent en position finale. Pour [*a*], nous disposons de trois correspondances: [*a*]-[*a*]-[*a*]-[*ø*] en position finale, [*a*]-[*a*]-[*e*]-[*a*] devant nasale et [*a*]-[*a*]-[*a*]-[*a*] en toute autre position. Pour toutes ces correspondances, nous poserons comme segment originel *[*a*], avec deux changements phonétiques: en L3, [*a*] est devenu [*e*] devant nasale, et, en L4, [*a*] est tombé en position finale.

Ainsi en possession d'une correspondance pour chaque segment et d'une reconstruction pour chaque correspondance, nous résumerons ces acquis en les présentant sous forme de mots reconstruits de la proto-langue:

P-L	L1	L2	L3	L4	Sens
*[*tuka*]	*tuha*	*duga*	*tuka*	*tuk*	"arbre"
*[*rifu*]	*rifu*	*rifu*	*lifu*	*rif*	"mouche"
*[*wani*]	*wani*	*wani*	*weni*	*wan*	"ciel"
*[*rana*]	*rana*	*rana*	*lena*	----	"pierre"
*[*kaka*]	*kaha*	*gaga*	*kaka*	*kak*	"abri"
*[*puti*]	*puti*	*budi*	*puti*	*put*	"éclair"
*[*niri*]	-------	*niri*	*nili*	*nir*	"lance"
*[*fanu*]	*fanu*	*fanu*	*fenu*	*fan*	"ruisseau"
*[*napu*]	*napu*	*nabu*	*napu*	*nap*	"flèche"
*[*mita*]	*mita*	*mida*	*mita*	*mit*	"océan"

Il conviendrait, pour finir, de rassembler pour chaque langue les altérations phonétiques postulées: en L1, [*k*] devient [*h*] entre voyelles; en L2, [*t*] devient [*d*], [*k*] devient [*g*], [*p*] devient [*b*]; en L3, [*r*] devient [*l*], [*a*] devient [*e*] devant nasale; en L4, [*u*], [*i*], [*a*] tombent en position finale.

La tâche qui consiste à opérer la comparaison des langues dans la perspective d'une reconstruction historique est évidemment loin d'être aussi aisée que l'esquisse qu'on vient d'en donner sur un matériau rigoureusement hypothétique pourrait le faire croire. Dans la pratique, le travail s'effectue sur un nombre de segments infiniment plus grand (qu'on songe à ces exemples limites que sont les quatre-vingts consonnes de certaines langues caucasiennes), les formes qu'il convient de rapprocher ne le sont pas de manière aussi évidente, les emprunts sont plus difficiles à identifier, les changements phonétiques plus nombreux et plus complexes et il faut également faire face aux problèmes posés par l'existence de sous-groupes qui permettent souvent des solutions concurrentes. Pourtant, les principes que nous avons rencontrés chemin faisant restent valables pour traiter les données fournies par les langues, dès lors qu'on entend les utiliser pour résoudre les problèmes posés par leur histoire et leur évolution en retraçant leurs relations génétiques.

Pour en savoir plus: outre les travaux d'A. Meillet, J. Manessy-Guitton, P. Thieme, T. Bynon, déjà cités dans la note bibliographique précédente, voir H.H. Hock *Principles of Historical Linguistics* Berlin/New York/Amsterdam 1986, 532-626; W.P. Lehmann *Historical Linguistics: an Introduction* Londres 1992, 141-161; R. Anttila *An Introduction to Historical and Comparative Linguistics* New York 1972, 229-285; R.J. Jeffers I. Lehise *Principles and Methods for Historical Linguistics* Cambridge MA. 1982, 17-54; L. Bloomfield *Language* Chicago 1933, 297-321; T. Crowley *An Introduction to Historical Linguistics* Oxford 1992, 90-132; L. Hjelmslev *Le Langage* [trad. fr.] Paris 1966, 31-54; R.W. Langacker *Language and its Structure. Some fundamental Linguistic Concepts* New York 1968, 199-232.

III. Les découvertes de la comparaison indo-européenne.

Dans l'histoire du développement de la linguistique, c'est d'abord sur le domaine des langues finno-ougriennes[33] que, de manière encore embryonnaire, au cours du dix-huitième siècle et sans qu'il soit question - sinon implicitement - ni de correspondances ni de changement phonétique régulier, le travail comparatif a donné ses premiers résultats. Il faut attendre les premières années du dix-neuvième siècle pour que des allemands, comme Franz Bopp et les frères Jacob et Wilhelm Grimm, et, figure plus isolée et moins connue, le danois Rasmus Rask, abordent un domaine jusque-là inexploré, celui des rapports qu'offrent à l'observateur

33. Dans les travaux de Sajnovics et de Gyarmathi sur les rapports du finnois et du lapon puis du finnois et du hongrois.

attentif les ressemblances manifestées entre elles par les langues classiques (grec, latin), les langues germaniques, les langues slaves, le persan et surtout la langue ancienne de l'Inde, le sanscrit, dont la connaissance se répand alors. Cet ensemble de langues s'est progressivement adjoint les langues celtiques, les langues baltes, l'arménien, l'albanais, puis, dans notre siècle, hittite et tokharien, pour constituer la famille de langues d'abord baptisée indo-germanique, puis indo-européenne.

Ce domaine a donné naissance à une somme d'études considérable et forme aujourd'hui l'ensemble sur lequel nos connaissances sont les plus abouties: les techniques de comparaison utilisées de nos jours pour étudier les relations génétiques éventuelles d'un groupe de langues sont issues des progrès effectués lors de la reconstruction de l'indo-européen. Les "lois phonétiques" en particulier, qui constituent le pivot du travail comparatif, dans ce qu'il a de plus de plus fiable, en sortent directement.

LA LOI DE GRIMM.

R. Rask, puis J. Grimm avaient d'abord remarqué une série de correspondances unissant les langues germaniques au grec, au latin et, plus marginalement, au sanscrit. Sur la base de plusieurs équations, comme celles du nom du "pied" en gotique, *fotus*, avec le latin *pedis*, le grec *podos*, le sanscrit *padas*, "pied", ou, pour les mêmes langues, au sens de "dix": avec *taihun* en got. par rapport à *decem*, *déka*, *das'a*, ces linguistes avaient reconnu que les langues germaniques offraient avec systématicité un *f* là où grec et latin avaient un *p*, et de même, un *p* en face d'un *b*, un *th* en regard d'un *t*, un *t* pour un *d*, et ainsi de suite. J. Grimm[34] proposa alors ce tableau d'ensemble:

Grec	*p*	*b*	*f*	*t*	*d*	*th*	*k*	*g* *ch*
Gotique	*f*	*p*	*b*	*th*	*t*	*d*	*h*	*k* *g*
Haut all.	*b(v)*	*f*	*p*	*d*	*z*	*t*	*g*	*ch* *k*

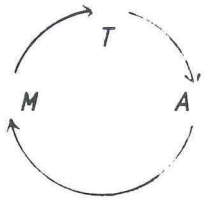
De plus, en définissant la série *p*, *t*, *k* comme *Tenues*, *ph*, *th*, *kh*, et aussi *f*, *th*, *ch* comme *Aspiratae* et enfin *b*, *d*, *g* comme *Mediae*, il donnait de ces changements l'interprétation suivante qui manifestait une régularité remarquable:

---

34. Dans la seconde édition du premier tome de sa *Deutsche Grammatik* en 1822.

Grec	*T*	*M*	*A*
Gotique	*A*	*T*	*M*
Haut all.	*M*	*A*	*T*

Il allait même plus loin en interprétant l'ensemble des développements connus par les langues comme obéissant à un mouvement d'évolution constant. Le changement opéré du grec au germanique le plus ancien se reproduit du gotique à l'ancien haut-allemand: la Tenuis (*T*) devient Aspirée (*A*), puis Médiane (*M*), ce qui peut être figuré de manière circulaire, et surtout considéré comme une loi:



L'APPROCHE NEO-GRAMMAIRIENNE: LES "LOIS PHONETIQUES".

De nos jours, le terme de loi est dénoncé comme abusif pour désigner de telles correspondances et n'est plus conservé que pour identifier les diverses découvertes réalisées par les linguistes. Mais la régularité mise au jour par Grimm produisit un effet considérable sur ses contemporains. Elle ouvrait des perspectives insoupçonnées sur la possibilité d'observer l'évolution historique des langues au moyen de la comparaison. Très vite, pourtant, on trouva à la généralisation hardie avancée par Grimm de nombreuses exceptions, mais l'effort de ses successeurs pour réduire les anomalies rencontrées ne fit que renforcer leur conviction que les changements phonétiques, à condition d'observer finement l'ensemble des données, ne souffraient aucune exception.

On a vu que, selon la loi de Grimm, à un *t* du grec, du latin et du sanscrit correspondait, en gotique, un *th*[35]. Or, dans certains mots apparemment reliés par le sens et la forme, les langues germaniques offrent parfois la même consonne que les autres langues, ainsi pour "il est", on a en sanscrit *asti*, en grec *esti*, en latin *est*, et le gotique *ist* conserve son *t*. L'accumulation d'autres exemples (got. *hafts*, "marié" / lat. *captus*; *fisks* / *piscis*, "poisson"; *nahts* / *nox*, *noctis*, "nuit") permet de réduire cette apparente disparité de traitement en montrant qu'une occlusive sourde

35. Par exemple, pour des mots au sens de "sec, sécher, soif" on a got. *þaursus*, "sec", lat. *torrus torridus*, "sec", gr. *térsesthai*, "devenir sec", scr. *tarsah*, "soif".

se maintient après une fricative germanique. Par observation soigneuse de l'environnement et des caractéristiques phonétiques, l'entorse à la loi de Grimm est vite apparue comme obéissant en fait à une règle particulière.

Une autre anomalie était observable dans des correspondances comme celle qu'offrait, pour le verbe "lier", le gotique *bindan* en face du sanscrit *bandh-*, et, pour "fille", le gotique *dauhtar* par rapport au sanscrit *duhitar*. Si, dans ces exemples, got. *-d-* / scr. *-dh-* est bien conforme à ce que fait attendre la loi de Grimm, les correspondances got. *b-* / scr. *b-* et got. *d-* / scr. *d-* viennent la démentir, car au *b-* du sanscrit devrait répondre un *p-* en germanique, tout comme est attendu un *t-* germanique en regard d'un *d-* du sanscrit. En l'occurrence, la disparité s'explique par le fait que le germanique est un témoignage plus fidèle de l'indo-européen que le sanscrit. Dans un mot gotique, la séquence *b - d* reflète fidèlement la présence dans deux syllabes successives de deux aspirées indo-européennes **bh - dh*. Et c'est plutôt grec et sanscrit qui doivent être taxés d'irrégularité pour avoir connu une dissimilation des aspirées: dans ces deux langues, lorsque deux aspirées se succédaient dans un mot, l'une perdait son aspiration. Ainsi, en sanscrit, le présent à redoublement d'une racine *dhâ-* "poser" est *da-dhâ-mi*, "je pose", et, en grec, avec la racine correspondante *thê-*, on a la forme verbale *ti-thê-mi*. Cette démonstration est à mettre au crédit de H. Grassmann en 1862[36].

Dans la seconde moitié du siècle fut également justifiée la différence de traitement constatée entre le nom du "frère" qui offre une correspondance régulière des consonnes: got. *brothar* / lat. *frater* (*b* / *f* et *th* / *t*) et le nom du "père", quant à la seconde consonne: got. *fadar* / lat. *pater*, où l'on a cette fois *d* / *t*. En 1875, le danois Karl Verner montra que cette correspondance d'un *d* gotique avec un *t* latin cessait de constituer une exception à la loi de Grimm dès qu'on prenait en compte la position de l'accent dans les mots sanscrits correspondants[37]. Il émit l'hypothèse que le sanscrit avait conservé la position originelle de l'accent de mot en indo-européen et que la mutation germanique avait eu lieu avant que l'accent ne se déplace sur la première syllabe du mot en germanique: les aspirées sourdes comme *þ* (ou *th*) (plutôt fricative sourde) (mais aussi *f* et *h*) résultant de la loi de Grimm, sont donc conservées

---

36. "Über die Aspiraten und ihr gleichzeitiges Vorhandensein im An- und Auslaute der Wurzeln", *KZ* 12, 1863, 81-138.
37. "Eine Ausnahme der ersten Lautverschiebung", *KZ* 23, 1876, 97-130.

quand l'accent frappe la syllabe immédiatement précédente: got. *brôPar* (sanscrit *bhrâ'tâ*), mais deviennent sonores dans les autres cas: got. *fadar* (sanscrit *pitâ'*). L'anomalie supposée devenait ainsi parfaitement régulière et sa réduction donna un poids considérable à l'affirmation d'un groupe de linguistes, appelés néo-grammairiens, selon laquelle les changements phonétiques se produisent selon des lois immuables et n'admettent de variation qu'en conformité avec de nouvelles lois.

Le concept de loi phonétique est seulement un postulat de la recherche qui n'est garanti que par une série impressionnante d'observations; il n'a donc finalement qu'un fondement inductif. Or on sait depuis les travaux de K. Popper qu'aucune hypothèse ne peut être vérifiée par induction, sa valeur tenant à ce qu'elle suscite par élargissement et expérience une falsification. Tant que rien de tel ne se produit, elle conserve sa validité. Et, de fait, si l'on tient compte de ce que, parallèlement au changement phonétique aveugle qui dans un espace géographique donné et dans une période donnée affecte un son précis dans un environnement phonétique lui aussi précis, on trouve des changements qui, parce qu'ils dépendent des associations analogiques dictées par le sens ou sont imputables à l'emprunt, viennent contrecarrer les lois phonétiques. Mais celles-ci - avec cette restriction, assurément majeure - peuvent être considérées comme opérant sans exception.

Même si la notion de correspondance phonétique a été élaborée, avec un grand raffinement de détails, sur les langues indo-européennes, sa validité ne se limite pas à ce domaine. Quand par exemple l'américain L. Bloomfield, travaillant sur les langues algonquines, compara les langues fox, cree, menomini et ojibwa, il posa, sur la base d'une seule correspondance - une racine signifiant "rouge", qui offrait en fox et ojibwa *shk*, et en cree et menomini *hk* - l'existence en proto-algonquin d'un groupe *çk*, dans la mesure où, dans chacune de ces langues, le groupe en cause reflétait déjà un ou plusieurs autres groupes du proto-algonquin. Or, étudiant un peu plus tard un dialecte cree jusque-là non décrit, il y découvrit pour la même racine "rouge", le groupe *htk* qui n'était, cette fois, le représentant d'aucun groupe déjà posé en proto-algonquin. Du coup il ne s'agissait plus seulement d'une correspondance isolée et exceptionnelle, mais bien de la conservation, en un seul dialecte, d'un groupe du proto-algonquin qui s'était maintenu là différent des autres. Bloomfield y vit confirmée la productivité du postulat de la régularité du

changement phonétique. Plus tard encore, il reconnut que son informateur ojibwa prononçait en réalité la racine "rouge" avec *sk* et non *shk*: cet *sk* ne pouvant être le reflet d'aucun autre groupe proto-algonquin, le dialecte cree n'était plus seul à avoir conservé le *çk* originel[38].

L'APPROCHE STRUCTURALE: LES LARYNGALES.

La démarche comparative est ainsi toujours le produit d'une interaction entre, d'une part, une pratique fondée sur l'existence de correspondances d'où sont tirées des reconstructions visant au réalisme, car la langue commune que l'on s'efforce de restituer doit répondre aux caractéristiques reconnues aux langues naturelles, et d'autre part une reconstruction interne guidée par des exigences d'ordre théorique. On peut illustrer ce va-et-vient constant par l'histoire de l'émergence des laryngales en indo-européen.

Traditionnellement, étaient reconnus, sur la base de plusieurs correspondances phonétiques, un certain nombre de sons de l'indo-européen commun.

	Indo-ir.	Slave	Grec	Latin	Germanique
1:	*a*	*o*	*a*	*a*	*a*
2:	*a*	*e*	*e*	*e*	*e*
3:	*a*	*o*	*o*	*o*	*a*
4:	*i*	*o*	*a/e/o*	*a*	*a*

Au lieu d'adopter une simple notation chiffrée ne fournissant aucune indication sur la nature probable des sons postulés avoir ces différents reflets dans les langues observées, on peut, dans le symbole choisi pour figurer le son reconstruit, tenir compte des probabilités. Pour 1, on optera pour **a*, puisque de cette manière nous n'avons à envisager un changement que pour le slave; pour 2, ce sera **e*, le changement ne se produisant qu'en indo-iranien, et au lieu de 3, ce sera **o*, ce qui n'implique de changement que pour l'indo-iranien et le germanique.

Notre reconstruction obéit donc à la loi du moindre effort, et, en outre, nous tirons argument du fait que notre **e* produit sur certains sons précédents (comme **k*^{w} et **g*^{w}) des effets dits de "palatalisation" qui sont analogues à ceux produits par le **i*; de même, notre **o* a des conséquences analogues à celles du **u*: nous sommes donc légitimés à reconstruire pour 2

38. C. F. Hockett "Implications of Bloomfield's Algonquian Studies" in D. Hymes. ed. *Language in Culture and Society. A Reader in Linguistics and Anthropology* New York 1964. 599-611.

une voyelle palatale et pour 3 une non palatale. De plus, un **i* et un **u* ayant déjà été reconnus dans le système vocalique reconstruit, nous pencherons respectivement pour le choix d'un **e* et d'un **o*, au lieu de **i* et **u* pour ces séries. Mais la dernière série (4) exige de reconstruire un son qui n'est conservé nulle part, puisque aucune langue ne laisse apparaître dans cette correspondance un son qui ne soit déjà présent dans une autre série. Le symbole *schwa*[39] a été choisi pour indiquer une voyelle qui, comme *i* et *u*, ne peut porter l'accent de hauteur et peut apparaître comme second élément de diphtongue. Pour définir plus précisément sa couleur il convient de tenir compte du grec qui, en offrant trois correspondances différentes *a* / *e* / *o* pour *schwa, a préservé une distinction originelle.

Mais plus intéressant est le fait qu'une approche de type structural est nécessaire pour cerner de plus près ce son dont la nature dépend tout entière de sa position exacte dans le système phonétique de l'indo-européen. Depuis longtemps avait été reconnue l'existence d'une alternance vocalique à fonction morphologique. En grec par exemple, pour des racines à voyelle brève, *legô*, "je dis", et *logos*, "mot" sont des exemples de degré *e* et de degré *o* (Cf. le latin *tego*, "je couvre" / *toga*, "toge"). Aux côtés de ces deux degrés existait aussi un degré zéro dans lequel la voyelle disparaît complètement; ainsi la racine grecque *pet-*, "voler" se trouve sous la forme *pt-esthai* (infinitif aoriste) ou *pt-eron*, "plume", et la racine "laisser" offre les trois degrés *leip- / loip- / lip*.

La structure des racines indo-européennes reconstruites se présente, pour l'essentiel, sous la forme *CVC* (consonne-voyelle-consonne: ex. **pet-*[40]), ou *CVSC* (consonne-voyelle-sonante-consonne: ex. **leip-*[41]), la forme de base de la voyelle dans les deux types étant *e*. Dans des conditions morphologiques précises, ces racines peuvent apparaître au degré zéro, c'est-à-dire sans voyelle *e*, soit *CC* (**pt-*[42]) et *CSC* (**lip-*). On voit que, dans ce dernier cas, la sonante qui était second élément de diphtongue au degré plein (**leip-*) prend alors sa forme vocalique entre deux consonnes (**lip-*). Or, dans des racines montrant au degré plein des diphtongues dont

---

39. Le terme a été emprunté à la grammaire de l'Hébreu où il désigne la voyelle d'appui entre deux consonnes.

40. En sanscrit à la troisième personne du singulier du présent de l'indicatif actif *pat-ati*.

41. On désigne par sonantes les formes variées que prennent, selon leur position, les phonèmes *y, w, r, l, m, n*.

42. A la troisième personne du singulier de l'injonctif en sanscrit, on a ainsi *pa-pt-at*.

le second élément est *m*, *n*, *r*, *l*, il en est de même: au degré zéro, la sonante prend également sa forme syllabique[43] (*m̥*, *n̥*, *r̥*, *l̥*) avec des traitements différents selon les langues. Le phénomène s'éclaire aisément avec quelques données sanscrites[44]:

√ i-e.	Normal	Zéro	Zéro (entre cons.)
**mei-* "fixer":	*mi-may-a*	*mi-my-e*	*mi-ta*
**deik-* "pointer":	*di-des'-a*[45]	*di-dis'-e*	*dis.-ta*
**bher-* "porter:	*ba-bhar-a*	*babhr-e*	*bhr.-ta*
**g^{w}em-* "aller":	*ja-gam-a*	*ja-gm-e*	*ga-ta*[46]

Mais, à ce patron général, un certain nombre de racines bien attestées semblent faire exception: d'une part elles n'ont qu'une consonne, de l'autre, la voyelle de la racine est longue dans des emplois qui sont ceux du degré normal et de forme *ê*, *â*, *ô*. En revanche, dans des contextes morphologiques qui seraient, pour d'autres racines, de degré zéro, on voit apparaître la voyelle brève qui a été reconstruite sous forme de *schwa[47]. On peut illustrer cette alternance avec des exemples sanscrits et grecs:

√ Normal / Zéro i-e.	Scr. et Gr.[48]	Scr. et Gr.[49]
**dhê/dh&* "placer":	*da-dhâ-mi*; *ti-thê-mi*	*hi-ta*; *the-to*
**stâ/st&* "être debout":	*ti-s.thâ-mi*; *hi-stâ-mi*	*sthi-ta*; *sta-to*
**dô/d&* "donner":	*da-dâ-mi*; *di-dô-mi*	*di-ta*; *do-to*

En 1879, le linguiste genevois F. de Saussure fit la suggestion suivante. Selon lui, cette classe de racines offrant une alternance entre voyelle longue et schwa présentait en fait une cohérence structurelle avec le type canonique présenté plus haut, les changements phonétiques ayant ensuite obscurci l'identité originelle. Il proposa en conséquence de réécrire les alternances **â/&*, **ô/&*, **ê/&*, en se conformant au patron précédent. Puisqu'il suffit dans ce cas, pour obtenir le degré normal, d'ajouter la voyelle **e* au degré zéro (ex: *lip-/leip-*), avec des alternances du type: *e / ø*; *ei / i*; *eu / u*; *er / r̥*; *el / l̥*; *em / m̥*; *en / n̥*,

---

43. Des correspondances obligent en effet à poser par ailleurs l'existence de tels sons. En latin, grec, sanscrit et gotique, les numéraux "sept": *septem / hepta / sapta / sibun*; "neuf": *novem / enne(w)a / nava / niun*; "cent": *centum / hekaton / s'atam / hund* qui font reconstruire **septm̥*, **newm̥*, **kntom̥*.

44. Successivement à la première personne du singulier du parfait indicatif actif (degré plein), à la première personne du singulier du parfait de l'indicatif médio-passif et participe passé (zéro).

45. Avec **ei* > *e*.

46. Avec **m̥*. > *a*.

47. Ici représentée par &, pour des raisons de pénurie de caractères typographiques, mais normalement figurée par un *e* renversé.

48. A la première personne du singulier du présent de l'indicatif actif.

49. Au participe passif.

alors, si on suit ce schéma, pour les alternances jusque-là posées sous la forme **â/&*, **ô/&*, **ê/&*, il faut substituer **e&/&*. La forme originelle de l'alternance était **e&/&* et ce *&* a disparu au degré plein entraînant un allongement compensatoire de la voyelle précédente, d'où **e&* > *ê*. Mais comme une seule diphtongue originelle ne peut avoir eu trois traitements distincts, Saussure postula plus d'un *&* pour rendre compte de la différence de qualité des voyelles longues, et, baptisant ces trois segments reconstruits "coefficients sonantiques", préconisa d'opérer avec **e&1* > *ê*, **e&2* > *â*, et **e&3* > *ô*, chacun ayant une propriété différente: *&1* allonge un *e* précédent sans affecter sa qualité; *&2* l'allonge et l'altère en *a*; *&3* l'allonge de même mais l'altère en *o*[50].

La consécration définitive de cette analyse structurale fut obtenue en 1927 lorsque la langue hittite[51], fut examinée de près par le linguiste polonais Jerzy Kurylowicz. Il découvrit que certains morphèmes hittites clairement analogues à ceux d'autres langues indo-européennes présentaient un phonème (transcrit par *hh*) là où l'hypothèse de Saussure faisait attendre un coefficient sonantique, comme dans le hittite *lahh-u* comparé au grec *lâ-os*, "peuple". La preuve acquise en hittite donna du poids à l'hypothèse formulée auparavant selon laquelle les schwas étaient en fait des sons produits au niveau du larynx ou du pharynx, d'où le nom de théorie des laryngales donné à cette analyse et l'adoption d'une notation rendant compte de cette prononciation laryngale: au lieu de **&1*, **&2*, **&3*, **Hə*, **Ha*, **Ho*, ou encore, plus couramment, **H1*, **H2*, **H3*. Cette conclusion est évidemment exceptionnelle en ce qu'une reconstruction uniquement fondée sur des considérations théoriques et sur une analyse interne a fini par trouver une confirmation dans des données factuelles ultérieurement découvertes. On pense évidemment aux anticipations de Leverrier quant à la position de la planète Neptune sur la base de purs calculs, finalement repérée par les observations.

---

50. Accessoirement, cette analyse ouvre de nouvelles perspectives pour analyser des exceptions apparentes. C'est ainsi que le dérivé sanscrit du degré zéro postulé **ste&2tos* est *sthitas*, qui manifeste une aspiration inexpliquée du *t* précédant **&2*, nous pouvons, dans la mesure où ce phénomène n'est pas unique ajouter un nouveau trait phonétique à **&2*, outre sa coloration en *a*, il aspire un *t* précédent.

51. Découverte en Anatolie, non loin d'Ankara, en 1905, puis déchiffrée par le linguiste hongrois B. Hrozny dix ans plus tard et aussitôt identifiée par lui comme langue indo-européenne.

Pour en savoir plus: A. Meillet "Aperçu du développement de la Grammaire comparée" in *Introduction à l'étude comparative des langues indo-européennes* Paris 1937 [reprint Alabama 1964] 453-483; R.H. Robins *A short history of Linguistics* Londres 1987; B. Malmberg *Histoire de la Linguistique de Sumer à Saussure* Paris 1991; W.P. Lehmann *Historical Linguistics: an Introduction* Londres 1992, 23-45; W.P. Lehmann *Theoretical Bases of Indo-European Linguistics* Londres 1993, 1-27; H. Pedersen *The Discovery of Language. Linguistic science in the Nineteenth Century* [reprint Bloomington 1959], 240-310; O. Szemerényi *Einführung in die vergleichende Sprachwissenschaft* Darmstadt 1980, 13-27; W.B. Lockwood *Indo-European philology. Historical and Comparative* Londres 1969, 23-33; N.E. Collinge *The Laws of Indo-European* Amsterdam 1985.

Sur la base des principes exposés plus haut et progressivement constitués en une méthode aux procédures éprouvées, au long du dix-neuvième siècle, des résultats considérables ont été obtenus. Ils constituent, surtout pour ce qui touche à la famille de langues indo-européenne, une des plus belles découvertes de l'esprit humain, car ils autorisent, quant à l'origine de ces langues, des conclusions qu'aucune investigation d'ordre archéologique n'aurait pu susciter.

On peut tenter de donner une idée des regroupements que la méthode comparative a permis d'effectuer entre les langues du point de vue de leur affiliation génétique. Dans ce qu'on sait aujourd'hui, ou plutôt dans les hypothèses que l'on peut faire sur ces diverses familles, la part de la méthodologie qui revient à la comparaison généalogique est prépondérante, mais bien des acquis dépendent aussi de la reconstruction interne, ou encore d'inférences tirées de la typologie [VOIR EXCURSUS: TYPOLOGIE ET UNIVERSAUX.]. C'est le cas même pour l'indo-européen, mais plus encore à mesure qu'on s'éloigne de ce noyau dur de quasi certitudes et qu'on aborde d'autres ensembles linguistiques où un matériel moins abondant, voire d'une nature différente, impose de recourir à d'autres techniques. Passer en revue quelques-uns des regroupements auxquels est présentement parvenue la classification des langues nous fournira l'occasion d'examiner, chemin faisant, d'autres techniques de comparaison que celles élaborées pour le domaine indo-européen.

TYPOLOGIE ET UNIVERSAUX.

La comparaison de plusieurs langues dans une perspective non historique et du point de vue de certaines caractéristiques telles que l'inventaire phonologique, la structure morphologique, les procédés syntaxiques et l'ordre de leurs éléments, peut s'effectuer selon deux directions. On peut soit s'efforcer de montrer en quoi une langue donnée est comme

certaines autres, c'est la démarche typologique classiquement illustrée par la recherche de types morphologiques. Ou encore essayer de montrer en quoi une langue est comme toutes les autres, dans ce cas on tente d'établir des universaux. Cette dernière stratégie a été formalisée et développée par J.H. Greenberg depuis un article fondateur de 1963.

Les universaux linguistiques sont des assertions concernant l'ensemble des langues naturelles et impliquant des hypothèses sur les propriétés universelles du langage humain. Greenberg distingue quatre types d'universaux:

	inconditionnels	implicationnels
absolus	Type A	Type C
statistiques	Type B	Type D

Les universaux du Type A sont des affirmations de forme: "toutes les langues ont la propriété X". Par exemple: "toutes les langues distinguent les voyelles et les consonnes", ou "ont des questions par oui ou par non". Ceux du type B prennent la forme: "presque toutes les langues ont la propriété X"; ex. "99% de toutes les langues ont au moins deux voyelles distinctes". A la place du terme "statistique", on parle souvent dans ce cas de tendances. En dépit des apparences, ces quasi généralisations sont aussi fécondes que les précédentes, dans la mesure où des déviations radicales par rapport à une distribution purement aléatoire présupposent l'existence d'un facteur déterminant à découvrir.

Les universaux de type C, eux, prennent la forme: "pour toute langue, si une langue a la propriété P, elle a aussi la propriété Q.". On définit par là une relation universellement valide entre P et Q. Il n'est pas question de poser que toutes les langues ont P, ni que toutes ont Q, mais seulement ceci: si elles ont P, elles doivent aussi avoir Q. On peut donc prédire la présence de Q à partir de la présence de P. Ce qui signifie qu'un tel universel exclut les langues avec P, mais sans Q, mais que sont autorisées les autres possibilités logiques. D'où:

P	Q	
+	+	oui
-	-	oui
-	+	oui
+	-	non

Un tel universel est, par exemple, "si une langue a le phonème /m/, elle a aussi /n/." S'il est correct, on inférera la présence de /n/ de la présence de /m/, puisque /n/ peut apparaître sans /m/, mais pas /m/ sans /n/. Le phonème /n/ est donc le phonème de base, le plus fondamental des deux.

/m/	/n/	
+	+	oui
-	-	oui
-	+	oui
+	-	non

Un autre universel du même type est "si une langue a comme ordre des mots fondamental VSO (Verbe-Sujet-Objet), alors elle aura des prépositions". Pour qu'il soit valide, il suffit

que l'on ne puisse produire de contre-exemple, ce que serait une langue avec l'ordre VSO mais des postpositions. Enfin, on a les universaux de type D qui sont de même forme, mais leur validité n'est plus absolue; ils constituent de simples tendances portant sur les relations de cooccurrence entre deux propriétés, P et Q. Ainsi, "si une langue a comme ordre des mots fondamental SOV (Sujet-Objet-Verbe), alors elle aura probablement des postpositions". Mais le persan, par exemple, est SOV avec des prépositions.

Plus intéressant, dans certains cas, les universels implicationnels peuvent être ordonnés en séquence. Avec (1) "si une langue a P, elle a aussi Q" et (2) "si une langue a Q, elle a aussi R", on peut ordonner les universaux en série: P implique Q qui implique R. Dans un tel cas R est à considérer comme la propriété la plus fondamentale des trois, et l'on a R > Q > P, où > se lit "plus central que", moins conditionné que", "plus fréquent dans une langue que". On a ainsi une telle hiérarchie entre /n/ > /m/ > /ñ/ quand une suite de propriétés est telle qu'une propriété précédente peut apparaître sans les propriétés qui suivent, mais non l'inverse.

Pour en savoir plus: J.H. Greenberg "Some Universals of grammar with particular reference to the order of meaningful elements" in Greenberg ed. *Universals of language* Cambridge MA. 1963, 75-113; B. Comrie *Language universals and Linguistic Typology* Chicago 1981, 19-22; W. Croft *Typology and Universals* Cambridge 1990, 44-63.

IV. Le groupe des langues indo-européennes: le triomphe de la grammaire comparée.

La famille indo-européenne, dont on donnera en fin de volume une représentation géographique sommaire [voir annexe: les langues indo-europeennes], occupe la majeure partie de l'Europe, et, au-delà du Caucase, où elle est encore représentée, elle couvre l'essentiel de l'Iran, de l'Afghanistan, le nord et le centre de la péninsule indienne, et s'étend jusqu'à Sri Lanka et aux îles Maldives. En outre, la colonisation lui a fait conquérir de nouveaux espaces: le continent américain tout entier, l'Australie, la Nouvelle-Zélande, le sud de l'Afrique, enfin, par la colonisation russe, la Sibérie, le nord de l'Asie. Aujourd'hui une personne sur deux dans le monde parle une langue indo-européenne. On distingue dans cette famille plusieurs groupes.

De tous, l'indo-iranien est celui qui a joué le rôle le plus fondamental. Il a permis de constituer la famille tout entière et de forger la méthode comparative, car c'est lui qui a le moins altéré l'aspect général de la morphologie indo-européenne, préservant par exemple le rôle des racines et la distinction ancienne des huit cas. Aussi le seul rapprochement de l'indo-iranien avec le grec, le latin et le germanique a-t-il suffi pour débloquer la comparaison et édifier la grammaire comparée au siècle dernier. L'indo-iranien comprend deux branches.

Dans la branche indienne, les premiers hymnes en langue védique ont été composés aux alentours de 1.000 avant J.C. Mémorisés et transmis oralement, mais avec une exactitude totale, ils n'ont été notés qu'au dix-huitième siècle de notre ère. La langue qui en est issue, passablement artificielle, codifiée et standardisée par les grammairiens, est le sanscrit qui a fait l'objet d'une description exhaustive vers 400 av. J.C. Cette grammaire, oeuvre de Panini est, par sa précision, son degré d'élaboration et sa concision, un véritable prodige de la création intellectuelle. La qualité de ses descriptions phonétiques, par exemple, ne sera rejointe en Occident qu'avec les recherches expérimentales sur l'articulation qui datent de cent ans à peine. Cette exigence s'explique par des impératifs religieux: une prière mal récitée, par une simple erreur de longueur dans une syllabe, rendait inopérant le sacrifice tout entier[52]. Le sanscrit est ainsi la langue des grandes épopées comme le Ramayana ou le Mahabharata. De nos jours encore, en Inde, plusieurs milliers de locuteurs brahmanes déclarent dans les recensements avoir le sanscrit comme langue maternelle. C'est du sanscrit que, par divers relais, sont issues toutes les langues modernes de l'Inde non dravidienne.

L'autre branche est l'iranienne. Les premiers textes, eux aussi transmis oralement et notés tardivement (à l'époque sassanide) en sont fournis par l'Avesta, livre sacré de la religion zoroastrienne qui remonte à 1000 ans av. J.C. Un autre dialecte, le vieux perse, est celui que nous ont conservé les inscriptions de Darius (roi de 522 à 486 av. J.C.) sur le falaises de Behistan en Iran. Trilingues (avec des versions aussi en akkadien et élamite), elles ont permis le déchiffrement du cunéiforme. Enfin une autre branche n'est représentée que par un parler moderne des montagnes du Caucase, l'ossète.

L'arménien constitue un autre groupe. Parlée dans les montagnes du Caucase et en Turquie, cette langue dont les matériaux les plus anciens datent du cinquième siècle après J.C., au moment de la christianisation, a longtemps été prise pour un simple dialecte de l'iranien. Elle a en effet emprunté de nombreux mots aux parlers parthes et perses du groupe iranien.

Un autre est formé par l'albanais qui n'est connu qu'à partir du quinzième siècle. Cette langue est néanmoins d'utilisation difficile pour

---

52. Or, certains, tels l'aśvamedha ou sacrifice du cheval, duraient une année entière et mobilisaient des moyens matériels considérables.

la comparaison, car le vocabulaire a subi des influences considérables de la part du latin, du grec, du slave, du turc et de l'italien.

Le baltique et le slave sont peut-être à regrouper car ils exhibent de nombreuses ressemblances et ont gardé un aspect archaïque. En particulier, la flexion est restée très riche (on trouve presque autant de cas qu'en indo-iranien), si bien que la phrase a conservé un type ancien. Mais la question reste ouverte de savoir si le développement du baltique et du slave n'a pas plutôt été parallèle avec des influences mutuelles durant leur période de contact.

Dans le (sous-?)groupe baltique, on trouve le vieux prussien, connu par des traductions du seizième siècle, mais qui s'est éteint au début du dix-huitième. Le lituanien dont les premières traductions datent de 1547 est remarquable par ses traits de conservation indo-européens. De nos jours encore bien des formes sont exactement parallèles à des formes sanscrites ou homériques: *esti*, "il est" (Cf. scr. *asti*, gr. *esti*), *gyvas*, "vivant" (Cf. scr. *jivah.*, lat. *vivos*), *sunus*, "fils", *eiti*, "il va" etc. En présentant des structures de type ancien, cette langue fournit des enseignements dont latin et gotique n'offrent pas l'équivalent. Le lette, autre langue du même groupe, se rencontre sous une forme plus altérée.

Le Slave, qui était parlé au sud-est de la Pologne et en Russie de l'Ouest à époque romaine, ne fournit de documents qu'à partir de la christianisation, quand les apôtres Cyrille et Méthode traduisirent l'Evangile dans une langue appelée vieux slave ou vieux bulgare, très proche du slave commun que l'on peut reconstruire. Cette langue, maintenue par l'église russe comme langue officielle, joue le même rôle que le latin pour l'église romaine. On distingue actuellement un sous-ensemble méridional (bulgare, serbo-croate, et slovène), un autre, occidental, (tchèque, slovaque, polonais, et wende) et un oriental (grand-russe, russe blanc, et ukrainien).

La langue grecque constitue à elle seule un groupe d'importance toute particulière, car elle fournit des documents de manière continue sur près de 3500 ans. Les peuples de langue hellène ont commencé à se répandre en Grèce, dans les îles, et en Asie mineure, vers 2000 avant J.C. Jusqu'à une date récente les premiers documents datables ne remontaient qu'au septième siècle avant notre ère, les poèmes homériques étant difficiles à dater, mais le déchiffrement, par des techniques d'inspiration

cryptographique, de milliers de tablettes en écriture syllabique[53] découvertes en Crète et en Grèce au début de ce siècle a permis, en 1952, d'y reconnaître une forme de grec archaïque, dit grec mycénien. Des divers groupes dialectaux, le plus important est bien sûr l'attique, dont l'emploi se généralisa à l'époque hellénistique comme langue commune ou *koinè*, les parlers du grec moderne en étant pratiquement tous issus. Le grec, qui nous est aussi anciennement connu que l'indo-iranien, a certes moins bien conservé la morphologie indo-européenne, mais, en revanche, il a mieux préservé le vocalisme, et pendant longtemps, pour l'essentiel, la reconstruction de l'indo-européen n'a dépendu que du grec homérique et du sanscrit védique.

Le groupe des langues italiques a été introduit dans la péninsule sensiblement plus tard que le grec ne l'a été en Grèce. L'ombrien et l'osque ne nous sont connus que par tablettes et inscriptions antérieures à l'ère chrétienne, mais le latin, attesté par l'inscription dite de Préneste[54] vers 600 av. J.C., a graduellement éliminé tous les autres parlers et par l'effet de la conquête militaire s'est répandu dans tout l'empire romain. Sous sa forme parlée, le latin vulgaire, il forme la base d'où se sont développées toutes les langues romanes. Le latin est d'un type ancien, comme le grec et l'indo-iranien, mais son développement est plus avancé: les systèmes phonétique et morphologique en particulier offrent de nombreux traits nouveaux.

Le groupe celtique présente plusieurs caractéristiques communes avec l'italique (génitif en *-ī* des thèmes en *-o-*, formes du passif et du déponent) et peuvent représenter un même dialecte indo-européen. Nous ne pouvons cependant pas l'affirmer, car rien ne nous a été conservé avant notre ère. Les langues celtiques sont classées en deux sous-groupes, selon leur traitement de la labio-vélaire indo-européenne **k*^{w}: soit en *p* en privilégiant l'arrondissement labial, d'où le nom de P-celtique ou brittonique, soit comme vélaire *q* d'où Q-celtique ou goïdélique. Du gaulois, répandu en Gaule, en Italie du Nord, et jusqu'en Asie mineure, aucun texte étendu n'a survécu. Des autres dialectes brittoniques, le

---

53. Ce qui signifie que le grec est restitué de manière altérée, les groupes consonantiques étant décomposés ou simplifiés: *basileus*, "roi", est figuré par des symboles syllabiques que nous translittérons *qa-si-re-u*; *hiereus*, "prêtre", devient *i-je-re-u*; *knapheus* est *ka-na-pe-u*.
54. MANIOS MED FHEFHAKED NUMASIOI = "Manius m'a fait pour Numérius", mais c'est peut-être l'oeuvre d'un faussaire de génie du siècle dernier.

gallois et le breton[55] sont attestés par des gloses à partir du IXe siècle, et, pour le premier, des textes littéraires à partir du douzième siècle, tandis que le cornique, moins bien connu, s'est éteint au dix-huitième siècle. Du côté gaélique, l'irlandais attesté depuis le quatrième siècle par des inscriptions et qui a connu après le onzième une riche littérature, a du mal à survivre face à l'anglais, mais, sous ses formes les plus anciennes, c'est la seule langue celtique à avoir conservé une flexion riche et archaïque.

Le groupe germanique est, traditionnellement, réparti en trois sous-groupes. Le gotique nous est connu par une traduction de la Bible d'un évêque wisigoth, Wulfila, au IVe siècle[56]. Par son archaïsme et la transparence de sa structure grammaticale, il est d'une importance décisive pour la comparaison. Du germanique nordique nous avons des inscriptions runiques au quatrième siècle et, à partir de l'époque de l'expansion scandinave, on distingue plusieurs dialectes: islandais, norvégien, danois, et suédois, ces trois dernières langues étant mutuellement intelligibles. Le sous-groupe occidental comprend plusieurs dialectes: le haut-allemand, attesté à partir du VIIIe siècle et d'où est sorti l'allemand littéraire, le bas-allemand, représenté aujourd'hui par le néerlandais et le flamand, le frison, et l'anglais dont la forme ancienne, attestée depuis le neuvième siècle, est désignée par le terme de vieil anglais ou anglo-saxon.

Tous ces groupes étaient jusqu'au début du siècle les seuls membres de la famille indo-européenne à avoir survécu. Depuis lors ont été d'abord découverts dans le Turkestan chinois des écrits bouddhiques du VIIe siècle écrits dans une langue baptisée tokharien qui, par sa morphologie et son vocabulaire, est clairement indo-européenne, avec des traits archaïques.

Enfin, des fouilles effectuées de 1905 à 1907, en Turquie, à Bogaz-köy ont livré les archives de l'empire hittite (1900-1200 av. J.C.) sous forme de milliers de tablettes en cunéiforme syllabique. Parce qu'elle s'accompagnait souvent d'akkadien et de sumérien la langue hittite a pu être identifiée comme indo-européenne. A cause de ses archaïsmes, en particulier de l'existence de sons transcrits comme $\underset{\sim}{h}$, identifiés aux reconstructions de Saussure, on a suggéré que le hittite était une langue-

---

55. Le breton, contrairement à une conviction ancienne, n'est pas une survivance du gaulois, puisqu'il a été importé sur le continent en provenance de Grande-Bretagne au moment de la conquête saxonne.

56. On retrouve cette langue parlée en Crimée au seizième siècle.

soeur plutôt que fille de l'indo-européen, ce qui conduirait à poser une famille Indo-hittite. Mais ces traits archaïques peuvent aussi s'expliquer par une séparation précoce des locuteurs hittites de la communauté indo-européenne, suivie par une très longue période de séparation. De plus, la plus grande simplicité de la morphologie hittite par rapport à l'indo-iranien et au grec peut amener à considérer que les systèmes verbaux complexes de ces langues sont moins un héritage qu'une innovation.

Pour en savoir plus: J. Manessy-Guitton "L'indo-européen" in A. Martinet éd. *Le Langage* Paris 1968, 1240-1287; P. Baldi *An Introduction to the Indo-European Languages* Carbondale Ill. 1983; B. Comrie et al. *The World's Major Languages* New York/Oxford 1990, 31-565; W.B. Lockwood *Indo-European Philology. Historical and Comparative* Londres 1969; W.B. Lockwood *A Panorama of Indo-European Languages* Londres 1972; A. Meillet et M. Cohen *Les Langues du Monde* Paris 1952, 3-80; W. Cowgill "Indo-european languages" *Encyclopaedia Britannica* 1974 Vol.9, 431-438; A. Martinet *Des steppes aux océans. L'indo-européen et les "Indo-européens"* Paris 1986; A. Meillet *Introduction à l'étude comparative des langues indo-européennes* Paris 1937 [reprint Alabama 1964], 51-81; W.P. Lehmann *Historical Linguistics: an Introduction* Londres 1992, 67-83; B. Comrie "Languages of the World: Who speaks what" in N.E. Collinge ed. *An Encyclopaedia of Language* Londres 1990, 956-983; M. Ruhlen *A Guide to the World's Languages, Vol. I: Classification* Stanford CA. 1987, 24-64.

Pour caractériser la nature des relations entre ces langues fut introduit par A. Schleicher en 1871 le concept d'arbre généalogique qui reflétait l'intérêt du temps pour la théorie évolutionniste. Il en applique l'hypothèse touchant au développement des différentes espèces à l'évolution des langues-filles à partir d'un ancêtre commun: des séparations se produisent à des états premiers de manière relativement homogène suivies de périodes de développement pendant lesquelles des changements se produisent occasionnant de nouvelles séparations. Ce modèle suscita la critique, essentiellement parce qu'il adoptait comme principe implicite que si une langue diverge en deux ou plusieurs langues-filles, les locuteurs poursuivent leur chemin sur le plan linguistique et physique de manière isolée, sans contact avec les autres langues. Or, des relations linguistiques transversales étaient nettement visibles entre toutes les langues concernées.

DIALECTES ET ISOGLOSSES DE L'INDO-EUROPEEN.

Très tôt, on a tenté de pratiquer entre les différents groupes ainsi bien identifiés qu'on vient de passer en revue des regroupements plus vastes, censés refléter une parenté spécifique unissant plus étroitement

certains d'entre eux. La principale est celle qu'on a cru pouvoir poser sur la base du traitement des occlusives vélaires entre deux grands groupes. D'un côté en effet, pour le sanscrit, l'avestique, l'arménien, le vieux slave et le lituanien, on trouve des spirantes, là où grec, latin, irlandais, et gotique ont conservé des vélaires. Ainsi, pour "dix", on a sc. *das'a* / av. *dasa* / arm. *tasn* / v.sl. *deseti* / lit. *dêsimt*, en face de gr. *déka* / lat. *decem* / v. ir. *deich* / got. *taihun*; de même, pour "cent", on constate sc. *s'atam* / av. *satem* / v.sl. *suto* / lit. *simtas*, qui répondent à gr. *hékaton* / lat. *centum* / v. ir. *cêt* / got. *hund*. Les langues orientales ont été alors baptisées *satem*, d'après la forme avestique pour "cent", tandis que les langues occidentales étaient étiquetées langues *centum*. On pensait que, préalablement à la différenciation dialectale déjà retracée, les locuteurs indo-européens s'étaient d'abord séparés en deux groupes, dont l'oriental avait connu un changement phonétique absent du côté occidental. Cette inférence n'est plus actuellement avancée car la distinction ainsi posée devrait se retrouver dans d'autres caractéristiques différenciatrices affectant globalement chacun des deux groupes putatifs, ce qui ne se produit pas. Il ne s'agit que d'une isoglosse parmi d'autres et toutes celles qui existent offrent plutôt l'image d'un entrecroisement de réseaux tous partiellement superposables seulement. En outre, le tokharien a préservé les palatales devant les voyelles d'arrière alors que les autres langues orientales les ont assibilées, ce qui a achevé de ruiner la dichotomie qu'on avait cru, un temps, pouvoir poser.

Malgré ces doutes, l'existence d'isoglosses contredisait l'organisation des relations entre les langues sous la seule forme d'un arbre aux ramifications successives et il fallut élaborer des modèles plus complexes pour rendre compte du développement génétique [VOIR EXCURSUS: LA THEORIE DES ONDES].

LA THEORIE DES ONDES.

Lorsqu'on a découvert, à la fin du dix-neuvième siècle, que des isoglosses pouvaient redécouper des limites linguistiques bien établies, l'une des propositions, de J. Schmidt (*Die Verwandtschaftsverhältnisse der indogermanischen Sprachen* Weimar 1872) fut de substituer à la théorie de l'arbre la théorie des ondes. Sa thèse partait de l'observation suivante: quand on compare des langues apparentées sur des traits spécifiques, on constate généralement une corrélation entre la distance géographique et la distance linguistique . Le terme d'onde se réfère au fait que les innovations linguistiques se répandent dans une aire géographique donnée, comme des vagues créées par une pierre jetée dans l'eau, de leur point d'origine à la

périphérie. Le changement perd lentement sa force d'expansion: plus il va loin, moindre est son effet, car il vient rencontrer les vagues créées par d'autres innovations. Dans l'ensemble des locuteurs indo-européens, des traits nouveaux apparaissent, se développent et se répandent dans des aires particulières, par vagues qui touchent certaines autres langues et, plus rarement toutes les autres. Ces ondes dialectales, roulant en toutes les directions, à partir de différents centres, expliquent que, sur un territoire donné on puisse trouver des caractéristiques communes avec un groupe voisin, d'un côté, et avec un autre groupe voisin, d'un autre côté. C'est ce qui permet d'expliquer pourquoi les "branches" de l'indo-européen forment, d'une certaine façon, une chaîne ininterrompue L'image qui en résulte n'est plus celle de l'arbre, mais de cercles imbriqués les uns dans les autres en intersections multiples. Elle a le mérite de traiter de manière plus flexible les interrelations entre langues, au travers de ressemblances spécifiques, et les changements communs qui les ont affectées. Mais ce modèle est surtout une description synchronique qui fournit une photographie des réalités géographiques, sans préciser le moment où tel ou tel changement est intervenu, il échoue donc à offrir aussi une perspective historique.

Pour en savoir plus: W. P. Lehmann *Historical Linguistics: An Introduction* Londres 1992, 119-125; T. Bynon *Historical Linguistics* Cambridge 1977, 192-195; N.E. Collinge "Language as it evolves: Tracing its Forms and Families" in Collinge ed. *An Encyclopaedia of Language* Londres 1990, 876-916; J.P. Mallory *In Search of the Indo-Europeans. Language, Archaelogy and Myth* Londres 1989, 14-22; R. Anttila *An Introduction to Historical and Comparative Linguistics* New York 1972, 300-309.

On considère aujourd'hui que les deux modèles, celui, historique, de l'arbre et celui, spatial, des ondes, illustrent des propriétés présentes simultanément à tous les niveaux d'une classification génétique, les deux hypothèses, complémentaires, ne sont pas mutuellement exclusives.

LE PROTO-INDO-EUROPEEN.

Le travail patient des comparatistes a débouché sur une reconstitution partielle de la proto-langue posée au départ des langues indo-européennes, dont on peut présenter quelques caractéristiques.

Le système phonétique traditionnellement reconstruit était riche en consonnes occlusives: une série non-voisée *p*, *t*, *k*, *k*^{w}, une autre voisée *b*, *d*, *g*, *g*^{w} et une série de voisées aspirées *bh*, *dh*, *gh*, *g*^{w}*h* (prononcées comme les voisées mais suivies d'une expiration). Sur la base de certaines formes on a suggéré de postuler l'existence d'une série d'aspirées non-voisées *ph*, *th*, *kh*, au moins pour le versant dialectal d'où sont issus grec, arménien, et indo-iranien. En revanche, la langue était pauvre en fricatives, n'offrant que *s*, voisé en *z* devant les consonnes occlusives voisées.

Ce système reconstruit a été soumis à des critiques diverses. La plus décisive s'appuie sur le fait qu'il semble venir en conflit avec les modèles phonologiques connus. En particulier, la présence d'occlusives voisées aspirées mais non d'une série correspondante d'aspirées non-voisées viole le principe selon lequel les systèmes reconstruits doivent être caractérisés par les mêmes régularités que celles manifestées dans les systèmes historiquement attestés

Pour pallier cette difficulté, et d'autres encore, on a alors proposé un système d'occlusives distinguant des occlusives non voisées / voisées / glottalisées; les occlusives glottalisées ou éjectives (*p'*), *t'*, *k'*, *k'*^{w} remplacent les occlusives traditionnelles simples voisées (*b*), *d*, *g*, *g*^{w}. Les occlusives voisées et non voisées ont chacune des allophones aspirées, chaque langue indo-européenne ayant généralisé l'un ou l'autre des allophones.

Il faut ajouter à ce système la reconstruction de trois laryngales partiellement préservées en hittite[57]. Ailleurs, leur présence ne peut être que déduite par des preuves indirectes. Le proto-indo-européen avait en outre deux nasales, *n* et *m*, deux liquides, *r* et *l*, et les glides *w* et *y*. Ces sons pouvaient fonctionner aussi bien comme consonnes que comme voyelles, ce dernier rôle étant symbolisé par des *r̥*, *l̥*, *m̥*, *n̥*, sous-pointés[58] ; dans le cas de *y* et *w*, les contreparties vocaliques étaient *u* et *i*. De même, les laryngales, on l'a vu à propos des trouvailles de Saussure, fonctionnaient à la fois avec la valeur consonantique de sons comme *h* et, vocalique, de variétés de *schwa*. Les autres voyelles étaient *e*, *a*, *o*, qui, comme *i* et *u*, pouvaient être longues ou brèves. Dans la mesure où l'on peut reconstruire des étapes chronologiques, les voyelles longues de l'indo-européen tardif peuvent être considérées comme résultant de la contraction de voyelles brèves de l'indo-européen primitif avec un *schwa*. Déjà en proto-indo-européen, deux des trois laryngales avaient la propriété de colorer une voyelle fondamentale *e* adjacente. La racine **pâ-*, "protéger", est donc contractée d'un plus ancien **pa&-*, avec un *schwa* colorant en *a*; la racine **dô-*, "donner", est contractée d'un plus ancien **do&-* avec *schwa* colorant en *ô*, et enfin, la racine **dhê-*, "poser", de **dhe&-*, avec un *schwa* sans coloration[59] ; la voyelle fondamentale, dans tous ces exemples, étant

---

57. Par exemple dans *pah(s)-*, "protéger" de **pa&(s)-*, qui se retrouve dans le latin *pastor*.

58. Le son devait être analogue à celui de la syllabe finale de l'anglais *bottle, butter, bottom, button*.

59. On symbolise la laryngale non colorante par *&1* (ou *h1*), donc **dhe&1*, celle qui colore en *a*

originellement *e*. Ajoutons enfin que le ton jouait un rôle grammatical important.

Les alternances vocaliques, encore appelées apophonie ou *ablaut*, étaient une des caractéristiques fondamentales du fonctionnement morphologique de l'indo-européen. Dans ces changements de voyelle interne, la forme fondamentale était en *e*, mais pouvait apparaître comme *o* dans certaines conditions, tandis que, dans d'autres conditions, *e* et *o* pouvaient disparaître entièrement: on parle alors de degré *e* ou degré plein, de degré *o*, et de degré *zéro*. De plus, *e* et *o* pouvaient apparaître au degré long, sous la forme *ê* et *ô*. Ainsi, la racine **ped-*, "pied", apparaît au degré *e* dans le latin *ped- (pes, pedis)*; au degré *o* dans le grec *pod- (pous, podos)*, tandis que le germanique **fotuz* (d'où est issu l'anglais *foot*) reflète le degré *o* long, **pôd-*; le degré zéro, lui, qui ne fait apparaître aucune voyelle (**pd-*) se trouve en sanscrit. Quand le degré zéro affectait une racine avec l'un des sons *m*, *n*, *r*, *l*, *w*, *y*, la sonante apparaissait sous sa forme vocalique et formait une syllabe: on a par exemple pour la racine **seng^{w}h-* (> angl. *sing*), le degré *o *song^{w}h* (> *sang*) et le degré zéro **sn̥g^{w}h* (> *sung*).

Le proto-indo-européen ou PIE était une langue offrant un haut degré de flexion: les relations grammaticales et les fonctions syntaxiques étaient indiquées par des variations dans les finales des mots. Les noms avaient des désinences différentes pour les différents cas (au nombre de huit: outre nominatif, génitif, datif et accusatif, il y avait un instrumental, un locatif, un ablatif, un accusatif et un vocatif), et les différents nombres (singulier, pluriel, et un duel - déjà en déclin - pour les objets se présentant par paire). Une même désinence pouvait exprimer plusieurs catégories. Les adjectifs offraient la même morphologie que les noms et s'accordaient avec eux.

Les pronoms distinguaient une première, deuxième et troisième personne: la première et la deuxième offraient singulier duel et pluriel mais sans distinction de genre, tandis que la troisième distinguait masculin, féminin et inanimé, au singulier et au pluriel. Ce système s'est, apparemment, développé à partir d'un état antérieur où s'opposaient seulement animé et inanimé.

---

par *&2* (ou *h2*), donc ***pa&2***, et celle qui colore en *o* par *&3* (ou *h3*), donc ***do&3***.

Les verbes avaient des désinences différentes pour les différentes personnes (première, deuxième, troisième) et les nombres (singulier, pluriel, duel), pour les diathèses active et médio-passive, ainsi que des suffixes pour une grande variété de distinctions aspecto-temporelles, quatre modes (ind. subj. cond. et impér.) et des catégories telles que causatif-transitif et statif-intransitif. Un aspect particulier de la formation des verbes tenait au changement vocalique radical pour marquer les différentes formes grammaticales (*ablaut*), bien conservé dans les langues germaniques (verbes anglais dits irrégulier et verbes forts de l'allemand). La liberté syntaxique était très grande et l'ordre des mots libre.

La structure de tous les mots fléchis était la même: racine + un ou plusieurs suffixes + désinence: **ker-wo-s*, "cerf" (> lat. *cervus*) avec la racine **ker-* désignant la corne, le suffixe nominal *-wo-* et la désinence de nominatif singulier *-s*. L'ensemble racine + suffixe constitue le thème, et les thèmes forment le stock lexical de base de l'indo-européen. La racine contient la notion sémantique de base que le suffixe peut faire varier en déterminant la partie du discours du mot. La racine **prek-*, "demander", peut former, selon le suffixe, un verbe, **prk-sko-*, "demander" (> lat. *poscere*) un nom, **prek* (> lat. *precês*, "prière") et un adjectif **prok-o-*, "qui demande" (> lat. *procus*). Les suffixes avaient des valeurs diverses et encore plus spécifiques, pour transformer les noms en verbes, pour marquer différents types d'actions (transitive et intransitive), pour former des noms d'agent, des noms abstraits, des noms verbaux, des adjectifs verbaux etc.

Pour en savoir plus: J. Haudry *L'indo-européen* QSJ n° 1798, Paris 1980; J. Manessy-Guitton "L'indo-européen" in A. Martinet éd. *Le Langage* Paris 1968, 1240-1287; P. Baldi "Indo-European Languages" in B. Comrie et al. *The World's Major Languages* New York/Oxford 1990, 33-67; W.B. Lockwood *Indo-European Philology. Historical and Comparative* Londres 1969; A. Martinet *Des steppes aux océans. L'indo-européen et les "Indo-européens"* Paris 1986; A. Meillet *Introduction à l'étude comparative des langues indo-européennes* Paris 1937 [reprint Alabama 1964]; W.P. Lehmann *Theoretical Bases of Indo-european Linguistics* Londres 1993; O. Szemerényi *Einführung in die vergleichende Sprachwissenschaft* Darmstadt 1980; M. Ruhlen *A Guide to the World's Languages, Vol. I: Classification* Stanford CA. 1987, 24-64.

LES "INDO-EUROPEENS".

La reconstruction d'une protolangue, parce qu'elle porte à la fois

sur les phonèmes, la morphologie, le vocabulaire essentiel, peut être poussée plus ou moins loin. Dans certains cas la confiance des auteurs dans la validité de leurs triangulations les a conduits à écrire en proto-indo-européen une fable intitulée "le mouton et les chevaux" [VOIR EXCURSUS: UNE FABLE INDO-EUROPEENNE]. Mais, surtout, dans une autre perspective, une telle langue implique avec elle l'existence d'une société préhistorique qui la parle. La connaissance du lexique indo-européen nous introduit à une compréhension de la culture des locuteurs de cette langue qui va au-delà de ce que pourraient nous apprendre, si nous en disposions, des témoignages archéologiques, en ce que le vocabulaire ne se borne pas aux réalités matérielles.

UNE FABLE EN INDO-EUROPEEN.

Version d'A. Schleicher en 1868:

Avis akvasas ka

Avis, jasmin varna na a ast, dadarka akvams, tam, vagham garum vaghantam, tam, bharam magham, tam manum aku bharantam. Avis akvabhjams a vavakat; kard aghnutai mai vidanti manum akvams agantam.

Akvasas a vavakant krudhi avai, kard aghnutai vividvant- svas: manus patis varnam avisams karnauti svabhjam gharmam vastram avibhjams ka varna na asti.

Tat kukruvants avis agram a bhugat.

Version de W. Lehmann et L. Zgusta en 1979:

[G^{w}&rei] owis, k^{w}esyo wlhna ne est, ekwons espeket, oinom ghe g^{w}rum woghom weghontm, oinomk^{w}e megam bhorom, oinomk^{w}e ghmenm oku bherontm. Owis nu ekwobh(y)os ewewk^{w}et: Ker aghnutoi moi ekwons agontm nerm widntei.

Ekwos tu ewewk^{w}ont: Kludhi, owei, ker aghnutoi nsmei widntbh(y)os: ner, potis, owiom r wlhnam sebhi g^{w}hermom westrom k^{w}rneuti. Neghi owiom wlhna esti.

Tod kekluwos owis agrom ebhuget.

Traduction littérale

(Le) mouton et (les) chevaux

[Sur (une) colline] (un) mouton sur lequel laine pas était, vit (des) chevaux, (l')un (un) chariot lourd tirant, un (autre) (une) charge grande, un (autre) (un) homme rapidement portant. (Le) mouton aux chevaux dit: coeur souffre moi voyant (un) homme (des) chevaux conduisant.

(Les) chevaux au mouton dirent: écoute mouton, coeur souffre nous voyant homme, (le) maître, (qui) laine du mouton fait pour soi (un) chaud manteau et au mouton laine pas est.

Ceci ayant entendu, (le) mouton dans la plaine s'enfuit.

Pour en savoir plus: W. Lehmann et L. Zgusta "Schleicher's Tale after a Century" in Bela Brogyanyi éd. *Festschrift for Oswald Szemerényi on the Occasion of his 65th Birthday* Amsterdam 1979, 455-466.

Que l'on ait pu reconstruire les mots du PIE[60] qui signifiaient "jour", "année", "hiver", "printemps", "été", "automne", "mois / lune", "soleil" "étoile" ne nous dit rien sur les coutumes des locuteurs. En revanche, le fait qu'il y ait eu un mot pour la neige (*sneigwh-) nous dit quelque chose sur le climat de la région où la proto-langue était parlée et des mots pour "hêtre", "pomme", "cerise" révèlent aussi une origine nordique. De même "loup", "castor", "souris", "saumon", "anguille". Les noms de plusieurs animaux domestiques nous renseignent sur le mode de vie ("vache", "mouton", "agneau", "chèvre", "porc", "chien", "cheval"). Des pratiques agricoles nous sont suggérées par des mots pour "grain", et un autre qui peut avoir signifié "blé", deux mots pour "moudre", d'autres pour "sillon", "charrue", "joug", "faucille", "cueillir".

Un mot signifiant "cuivre" et d'autres, moins sûrs, pour "bronze", "or", "argent" ont été attribués au peuple parlant le proto-indo-européen et impliquent une connaissance de la métallurgie. Mais, faute d'un mot pour le fer on a vu en ce peuple des fermiers de l'âge du Bronze ignorant encore l'usage du fer (entre -3000 et -2000). L'existence d'un mot pour "axe, essieu" qui semble tiré du nom du "nombril", et pour "roue", tiré d'une racine signifiant "tourner" fait penser que les véhicules à roues étaient une innovation récente.

Des termes de parenté peuvent également être reconstruits, en particulier du point de vue de la femme: ceux des parents par alliance avec son époux, comme "père, mère, frère, soeur du mari", "femme du frère du mari". Ce n'est pas le cas pour les parents par alliance de la femme, ce qui implique l'existence d'une famille élargie, patriarcale et patrilocale, la fille intégrant par mariage la famille de son mari qui, lui, reste sans lien avec la famille de son épouse.

Le langage reflétant les concepts essentiels autour desquels tournent les relations socio-économiques et l'idéologie sous-jacente à la structure interne de la société, il est possible de décrire, comme E.

60. On trouvera rassemblées sous une forme commode, les racines indo-européennes désignant ces divers concepts dans l'introduction de Calvert Watkins à son dictionnaire des racines (voir bibliographie de la section).

Benveniste l'avait entrepris, le système des relations personnelles en jeu dans la structure familiale, clanique et tribale des Indo-européens, de définir le système de parenté, de spécifier les obligations contractuelles impliquées dans l'échange des services et des biens, de poser les droits et devoirs des gouvernants, de retracer les manières d'honorer les dieux. De même G. Dumézil a reconnu la hiérarchisation de la société en trois niveaux fonctionnels (du prêtre, du guerrier, et du producteur) qui éclaire la mythologie indo-européenne.

Tout autre est la stratégie qui consiste à relier ces données linguistiques et les concepts qu'elles impliquent avec des realia, des données de fait touchant à la géographie ancienne ou à l'archéologie préhistorique [VOIR EXCURSUS: LA PALEONTOLOGIE LINGUISTIQUE]. En relève en particulier ce pan de la recherche qui entend, sur la base des reconstitutions auxquelles le lexique peut donner lieu, restituer l'emplacement occupé par les indo-européens primitifs, en entendant par là la communauté parlant la langue indo-européenne commune antérieurement à sa dispersion.

LA PALEONTOLOGIE LINGUISTIQUE

Elle consiste à faire des déductions sur la culture, le mode de vie, et la localisation géographique des peuples ayant parlé la langue supposée commune d'où sont issues les langues indo-européennes, à partir de preuves linguistiques fournies par sa reconstitution.

L'idée de base est que si on peut prouver que des plantes ou des animaux précis avaient une extension aux contours précis et limités, lorsqu'il est possible de reconstruire un mot se référant à telle plante ou animal particulier, il sera possible de désigner avec précision l'aire où vivaient les locuteurs du proto-indo-européen. Si l'on pose que, selon toute probabilité, la proto-langue était parlée dans une des aires aujourd'hui occupée par l'une des langues indo-européennes, il est clair que nous éliminerons d'emblée celles dont les plantes et les animaux les plus caractéristiques n'ont pas de désignation commune qui se puisse reconstruire à partir d'un grand nombre de langues. C'est le cas de l'Inde (rien en PIE pour le tigre, l'éléphant, le singe) de l'Iran (rien pour le chameau, l'âne, le lion), les régions méditerranéennes (rien pour l'âne, le lion, l'olive, le vin, le cyprès).

On est conduit à penser au sud de la mer Baltique ou au nord de la mer Noire, d'autant plus que cette région offre à période historique la plus forte concentration de langues indo-européennes les plus diverses (germanique, balte, slave). Or le système lexical pour arbres et plantes que nous pouvons reconstruire s'accorde avec cette région du sud de la mer Baltique: "orge", "bouleau", "hêtre", "tremble", "chêne", "if", "saule", "sapin", "épicéa", "aulne". Le hêtre par exemple ne franchit pas une ligne qui va de Königsberg à

Odessa. Le seul arbre qui se trouve en Inde, le bouleau, y porte un nom indo-européen, *bhûrja*, ce qui n'est le cas d'aucun autre arbre figurant en sanscrit. Quant aux animaux dont le nom est reconstruit, le tableau est de même dépourvu de contradictions. Parmi les animaux domestiques ou de proie, aucun ne fait défaut de ceux dont nous avons par ailleurs la preuve qu'ils ont vécu à l'âge de Bronze dans les plaines de la Baltique et du nord de l'Allemagne.

Un cas particulièrement intéressant est celui du saumon. On le trouve comme le bouleau le hêtre et la tortue dans les rivières qui se jettent dans la Baltique. Dans la mesure où le mot **laks* signifie "saumon" en germanique et en baltique et "poisson" en tokharien, cette distribution suggère une origine nordique. Or, en sanscrit, une forme *laks*, au sens de "cent mille", peut être interprétée comme une extension du caractère innombrable des bancs de saumon. Le fait que le mot se trouve en tokharien et dans le domaine indo-iranien où ce poisson est inconnu prouve selon Thieme une origine germano-baltique. L'argument est le même pour le nom du hêtre ou de la tortue: si l'on trouve ces racines dans des langues indo-européennes où ces données naturelles font défaut la région d'origine est là encore pointée de manière plausible.

On localise ainsi par recoupements l'espace où était parlé l'indo-européen: à l'intérieur du domaine des rivières à saumon, à l'ouest de la ligne du hêtre, et hors de la Scandinavie à cause de la présence de la tortue, soit dans la région de la Vistule, de l'Oder et de l'Elbe, au point où des représentants de langues indo-européennes orientales (balto-slave) et occidentales (germanique) viennent se rencontrer.

Mais les opposants à cette construction rétorquent que l'existence d'une ligne du hêtre il y a 5000 ans n'a rien de vraiment probant. De plus, il est courant que des noms anciens soient transférés sur d'autres objets à l'occasion d'un contact avec un environnement différent; **bhâgos* (reconstruit à l'aide du grec du latin et du germanique) peut très bien avoir désigné initialement d'autres arbres que le hêtre qu'il désigne dans certains dialectes où, loin de constituer un archaïsme il constituerait plutôt une innovation commune par glissement sémantique ou encore un emprunt lexical à date ancienne.

Pour en savoir plus: P. Thieme: "The Indo-European Language" *Scientific American* Oct. 1958 63-74; H.H. Hock *Principles of Historical Linguistics* Berlin/New York/Amsterdam 1986, 573-578; P. Baldi "Indo-European Languages" in B. Comrie et al. *The World's Major Languages* New York/Oxford 1990, 63-66; A. Martinet *Des steppes aux océans. L'indo-européen et les "Indo-européens"* Paris 1986.

Outre flore, faune, type d'habitat et usages religieux, on a voulu déduire aussi l'emplacement du point de départ de la dispersion, sur la base de certains mots ("bouleau" ou "saumon") supposés désigner une espèce particulière. mais la possibilité que celle-ci corresponde spécifiquement à une aire géographique stable à travers le temps et donc projetable sur l'époque préhistorique est bien sujette à caution. C'est pourquoi, au gré des théories, ce foyer a été placé en Inde, en Asie, au sud de la Russie, en Scandinavie, Lituanie, Hongrie, Allemagne du Nord et en Afrique.

Parmi toutes les théories rivales, une vision séduisante est celle de la linguiste d'origine lituanienne Marija Gimbutas, pour qui le groupe ethnique proto-indo-européen correspond à la civilisation pastorale et patriarcale dite des Kourganes[61] que l'on trouve dans les steppes du Pont et de la Volga, au sud de la Russie, à l'est du Dniepr, au nord du Caucase et à l'ouest de l'Oural. Pour son auteur, les témoignages archéologiques d'une domestication du cheval, les véhicules, les modes d'habitat, la structure sociale et la religion s'accordent avec ce qu'on peut reconstruire du proto-indo-européen. Par vagues successives, cette culture se serait d'abord (4000-3500) répandue progressivement au nord du Pont, à l'ouest de la mer Noire en Ukraine, Roumanie, Yougoslavie et Hongrie, introduisant dans cet espace jusque-là néolithique une culture nouvelle particulièrement mobile et agressive, puis (3500-3000), elle aurait atteint l'Europe centrale les Balkans, et par une autre route, la Transcaucasie, l'Anatolie et le nord de l'Iran; enfin le nord de l'Europe, la partie au nord du Danube, la Grèce, l'Anatolie occidentale et l'est de la Méditerranée. Mais de ces mouvements des peuples des Kourganes, seul le premier est compatible avec une date plausible pour le PIE, suffisamment haute pour permettre à une seule langue de se développer sous des formes aussi divergentes que le grec mycénien et le hittite, l'un et l'autre attestés au milieu du second millénaire.

Pour des raisons diverses[62], des linguistes soviétiques V.V. Ivanov et T.V. Gamkrelidze ne voient dans cette localisation que le point de départ de la diversification du stock d'Indo-européens tardifs en Baltes, Slaves, Celtes, Germaniques, et Italiques, mais ils posent un ensemble antérieur (au cinquième et quatrième siècle), en Asie, quelque part entre la Turquie de l'est, le nord de l'Irak et le sud du Caucase, séparé en Hittites, Arméniens, Hellènes et Indo-iraniens. A quoi M. Gimbutas répond que le mouvement ainsi décrit peut très bien avoir été inverse.

La question reste pendante de savoir pourquoi un peuple parmi ceux de l'âge du Bronze, qui occupait une position périphérique, a pu imposer sa langue sur une aussi vaste étendue. De quels pouvoirs particuliers pouvait il disposer?. C. Renfrew, sur la base de l'archéologie et d'un modèle démographique de diffusion, voit un point de départ en Anatolie au septième

---

61. C'est le nom russe pour les tombes à tumulus, caractéristiques de cette culture.
62. Les liens de l'i.e. et des langues du sud du Caucase, la domination hittite dans le centre de l'Anatolie, l'extension de certains mots comme ceux pour le vin et la chèvre.

millénaire, répandu en Europe en y introduisant les techniques agricoles. Dans cette hypothèse, c'est la culture qui aurait donné à ces peuples un avantage technologique tel qu'eux et leur langue auraient pu absorber ou déplacer des populations préexistantes par un mouvement vers le nord et l'ouest à travers l'Europe en même temps que la culture.

Pour beaucoup de linguistes, 6000 av JC paraît cependant une date trop haute pour l'ancêtre commun des langues indo-européennes, car les premiers fermiers ne connaissaient pas le cuivre, ni la roue, ni les chevaux et ne disposaient pas de mots pour les désigner. Il faudrait alors supposer que ces mots ne feraient pas partie du vocabulaire indo-européen originel mais se seraient répandus par emprunt ultérieur, conformément à la théorie des ondes. Un archéologue anglais, J. P. Mallory, plus récemment, opte pour une aire Ponto-Caspienne.

Pour en savoir plus: J. Haudry *Les indo-européens* QSJ n°1965, Paris 1992; A. Martinet *Des steppes aux océans. L'indo-européen et les "Indo-européens"* Paris 1986; P. Thieme "The Comparative Method for Reconstruction in Linguistics" in D. Hymes *Language in Culture and Society. A Reader in Linguistics and Anthropology* New York 1964, 585-598; W.P. Lehmann *Theoretical Bases of Indo-european Linguistics* Londres 1993, 258-288; J.P. Mallory *In Search of the Indo-Europeans. Language, Archaelogy and Myth* Londres 1989; E. Benveniste *Le Vocabulaire des Institutions indo-européennes. I. Economie, parenté, société. II. Pouvoir, droit, religion.* Paris 1969-73; G. Dumézil *L'Idéologie tripartite des Indo-Européens* Bruxelles 1959; A. Meillet *Introduction à l'étude comparative des langues indo-européennes* Paris 1937 [reprint Alabama 1964], 378-417; C. Renfrew *Archaeology and Language* Londres 1987; C. Watkins "Indo-European and the Indo-Europeans" in *The American Heritage Dictionary of Indo-European Roots* Boston 1985, Xi-XXiV.

V. Les autres familles de langues: difficultes et limites de la comparaison.

Il ne peut être question de passer en revue toutes les autres familles qui ont pu être identifiées hors du domaine indo-européen. On se contentera d'en mentionner certaines en se préoccupant surtout des principes qui ont été invoqués pour constituer ces regroupements.

Pour en savoir plus: B. Comrie et al. *The World's Major Languages* New York/Oxford 1990; M. Ruhlen *A Guide to the World's Languages, Vol. I: Classification* Stanford CA. 1987; A. Meillet et M. Cohen *Les Langues du Monde* Paris 1952; M. Sala et I. Vintila-Radulescu *Les Langues du Monde, petite encyclopédie* Bucarest/Paris 1984. Articles divers in *The New Encyclopaedia Britannica, Macropaedia* Chicago 1974.

Dès que l'on quitte en effet l'aire indo-européenne où la méthode

comparative a été élaborée, on est bien souvent obligé de recourir à d'autres stratégies, faute de la profondeur historique exceptionnelle dont on a eu la chance de disposer pour la restituer. Ce point doit toujours être gardé à l'esprit. Pour n'en donner qu'un exemple, si nous n'avions la chance de pouvoir, via germanique et latin, produire tous les chaînons intermédiaires qui constituent la preuve de leur parenté, qui oserait rapprocher deux mots aussi différents que le français *chef* et l'anglais *head*? De même, le passage du latin *aqua* au français *eau* [o] doit inciter à la modestie toute tentative de remonter le temps, en l'absence d'autres attestations que contemporaines pour une langue donnée.

Pour deux autres familles néanmoins, on dispose de documents sur une période sensiblement aussi longue que l'indo-européen: celle qui est dite aujourd'hui afro-asiatique et sino-tibétaine. Mais, à la différence de l'indo-européen, dont sur ce point le statut reste parfaitement unique, ces familles incluent des langues attendant encore d'être décrites, leurs proto-langues n'ont pas été posées d'abord sous forme de sous-groupes, ni a fortiori établies comme telles.

LE CHAMITO-SEMITIQUE OU AFRO-ASIATIQUE.

La famille afro-asiatique (anciennement chamito-sémitique) comprend cinq rameaux: égyptien, dont les premiers témoignages remontent à -3000 av. JC., et que le copte continue encore comme langue liturgique des Chrétiens monophysites d'Egypte; berbère, couchitique et tchadien, tous trois seulement attestés à époque moderne et parlés sur une large bande au Nord de l'Afrique; enfin le sémitique, lui-même réparti en plusieurs groupes où figurent l'akkadien (-3000), le phénicien, l'hébreu (apparu vers 1100 av. JC et que l'araméen remplace vers le sixième siècle av. JC.), l'arabe, largement répandu par la conquête islamique, l'amharique, en Ethiopie.

Compte tenu de l'extrême diversité de ces langues, il n'y a que très peu de traits communs partout repérables, comme la distinction de genre avec *t* comme marque de féminin, *k* pour désigner la deuxième personne, et de rares unités lexicales, telle la racine **mut*, "mourir". Pour l'essentiel, le regroupement de toutes ces langues en une famille unique se fonde sur des attendus d'ordre typologique, principalement ceux qui tournent autour d'une répartition des langues an-historique, selon leur fonctionnement morphologique [VOIR EXCURSUS: LA TYPOLOGIE MORPHOLOGIQUE].

LA TYPOLOGIE MORPHOLOGIQUE

C'est sans doute l'aspect de la recherche typologique qui est le mieux connu. Ancienne, son élaboration lente s'est poursuivie tout au long du siècle dernier, parallèlement au travail comparatif proprement dit. Toujours présente dans les manuels, la partition des langues en fonction de leurs caractéristiques morphologiques mérite d'être esquissée brièvement.

Traditionnellement, trois types de langues étaient reconnus: isolant, agglutinant et fusionnel, auxquels était souvent ajouté un quatrième dit polysynthétique ou incorporant.

Une langue isolante est une langue qui n'a pas de morphologie, offrant, dans son type pur, une correspondance bi-univoque entre mots et morphèmes. La phrase vietnamienne suivante permet de l'illustrer:

Khi tôi dèn nha ba.n tôi, chung tôi bâtdâu làm bài

quand moi venir maisonami moi PLURIEL moi commencer faire leçon

"quand je suis arrivé dans la maison de mon ami, nous avons commencé à faire les leçons".

Chaque mot y est invariable, même le pluriel est obtenu non par modification morphologique, mais par addition d'un mot séparé, et ne consiste qu'en un seul morphème: l'exception apparente de "commencer" est contredite par la décomposition de l'unité sémantique en deux morphèmes *bat*, "prendre" et *dâu*, "tête". Pour cette raison, les langues isolantes ont souvent été appelées monosyllabiques, mais le seul trait pertinent de ce type de langues est bien l'absence de morphologie. Chaque mot ou morphème peut, en principe, être éventuellement formé de plusieurs syllabes, pourvu qu'on ne puisse leur assigner une signification particulière. A l'inverse une langue monosyllabique pourrait très bien manifester des variations morphologiques. Le chinois, le birman, les langues soudanaises sont ainsi exemplaires du mode de fonctionnement isolant.

Une langue agglutinante est une langue où le mot consiste en un ou plusieurs morphèmes qui ont comme caractéristique d'être toujours parfaitement identifiables avec une forme, sinon invariable, du moins aux contours phonétiques nettement délimités. Comme son nom l'indique (lat. *gluten*, "colle"), tout se passe comme si les affixes étaient simplement collés les uns aux autres. Ainsi en turc, où les noms varient en nombre (2) et en cas (6), il est toujours possible de découper un mot en radical lexical + affixe de nombre (*ø / lar*) + affixe de cas (*ø / -i / in / -a / -da / -dan*):

	Singulier	Pluriel
Nominatif	*adam*	*adam-lar*
Accusatif	*adam-i*	*adam-lar-i*
Génitif	*adam-in*	*adam-lar-in*
Datif	*adam-a*	*adam-lar-a*
Locatif	*adam-da*	*adam-lar-da*
Ablatif	*adam-dan*	*adam-lar-dan*

Cette description pourrait peu ou prou s'appliquer aussi au coréen, au tamoul et autres

langues dravidiennes, au swahili, et à un moindre degré au hongrois et au finnois. L'espéranto en offre des exemples frappants. Soit la racine invariable *san*, affectée d'un suffixe *-a* qui la rend adjectivale, l'ensemble *sana* signifiera "sain", et si l'on dispose des affixes suivants:

mal = OPPOSE / *ar* = COLLECTION / *ig* = CAUSATIF / *in* = FEMININ / *ul* = CARACTERISANT / *ge* = EPICENE / *ist* = AGENT / *edz* = "marié" / *a* = NOMINAL, la signification de *gemalsanularoj*, *sanigistino*, *sanigistedzino*, sera facile à reconstituer: respectivement, "tous les patients hommes et femmes", "femme docteur", "femme d'un docteur".

Une langue fusionnelle est caractérisée par le fait que l'expression de plusieurs catégories attachées à un mot est fondue en un seul morphème non segmentable. Comme le montre l'exemple du latin, les distinctions de nombre et de cas se combinent en un affixe unique assurant l'expression de ces deux catégories. Même ces morphèmes portemanteau n'ont pas une forme invariable: d'une déclinaison à l'autre des affixes différents peuvent être utilisés pour exprimer par exemple le datif singulier (*aquae* / *dominô* / *militi*); de plus, une même marque phonétique peut, selon qu'elle prend place dans tel ou tel paradigme, assurer l'expression de catégories toutes différentes (*-um* : acc. sg. de la deuxième décl. dans *dominum*, de *dominus*, *domini* / *-um* : gén. pl. de la troisième dans *nomin-um*, de *nomen*, *nominis*). Le terme fusionnel tend à remplacer celui de flexionnel employé traditionnellement, car les langues agglutinantes comme les langues fusionnelles ont des flexions.

Le quatrième type morphologique est communément appelé polysynthétique ou incorporant, indifféremment. Pourtant, l'incorporation se réfère à la possibilité de combiner ensemble des morphèmes lexicaux en un seul mot. Cette possibilité existe en français, mais de manière limitée (*chou-fleur*, *pousse-pousse*, *timbre-poste*), tandis que dans d'autres langues elle est extrêmement productive, donnant naissance à des mots extrêmement longs qui se traduiront en français par une phrase complète. Un seul mot de la langue chukchi est décomposable en "marque de première personne" - "grand" - "tête" - "mal" - "marque d'aspect imperfectif" et équivaut à "j'ai un grand mal de tête".

La polysynthèse s'en distingue en ce qu'elle consiste à combiner un grand nombre de morphèmes lexicaux ou grammaticaux en un seul mot qui correspond à une phrase du français. En esquimau par exemple un mot donné ne contient qu'un seul morphème lexical, tous les autres étant grammaticaux: *angya-ghlla-ng-yug-tuq*, "il veut acheter un grand bateau", s'analyse en "bateau - AUGMENTATIF - ACQUERIR - DESIDERATIF - 3SING.", ce qui fait de l'incorporation un cas particulier du polysynthétisme. Les deux types de langues peuvent en outre recourir aussi bien à l'agglutination qu'à la fusion.

Cette approche classique en quatre types est aujourd'hui remplacée par une autre qui recourt à deux paramètres. L'un tient au nombre de morphèmes par mots, avec comme pôles langue isolante / langue polysynthétique: dans les premières, chaque mot consiste en un seul morphème, dans les secondes, chaque phrase consiste en un seul mot qui lui-même comprend un grand nombre de morphèmes. Le second paramètre concerne la capacité qu'ont les morphèmes à être segmentés dans le mot: les deux extrêmes sont alors constitués par l'agglutination d'un côté et la fusion de l'autre. En fonction de ces deux indices, de synthèse et de fusion, on peut voir que les langues polysynthétiques sont des langues à haut indice de synthèse mais que leur indice de fusion est variable; les langues agglutinantes ont un faible indice de fusion et un indice moyen de synthèse; les langues fusionnelles ont un haut indice de fusion, cet indice étant,

par définition, non-pertinent pour les langues isolantes.

Dès lors, plutôt que de fonctionner de manière discrète avec quatre types seulement auxquels la majorité des langues ne correspondent jamais exactement, il est possible d'utiliser une typologie continue et de placer chaque langue dans un continuum défini par l'index de synthèse et l'index de fusion.

Néanmoins les problèmes posés par l'établissement de chaque indice ne sont pas minces et surgissent à chaque instant. Pour déterminer le degré de synthèse, il est souvent difficile de décider si l'on a affaire à un morphème ou à un mot: dans "je le vois", "le" est-il une forme libre, prononçable isolément? Selon l'option retenue, l'indice variera considérablement. Doit-on inclure comme morphèmes les morphèmes zéro, ou encore compter comme plusieurs morphèmes les morphèmes portemanteau au prétexte qu'ils exprimeraient plusieurs catégories? De même, quand il s'agit du degré de fusion, la notion de segmentabilité est une affaire de degré et non de tout ou rien. Celle d'invariance est tout autant de l'ordre d'un continuum avec, à un bout, les cas de supplétisme, en passant par les variations prévisibles en fonction d'une règle phonologique jusqu'à l'invariance absolue.

Pour en savoir plus: B. Comrie *Language universals and Linguistic Typology* Chicago 1981, 39-50; W. Croft *Typology and Universals* Cambridge 1990, 39-43; E. Sapir *Language. An Introduction to the Study of Speech* New York 1921 [trad. fr. Paris 1953]; W. P. Lehmann *Historical Linguistics. An Introduction* Londres 1992, 175-182.

Parmi ces traits typologiques décisifs, l'utilisation systématique des racines [VOIR EXCURSUS: LA RACINE EN SEMITIQUE], qui dans une armature consonantique offrent des alternances vocaliques, ou flexion interne, pour assurer l'expression des temps, des modes et d'autres catégories[63].

LA RACINE EN SEMITIQUE

Dans les langues sémitiques et, plus largement, quoique à un moindre degré, dans les langues d'abord dénommées chamito-sémitiques puis afro-asiatiques, le centre du mot est la racine. Dans presque toute forme se laisse reconnaître une suite d'éléments phonétiques définissant la base lexicale, et identifiable dans un grand nombre d'autres formes. Cette racine est constituée d'une suite de phonèmes, dont le nombre - le plus souvent trois - la nature - généralement consonantique - et l'ordre sont constants. Cette structure trilitère discontinue est fondamentale; les éléments radicaux exprimant l'idée lexicale restent apparents dans tous les mots formés sur une même racine et reliés plus ou moins étroitement à l'expression d'un concept. Toutes les combinaisons de trois consonnes ne sont pas utilisées: une sélection s'est opérée qui tend à éviter la contiguïté de consonnes d'articulation proche, en sorte que le plus souvent, les consonnes d'une racine sont différentes par leur point et leur mode articulatoires. Pour reprendre un exemple canonique, fourni par la langue arabe, à partir du squelette consonantique, encore non actualisé, que constituent les trois phonèmes *K-*

63. Ce mode de fonctionnement est très vivace en sémitique, il l'est moins en berbère, et n'existe plus qu'à l'état de traces en couchitique et en tchadien.

T-B, exprimant la notion "écrire", on peut ainsi dériver un grand nombre de noms, de formes et de temps verbaux: *KaTaBa*, "il a écrit" / *KuTiBa*, "il a été écrit" / *yaKTuBu*, "il écrira" / *?aKTiB*, "j'écrivis" / *KiTâB*, "livre" / *KuTuB*, "livres" / *KuTuBî*, "libraire" / *KuTTâB*, "école coranique" / *KuTayyiB*, "livret" / *KâTiB*, "écrivain" / *maKTûB*, "écrit, lettre, destin" / *KiTâBa*, "écriture" / *maKTaBa*, "bibliothèque" / *maKTaB*, "bureau" / *miKTâB*, "machine à écrire" / *muKâTaBa*, "correspondance" / *?iKtiTâB*, "enregistrement" / *?istiKTâB*, "dictée" etc.

On voit que, pour former un mot, les consonnes radicales doivent se voir complétées par l'insertion de voyelles en des suites de combinaisons variables dont chacune est porteuse d'une signification dérivationnelle ou grammaticale précise. Les divers schèmes vocaliques ainsi infixés et éventuellement accompagnés de consonnes (préfixées, suffixées, plus rarement infixées) caractérisent donc à leur tour les formes verbales ou les différentes catégories nominales: la succession *a - a - a*, par exemple, est un morphème discontinu qui inséré dans les éléments radicaux (*C1aC2aC3a*) sera toujours l'expression d'une troisième personne du singulier du perfectif, se retrouvant pour les racines *H-M-L*, "porter", *K-S-R*, "briser", *Q-T-L*, "tuer", dans leurs formes *HaMaLa*, "il a porté", *KaSaRa*, "il a brisé", *QaTaLa*, "il a tué". L'arabe classique comprend environ 150 schèmes ou patrons de ce type et, finalement, tout mot, à l'exception des monèmes grammaticaux, pourra s'analyser comme issu de l'entrelacement d'une racine et d'un schème, la racine étant comme une sorte de coquille ou de moule dans laquelle se coule un contenu variable de voyelles.

Ce triconsonantisme des racines est un trait fondamental des langues chamito-sémitiques, mais il n'est nulle part aussi systématique qu'en arabe classique. Un tel système appartenait-il déjà au proto-sémitique ou au proto-chamito-sémitique avant de s'affaiblir progressivement en certaines langues ou bien n'était-il qu'à l'état naissant dans la proto-langue pour trouver en arabe son développement le plus achevé, il est impossible de trancher.

Pour en savoir plus: J. Cantineau "Racines et schèmes" in *Mélanges W. Marçais* Paris 1950, 119-124; A. Meillet et M. Cohen *Les Langues du Monde* Paris 1952, 81-181; D. Cohen "Les langues chamito-sémitiques" in A. Martinet et al. *Le langage* Paris 1968, 1288-1329; D. Cohen *Les Langues dans le monde ancien et moderne: langues chamito-sémitiques* Paris 1989; H. Fleisch *L'Arabe classique: esquisse d'une structure linguistique* Beyrouth 1956.

Un autre trait commun est constitué par un système de dérivation verbale par affixation et fondé sur les distinctions aspectuelles (processif / statif-duratif) plutôt que temporelles.

Pour en savoir plus: I.M. Diakonoff "Hamito-Semitic languages" *Encyclopaedia Britannica* 1974 Vol.8, 589-598; S. Moscati et al. *An Introduction to the Comparative Grammar of the Semitic Languages* Wiesbaden 1964; D. Cohen *Les Langues dans le monde ancien et moderne: langues chamito-sémitiques* Paris 1989.

LE SINO-TIBETAIN.

L'autre famille d'importance majeure, qui, compte tenu de

l'antiquité de ses témoignages, de l'impact de son influence culturelle, et de la masse de ses locuteurs, ne le cède en rien à la famille indo-européenne, est la famille sino-tibétaine. Bien des langues de ce groupe attendent encore d'être décrites et sans doute aussi d'être découvertes tout court et les rapports existant précisément entre les deux principaux sous-groupes, tout comme les regroupements à effectuer à l'intérieur de chacun d'eux pour établir les relations génétiques, demeurent encore largement matière à controverse. On peut néanmoins poser avec assurance l'existence de deux branches, tibéto-birmane et sinitique, et exclure désormais thaï et vietnamien qui y furent longtemps intégrés sur la base d'une ressemblance typologique avec le chinois, alors que leurs similitudes structurelles sont dues à des convergences de type aréal. Beaucoup de ces langues sont monosyllabiques et isolantes.

Les langues de la branche tibéto-birmane qui se comptent par centaines s'étendent du Tibet à la péninsule malaise et du Pakistan au Vietnam. La branche sinitique est parlée par plus d'un milliard d'individus et comprend, outre le chinois mandarin, attesté dès le deuxième millénaire avant JC., un grand nombre de variétés. Mais l'utilisation d'un système unique de notation idéographique susceptible d'être "lu" par toute personne éduquée a pu donner une fausse impression d'unité, encore renforcée par l'appellation abusive de "dialectes" pour désigner les six autres ensembles repérables en Chine (Wu, Min, Gan, Hakka, Xiang, Yuè), alors qu'ils ne sont pas mutuellement intelligibles sous leur forme parlée.

Pour en savoir plus: S.C. Egerod "Sino-Tibetan languages" *Encyclopaedia Britannica* 1974 Vol.16, 796-806.

L'OURALO-ALTAIQUE?

En repartant de l'Europe, on trouve le groupe des langues finno-ougriennes, incluant d'une part finnois, estonien, live, tchérémisse, et de l'autre hongrois, ostiak. Si l'on y adjoint, d'un côté, le lapon, et, sur le Yénisséi, les langues samoyèdes, on obtient une famille plus vaste celle des langues ouraliennes. Ces langues sont agglutinantes et offrent des traits tels que la position finale du verbe et l'harmonie vocalique, sorte d'assimilation à distance modifiant la voyelle du (ou des) suffixe(s) pour l'assimiler au moins partiellement à la voyelle radicale. Si celle-ci est d'avant (*e, i, ö*) les voyelles figurant dans les différents affixes seront

aussi nécessairement des voyelles frontales; et de même pour les voyelles d'arrière (*a, o, u*): ainsi, en hongrois, *haz-ban*, "dans la maison", est formé à l'aide de la racine *haz* et de l'affixe *-ban*, mais, pour *kéz*, "main", on aura *kéz-ben*, "dans la main".

Or, cette caractéristique qui ne permet la coexistence dans un seul mot que de certaines sous-classes de voyelles se retrouve dans un groupe de langues dénommées altaïques, consistant en trois ensembles, le groupe turc (turc, ouzbek, tatar, kazakh), mongol (mongol, bouriate, kalmouk) et mandchou-tongouse. La plupart de ces langues présentent l'harmonie vocalique: le turc sélectionne certains sons tout au long du mot, en fonction de ceux qui figurent dans le radical: le morphème de pluriel est *-ler/-lar*, d'où *Turkler*, "les Turcs", mais on a *çocuklar*, "les enfants". Ce trait typologique, parmi d'autres, et en dépit de l'absence d'un vocabulaire commun et de correspondances phonétiques, a amené certains linguistes à concevoir un grand groupe ouralo-altaïque, alors que le caractère très récent des textes disponibles en ces langues interdit de reconstruire des proto-langues par sous-groupes puis de remonter au proto-altaïque et encore moins au proto-ouralo-altaïque. Quand elles sont aussi limitées les similarités constatables entre des groupes de langues peuvent toujours n'être que les vestiges de contacts remontant à des périodes si anciennes qu'il n'est plus possible de faire, à leur endroit, une distinction valide entre faits de convergence, faits hérités et emprunts divers.

Pour en savoir plus: R. T. Harms "Uralic languages" *Encyclopaedia Britannica* 1974 Vol.18, 1022-1032; G. Hazai "Altaic languages" *Encyclopaedia Britannica* 1974 Vol.1, 635-639; R. Austerlitz "L'ouralien" in A. Martinet éd. *Le Langage* Paris 1968, 1331-1387.

Dans le cas de l'ouralo-altaïque, et a fortiori quand on entend y joindre en outre coréen et japonais (sans doute unis par une lointaine parenté), les preuves invoquées n'ont plus rien à voir avec celles auxquelles on recourait naguère pour circonscrire le domaine indo-européen. En matière de classification, les linguistes sont loin de s'accorder sur ce qui peut constituer une preuve de parenté génétique. Il est déjà impossible d'apporter la preuve que deux langues ne sont pas apparentées, mais surtout les principes peuvent être plus ou moins rigoureux. Entre A. Meillet qui considérait qu'il n'y a de comparaison probante que celle que fournissent les morphèmes grammaticaux ("les concordances grammaticales prouvent, et

elles seules prouvent rigoureusement")[64] et des comparaisons extensives pratiquées à l'échelle de tout un continent à l'aide du seul lexique; entre le souci de Meillet de commencer par poser la grammaire comparée de chaque famille pour comparer ensuite entre eux des ensembles déjà bien constitués et la stratégie de J.H. Greenberg qui se contente de prendre des mots dans une langue de chaque famille pour confronter les familles entre elles, il y a un monde dans la rigueur impliquée.

DE L'ASIE A L'AMERIQUE.

Quand on élargit un peu plus le cercle géographique, les ensembles constitués le sont presque toujours sur d'autres bases que la comparaison génétique de type indo-européen, ils dépendent de stratégies sensiblement éloignées des techniques traditionnelles.

C'est le cas de la famille dite austrique sous laquelle on rassemble un très grand nombre de langues du Sud-Est asiatique et du Pacifique en trois sous-groupes: l'austro-asiatique, s'étendant de l'est de l'Inde au Vietnam (langues munda, mon-khmer), le Tai-Kadai en Birmanie, Chine du sud et Thaïlande; enfin, le plus important, l'austronésien, déjà reconnu par W. v. Humboldt et anciennement appelé malayo-polynésien, qui va de Madagascar à l'île de Pâques et de Hawaii à la Nouvelle-Zélande. Pour établir de nouveaux sous-groupes pour les 500-700 langues environ que compte l'austronésien, et avec des difficultés qui tiennent à l'existence d'un grand nombre de langues très proches pratiquées par de très petits groupes humains, on a dû recourir à la méthode de la glottochronologie [VOIR EXCURSUS: LA GLOTTOCHRONOLOGIE].

LA GLOTTOCHRONOLOGIE

Le postulat de base de la glottochronologie est le suivant: s'il est vrai que dans toute langue, au fil du temps, des mots disparaissent et sont remplacés par d'autres, d'une part certaines portions du vocabulaire sont moins sujettes à changer que d'autres, et l'on peut définir l'aire du lexique qui est ainsi la plus résistante au changement, de l'autre, ce vocabulaire de base est le même pour toutes les langues. En toute langue les pronoms, les numéraux, des adjectifs (*grand, long, petit*), des termes de parenté (*mère, père*), les désignations des parties du corps (*oeil, oreille, tête*), d'événements ou objets naturels (*pluie, pierre, étoile*), des états et des actions élémentaires (*voir, entendre, venir, donner*)

---

64. Même si, à l'occasion, il évoque "le fonds principal du vocabulaire courant" (*Linguistique historique et linguistique générale I.* Paris 1938 [rééd. Paris 1958], 91.

sont moins susceptibles d'être remplacés par des mots empruntés à d'autres langues. Le vocabulaire général de l'anglais est pratiquement pour moitié emprunté à d'autres langues, mais le chiffre chute à 6 % si l'on ne tient compte que de ce noyau dur du lexique.

La seconde affirmation de la méthode lexicostatistique est que le taux d'usure, de déperdition de ce vocabulaire fondamental est à peu près immuable, les mots disparaissent et sont remplacés selon un rythme sensiblement constant, alors que pour le vocabulaire périphérique, celui qui est intimement lié à des données plus proprement culturelles, le taux de changement est variable en fonction des contacts éprouvés par les locuteurs.

Les conséquences sont de deux ordres. Une fois qu'on aura fait la liste des mots qui constituent le noyau du vocabulaire, et qu'on aura déterminé ce taux à partir d'exemples historiquement attestés d'une langue qui a connu une telle évolution lente, comme c'est le cas par exemple du latin aux langues romanes, il sera possible pour toute langue qu'on comparera à une autre à l'aide de ce même lexique de base d'obtenir un pourcentage de rétention. Ce chiffre constituera d'abord une mesure du degré de similitude entre les deux langues indiquant leur degré de relation du point de vue génétique: plus la ressemblance des vocabulaires sera grande et plus le groupe auquel appartiendront les deux langues sera de niveau inférieur dans l'arbre figurant les parentés. En corollaire, ce pourcentage traduisant le degré de ressemblance entre deux langues peut servir à mesurer le temps écoulé depuis le moment de leur divergence.

L'examen de treize langues (la plupart indo-européennes), fournissant des témoignages écrits sur une longue période, a permis de calculer un taux de perte de 80-85 % pour une durée de 1.000 ans. Si on adopte le premier chiffre pour la commodité des calculs, on en déduira que le français moderne a conservé 80 % du vocabulaire de la langue parlée en 993; 64 % de celle précédant de peu notre ère, et remonter jusqu'à 3.000 ans donnerait un chiffre de 51 %, de 40 % pour 4.000 ans. Enfin, si l'on date l'indo-européen commun de - 3.000 ans av. J.C., il faut s'attendre à ce que le français ait approximativement conservé 30 % de son vocabulaire.

Pour la comparaison de deux langues, sur la base d'un taux de conservation du vocabulaire de 85 % pour chaque langue, et étant entendu que ce ne sont évidemment pas les mêmes 85 % qui se seront maintenus dans les deux langues, une formule a été élaborée permettant de calculer la profondeur temporelle séparant les deux langues du moment de leur divergence, elle est égale au logarithme du pourcentage de mots apparentés divisé par le double du logarithme du pourcentage postulé de mots maintenus après une séparation de mille ans (r = 85 %), et ce, avec une marge d'erreur calculable.

On obtient la formule suivante: $t = \frac{\log C}{2 \log r}$

Ont été ainsi élaborées des listes de 100 ou 200 mots, mais d'emblée des objections ont été soulevées. Déjà la notion de vocabulaire de base non soumis à l'emprunt et valable pour toutes les cultures ne va pas de soi. Des objets aussi naturels d'apparence que le soleil peuvent relever du vocabulaire religieux (c'est le cas dans le sud de l'Asie), et, du coup, voir leur désignation faire l'objet d'un emprunt, de même le vocabulaire de base peut très bien devenir l'objet de tabous et être importé d'une langue voisine pour pallier l'interdiction; certains mots de la liste font parfois défaut (on ne trouve pas de glace sous

les tropiques). En outre, la segmentation conceptuelle variant de langue à langue, il se peut qu'à un seul mot du lexique anglais ou français correspondent, dans une langue non européenne, des affixes et non des mots, ou bien encore quatre ou cinq mots entre lesquels il faudra choisir, ce qui rendra la comparaison ultérieure plus arbitraire. Inversement "épouse" et "femme" peuvent être confondus en un seul concept. Il est enfin très peu probable que le taux de rétention soit constant pour toutes les langues, et à toutes les époques. Dans des conditions particulières tenant à l'isolement de la communauté, à sa cohésion sociale, au respect éventuel d'une norme religieuse, voire littéraire, ce taux pourra varier de manière significative - on en a, en Europe même, un exemple frappant avec la stabilité de l'islandais - assez en tout cas pour invalider partiellement la méthode en ruinant sa prétention à l'universalité.

En second lieu, dans l'identification des termes à déclarer apparentés, de nouveaux problèmes surgissent. Dans la mesure où on ne recourt à la technique de la lexicostatistique qu'en désespoir de cause, quand on travaille sur une aire géographique très large et des centaines de langues pour lesquelles l'information est très lacunaire, avec des descriptions partielles et surtout récentes, il est alors exclu, faute de matière, de poser des correspondances phonétiques régulières. Dès lors, l'élimination du vocabulaire emprunté qui se fonde sur la maîtrise préalable de ces régularités est rendue d'autant plus difficile. En conséquence, l'identification du seul vocabulaire effectivement apparenté, et donc authentiquement hérité en parallèle, se trouve très menacée. A quoi s'ajoute que ce repérage est toujours aléatoire: qui oserait rapprocher *chef* de l'anglais *head*, pourtant de même origine? qui, au rebours, éliminerait lat. *dies* / angl. *day* ou lat. *habere* / all. *haben*, pourtant non pertinents?

Néanmoins, la méthode peut assurément rendre des services, lorsque la stratégie comparative classique ou la reconstruction interne sont inapplicables. Quand, par exemple, on ne dispose que de listes souvent très incomplètes de vocabulaire. Ainsi, pour les langues austronésiennes dont on a des matériaux pour près de mille d'entre elles, ou celles d'Australie (environ 250) c'est-à-dire des groupes très larges seulement attestés à époque moderne, dans ce cas elle fournit une première esquisse de répartition en groupes et en sous-groupes, assez proche finalement des répartitions qu'autorise l'examen des identités manifestées par l'ADN pour les différentes formes de vie. Sans jamais pouvoir s'y substituer, la glottochronologie sera, au mieux, pour la véritable investigation historique, un point de départ plausible, une invite à poursuivre classification et reconstruction.

Pour en savoir plus: T. Bynon *Historical Linguistics* Cambridge 1977, 266-72; R. J. Jeffers et I. Lehiste *Principles and Methods for Historical Linguistics* Cambridge MA. 1982, 133-37; W. P. Lehmann *Historical Linguistics: An Introduction* Londres 1992, 175-182; T. Crowley *An Introduction to Historical Linguistics* Oxford 1992, 168-190; S. Gudschinsky "The ABC's of Lexicostatistics (Glottochronology)" in D. Hymes ed. *Language in Culture and Society. A Reader in Linguistics and Anthropology* New York 1964, 612-623; K. Bergsland et H. Vogt "On the validity of Glottochronology" in *Current Anthropology* 3 1962 115-153; T. Penchoen "La Glottochronologie" in A. Martinet et al. *Le Langage* Paris 1968.

On a procédé à l'aide de la même technique pour classer les 200

langues d'Australie en 28 sous-groupes. Mais pour les 600 langues de Nouvelle-Guinée, encore insuffisamment décrites, un tel objectif reste, pour l'heure, hors de portée.

Pour en savoir plus: G. Diffloth "Austro-asiatic languages" *Encyclopaedia Britannica* 1974 Vol.2, 480-484; A. Pawley "Austronesian languages" *Encyclopaedia Britannica* 1974 Vol.2, 484-494; S.A. Wurm "Australian Aboriginal languages" *Encyclopaedia Britannica* 1974 Vol.2, 430-431; S.A. Wurm "Papuan languages" *Encyclopaedia Britannica* 1974 Vol.13, 977-978.

Quant aux langues indiennes d'Amérique, un premier travail de classification avait été effectué au cours du dix-neuvième siècle, avec les méthodes expérimentées sur l'indo-européen, qui avait conduit à poser de très nombreuses familles: 54 pour l'Amérique du Nord, 23 pour le Mexique et la partie centrale et 75 pour le sud du continent. En 1929, Edward Sapir, en s'appuyant davantage sur la similitude des structures, réduisit cette mosaïque à 6 souches seulement pour l'Amérique du Nord jusqu'au Mexique: Eskimo-aléoute; Na-Déné; Algonquin-Wakashan, Hokan Sioux; Pénutien; Aztèque-Tanoen. Plus récemment encore, Joseph Greenberg, en recourant à une technique inédite a proposé l'existence de trois familles seulement pour tout le continent: Eskimo-Aléoute, Na déné et l'Amérindien pour tout le reste.

Pour en savoir plus: J.H. Greenberg *Language in the Americas* Stanford 1987; W. Bright "North American Indian languages" *Encyclopaedia Britannica* 1974 Vol.13, 208-213; T. Kaufman "Meso-American Languages Indian languages" *Encyclopaedia Britannica* 1974 Vol.11, 956-963; J.A. Suarez "South American languages" *Encyclopaedia Britannica* 1974 Vol.17, 105-112.

La procédure préconisée par J. Greenberg n'est pas d'ordre typologique, à la différence des regroupements que nous avons vu pratiquer pour l'altaïque, car elle se fonde sur l'identification de caractéristiques précises qui sont des unités lexicales et morphologiques [VOIR EXCURSUS: LA COMPARAISON DE MASSE].

LA COMPARAISON DE MASSE SELON JOSEPH H. GREENBERG

Pour ce linguiste américain, issu de l'anthropologie, qui, après avoir entrepris une classification complète des langues d'Afrique (1963), puis d'Amérique (1987), s'attaque aujourd'hui à l'ensemble Eurasiatique dans lequel est inclus l'indo-européen, un classement génétique des langues est moins le produit d'un travail comparatif rigoureux qu'il n'en est

une exigence préalable. Loin d'y voir un moyen d'arriver à une classification des langues, il considère que la méthode comparative doit plutôt servir à prouver une classification déjà posée à titre d'hypothèse et, à l'appui, il rappelle que c'est ce qui s'est produit pour le domaine indo-européen, où l'on a commencé par postuler des regroupements bien avant d'être en mesure d'en apporter techniquement la preuve, grâce aux lois phonétiques. A la mauvaise question: "quand des langues sont-elles génétiquement apparentées?" il faut donc, selon lui, en substituer une autre: "comment classer génétiquement les langues?". Il entend ainsi effectuer un travail purement taxinomique, sans vouloir prouver que telle famille de langues est reliée à telle autre, mais en se bornant à classer langues et familles: les relations posées seront donc des propriétés dérivées de la classification, et non l'inverse.

Greenberg présente alors comme une véritable révolution méthodologique sa technique de comparaison dite "multilatérale" qui consiste, non pas à examiner quelques langues sur un grand nombre de mots, mais à passer en revue un grand nombre de langues au travers de quelques mots. Si on découvre dans n'importe quel membre d'une famille une unité qui soit comparable avec une unité présente dans un membre quelconque d'une autre famille, on sera en droit de les considérer comme héritées et donc de poser une relation originelle des familles où on les rencontre. La méthode consiste à collecter des listes de mots, quelques centaines, pour autant de langues que possible et leur comparaison conduit inévitablement à regrouper les langues en ensembles génétiquement valides.

Alors que les spécialistes des langues indiennes postulaient l'existence de plus de deux cents familles indépendantes, Greenberg outre les deux familles déjà reconnues avant lui (Esquimo-Aléoute et Na-Déné) a posé une seule famille amérindienne avec 11 sous-groupes, sur la base de plus de 300 étymologies. Un scepticisme général a accueilli sa tentative. D'abord on lui reproche son incapacité à établir des correspondances phonétiques régulières qui forment la seule véritable preuve d'une relation génétique. Mais surtout, l'argument décisif est que, dans la comparaison de deux familles, c'est l'existence de correspondants dans les langues-soeurs de chacune qui confère à chaque étymon son antiquité à l'intérieur de sa propre famille et lui donne donc une légitimité à prétendre à une antiquité potentiellement encore plus grande pour une comparaison plus lointaine. Des correspondances fondées sur des unités repérables dans une seule langue ou un seul sous-groupe n'ont pas, avec une distribution aussi limitée, la même force, car elles peuvent aussi bien être de simples innovations, résulter du changement lexical, d'emprunts, de formations symboliques, de reformulations analogiques. Finalement, pour tout mot sélectionné au hasard, il doit y avoir un nombre de formes accidentellement semblables par le son et la signification dans les lexiques des différentes langues comparées. Le nombre de paires dues au hasard doit croître proportionnellement au nombre de langues impliquées.

Pourtant, pour Greenberg, un point fort de son argumentation demeure la présence, depuis la Colombie britannique jusqu'au Chili, et ce dans chaque sous-groupe de l'Amérindien, d'un système pronominal caractérisé par *n-* pour la première personne et *m-* pour la deuxième. Le phénomène qui avait été remarqué depuis longtemps, était effectivement attribué à une commune descente par A. Trombetti et E. Sapir, ou imputé par F. Boas à d'obscures raisons psychologiques; plus récemment encore on y voyait des restes de processus d'usure parallèles mais sans rapports historiques dans des langues séparées. Selon Greenberg, dans la mesure où cette configuration systématique ne se retrouve pas dans les autres familles de langues du monde, sa distribution ne peut être attribuée ni à l'emprunt (très rare pour les marques de ce type) ni à la chance. Greenberg tire également argument de la biologie génétique, car sa

classification correspond à la classification génétique des biologistes et aux preuves fournies par les dents fossilisées.

Pour en savoir plus: M. Ruhlen An Overview of genetic Classification in J.A. Hawkins M. Gell-Mann ed. *The Evolution of Human Languages* Reading MA. 1992, 159-189; M. Ruhlen *A Guide to the World's Langages Vol. I: Classification* Stanford 1987, 120-124, 220-227; J.H. Greenberg *Language in the Americas* Stanford 1987; L. Campbell Review of "Language in the Americas" *Language* 64 N° 3 1988, 591-615; J.H. Greenberg Classification of American Indian languages A reply to Campbell *Language* 65 N° 1 1989, 107-114; J. Matisoff On megalocomparison *Language* 66, N° 1 1990, 106-120.

L'effort pour constituer des ensembles de langues de grande extension sur la base de techniques de comparaison qui ne mettent plus au premier plan le strict jeu des correspondances phonétiques a conduit Greenberg à poser, pour toutes les langues du monde une répartition en quelques grands groupes incluant chacun plusieurs familles [VOIR EXCURSUS: LES LANGUES DU MONDE]

CATALOGUE DES FAMILLES DE LANGUES DU MONDE D'APRES J. GREENBERG.

1. Khoisan (48 l. Afrique australe; ex. boschiman, hottentot)

2. Nigéro-Kordofanienne (Afrique)

   a. Kordofanienne (32 l. au Soudan)

   b. Nigéro-Congolaise (+ de 1000 l. Afrique ocidentale et sud de l'Equateur; ex. kongo, peul, swahili, wolof, zulu)

3. Nilo-Saharienne (± 145 l. Afrique, au sud de la famille afro-asiatique).

4. Australienne (± 300 l. Australie)

5. Indo-pacifique (± 700 l.Nouvelle-Guinée, Iles Salomon, Andaman)

6. Austrique

   a. Austroasiatique (+ de 150 l. Asie du Sud-Est; ex. vietnamien, khmer)

   b. Miao-Yao (4 l. Asie du Sud-Est)

   c. Daïc (± 50 l. Chine, Laos, Thaïlande; ex. laotien, thaï)

   d. Austronésienne (± 1000 l. Océanie, Asie du Sud-Est; ex. malais, tagalog, malgache)

7. Déné-Caucasienne

   a. Basque (1 l. France, Espagne)

   b. Nord-Caucasienne (38 l. Caucase; ex. oubykh)

   c. Bourouchaski (1 l. Pakistan, Inde)

   d. Nahali (1 l. Inde centrale)

   e. Sino-Tibétain (+ de 300 l. Asie du Centre et du Sud-Est; ex. chinois, birman, tibétain)

   f. Iénisséien (1 l. URSS)

g. Na-Déné (34 l. Alaska, Canada, USA; ex. navajo, apache)

8. Afro-asiatique (241 l. Arabie et moitié nord de l'Afrique; ex. amharique haussa hébreu arabe)

9.Kartvélienne (4 l. Caucase, Turquie, Iran; ex. géorgien)

10.Dravidienne (28 l. Inde du Sud et de l'Est; ex. telougou, tamoul)

11.Eurasiatique

a. Indo-européenne (± 150 l. Europe, Asie)

b. Uralique-Yukaghir (24 l. Finlande, Estonie, Hongrie URSS; ex. finnois, hongrois, estonien)

c. Altaïque (63 l. Asie, Asie mineure; ex. turc, mongol, mandchou)

d. Coréen-Japonais-Aïnou (4 l.)

e. Gilyak (1l. bouches de l'Amour, Ile Sakhaline)

f. Tchouktche-Kamtchatka (5 l. Nord-Est de la Sibérie)

g. Esquimo-Aléoute (9 l. USA, Canada, Groenland, Nord-Est de l'URSS; ex. esquimau)

12. Amérindien (583 l. Nord Centre et Sud de l'Amérique)

Pour en savoir plus: M. Ruhlen *A Guide of the World Languages. Vol I Classification* Stanford 1987 275-378; M. Ruhlen "An Overview of Genetic Classification" in J. Hawkins M. Gell-Mann *The Evolution of Human Languages* Reading MA. 1992, 159-189; M. Sala, I. Vintila-Radulescu *Les Langues du Monde* Bucarest/Paris 1984; J.H. Greenberg *Indo-European and its Closest Relatives: The Eurasiatic Language Family.* Stanford, à paraître.

INDO-EUROPEEN, NOSTRATIQUE ET EURASIATIQUE.

Il était inévitable, la liste ci-dessus en a donné un premier aperçu, que cette technique et cet effort finissent par toucher l'indo-européen lui-même et qu'on s'efforce de l'inclure dans un ensemble plus vaste. Au début du vingtième siècle déjà, un linguiste danois, Holger Pedersen, avait émis l'hypothèse que l'indo-européen pouvait être mis en relation non pas avec une seule famille - comme cela avait été le cas jusque-là, le sémitique étant le meilleur candidat pour un tel rapprochement - mais avec plusieurs. Il intitulait Nostratique cette "hyper-famille" de langues, y incluant les familles indo-européennes, chamito-sémitiques, ouralienne, altaïque, samoyède, yukaghir, esquimau, voire même le basque[65]. Dans les années soixante, des linguistes soviétiques ont entrepris de montrer que l'indo-européen était à relier à l'afro-asiatique, kartvélien, ouralien, altaïque, dravidien, et ont produit, à l'appui, plus de 700 étymologies [VOIR EXCURSUS: LE NOSTRATIQUE].

65. H. Pedersen *The Discovery of Language. Linguistic Science in the Nineteenth Century* Bloomington 1931, 335-339.

LE NOSTRATIQUE

Deux démarches sont possibles pour venir étayer cette hypothèse. L'une consiste à poser pour chaque famille à comparer une unité reconstruite en proto-langue et, sur cette base, à poser ensuite de nouvelles unités hypothétiques encore plus anciennes. La validité de l'entreprise repose alors sur la fiabilité des proto-formes de chaque famille. Or, on sait que de telles reconstructions font par exemple défaut pour l'afro-asiatique, que l'existence même de l'altaïque comme famille est discutée par les spécialistes de ce domaine. Seul l'indo-européen fournit de tels matériaux et encore est-ce l'objet de multiples affrontements Il n'est donc pas surprenant que les conclusions auxquelles aboutissent ceux qui reconstruisent le Nostratique soient très divergentes.

L'autre stratégie revient à isoler et à comparer des listes d'éléments supposés relever de l'équipement minimum et universel de toute langue et exprimant les significations les moins sujettes à variation. Par exemple Dolgoposky en a sélectionné quinze: "je/moi", "deux/paire", "tu/toi", "qui/quoi", "langue", "nom", "oeil", "coeur", "dent", la négation verbale, "doigt/ongle", "pou", "larme", "eau", "mort". Il cherche ensuite des apparentements dans toute langue des familles concernées: "nom" est bien attesté en indo-européen (par exemple dans le sanscrit *nâman*), pour le chamito-sémitique, il cite l'arabe *?ism*, pour le sumérien *inim*, etc. La ressemblance est donnée comme suffisante pour permettre de poser une relation entre ces unités et à travers elles, entre les familles de langues impliquées. On voit qu'à la différence de la méthode comparative classique, il ne s'agit pas de comparer les unités potentiellement similaires que seraient les correspondants à tous les autres mots sanscrits commençant également par *n-*, ou ceux qui offriraient en arabe le même groupe *-sm-*, ni de prendre en considération toutes les unités lexicales signifiant "nom" dans chacune des langues tenues pour nostratique.

C'est aussi oublier l'importance parfois décisive du tabou: tous les groupes humains substituent des mots pour des objets ou des actions qui ne doivent pas être mentionnés, sauf dans des circonstances particulières, si bien que le mot finit par se perdre. On le constate en indo-européen même, où, à côté du latin *lingua*, on trouve le grec *glôssa*. Les proto-indo-européens tabouisaient les noms d'animaux dangereux comme l'ours, et ceux de parties importantes du corps comme la main et la langue. Dans notre culture occidentale moderne l'interdit porte sur le sexe ou la race, le français *nègre* ou l'anglais *negro* voient leur emploi se raréfier et risquent de disparaître. De même, le mot original pour "langue" a été perdu en grec, un autre terme s'y substituant et demeurant même après la fin du tabou. En dernière analyse, en dehors de quelques linguistes soviétiques, ces tentatives n'ont trouvé qu'un accueil négatif chez les spécialistes.

Pour en savoir plus: M. Kaiser V. Shevoroshkin "Nostratic" *Annual Review of Anthropology* 17 1988 309-329; W.P. Lehmann *Theoretical Bases of Indo-European Linguistics* Londres/New York 1993, 35-37, 43-45; R. Antilla *An Introduction to Historical and Comparative Linguistics* New York 1972, 320-321.

Greenberg, lui, sur la base de 64 étymologies grammaticales, et de 500 étymologies lexicales, propose l'existence d'une famille, baptisée

Eurasiatique, où il fait entrer l'indo-européen, l'ouralien, l'altaïque, l'ensemble coréen-japonais-aïnou, le gilyak[66], le tchouktchi-kamtchatka et l'esquimau-aléoute. Un exemple de ces correspondances est la généralisation d'un fait bien connu pour l'indo-européen, la désignation de la première personne par *m-* et de la seconde personne par *t-*. Greenberg s'efforce de montrer que ces pronoms sont d'origine eurasiatique qui distinguait deux pronoms de première personne, *m-* (ergatif et actif), et d'autre part *k-* (absolutif, passif ou statif), certaines familles maintenant la distinction. L'Eurasiatique possédait deux pronoms interrogatifs, *ki* et *ja*, le premier impersonnel (quoi?), le second personnel (qui?). Le système démonstratif incluait deux pronoms, avec une opposition de proximité, *ku*, "ceci", *to* "cela". Enfin, on constate, toujours selon Greenberg, la présence de la marque de duel *-k* et de pluriel *-t* [VOIR ANNEXE: QUELQUES RAPPROCHEMENTS EURASIATIQUES].

D'autres travaux encore plus grandioses - sinon plus délirants - tentent de reconstruire rien moins que la langue-mère originelle parlée peut-être il y a 100.000 ans. Tel auteur comme Shevoroshkin postule qu'en cette langue "je" se disait **ngai*, que **waru* y signifiait "brûler, feu, être chaud", **pari*, "ongle", **wina*, "oreille, entendre". A l'évidence ce type de spéculation monogénétique, auprès de laquelle les tentatives de reconstitution de la langue primitive avancées jusqu'au dix-huitième siècle environ paraissent du coup plutôt mesurées et singulièrement rationnelles, sort du domaine de la linguistique au sens strict, pour entrer dans la sphère accueillante de la fiction héroïque, rejoignant les thèses créationnistes si en vogue outre-atlantique. Elles ont pu, un temps, trouver un appui du côté des travaux effectués en génétique, avec l'hypothèse de l'Eve primitive, selon laquelle, en se fondant sur la distribution de l'ADN mitochondrial dans des échantillons de la population du globe, tout être humain a eu comme ancêtre commun une femme vivant dans l'Est africain, il y a 100.000 ou 200.000 ans, hypothèse depuis peu abandonnée.

Une autre tentative tout aussi sujette à caution consiste à poser des étymologies pour toutes les familles de langues du monde sur la base de la technique de comparaison multilatérale. Une telle étymologie globale a été en particulier proposée pour deux racines, l'une signifiant "doigt" ou "un", l'autre pour "deux" [VOIR EXCURSUS: UNE ETYMOLOGIE GLOBALE?].

---

66. Une langue paléo-sibérienne isolée, parlée sur les bouches de l'Amour.

UNE ETYMOLOGIE GLOBALE?

Famille:	Forme(s):	Sens:
Nilo-Saharien	*tok ≈ tek ≈ dik*	"un"
Kartvélien	*tit-i*	"doigt"
	tito	"seul"
Indo-européen	**d(e)ik*	"pointer"
Ouralien	*ik ≈ odik*	"un"
Aïnou	*tek*	"main"
Japonais	*te*	"main"
Esquimau-Aléoute	*tik(-eq)* Esqu.	"index"
	tik(-laq) Al.	"médius"
Sino-Tibétain	**tik*	"un"
Yénisséien	**tok*	"doigt"
Miao-Yao	**nto?*	"doigt"
Austroasiatique	**ti?*	"main, bras"
Austro-TaïØ	**dian*	"doigt, point"
Indo-pacifique	*tong ≈ tang ≈ teng*	"doigt, main, bras"
Na-Déné	*t'ek≈tikhi≈laq, (ka-)tleek*	"un"
	tl?eq ≈ (ka-)tliki	"doigt"
Amérindien	*tik*	"doigt"

Pour en savoir plus: M. Ruhlen *A Guide of the World Languages. Vol I Classification* Stanford 1987, 261; M. Ruhlen An Overview of Genetic Classification in J. Hawkins M. Gell-Mann *The Evolution of Human Languages* Reading MA. 1992, 178-181.

Malgré ces cas limites par lesquels se clôt notre exposé et les doutes qu'ils suscitent, les acquis sont considérables. On dispose aujourd'hui de méthodes permettant de montrer que certaines langues sont apparentées génétiquement, c'est-à-dire qu'elles exhibent des ressemblances assez régulières pour autoriser l'hypothèse selon laquelle, à un moment de leur développement, et même si les modalités en sont peu claires, elles ont dû connaître une forme d'unité, car les similitudes constatées ne sont expliquables que si un héritage commun a été suivi d'une dispersion dialectale.

DEUXIEME PARTIE: DE LA COMPARAISON A L'INTERCOMPREHENSION.

Peut-on dès lors exploiter dans l'apprentissage des langues, ce fonds commun d'identité, même altéré par l'usure temporelle, qui relie toutes les langues indo-européennes? Si l'on néglige les langues celtiques, rarement enseignées hors de leur frontières et dangereusement menacées par l'anglais, et aussi, plus lointaines, mais à la limite non moins mobilisables, les langues du domaine indo-iranien, persan, hindi, bengali, ce sont en effet toutes les langues d'Europe, qui, à l'exception du finnois, du hongrois, et du basque, sont concernées par cette parenté. Pour un français désireux de les apprendre, ou, au rebours, dans l'apprentissage du FLE du point de vue d'un locuteur parlant une de ces langues européennes, comment tirer parti de l'identité rémanente qui, quoique variable en intensité, et dans les limites de l'Europe, parcourt les langues romanes, slaves, germaniques, baltes, le grec moderne, l'albanais, l'arménien? Différentes stratégies sont possibles pour réactiver lors de l'apprentissage de ces langues leur histoire partagée. On se contentera ici d'esquisser quelques pistes.

I. Ouverture et sensibilisation a la perspective historique.

Tout d'abord, il importe de ne pas négliger l'intérêt que peut susciter une information sur ces matières historiques. Apprendre une langue étrangère à des adolescents est aussi affaire de séduction intellectuelle. Vouloir parler une langue qui n'est pas la sienne est une idée a priori absurde, car il n'est rien de plus artificiel que de s'abstenir de parler sa propre langue, tant que la dépense d'énergie qu'entraîne cette activité parfaitement gratuite ne paraît obéir à aucune nécessité impérieuse; c'est vouloir renoncer volontairement à un équivalent général pour revenir à une sorte de troc inégal bien peu satisfaisant.

Tout un pan de l'apprentissage sera donc consacré à préparer le futur apprenant à accueillir avec bienveillance une activité qui va lui coûter beaucoup d'efforts et l'obliger à puiser dans son énergie intellectuelle. Au préalable, il doit prendre conscience de ce qu'implique son existence de locuteur d'une langue naturelle et son inscription dans un contexte social, géographique, historique[67]. Le phénomène du langage, toujours frappé d'évidence première pour celui qui en est le lieu

67. C'est tout le sens du mouvement anglo-saxon dit de "conscience critique du langage" (voir E. Hawkins *Awareness of Language: An Introduction* Cambridge 1987).

inconscient, doit d'abord être éclairé dans toute sa complexité (existence de dialectes, de sociolectes etc.) pour que soit faite ensuite une place légitime à l'autre, à l'étranger irréductible qui parle une autre langue. L'exploration du passé de la langue maternelle débouche ainsi naturellement, d'abord sur la fascination de l'origine, bien connue par ceux qui ont fait part à des auditoires de données étymologiques, et surtout sur l'altérité, dès que l'exploration remonte assez avant pour rencontrer les langues parentes, à des degrés successivement plus lointains.

Sur ce premier plan, doit intervenir une sensibilisation à la dimension indo-européenne qui, à un niveau général, unit la langue étrangère concernée au français. Elle prendra d'abord la forme, pour discutable qu'en soit le fondement théorique, d'une présentation de l'arbre généalogique des langues i.e. Elle pourra se poursuivre par l'étude d'une racine reconstruite, soit qu'on montre comment s'est vraisemblablement effectuée la dérive sémantique dans un certain nombre de langues qui en ont conservé des représentants, soit que l'on offre un panorama dans les différentes langues i.e. de ces illustrations divergentes de sa descendance. On trouvera en annexe deux exemples d'une telle approche à partir d'un travail sur la racine i.e. AK- "être pointu". Enfin, on devra signaler la présence en français même d'une structuration sémantico-formelle héritée de l'indo-européen: on montrera comment une seule racine i.e., par la voie de l'héritage gréco-latin, se retrouve dans des mots du français que l'on ne songerait pas à rapprocher[68], les constituant en une famille parfois hétéroclite, mais bien fondée historiquement.

On pourra aussi opérer de manière strictement transversale en étudiant comment un concept est désigné dans les différentes langues i.e. - nous illustrerons ce point plus loin dans l'exposé - ou encore en présentant un étymon i.e. reconstruit, correspondant à un concept bien caractérisé, qui se retrouve dans un maximum de langues. Des exemples comme celui du nom de la mère mentionnés la partie consacrée à la comparaison sont ainsi utilisables et l'on fournit en annexe quelques unes des

---

68. A cet effet un travail est particulièrement précieux, celui de J. Picoche *Dictionnaire Etymologique du Français* Paris 1979, car il adopte comme principe de classement, chaque fois que possible les étymons de l'indo-européen. Pour l'anglais, on dispose d'un travail analogue, mais plus technique dans ses reconstructions, car il est l'oeuvre d'un grand indo-européaniste, C. Watkins *The American Heritage Dictionary of Indo-European Roots* Boston 1985. En italien, l'ouvrage de G. Devoto *Avviamento alla etimologia italiana. Dizionario etimologico* Florence 1968, est d'égal intérêt.

correspondances i.e. les plus classiques[69].

Il est également possible et souhaitable de focaliser l'exploitation d'une dimension historique commune en restreignant la portée des rapprochements à deux langues seulement.

II. Echanges et heritage dans un couple de langues europeennes.

Soit deux langues aussi peu immédiatement transparentes que le français et le norvégien. L'objectif est de faire prendre conscience aux apprenants norvégiens du français que l'écart entre les deux langues est réductible, de restituer une proximité invisible, sans recourir évidemment aux très nombreux emprunts faits au français dans le domaine culturel ou intellectuel. Cet objectif peut être rempli si on saisit toutes les occasions d'exhumer une circulation de formes et de significations à travers les mots. Il s'agira de montrer qu'en dépit des apparences, dans un très grand nombre de cas, la distance entre le mot français et le mot norvégien qui en est la traduction, peut, à condition de mobiliser un réseau lexical s'appuyant sur la dimension historique commune des deux langues, se voir considérablement réduite et s'ouvrir sur un nouvel espace de correspondances inédites. L'élève fixera plus aisément le mot français s'il l'associe à un mot de sa langue maternelle avec lequel il entre en résonance, selon des rapports de nature variable.[70]

Il faut au préalable esquisser une typologie des écarts existant entre les deux langues, en fonction des deux axes que sont la distance du point de vue de la forme et de la signification.

---

69. On pourra trouver nombre de matériaux dans l'ouvrage de X. Delamarre *Le Vocabulaire Indo-Européen. Lexique étymologique thématique* Paris 1984.

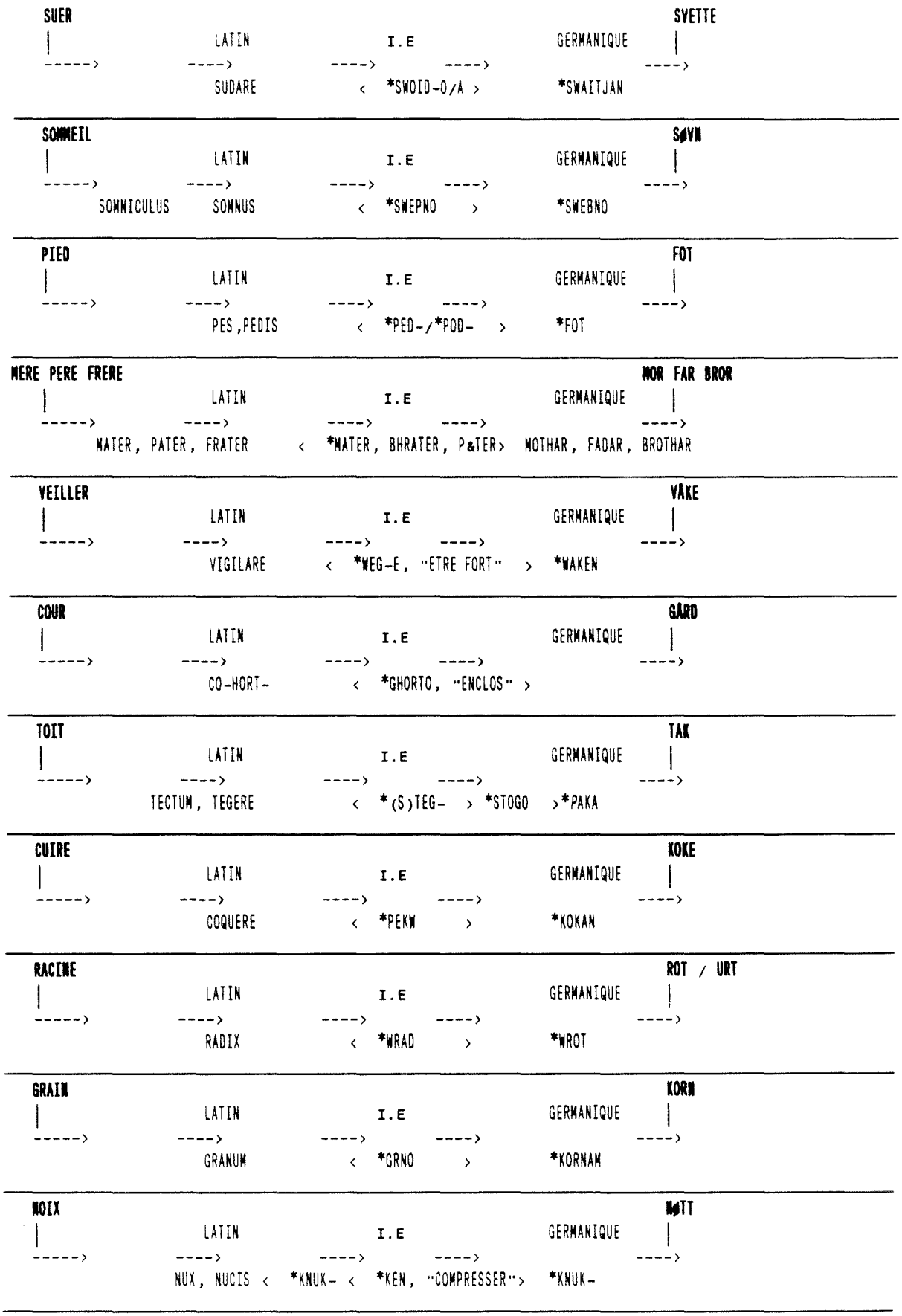
70. Pour ce travail d'ordre étymologique, on a recouru à H.S. Falk A. Torp *Norwegisch-Dänisches Etymologisches Wörterbuch* 2 Bd. Heidelberg 1910, et accessoirement à F. Kluge *Etymologisches Wörterbuch der deutschen Sprache* Berlin/Leipzig 1921 [nombreuses mises à jour ultérieures]; A. Ernoult A. Meillet *Dictionnaire étymologique de la langue latine; histoire des mots*. Paris 1967.

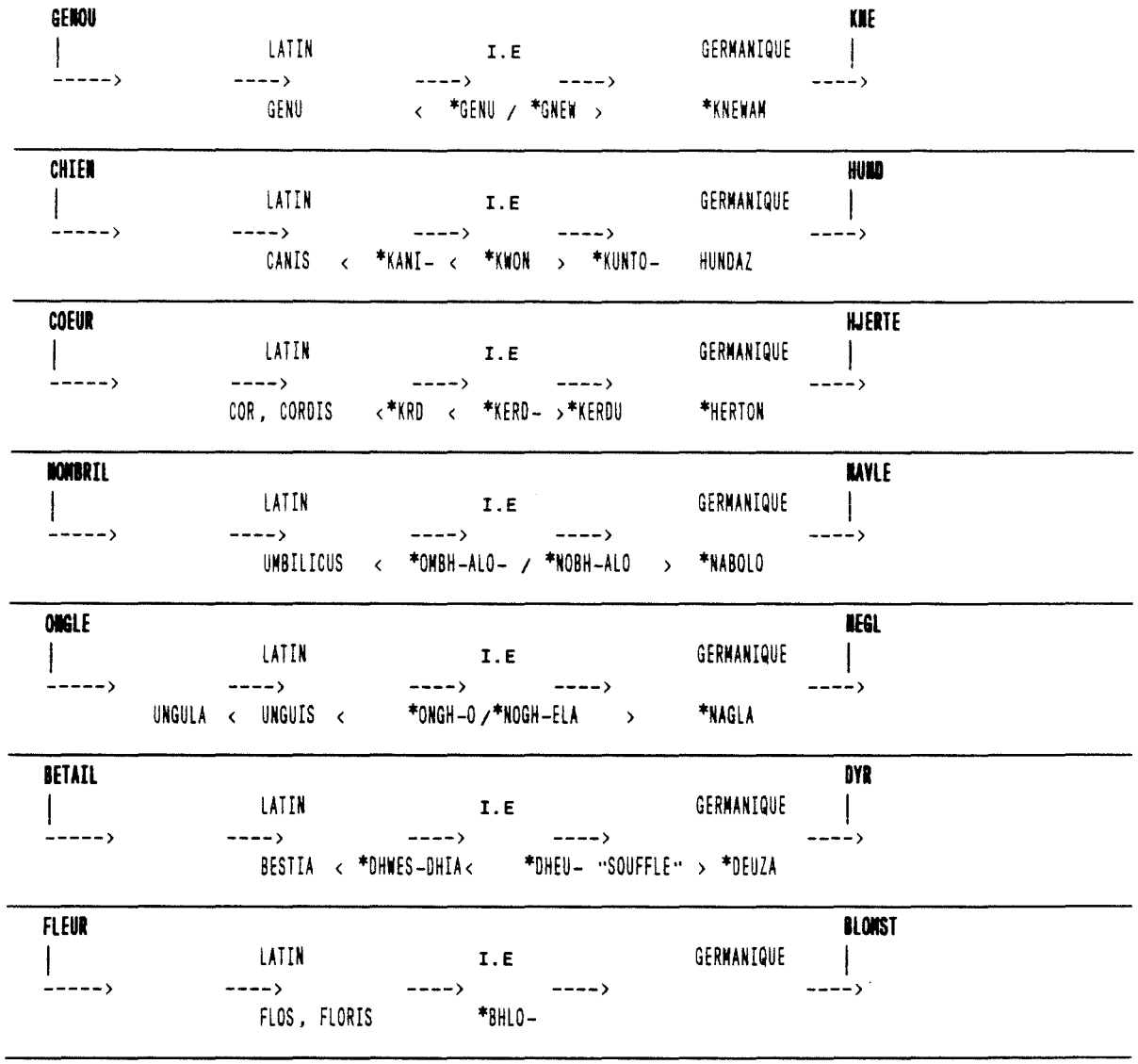
FORME +		-
+	I	VI	V
S	DESSERT > DESSERT	CHOU > KAAL	TOMBER > FALLE
E	NORD > NORD		
N			
S			
	II		
	RASER >		
	|		
	RASERE		
	"niveler"		
	III		
	MOT >		
	|		
	MOT		
-	"courage"		

Dans le cas I, le recoupement des mots des deux langues est parfait sur les deux plans. En II, à un mot français correspond un mot norvégien qui n'en est pas la traduction exacte, mais qui offre une divergence sémantique facilement récupérable. En III, l'écart de ce point de vue est tel entre le mot français et le mot norvégien formellement identique qu'il est inutilisable. En V, de même, la traduction norvégienne du mot français est a priori trop distante par la forme pour se prêter à un rapprochement et donc à une exploitation quelconque.

En revanche, dans le cas de figure VI, un mot français se traduit en norvégien par un mot qui certes manifeste une différence de forme mais celle-ci n'est pas telle qu'elle soit insurmontable. on peut, soit en mettant en évidence l'existence d'une correspondance phonétique plus ou moins régulière, soit au prix d'un rapprochement qui mobiliserait les identités plus nettes saisies au niveau du latin et du germanique commun, justifier l'écart formel. Donnons d'abord quelques exemples de ce dernier type de réduction:

FRANÇAIS					NORVEGIEN
BROSSE		< EMPRUNT <			**BØRSTE**
CHOU					**KÅL**
|	LATIN	(I.E)	GERMANIQUE		|
----->	---->	---->	---->		---->
	CAULIS				

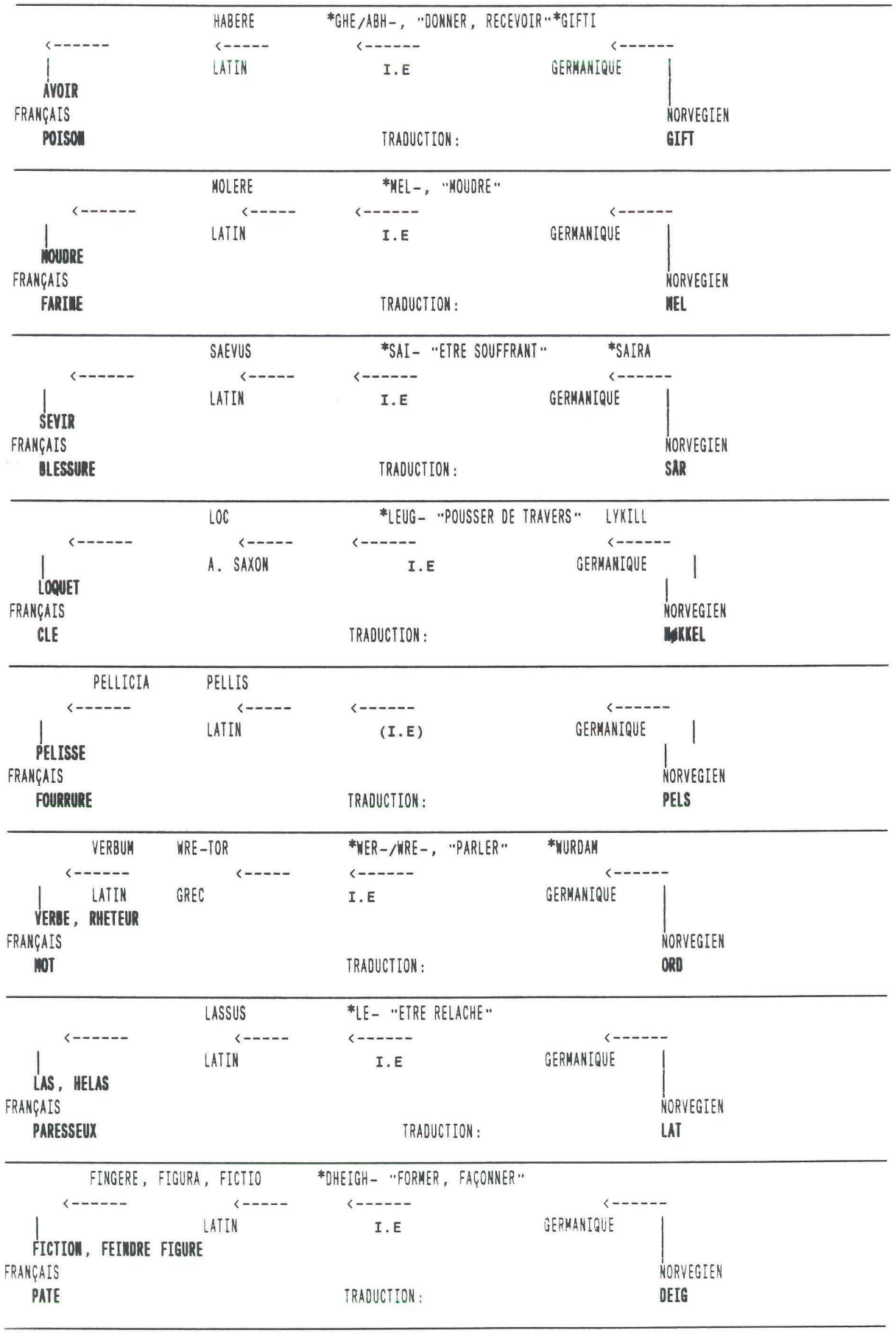

SUER
LATIN
I.E
GERMANIQUE
SVETTE
SUDARE
*SWOID-O/A
*SWAITJAN
SOMMEIL
SØVN
LATIN
I.E
GERMANIQUE
SOMNICULUS
SOMNUS
*SWEPNO
*SWEBNO
PIED
FOT
LATIN
I.E
GERMANIQUE
PES,PEDIS
*PED-/*POD-
*FOT
MERE PERE FRERE
MOR FAR BROR
LATIN
I.E
GERMANIQUE
MATER, PATER, FRATER
*MATER, BHRATER, P&TER
MOTHAR, FADAR, BROTHAR
VEILLER
VÅKE
LATIN
I.E
GERMANIQUE
VIGILARE
*WEG-E, "ETRE FORT"
*WAKEN
COUR
GÅRD
LATIN
I.E
GERMANIQUE
CO-HORT-
*GHORTO, "ENCLOS"
TOIT
TAK
LATIN
I.E
GERMANIQUE
TECTUM, TEGERE
*(S)TEG-
*STOGO
*PAKA
CUIRE
KOKE
LATIN
I.E
GERMANIQUE
COQUERE
*PEKW
*KOKAN
RACINE
ROT / URT
LATIN
I.E
GERMANIQUE
RADIX
*WRAD
*WROT
GRAIN
KORN
LATIN
I.E
GERMANIQUE
GRANUM
*GRNO
*KORNAM
NOIX
NØTT
LATIN
I.E
GERMANIQUE
NUX, NUCIS
*KNUK-
*KEN, "COMPRESSER"
*KNUK-

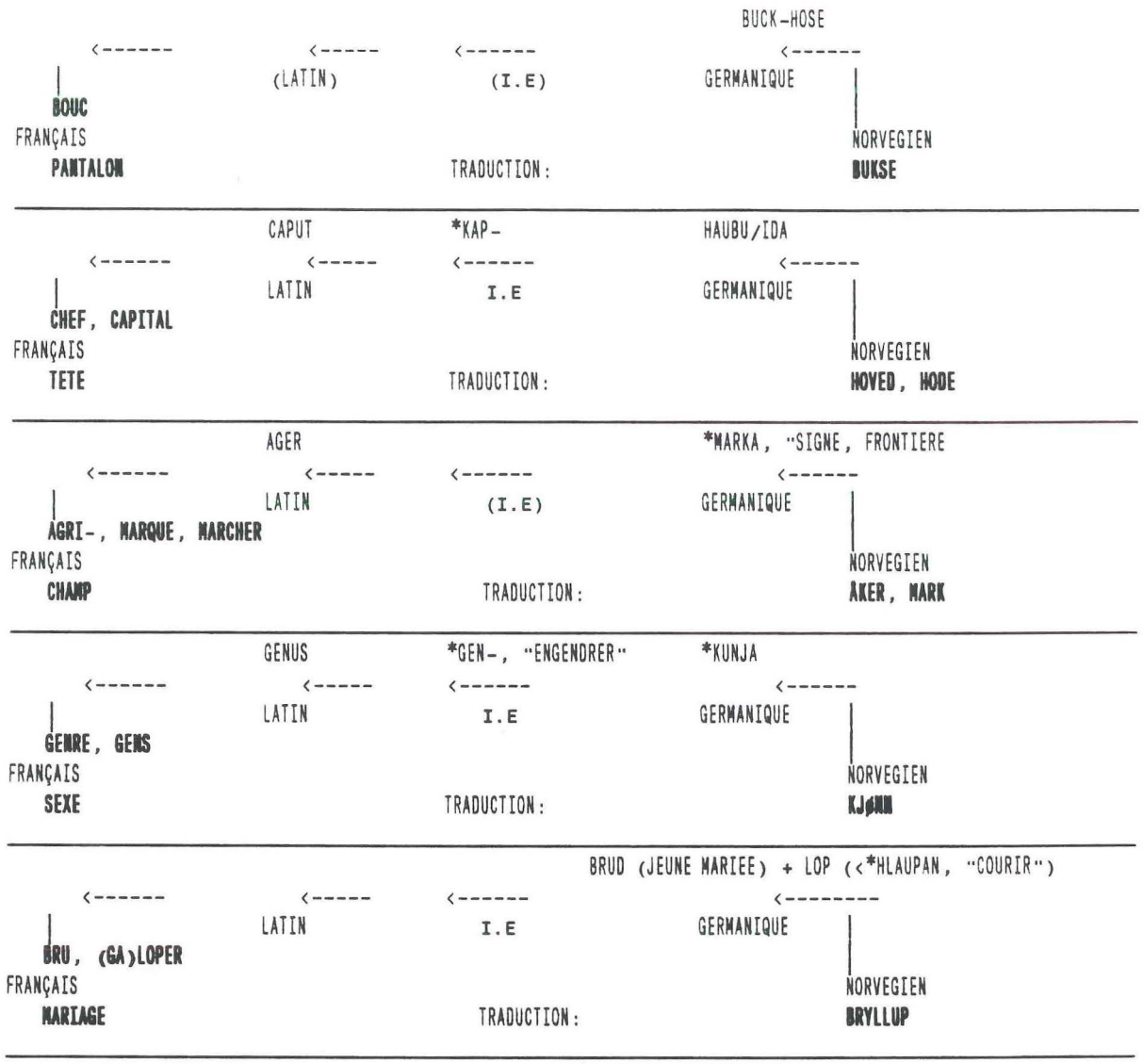


Tout aussi intéressant, et fécond, est le fait que, même si la traduction en norvégien d'un mot français donné livre un mot qui n'a, avec lui aucune similitude de forme plus ou moins réductible (comme dans les équations MOT - ORD, TOMBER - FALLE), il reste néanmoins possible, en recourant à la perspective historique, de resserrer les liens des deux langues. Car, même à partir de ces équations apparemment désespérées, au prix d'un déplacement de l'investigation, on peut montrer que les deux langues présentent une proximité insoupçonnée.

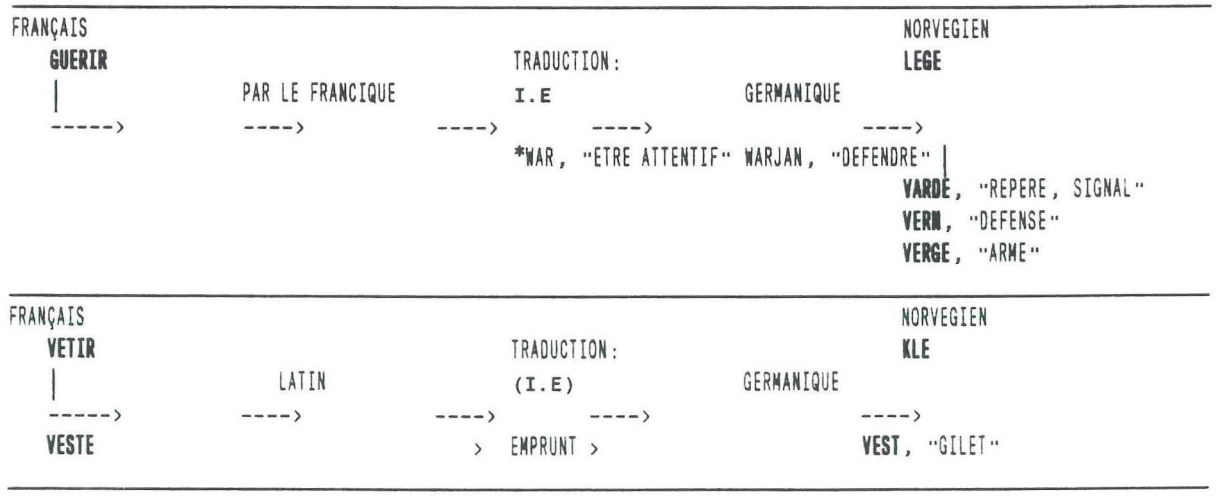
Trois cas de figure d'un tel élargissement appuyé sur la filiation historique sont envisageables.

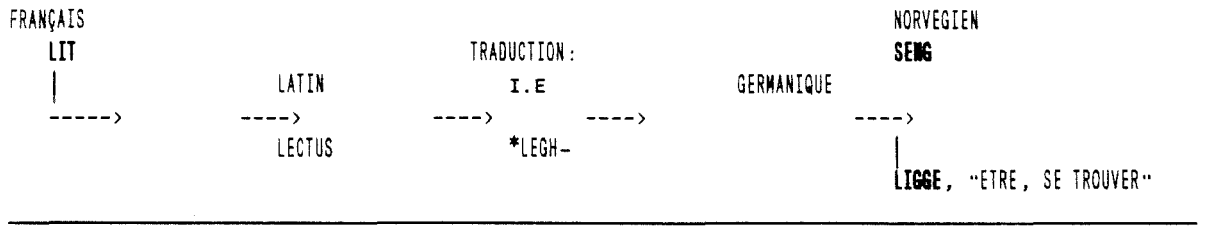
D'abord, le mot norvégien peut renvoyer à un autre mot français:

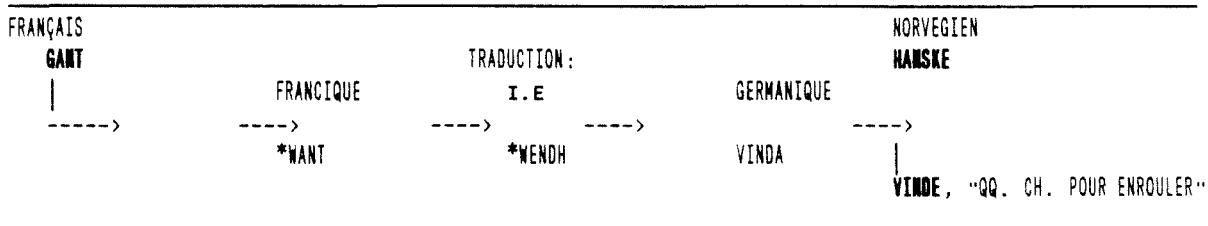

HABERE
*GHE/ABH-, "DONNER, RECEVOIR"
*GIFTI
LATIN
I.E
GERMANIQUE
AVOIR
FRANÇAIS
POISON
TRADUCTION:
NORVEGIEN
GIFT
MOLERE
*MEL-, "MOUDRE"
LATIN
I.E
GERMANIQUE
MOUDRE
FRANÇAIS
FARINE
TRADUCTION:
NORVEGIEN
MEL
SAEVUS
*SAI- "ETRE SOUFFRANT"
*SAIRA
LATIN
I.E
GERMANIQUE
SEVIR
FRANÇAIS
BLESSURE
TRADUCTION:
NORVEGIEN
SÅR
LOC
*LEUG- "POUSSER DE TRAVERS"
LYKILL
A. SAXON
I.E
GERMANIQUE
LOQUET
FRANÇAIS
CLE
TRADUCTION:
NORVEGIEN
NØKKEL
PELLICIA
PELLIS
LATIN
(I.E)
GERMANIQUE
PELISSE
FRANÇAIS
FOURRURE
TRADUCTION:
NORVEGIEN
PELS
VERBUM
WRE-TOR
*WER-/WRE-, "PARLER"
*WURDAM
LATIN
GREC
I.E
GERMANIQUE
VERBE, RHETEUR
FRANÇAIS
MOT
TRADUCTION:
NORVEGIEN
ORD
LASSUS
*LE- "ETRE RELACHE"
LATIN
I.E
GERMANIQUE
LAS, HELAS
FRANÇAIS
PARESSEUX
TRADUCTION:
NORVEGIEN
LAT
FINGERE, FIGURA, FICTIO
*DHEIGH- "FORMER, FAÇONNER"
LATIN
I.E
GERMANIQUE
FICTION, FEINDRE FIGURE
FRANÇAIS
PATE
TRADUCTION:
NORVEGIEN
DEIG

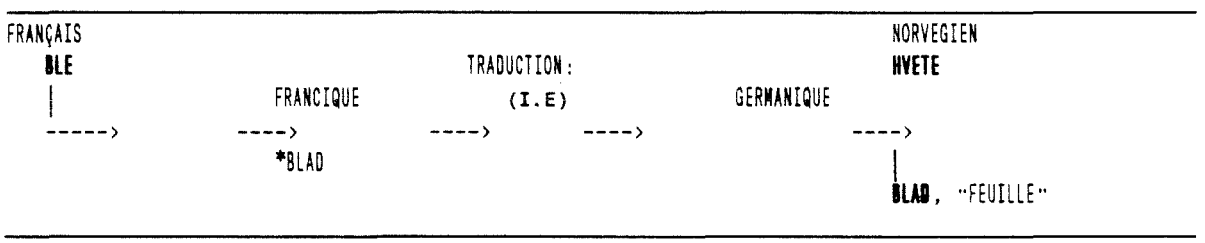


Ensuite, le mot français peut conduire à un autre mot norvégien que sa traduction:

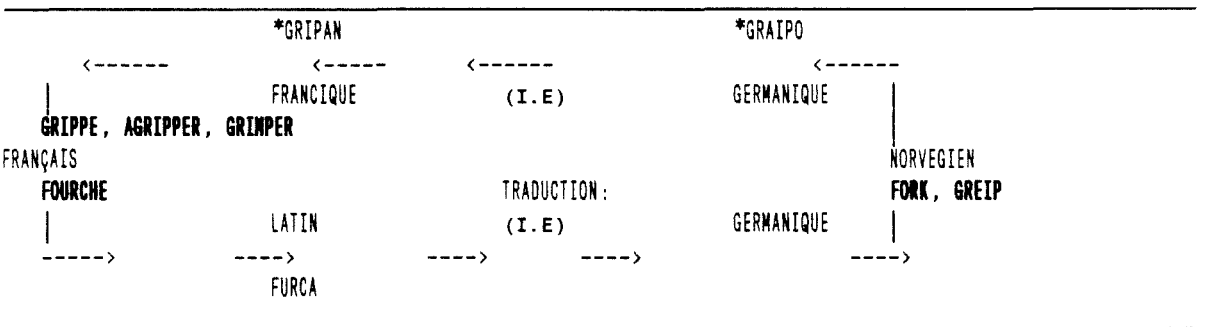


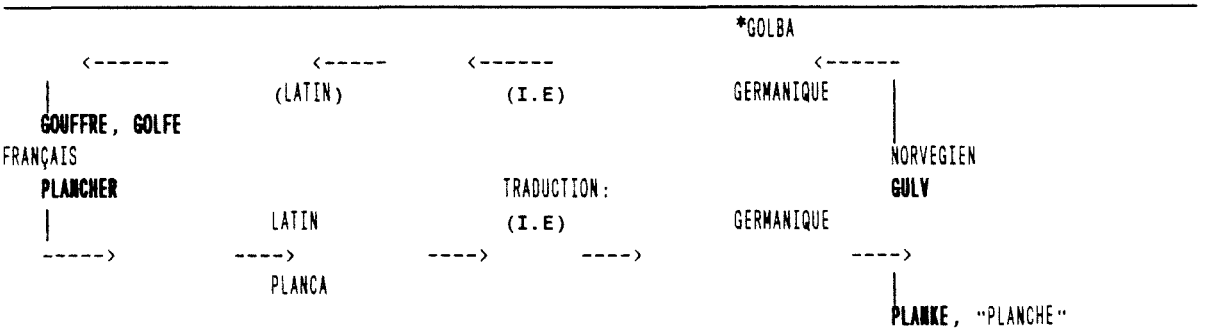


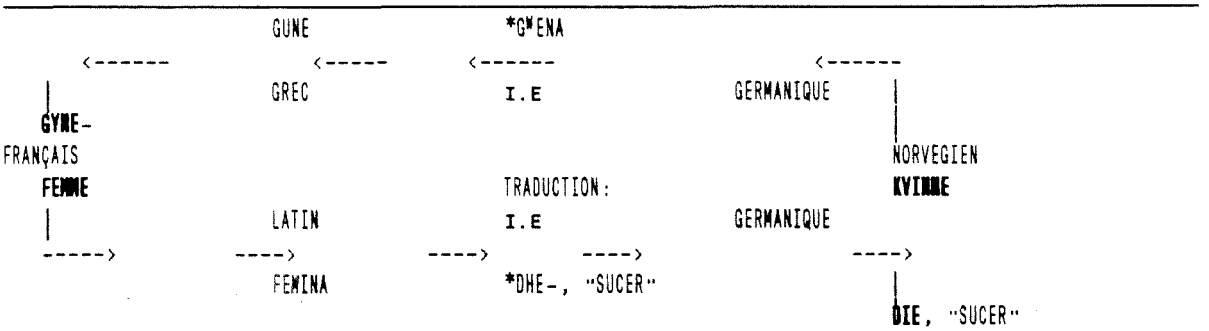


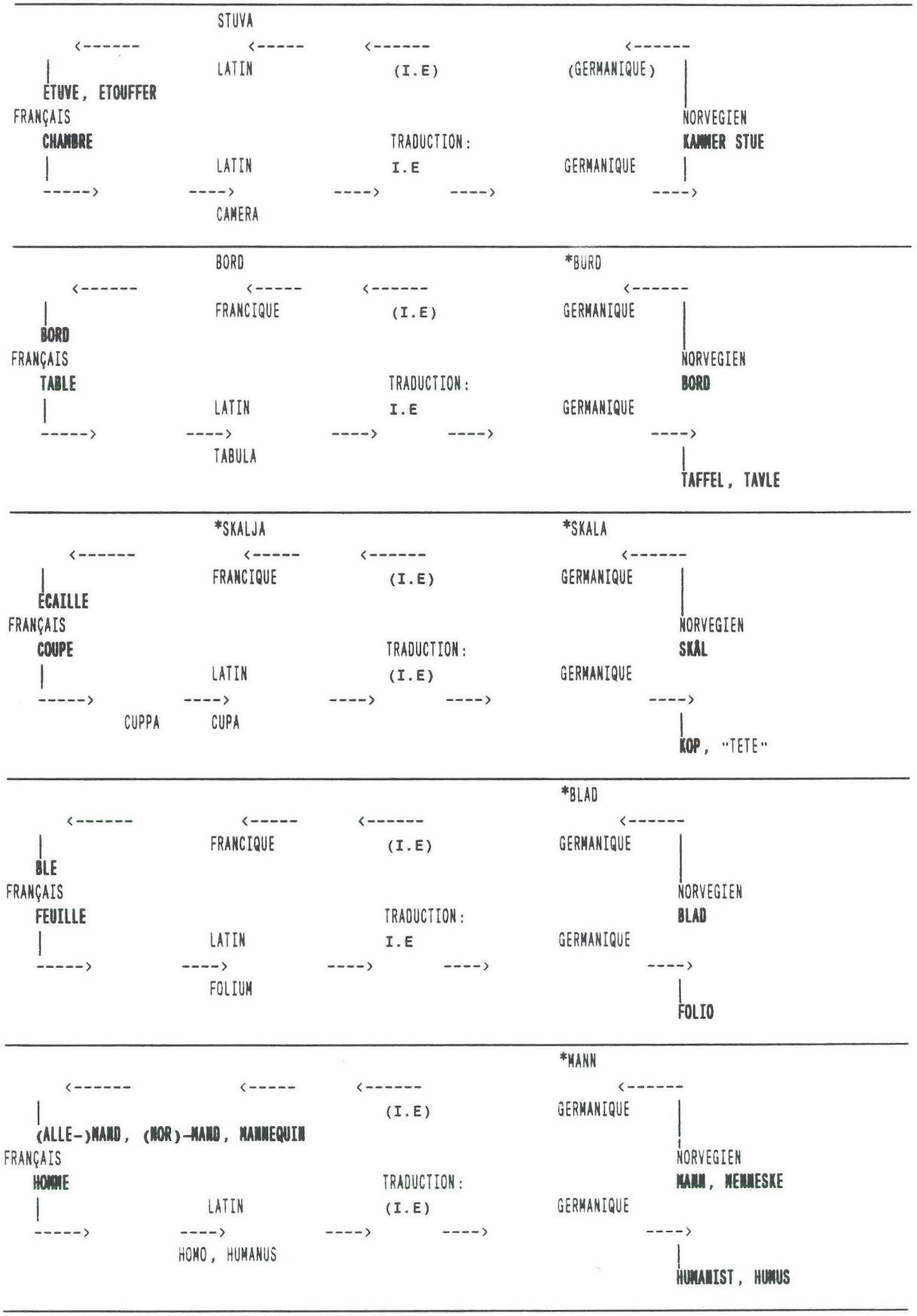


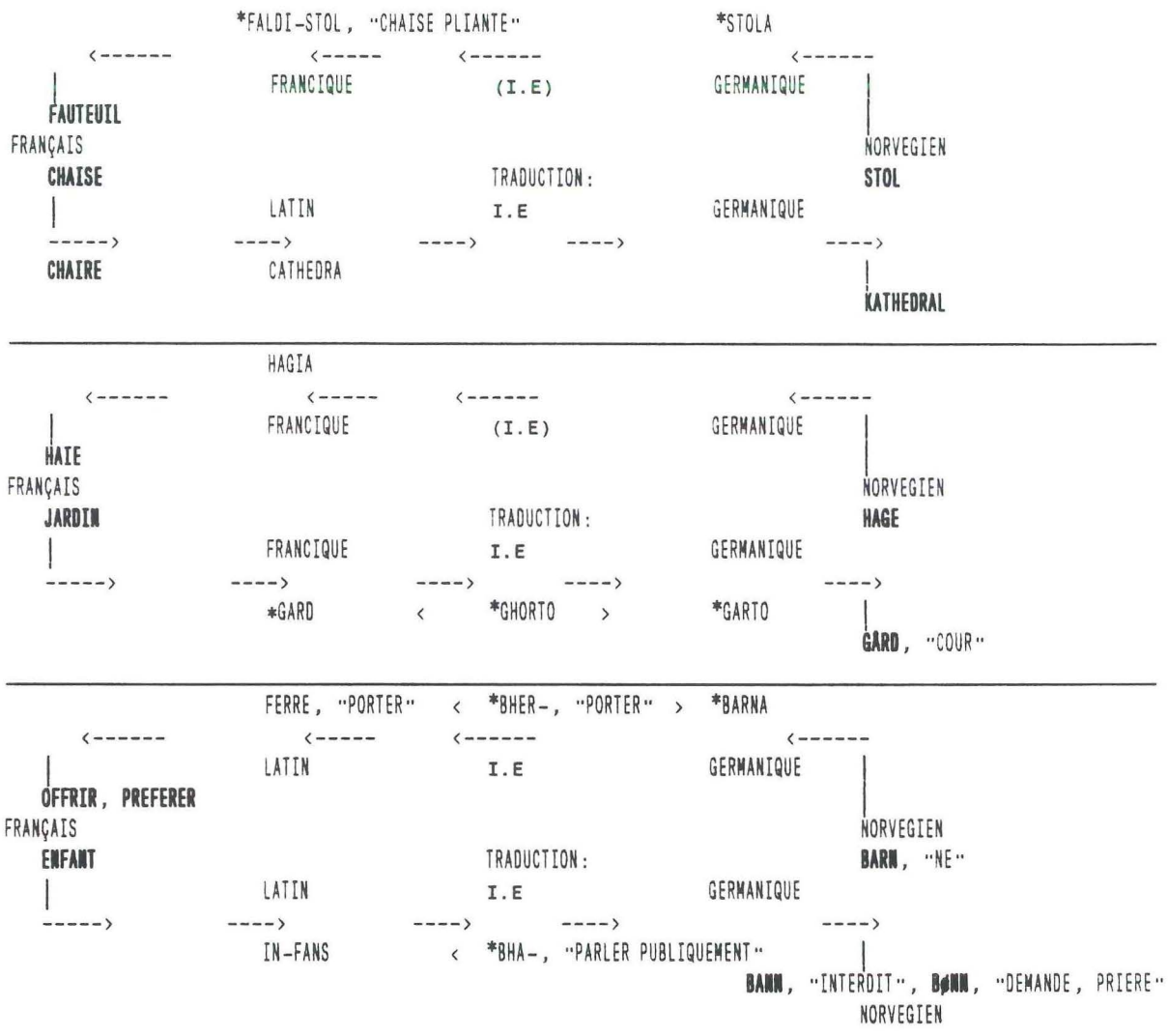
Enfin, les deux approches précédentes peuvent se combiner, selon un dispositif encore un peu plus complexe:








STUVA
LATIN
(I.E)
(GERMANIQUE)
ETUVE, ETOUFFER
FRANÇAIS
NORVEGIEN
CHAMBRE
TRADUCTION:
KAMMER STUE
LATIN
I.E
GERMANIQUE
CAMERA
BORD
*BURD
FRANCIQUE
(I.E)
GERMANIQUE
BORD
FRANÇAIS
NORVEGIEN
TABLE
TRADUCTION:
BORD
LATIN
I.E
GERMANIQUE
TABULA
TAFFEL, TAVLE
*SKALJA
*SKALA
FRANCIQUE
(I.E)
GERMANIQUE
ECAILLE
FRANÇAIS
NORVEGIEN
COUPE
TRADUCTION:
SKÅL
LATIN
(I.E)
GERMANIQUE
CUPPA
CUPA
KOP, "TETE"
*BLAD
FRANCIQUE
(I.E)
GERMANIQUE
BLE
FRANÇAIS
NORVEGIEN
FEUILLE
TRADUCTION:
BLAD
LATIN
I.E
GERMANIQUE
FOLIUM
FOLIO
*MANN
(I.E)
GERMANIQUE
(ALLE-)MAND, (NOR)-MAND, MANNEQUIN
FRANÇAIS
NORVEGIEN
HOMME
TRADUCTION:
MANN, MENNESKE
LATIN
(I.E)
GERMANIQUE
HOMO, HUMANUS
HUMANIST, HUMUS



On voit, au travers de ces quelques exemples, non commentés pour ne pas alourdir l'exposé, le profit que tirerait éventuellement d'une exploitation de la perspective historique l'enseignant norvégien de français, au gré de ses intérêts et de ses compétences, pour rendre la langue française plus familière. Montrer que, même si équivalences et transparences entre les deux langues sont rares, hormis le vocabulaire scientifique ou culturel, il reste de toute part des points de contact, des échanges, des emprunts, des héritages (en particulier via l'apport germanique considérable au français produit par la conquête franque), doit permettre de renforcer le sentiment d'une parenté invisible au premier regard, condition d'un éveil de l'intérêt intellectuel, et susciter l'envie de circuler par des voies détournées dans un réseau de significations qui facilite l'exercice de la mémoire.

La grande loi régissant l'activité mentale impliquée dans

l'acquisition du lexique est celle de l'analogie. Tout ethnologue, tout linguiste de terrain a fait cette expérience première d'avoir retenu immédiatement tel ou tel mot d'une langue exotique, simplement en fonction d'un lien, si ténu soit-il, avec un mot de sa langue maternelle ou d'une autre, plus familière. Dans une approche de l'apprentissage qui se donnerait comme objectif la seule compréhension passive d'une langue, on s'appuie nécessairement sur les mots qui sont d'emblée transparents. Mais ce ne sont pas ceux-là que vise à réduire puis à fixer une démarche qui recourt aux relations historiques.

III. La reduction d'une langue au connu.

On peut tenter d'illustrer la stratégie proposée à l'aide d'un exemple précis. Supposons un locuteur français désireux de connaître assez de tchèque pour lire de la documentation technique, des écrits scientifiques ou des textes relevant des sciences humaines. Il est probable qu'il a déjà une certaine familiarité avec l'anglais et sans doute l'allemand, qu'il est frotté de latin et/ou de grec, s'il est vrai que comme le disait un anglais bel esprit "pour être un gentleman, il ne faut pas savoir le latin et le grec, mais il est indispensable de les avoir oubliés", ou du moins il a conscience de la présence en français même de racines latines et grecques. Examinons dans quelle mesure il fera flèche de tout bois pour reconnaître s'il le peut, deviner s'il l'ose, et surtout retenir la signification des mots qu'il aura préalablement et péniblement cherchés dans un dictionnaire adapté à ses besoins.

Donnons-nous d'abord un corpus permettant de tester la validité à cet effet d'une approche historico-génétique. Choisissons, arbitrairement, un champ sémantique, celui des relations spatiales (localisation, forme, dimension). Pour ce faire, a été utilisée la liste donnée par un instrument de travail[71] qui, tout axé sur la comparaison qu'il soit, a le mérite de n'inclure que des concepts indispensables, au nombre de 62[72]. Ils nous

---

71. C'est l'outil fondamental que constitue le gros recueil de 1515 pp. de C.D. Buck *A Dictionary of selected Synonyms in the principal Indo-European Languages* Chicago 1949 [reprint 1988].

72. lieu, place / mettre, poser, placer / être couché / être debout / rester, demeurer, attendre / laisser / calme, tranquille / rassembler, recueillir / joindre, unir / séparer / ouvrir / fermer / couvrir / cacher / haut / bas / sommet, cime / fond / fin / point / bord, tranchant / côté / au milieu / droit / gauche / près / loin / est / ouest / nord / sud / forme / croître, pousser / mesurer / grand / petit / long / court / large / étroit / épais / dense / mince / clair, clairsemé / profond / peu profond / plat / creux / droit / tortueux / crochet / coin / croix / carré / rond / cercle / boule, globe / trou / égal / semblable / changer /

fournissent davantage encore de mots tchèques, qui, si l'on recherche systématiquement leur étymologie et donc les divers rapports de parenté entretenus avec les autres langues indo-européennes susceptibles d'être exploitées (latin, grec anglais allemand), se prêteraient aux explications historiques suivantes, certes un peu longues mais nécessaires néanmoins à l'intelligence de la démonstration.[73]

(1) LIEU, PLACE:
- *mjesto*: à rapprocher du lat. *meta*, "borne dans un cirque, lieu autour duquel tournent les chars" (on a aussi en tch. *meˆsto*, "ville").

(2) METTRE, POSER, PLACER:
- *postaviti*: causatif du verbe "être debout" (cf. 4).
- *polozˆiti*: causatif du verbe "être couché" (cf. 3).
- *klasti*: apparenté à l'angl. *to load* et à all. *laden*, "charger" (*hladan* en a. h. all.).

(3) ETRE COUCHE:
- *lezˆeti*: issu d'une racine i.e. **legh-* qui a fourni le lat. *lectus* (> fr. *lit*).

(4) ETRE DEBOUT:
- *stati*: d'une racine i.e. **stâ-* représentée par le lat. *stare*, avec de nombreux dérivés en fr. (*statue, station, constat*), l' all. *stehen*, l'angl. *to stand*.

(5) RESTER, DEMEURER, ATTENDRE:
- *zu°stati*: composé de *stati* (cf. 4).
- *cêkati*: d'origine obscure.

(6) LAISSER:
- *zu°staviti*: causatif de *zu°stati* (cf. 5).

(7) CALME, TRANQUILLE:
- *klidny*: d'origine obscure.
- *tichy*: sans correspondants utilisables.
- *pokojny*: se rattache à une forme verbale du v. sl. *po-cˆiti* dont le second élément a son correspondant dans le lat. *quies, quietis*, "repos" (cf. *quiétude, coi* en fr.).

(8) RASSEMBLER, RECUEILLIR:
- *sbirati*: d'une racine i.e. **bher-*, "porter", bien attestée par le gr. *pherô* et le lat. *fero* (avec de nombreux prolongements en fr.), ainsi que par l'angl. *to bear*.
- *shromazˆditi*: composé sur *hromazˆditi*, issu d'une racine **grem-* (cf. angl. *to cram*, "bourrer"), elle-même extension de **ger-*, "rassembler", qu'on trouve dans le gr. *agorâ*, "rassemblement, place", et le lat. *grex, gregis* (< **gre-g-s*), "troupeau" (d'où fr. *grégaire*).

(9) JOINDRE, UNIR:
- *spojiti*: composé sur *pojiti*, "lier", lui-même sans connexion.
- *slouc^iti*: composé sur *louc^iti* (cf. 10).

(10) SEPARER:
- *odlouc^iti*: composé sur *louc^iti*, "joindre" et "séparer", d'une racine i.e. **lenk-*, "fléchir, courber", d'où l'all. *entlang*, "le long de".
- *oddeliti*: composé sur le verbe du v. sl. *deliti*, "diviser", lui-même de *delu*, "portion,

---

signe.

73. Les accents circonflexes sont à lire inversés et affectent la lettre précédente.

part" (cf. all. *Teil*).

(11) OUVRIR:
- *otevr^iti*: composé où figure la racine i.e. **wer-*, "couvrir, fermer", présente dans le lat. *aperire* (< **ap-verire*), "ouvrir" (ce fr. mot en est issu, comme le savant *apéritif*).

(12) FERMER:
- *zavr^iti*: composé comme *otevr^iti* (cf. 11).

(13) COUVRIR:
- *(po)kryti*: se retrouve dans le gr. *kruptô*, "cacher" (cf. fr. *crypto-*).

(14) CACHER:
- *skryti*: (cf. (13).
- *schovati*: composé de *chovati*, "garder, préserver", à rapprocher du lat. *cavere*, "prendre soin de, faire attention à" (cf. *cave canem*, fr. *caution, cauteleux*).

(15) HAUT:
- *visoky*: d'un ancien slave **upsa*, élargissement de l'i.e. **upo*, dont le sens était "du bas vers le haut", d'où le gr. *hupo*, le lat. *sub*, "sous", mais aussi l'angl. *up*, "vers le haut".

(16) BAS:
- *nizky*: remonte au v. sl. *nizu*, "en bas", de l'i.e. **ni-/ nei-* que reflète aussi l'all. *nieder*.

(17) SOMMET, CIME:
- *vrch*: issu d'une racine qui se retrouve dans le lat. *verruca*, "éminence, excroissance" (> fr. *verrue*).

(18) FOND:
- *dno*: du v. sl. *duno* (< **dubno*) lui-même d'une racine i.e. **dheub-* qui, par ailleurs, a fourni l'angl. *deep*, "profond", *to dive*, "plonger", l'all. *Tief*, "profond".

(19) FIN:
- *konec*: mot général en slave mais sans connexions utilisables.

(20) POINT:
- *hrot*: apparenté à l'all. *Grat*, "crête d'une montagne".
- *s^pic^ka*: emprunt de l'all. *Spitze*, "pointe".

(21) BORD, TRANCHANT:
- *ostr^i*: d'une racine i.e. **ak*, dont le sens général est "pointu", et qui figure dans le gr. *akmê* (passé en français), le lat. *acies*, "pointe".
- *hrana*: de la même racine que *hrot* (cf. 20).
- *(o)kraj*: issu du v. sl. *-krojiti*, "couper", lui-même d'une racine i.e. **(s)ker-*, "couper", dont est aussi issu le lat. *curtus*, "mutilé" (> fr. *court*).

(22) COTE:
- *strana*: le sens premier en slave est "pays, région"; se rattache à une racine qu'on retrouve dans le lat. *sternere*, "s'étendre".
- *bok*: mot général en slave, mais d'origine douteuse.

(23) AU MILIEU:
- *str^edni*: en v. sl. *sreda*, issu, pour des raisons évidentes, de la racine i.e. **kerd-* "coeur", d'où le gr. *kardia*, le lat. *cor, cordis*, l'angl. *heart*.

(24) DROIT:
- *pravy*: proprement "ce qui est juste, en avant", issu de la racine i.e. **pro* "en avant,

devant", d'où gr. et lat. *pro*, "avant".

(25) GAUCHE :
- *levy*: au sens premier "courbé", correspond au lat. *laevus* (cf. fr. savant *lévogyre*).

(26) PRES :
- *blizko*: proprement "contigu, attenant, pressé contre", d'une racine dont le représentant lat. est *fligere*, "frapper" (cf. fr. *infliger*).

(27) LOIN :
- *daleko*: issu d'un radical i.e. **dlonghus* (cf. 37).

(28) EST :
- *vychod*: proprement "lever de soleil, ce qui monte, ce qui sort", formé sur le verbe *choditi*, "aller, venir", qui est issu de la même racine que le lat. *cedere*, "s'en aller" (cf. fr. *céder*).

(29) OUEST :
- *zapad*: formé, par référence à la chute derrière l'horizon, à partir de *za-*, "derrière", et d'un verbe v. sl. *padati*, "tomber", issu de même racine i.e. **ped-* que le gr. *pous, podos*, "pied".

(30) NORD :
- *sever*: aussi au sens de "vent du nord"; à relier à l'all. *Schauer*, "giboulée", et à l'angl. *shower*, "douche".
- *pu °lnoc*: proprement mi-nuit, composé de *pu °l*, "moitié", et *noc*, "nuit" (cf. lat. *nox, noctis*).

(31) SUD :
- *jih*: étymologie douteuse.

(32) FORME :
- *tvar*: dérivé du verbe v.sl. *tvoriti*, "faire", sans correspondant.
- *podoba*: composé qui, par le sens de "convenance", remonte à une racine i.e. **dhabh-* "adapter, mettre ensemble", qui a donné le lat. *faber*, "artisan"

(33) CROITRE, POUSSER :
- *ru °sti*: issu d'une forme sl. **orsto*, qui est à rattacher au lat. *arduus*, "haut" (> fr. *ardu*).

(34) MESURER :
- *me^r^iti*: issu de la racine i.e. **mê-*, de même sens, qui a donné le gr. *metron*, "mesure" (> fr. *mètre*).

(35) GRAND :
- *vel(i)ky*: issu de la racine i.e. **wel-*, "presser, comprimer" (d'ou une valeur intensive: v. sl. *velimi*, "très") qui, élargie en *-g*, a donné le lat. *vulgus*, "peuple", proprement "abondance de gens" (cf. fr. *vulgaire*).

(36) PETIT :
- *maly*: étymologie discutée; a été rapproché, par le biais du sens "moulu fin", de la racine i.e. **mel-*, "moudre", d'où le lat. *molere* (> fr. *moudre*).

(37) LONG :
- *dlouhy*: continue le radical i.e. **del&gho-*, représenté par le gr. *dolikhos* (cf. fr. *dolichocéphale*), tandis que le lat. *longus* sort du radical **dlongho-*.

(38) COURT :

- *kratky*: dérivé d'une forme verbale v. sl. *cˆriˆtu*, "couper", qui, comme *kraj* (cf. 21), remonte à l'i.e..

(39) LARGE:
- *sˆiroky*: mot général en slave mais d'étymologie inconnue.

(40) ETROIT:
- *uzky*: issu d'une racine i.e. **angh-*, "étouffer, oppresser", d'où le lat. *angere*, "serrer la gorge", l'all. *eng*, étroit".
- *tesˆny*: dérivé du v. sl. *tiskati*, "presser", sans correspondants utilisables.

(41) EPAIS:
- *tlusty*: sans doute issu d'une racine i.e. **teu-*, "se gonfler", qui a donné le gr. *tombos*, "tertre, tombeau", et le lat. *tumere*, "être enflé".

(42) DENSE:
- *husty*: d'un forme sl. ancienne **gong-to*, elle-même d'une racine i.e. **geng-*, "bloc, masse, pelote", d'où le lat. *gingiva*, "gencives" (cf. fr. *gingivite*).

(43) MINCE:
- *tenky*: issu de la racine i.e. **tenu-* (elle-même de **ten-*, "étendre") qui a donné le lat. *tenuis* (> fr. *ténu*).

(44) CLAIR, CLAIRSEME:
- *rˆidky*: du verbe v. sl. *oriti*, "dissoudre", lui-même d'une racine i.e. **erə-*, "lâche, séparé", représentée en lat. par *rarus* (> fr. *rare*).

(45) PROFOND:
- *hluboky*: sans doute à rattacher à un radical i.e. **glumbh-*, forme nasalisée d'une racine reflétée par le gr. *gluphô*, "creuser, sculpter" (cf. fr. savant *-glyphe*).

(46) PEU PROFOND:
- *melky*: issu, par l'intermédiaire de "finement moulu, fin", de la même racine que *maly* (36).

(47) PLAT:
- *plochy, plosky*: d'une racine i.e. **pelə- / plə-*, qu'on retrouve dans le lat. *planus* (avec élargissement nasal), d'où *plan, plain* en fr., le gr. *platus* (avec élargissement dental), d'où *plat* en fr. par le lat. vulgaire, et l'angl. *flat*, enfin dans l'all. *flach* (avec élargissement dorsal).

(48) CREUX:
- *duty*: se rattache, par l'idée de gonflement, à la racine i.e. **dheu-*, "souffler", repérable dans le lat. *fumus*, le gr. *thumos*, "âme, esprit" (fr. *cyclo-thymique*), *tuphô*, "souffler", (> fr. *typhus*), l'angl. *dust*, "poussière".

(49) DROIT:
- *rovny*: aussi au sens de "égal, plan, au niveau"; sort d'une racine qui se retrouve dans le lat. *rus*, "campagne, espace ouvert", l'angl. *room*, "place, pièce", l'all. *Raum*, "espace".
- *prˆimy*: d'origine obscure.

(50) TORTUEUX:
- *krivy*: issu de la même racine que le lat. *curvus* (> fr. *courbe*).

(51) CROCHET:
- *hak*: emprunté d'une forme germanique ancienne, en all. moderne *Haken*, en angl. *hook*.

(52) COIN:

- *roh*: aussi au sens de "corne", mot isolé dans le domaine i.e.
- *kout*: sans doute apparenté au lat. *canthus*, "bandage de roue", d'où le fr. *chant*, au sens de "côté le moins large d'un objet" (écrit *champ* dans *mettre qq. ch. de champ*).
- *uhel*: la forme en v. sl. est *agulu*, à rapprocher du lat. *angulus* (> fr. *angle*).

(53) CROIX:
- *krˆizˆ*: emprunté du lat. *crux, crucis*.

(54) CARRE:
- c*ˆtverec*: dérivé de *cˆtyrˆi*, "quatre", qui lui-même répond au lat. *quattuor*.
- *cˆtverhan*: composé de *cˆtyrˆi*, "quatre", et de *hrana*, "côté".

(55) ROND:
- *okrouhly*: dérivé de *kruh*, "cercle" (cf. 56).
- *kulaty*: dérivé de *koule*, "balle" (cf. 57).

(56) CERCLE:
- *kruh*: d'une racine i.e. **(s)ker-* qui se retrouve dans le lat. *circus* dont le diminutif a donné le fr. *cercle*.

(57) BOULE, GLOBE:
- *koule*: emprunté d'une ancienne forme all. qui a donné *Keule*, "massue", en all. moderne.

(58) TROU:
- *dira*: tiré du verbe du v. sl. *dirati*, "déchirer" qui correspond à l'angl. *to tear*.

(59) EGAL:
- *stejny*: dérivé du verbe *stati*, "se tenir debout, immobile, rester", d'où "stable, inchangé" (cf. 4).

rovny: voir (49).

(60) SEMBLABLE:
- *podobny*: composé sur *doba*, "occasion, ce qui convient" (cf. 32).

(61) CHANGER:
- *mˆeniti*: issu d'une racine i.e. **mei-/ moi-*, "changer, échanger", reflétée dans l'all *gemein*, le lat. *com-munis*, "commun".

(62) SIGNE:
- *znak, znameni*: dérivés du verbe du v. sl. *znati*, "savoir", issu d'une racine i.e. **gen&-/ gnô-*, également repérable dans le gr. *gignôskô*, le lat. *co-gnosco*, l'all. *kennen*, et l'angl. *to know*.

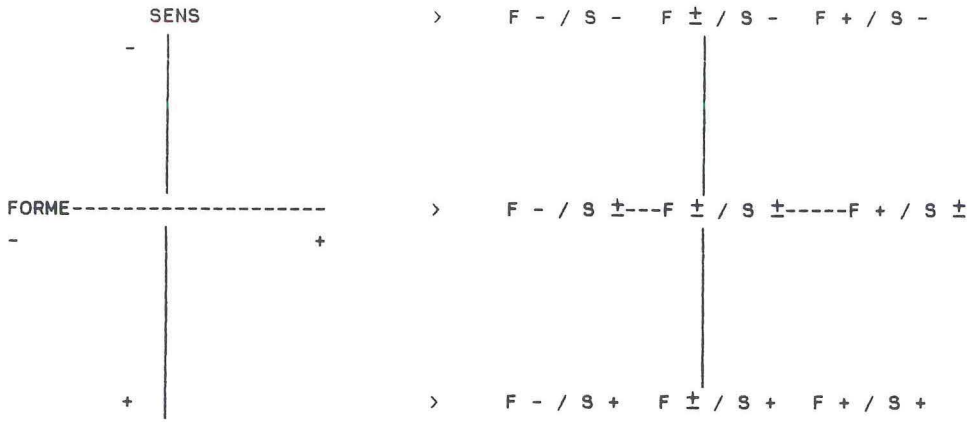
On voit qu'à une soixantaine de concepts correspondent 83 mots tchèques. Parmi ceux-ci, 12 n'offrent aucune prise à un quelconque rapprochement. De ceux qui restent, tous, soit immédiatement, soit après une décomposition en oeuvre à l'intérieur même de la langue tchèque, peuvent être reliés à un ou plusieurs mots des langues qu'on a supposées préalablement connues. Il peut s'agir d'emprunts, dans quelques cas, mais, le plus souvent, le rapprochement, s'effectuant du point de vue d'une parenté génétique, réclame un détour par une forme reconstruite au niveau indo-européen (radical ou racine), puisqu'il convient de repasser par cette instance assurant l'unité du groupe slave avec le groupe germanique (si

allemand ou anglais sont en cause), avec le latin, ou avec le grec.

Il est évident qu'il faut se passer d'explications aussi lourdes et aussi complexes que celles qui ont été données pour que le rapprochement soit utilisable par l'apprenant sans induire de nouvelles difficultés, sinon on se trouverait dans la situation paradoxale d'avoir à expliquer *obscurum per obscurius*. L'objectif n'est pas de faire du lecteur un Antoine Meillet de pacotille. On doit donc examiner la validité des rapprochements découverts par la perspective historique, du seul point de vue de leur efficacité brute, abstraction faite du cheminement étymologique qui les sous-tend. La vraie question devient: les "équations" posées, quelles que soient les médiations historiques impliquées facilitent-elles l'apprentissage?

Pour en juger, on recourra à deux critères appréciant les paires ainsi isolées, d'un mot tchèque avec un mot d'une langue plus familière. Ces critères portent, d'une part, sur le sens, de l'autre, sur la forme. Le rapport entre les significations des deux mots peut être tel que l'écart soit maximum, et donc pratiquement irréductible, mais il ira parfois aussi jusqu'à l'identité, en passant par tous les degrés de proximité. On a, par exemple, semble-t-il, une distance maximale entre *odlouc^iti*, "séparer" et all. *entlang*, "le long de", ou entre *husty*, "dense", et lat. *gingiva*, "gencives". A l'autre bout, la signification de *otevr^iti* et du *aperire* est la même, "ouvrir".

Si l'on adopte maintenant, mettant cette fois entre parenthèses la signification, le point de vue de la forme, il en va de même. Entre *uzky* et l'all. *eng*, la distance est considérable, ainsi qu'entre *visoky* et gr. *hupo*, lat. *sub*, angl. *up*, mais elle devient relativement acceptable entre *(za)-pad* et gr. *pod-os*, dans le cas de *kr^iz^* en face du lat. *crux, crucis*, comme entre *stati* et lat. *stat*, avec, de nouveau, tous les cas de figure intermédiaires. En matière de forme comme de sens, la difficulté est évidemment de positionner tous ces cas qui à considérer comme médians. On peut ensuite croiser les deux critères selon l'axe vertical de la proximité sur le plan du sens (du - au +) et, de même, selon celui, horizontal, de la proximité quant à la forme:



On obtient un tableau à double entrée, avec neuf possibilités seulement, en fonction de l'existence d'une zone d'indécidabilité dans les deux dimensions, mais qui devrait permettre de classer, dans une première approximation, les équations obtenues. Puisqu'il faut pourtant présenter les données de manière linéaire, faute de place, on accordera la prééminence au critère d'ordre formel, en lui subordonnant la variable de la proximité dans le domaine des significations. En effet, quelle que soit l'identité sur ce dernier plan, elle ne prêtera jamais son appui à la mémoire si l'éloignement des formes est trop grand, alors que, si les formes des mots rapprochés s'y prêtent, il sera toujours possible d'accepter une distance sémantique même maximale, car elle est toujours réductible. Il n'est, finalement, aucune signification qui ne puisse, au prix d'une construction intellectuelle plausible, s'articuler à une autre par une médiation quelconque. Car est alors en acte, surtout dans la volonté de comprendre mobilisée et comme tendue par l'apprentissage, l'esprit qui toujours relie, pour paraphraser Goethe en le détournant.

Voici donc les équations finalement posées entre "tchèque" et "fr. / lat. / gr. / all. / angl.", que livre l'enquête historique, mais réparties selon les neuf classes qui résultent de l'application des deux critères de similitude et enfin ordonnées en fonction de leur degré de ressemblance formelle.

CLASSE I. Forme: - Sens: - :
(10A) *od-louc^iti* SEPARER / all. *entlang*, "le long de".
(41) *tlusty* EPAIS / gr. *tombos*, "tertre, tombeau", lat. *tumere*, "être enflé".
(42) *husty* DENSE / lat. *gingiva*, "gencives".
(52A) *kout* COIN / lat. *canthus*, "bandage de roue".

CLASSE II. Forme: - Sens: ± :
(2) *klasti* METTRE, POSER, PLACER / angl. *to load*, all. *laden*, "charger".
(32) *po-doba* FORME / lat. *faber*, "artisan".

CLASSE III. Forme: - / Sens: + :

(11) *ote-vr^iti* OUVRIR / lat. *aperire.*
(15) *visoky* HAUT / gr. *hupo*, lat. *sub*, "sous", angl. *up*, "vers le haut".
(18) *dno* FOND / angl. *deep*, "profond", *to dive*, "plonger", all. *Tief.*
(40) *uzky* ETROIT / lat. *angere*, "serrer la gorge", all. *eng*, étroit".
(44) *r^idky* CLAIR, CLAIRSEME / lat. *rarus*, "rare".

CLASSE IV. Forme: ± / Sens: - :
(21c) *o-kraj* BORD, TRANCHANT / lat. *curtus*, "mutilé".
(26) *blizko* PRES / lat. *fligere*, "frapper".
(28) *vy-chod* EST / lat. *cedere*, "s'en aller".
(33) *ru°sti* CROITRE, POUSSER / lat. *arduus*, "haut".
(48) *duty* CREUX / lat. *fumus*, gr. *tuphô*, "souffler", angl. *dust*, "poussière".
(49) *rovny* DROIT / lat. *rus*, "campagne, espace ouvert", angl. *room*, "place, pièce", all. *Raum*, "espace".

CLASSE V. Forme: ± / Sens: ± :
(1) *mjesto* LIEU, PLACE / lat. *meta*, "borne".
(8B) *shromaz^diti* RASSEMBLER, RECUEILLIR / angl. *to cram*, "bourrer".
(30A) *sever* NORD / all. *Schauer*, "giboulée", angl. *shower*, "douche".

CLASSE VI. Forme: ± / Sens: + :
(7) *po-kojny* CALME, TRANQUILLE / lat. *quies*, *quietis*, "repos".
(8A) *sbirati* RASSEMBLER, RECUEILLIR / gr. *pherô*, lat. *fero*, angl. *to bear.*
(21A) *ostr^i* BORD, TRANCHANT / gr. *akmê*, lat. *acies*, "pointe".
(27) *daleko* LOIN / lat. *longus.*
(38) *kratky* COURT / lat. *curtus*, "mutilé".
(52B) *uhel* COIN / lat. *angulus*, "angle".
(56) *kruh* CERCLE / *circus*, "cirque".

CLASSE VII. Forme: + / Sens: - :
(21B) *hrana* BORD, TRANCHANT / all. *Grat*, "crête d'une montagne".
(29) *za-pad* OUEST / gr. *pous, podos*, "pied".
(36) *maly* PETIT / lat. *molere*, "moudre".
(46) *melky* PEU PROFOND / lat. *molere*, "moudre".

CLASSE VIII. Forme: + / Sens: ± :
(14) *s-chovati* CACHER / lat. *cavere*, "prendre soin de".
(17) *vrch* SOMMET, CIME / lat. *verruca*, "excroissance".
(20A) *hrot* POINT / all. *Grat*, "crête d'une montagne".
(22) *strana* COTE / lat. *sternere*, "s'étendre".
(23) *str^edni* AU MILIEU / gr. *kardia*, lat. *cor, cordis*, angl. *heart* "coeur".
(30B) *pu°l-noc* NORD / lat. *nox*, "nuit".
(35) *vel(i)ky* GRAND / lat. *vulgus*, "abondance de gens".
(58) *dira*, TROU, *dirati*, "déchirer" / angl. *to tear.*
(61) *m^eniti* CHANGER / all *ge-mein*, lat. *com-munis*, "commun".

CLASSE XI. Forme: + / Sens: + :
(20B) *s^pic^ka* POINT < all. *Spitze*, "pointe".
(51) *hak* CROCHET < all. *Haken*, angl. *hook.*
(57) *koule* BOULE, GLOBE < all. *Keule*, "massue".
(53) *kr^iz^* CROIX < lat. *crux.*
(3) *lez^eti* ETRE COUCHE / lat. *lectus.*
(4) *stati* ETRE DEBOUT / lat. *stare*, all. *stehen*, angl. *to stand.*
(10B) *od-deliti* SEPARER; *delu*, "portion, part" / all. *Teil.*
(13) *po-kryti* COUVRIR / gr. *kruptô*, "cacher".
(16) *nizky* BAS / all. *nieder.*
(24) *pravy* DROIT / gr. et lat. *pro*, "avant".
(25) *levy* GAUCHE / lat. *laevus.*
(34) *me^r^iti* MESURER / gr. *metron*, "mesure".
(37) *dlouhy* LONG / gr. *dolikhos.*
(43) *tenky* MINCE / lat. *tenuis*, "ténu".
(45) *hluboky* PROFOND / gr. *gluphô*, "creuser, sculpter".
(47) *plochy, plosky* PLAT / lat. *planus*, gr. *platus*, angl. *flat*, all. *flach.*
(50) *krivy* TORTUEUX / lat. *curvus*, "courbé".
(54) *c^tverec, c^tyr^i* CARRE / lat. *quattuor.*
(59) *stejny*, EGAL, *stati*, "se tenir debout" / lat. *stare.*
(62) *znak, znameni*, SIGNE, *znati*, "savoir" / gr. *gi-gnôskô*, lat. *co-gnosco*, all. *kennen*, angl. *to know.*

On ne reviendra pas sur l'arbitraire qui peut, aux yeux de chaque

lecteur, selon son idiosyncrasie, sa formation, sa fantaisie, sembler régner dans l'affectation du signe ± pour signaler une analogie sujette à caution dans l'ordre formel ou sémantique. Du moins peut-on s'accorder sur des pôles. Or, si l'on adopte le principe d'un primat de l'analogie de forme pour prêter main forte à l'exercice de la mémoire, à savoir si l'on considère que sont pertinents les rapprochements figurant dans les classes VII, VIII et IX, c'est près d'un mot tchèque sur deux qui par rapprochement avec un ou plusieurs mots d'une langue connue par le locuteur français va se trouver éclairé dès que l'équation est posée. Il est même possible, à condition de solliciter juste un peu plus les facultés de l'apprenant, de faire passer la limite entre les classes IV et V, d'inclure donc dans la catégorie des rapprochements éclairants également les classes V et VI. Dans ce cas, ce sont les deux tiers du vocabulaire de base de la langue tchèque qui pourraient être l'objet d'une approche utilisant des connaissances préalables dans son apprentissage. Ce pourcentage est tout à fait remarquable et - ajoutons-le - inattendu.

Le problème clé de tout apprenant consiste non pas à deviner le sens d'un mot rencontré pour la première fois dans un texte d'une langue qu'il est en train de découvrir - ce qui ne peut s'effectuer, et en courant le risque de tomber sur des faux-amis, que sur des langues unies par la parenté la plus étroite - mais, bien plutôt, après l'avoir identifié, de trouver un moyen de le garder en mémoire pour remobiliser ultérieurement cette connaissance chèrement acquise. Or, l'effort pour établir le lien entre une forme constatée et sa signification tirée du dictionnaire est grandement facilité si l'existence d'un mot dans une langue dont il a également connaissance lui permet de surmonter l'arbitraire du signe.

Dès lors, le problème majeur devient le suivant: comment apporter à celui qui est engagé dans le processus d'apprentissage d'une langue relevant du domaine indo-européen les informations adaptées à une stratégie d'apprentissage axée sur la compréhension textuelle et désireuse d'exploiter la dimension historique et comparative.

Une première démarche reviendrait à constituer des dictionnaires bilingues fournissant ce type de passerelles historiques qui sont autant de pilotis potentiels pour soulager le travail de fixation opéré par la mémoire. Il prendrait pour modèle ce que nous avions esquissé pour le tchèque au travers de la première liste que nous avions dressée. Les

informations ne devront pas être outrageusement réductrices, car il est toujours difficile de se faire une représentation des acquis et des capacités du lecteur, et, à condition d'être correctement organisées, on ne voit pas que l'abondance des données puisse constituer un obstacle à l'extraction de celles qui sont pertinentes pour chacun.

On peut cependant esquisser un programme minimum. En face d'un mot tchèque, le locuteur français devrait trouver son étymon en slave commun, accompagné de ses correspondants dans les autres langues slaves (russe, polonais, bulgare, serbo-croate), de là, il pourrait remonter, si on en dispose, vers la racine ou le radical indo-européen qui en est à la source; il redescendrait enfin vers un choix de reflets de cette même racine dans un certain nombre de langues, langues romanes et germaniques, supposées plus ou moins connues, via les étymons disponibles.

Il est évident que c'est du côté des moyens informatiques qu'il faudrait à terme se tourner pour exploiter au mieux ces données étymologiques, car dans ce cas, le lecteur pourrait, en fonction de ses connaissances particulières, sélectionner au préalable les langues qu'il considère comme seules pertinentes pour faciliter son approche du lexique visé et il filtrerait ainsi les informations requises par un choix plus restreint.

Dans une perspective plus ambitieuse encore, on peut envisager d'élaborer des dictionnaires comparatifs multilingues mettant en évidence le réseau de relations unissant les langues de l'espace indo-européen.

Deux formules peuvent être avancées pour leur réalisation.

La première partirait d'un classement par concepts et présenterait pour chacun leur réalisation dans les différentes langues retenues, renvoyant, pour entrer dans la profondeur historique de chaque terme, à des synthèses présentant les informations du point de vue indo-européen qui retraceraient l'engendrement des formes à partir des racines, sur le modèle de la présentation de la racine i.e. AK- esquissée plus haut. Le soin serait alors laissé au lecteur d'isoler dans ce champ ce qui lui paraîtrait utile à son objet, en fonction de son savoir antérieur.

Une autre, sans doute plus lourde d'exposition, mais plus immédiatement commode, consisterait à présenter d'abord de manière linéaire

les traductions d'un mot français en autant de langues qu'on voudra, puis, en arrière de cette masse de données fournies de manière synoptique, restituerait les regroupements, les parentés, les contacts, les origines communes, bref le tissu historique dans lequel sont pris les mots des langues vivantes de même signification, en les accompagnant de renvois lorsqu'un autre tableau peut élargir encore la compréhension d'un tel réseau. Bien évidemment, des index dressés par langues permettraient d'accéder au tableau procurant à la fois la traduction du terme cherché et l'information historique sur laquelle il ouvre éventuellement.

On trouvera en annexe trois esquisses d'un tel répertoire comparatif, l'un portant sur le concept de FIN, au sens temporel, l'autre sur PLEURER, le dernier sur ESPOIR. Voici les matériaux sur lesquels s'appuient les deux premiers et qui justifient les détails de la construction[74]; ils pourraient accompagner ces tableaux à titre de commentaire et d'explication.

FIN, au sens temporel.
Fr. *fin*, It. *fine*, Esp. *fin*, Port. *fim*, sont tous issus du latin *fînis* et également transparents, mais Roum. *sfîrsit* dérive du verbe *sfîrsi*, "finir", qui est emprunté au slave (voir la forme serbe). [Pok. 243]
Le groupe des langues germaniques offre en parallèle deux séries de formes: Angl. *end*; All. *ende*; Né. *einde*; Da. *ende*; Su. *ände*. La forme en gotique est *andeis* et doit être reliée au lat. *ante*, "avant", au grec *anti*, "contre, pour". [Pok. 48]
D'autre part on a All. *schluss*; Né. *slot*; Da. *slutning*; Su. *slut*, qui remontent à une forme verbale attestée en v.h.a. *sliozan*, "fermer", à rapprocher du latin *claudere* (> fr. *clore*), du grec *kleiô*, de même sens, qui sortent l'un et l'autre d'une racine i.e. **klau- / klâwi*, "crochet, cheville servant à fermer" (cf. lat. *clavus*, "clou", *clavis*, "clef"), avec cette nuance que les formes germaniques sont issues de la variante **skl-* (avec *s-* prothétique). [Pok. 604]
Gr. mod. *telos* < gr. ancien où le sens principal est "achèvement d'un cycle", d'où "perfection, fin, résultat", mais qui a comme sens premier "point central", sort d'une racine **k*^{w}*el-*, "tourner autour de, séjourner résider", la labio-vélaire initale ayant connu un traitement diversifié selon les langues (lat. *colere*, "cultiver, habiter") et en grec même où l'on a *telos* < *k^{w}el-e/os, mais aussi *polos*, "axe d'une sphère" < *k^{w}ol-o-, et *kuklos*, "cercle" < *k^{w}(e)-k^{w}l-o-. [Pok. 639]
Le Lit. *galas*, dont le sens est issu de celui de "pointe, pique", sort d'une racine **gwel-* "percer", d'où est sans doute issu aussi l'angl. *to kill*, "tuer". [Pok. 470]
L'autre mot lit. est *baiga*, verbe *baigti*, "finir", sans étymologie.
Dans le domaine slave, l'ensemble est homogène: Serbe *konac*; Tch. *konec*; Pol. *koniec*; Russe *konec*, formes dérivées de **konu* qui a le sens de "commencement", avec une dérivation sémantique concevant le début comme étant l'une des extrémités. Cette racine **ken-*, "surgir, commencer", est représentée dans le gr. *kainos*, "neuf", le lat. *recens*, "jeune, nouveau" (au sens propre, "proche de la naissance"), l'all. *hinter*, "derrière". [Pok. 563]
On a en outre le serbe *svrs^etak*, de *svrs^iti*, "finir", qui est un composé de *vrs^iti*, "accomplir, achever", lui-même déverbal de *vrh* (cf. Russe *verch*, "sommet", à rapprocher du latin *verruca*, "verrue"). [Pok. 1161]

PLEURER

---

74. Le principal instrument permettant ces mises au point est fourni par le recueil classique de Julius Pokorny *Indogermanisches Wörterbuch* 2 Vol. 1183 et 495 pp. Berne/Vienne 1969.

Fr. *pleurer*; Esp. *llorar*; Port. *chorar* sont issus du lat. *plorare* qui avait comme sens premier "crier fort" (cf. *im-plorare*, "implorer"), "pleurer bruyamment". [Pok. 831]
Mais It. *piangere* et Roum. *pînge* sortent du lat. *plangere*, "battre", d'où "se battre la poitrine (en signe de deuil)" et "se lamenter" (> fr. *plaindre*), issu d'une racine i.e. **plâg- / plâk-* "frapper", qui, par le même développement sémantique, a fourni le v. sl. *plakati*, "se lamenter, pleurer" et les autres formes slaves: Serbe et Tch. *plakati*; Pol. *plakac'*; Russe *plakat'*. [Pok. 832]
Le grec *klaiô*, d'étymologie douteuse, est probablement sans lien avec *kaleô*, "appeler".
Le groupe germanique offre, parfois en parallèle, des formes de provenance différente. D'une part Dan. *groede*; Suéd. *gråde*, d'une racine i.e. **ghrê-d-* qui, sous une forme causative, apparaît aussi dans l'anglais *to greet*, l'all. *grüssen*, mais au sens de "adresser la parole, saluer". [Pok. 439]
De l'autre, Angl. *weep* qui remonte à la racine i.e. **wâb-*, "appeler, crier", mais aussi *to cry*, qui est un emprunt au français *crier*. [Pok. 1109]
Enfin, le Néer. *weenen*; All. *weinen*, sont dérivés d'une interjection i.e. qu'on trouve en gotique, *wai*, "malheur!", en lat. *vae* (*vae victis!*). [Pok. 1110]
Le Lit. *verkti*, qui est à relier à la forme *urkti*, "grogner", est d'origine expressive.

Les différentes directions de recherche tournant autour d'une réappropriation de la dimension génétique dans une approche des langues européennes centrée sur leur acquisition du point de vue d'une compréhension passive (pour le tchèque), ou d'une étude du lexique (dans le cas du français pour un norvégien) demanderaient évidemment à être mises à l'épreuve pédagogique. De même, il faudrait s'atteler à l'élaboration systématique de recueils tabulaires multilingues. Ces perspectives ont néanmoins le mérite d'attirer l'attention sur un objet d'études aujourd'hui négligé. La comparaison des langues est la discipline fondatrice de la linguistique moderne en tant que science, tous les grands linguistes de notre siècle ayant d'abord été des comparatistes. Elle a obtenu des résultats considérables et même si elle se risque de nos jours dans des voies qui peuvent paraître aventurées, elle ne mérite certes pas l'oubli un peu condescendant qu'on lui porte souvent aujourd'hui. Ce sera peut-être à la didactique des langues d'en exploiter les principaux acquis et de lui redonner une légitimité nouvelle.

BIBLIOGRAPHIE GENERALE:

J. Aitchison *Language change: Progress or Decay?* Londres, 1981.

R. Anttila *An Introduction to Historical and Comparative Linguistics.* New York, 1972.

S. Auroux éd. *Histoire des idées linguistiques. T. 2: Le développement de la grammaire occidentale.* Liège, 1992.

R. Austerlitz "L'ouralien" *in* Martinet (1968), 1331-1387.

P. Baldi *An Introduction to the Indo-European Languages.* Carbondale Ill., 1983.

P. Baldi "Indo-European Languages" *in* Comrie (1990), 33-67.

E. Benveniste *Le Vocabulaire des Institutions indo-européennes. I. Economie, parenté, société. II. Pouvoir, droit, religion.* Paris, 1969-73.

K. Bergsland, H. Vogt "On the validity of Glottochronology", *Current Anthropology* 3, 1962, 115-153.

L. Bloomfield *Language.* Chicago, 1933.

W. Bright "North American Indian languages" *Encyclopaedia Britannica*, 1974, Vol.13, 208-213.

B. Brogyanyi éd. *Festschrift for Oswald Szemerényi on the Occasion of his 65th Birthday.* Amsterdam, 1979.

C.D. Buck *A Dictionary of selected Synonyms in the principal Indo-European Languages.* Chicago, 1949.

R. Burling *Patterns of language.* San Diego CA., 1992.

T. Bynon *Historical Linguistics.* Cambridge 1977.

L. Campbell Review of "Language in the Americas" *Language*, 64 N° 3, 1988, 591-615.

J. Cantineau "Racines et schèmes" *in Mélanges W. Marçais.* Paris, 1950, 119-124.

D. Cohen "Les langues chamito-sémitiques" *in* Martinet (1968), 1288-1329.

D. Cohen *Les Langues dans le monde ancien et moderne: langues chamito-sémitiques.* Paris, 1989.

M. Cohen *Le Langage (structure et évolution).* Paris, 1950.

N.E. Collinge *The Laws of Indo-European.* Amsterdam, 1985.

N.E. Collinge éd. *An Encyclopaedia of Language.* Londres, 1990.

N.E. Collinge "Language as it evolves: Tracing its Forms and Families" *in* Collinge (1990), 876-916.

B. Comrie *Language Universals and Linguistic Typology.* Chicago, 1981.

B. Comrie éd. *The World's Major Languages.* New York/Oxford, 1990.

B. Comrie "Languages of the World: Who speaks what" *in* Collinge (1990), 956-983.

W. Cowgill "Indo-european languages" *Encyclopaedia Britannica.* 1974, Vol.9, 431-438.

W. Croft *Typology and Universals.* Cambridge, 1990.

T. Crowley *An Introduction to Historical Linguistics.* Auckland/Oxford, 1992.

X. Delamarre *Le Vocabulaire Indo-Européen. Lexique étymologique thématique.* Paris, 1984.

G. Devoto *Avviamento alla etimologia italiana. Dizionario etimologico.* Florence, 1968.

I.M. Diakonoff "Hamito-Semitic languages" *Encyclopaedia Britannica*, 1974, Vol.8, 589-598.

G. Diffloth "Austro-asiatic languages" *Encyclopaedia Britannica*, 1974, Vol.2, 480-484.

G. Dumézil *L'Idéologie tripartite des Indo-Européens.* Bruxelles, 1959.

S.C. Egerod "Sino-Tibetan languages" *Encyclopaedia Britannica*, 1974, Vol.16, 796-806.

The New Encyclopaedia Britannica, Macropaedia. Chicago, 1974.

A. Ernoult, A. Meillet *Dictionnaire étymologique de la langue latine; histoire des mots.* Paris, 1967.

H.S. Falk, A. Torp *Norwegisch-Dänisches Etymologisches Wörterbuch.* 2 Vol. Heidelberg, 1910.

H. Fleisch *L'Arabe classique: esquisse d'une structure linguistique.* Beyrouth, 1956.

H. Grassmann Ueber die Aspiraten und ihr gleichzeitiges Vorhandensein im An- und Auslaute der Wurzeln *Kuhn's Zeitschrift* 12, 1863, 81-138.

J.H. Greenberg "Genetic Relationship among Languages" *in* Greenberg (1957), 35-45.

J.H. Greenberg *Essays in Linguistics.* Chicago, 1957.

J.H. Greenberg "Some Universals of grammar with particular reference to the order of meaningful elements" *in* Greenberg (1963), 75-113.

J.H. Greenberg éd. *Universals of language.* Cambridge MA., 1963.

J.H. Greenberg *Language in the Americas.* Stanford, 1987.

J.H. Greenberg Classification of American Indian languages A reply to Campbell *Language*, 65 N° 1, 1989, 107-114.

J.H. Greenberg *Indo-European and its Closest Relatives: The Eurasiatic Language Family.* Stanford, à paraître en 1993.

S. Gudschinsky "The ABC's of Lexicostatistics (Glottochronology)" *in* Hymes (1964), 612-623.

R. T. Harms "Uralic languages" *Encyclopaedia Britannica*, 1974, Vol.18, 1022-1032.

J. Haudry *Les Indo-européens*. QSJ n°1965, Paris, 1992.

J. Haudry *L'indo-européen.* QSJ n° 1798, Paris, 1980.

J.A. Hawkins, M. Gell-Mann éd. *The Evolution of Human Languages.* Reading MA., 1992.

E. Hawkins *Awareness of Language: An Introduction.* Cambridge, 1987.

G. Hazai "Altaic languages" *Encyclopaedia Britannica*, 1974, Vol.1, 635-639.

L. Hjelmslev *Le Langage* [trad. fr. de *Sproget.* Copenhague, 1963] Paris, 1966.

H.H. Hock *Principles of Historical Linguistics.* Berlin/New York/Amsterdam, 1986.

C.F. Hockett *A Course in Modern Linguistics.* New York, 1958.

C.F. Hockett "Implications of Bloomfield's Algonquian Studies" *in* Hymes (1964), 599-611.

D. Hymes éd. *Language in Culture and Society. A Reader in Linguistics and Anthropology.* New York, 1964.

R.J. Jeffers, I. Lehise *Principles and Methods for Historical Linguistics.* Cambridge MA., 1982.

C. Jones éd. *Historical Linguistics. Problems and Perspectives.* Londres, 1993.

M. Kaiser, V. Shevoroshkin Nostratic *Annual Review of Anthropology,* 17, 1988, 309-329.

T. Kaufman "Meso-American Languages Indian languages" *Encyclopaedia Britannica*, 1974, Vol.11, 956-963.

F. Kluge *Etymologisches Wörterbuch der deutschen Sprache.* Berlin/Leipzig, 1921.

R.W. Langacker *Language and its Structure. Some fundamental Linguistic Concepts.* New York, 1968.

W. Lehmann, L. Zgusta "Schleicher's Tale after a Century" *in* Brogyanyi (1979), 455-466.

W. P. Lehmann *Historical Linguistics: an Introduction*. Londres, 1992.

W.P. Lehmann *Theoretical Bases of Indo-European Linguistics*. Londres/New York, 1993.

W.B. Lockwood *A Panorama of Indo-European Languages*. Londres, 1972.

W.B. Lockwood *Indo-European Philology. Historical and Comparative*. Londres, 1969.

R. Lord *Teach yourself Comparative Linguistics*. Londres, 1966.

J.P. Mallory *In Search of the Indo-Europeans. Language, Archaelogy and Myth*. Londres, 1989.

B. Malmberg *Histoire de la Linguistique de Sumer à Saussure*. Paris, 1991.

J. Manessy-Guitton "La Parenté généalogique" *in* Martinet (1968), 814-864.

J. Manessy-Guitton "L'indo-européen" *in* Martinet (1968), 1240-1287.

A. Martinet *La prononciation du français contemporain. Témoignages recueillis en 1941 dans un camp d'officiers prisonniers*. Paris/Genève, 1945.

A. Martinet éd. *Le Langage*. Paris, 1968.

A. Martinet *Des Steppes aux Océans. L'indo-européen et les "Indo-Européens"*. Paris, 1986.

J. Matisoff On megalocomparison *Language*, 66, N° 1, 1990, 106-120.

A. Meillet *La Méthode Comparative en Linguistique historique*. Oslo, 1925 [reprint: Paris, 1970].

A. Meillet *Introduction à l'étude comparative des langues indo-européennes*. Paris, 1937 [reprint: Alabama, 1964].

A. Meillet "Aperçu du développement de la Grammaire comparée" *in* Meillet (1937), 453-483.

A. Meillet *Linguistique historique et Linguistique générale I*. Paris, 1938.

A. Meillet "Sur la Méthode de la grammaire comparée" *in* Meillet (1938), 19-43.

A. Meillet "Le problème de la parenté des langues", *in* Meillet (1938), 76-109.

A. Meillet et M. Cohen éd. *Les Langues du Monde*. Paris, 1952.

S. Moscati éd. *An Introduction to the Comparative Grammar of the Semitic Languages*. Wiesbaden, 1964.

A. Pawley "Austronesian languages" *Encyclopaedia Britannica*, 1974 Vol.2, 484-494.

H. Pedersen *The Discovery of Language. Linguistic science in the Nineteenth*

Century. [reprint Bloomington 1959].

T. Penchoen "La Glottochronologie" *in* Martinet (1968), 865-884.

J. Picoche *Dictionnaire étymologique du français*. Paris, 1979.

J. Pokorny *Indogermanisches Wörterbuch*. 2 Vol. Berne/Vienne, 1969.

C. Renfrew *Archaeology and Language*. Londres, 1987.

R.H. Robins *A short history of Linguistics*. Londres, 1987.

M. Ruhlen *A Guide to the World's languages. Vol.I Classification*. Stanford, 1987.

M. Ruhlen "An Overview of genetic Classification" *in* Hawkins, Gell-Mann (1992), 159-189.

M. Sala, I. Vintila-Radulescu *Les Langues du Monde, petite encyclopédie*. Bucarest/Paris, 1984.

E. Sapir *Language. An Introduction to the Study of Speech*. New York, 1921 [trad. fr. Paris 1953].

A.L. Schlözer *Öffentliches und privat-Leben, von ihm selbst geschrieben. Erstes fragment Aufenthalt und Dienste in Russland, vom J. 1761 bis 1765. LiterarNachrichten von Russland in jenen Jaren*. Göttingen, 1802.

J. Schmidt *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar, 1872.

H. Schuchardt *Brevier. Ein Vademekum der allgemeinen Sprachwissenschaft*. Halle, 1928.

J.A. Suarez "South American languages" *Encyclopaedia Britannica*, 1974, Vol.17, 105-112.

O. Szemerényi *Einführung in die vergleichende Sprachwissenschaft*. Darmstadt, 1980.

L. Tesnière *Table étymologique: Les mots russes classés d'après leur racine*. Paris 1970.

P. Thieme "The Comparative Method for Reconstruction in Linguistics" *in* Hymes (1964), 585-598.

P. Thieme: The Indo-European Language *Scientific American*. Oct. 1958, 63-74.

T. Vennemann "Language change as language improvement" *in* C. Jones (1993), 319-344.

K. Verner Eine Ausnahme der ersten Lautverschiebung *Kuhn's Zeitschrift* 23, 1876, 97-130.

H. Walter *Enquête phonologique et variétés régionales du français*. Paris 1982.

C. Watkins "Indo-European and the Indo-Europeans" *in The American Heritage Dictionary of Indo-European Roots.* Boston, 1985, Xi-XXiV.

U. Weinreich *Languages in contact: Findings and Problems.* La Haye, 1963.

S.A. Wurm "Australian Aboriginal languages" *Encyclopaedia Britannica*, 1974, Vol.2, 430-431.

S.A. Wurm "Papuan languages" *Encyclopaedia Britannica*, 1974, Vol.13, 977-978.

TABLE DES MATIERES:

PROLOGUE: LE RUSSE EN HUIT SEMAINES AU DIX-HUITIEME SIECLE?: 1

PREMIERE PARTIE: LA COMPARAISON GENETIQUE: 5

DE LA COMPARAISON EN LINGUISTIQUE A LA LINGUISTIQUE COMPAREE: 5

UN PEU D'HISTOIRE: 6

I. PRINCIPES ET CONDITIONS DE POSSIBILITE DE LA COMPARAISON: 9

ARBITRAIRE ET CONTINGENCE DU SIGNE LINGUISTIQUE: 9

LA COMPARAISON, ENTRE HASARD ET PERTINENCE: 11

LES EMPRUNTS: 13

L'HYPOTHESE GENETIQUE ET LE CHANGEMENT PHONETIQUE: 14

EXCURSUS: LES CAUSES DU CHANGEMENT PHONETIQUE: 15-17

EXCURSUS: LA LINGUISTIQUE DES AIRES: 19-20

II. METHODES ET PROCEDURES DE LA COMPARAISON: 21

DES CORRESPONDANCES A LA RESTITUTION: 21

TABLEAU GENEALOGIQUE DES LANGUES INDO-EUROPEENNES: 23

EXCURSUS: QU'EST-CE QUE L'INDO-EUROPEEN?: 23-24

LA RECONSTRUCTION INTERNE: 25

UN EXEMPLE DE LA METHODE COMPARATIVE: 26

L'ETAPE DE LA RECONSTRUCTION: 29

III. LES DECOUVERTES DE LA COMPARAISON INDO-EUROPEENNE: 32

LA LOI DE GRIMM: 33

L'APPROCHE NEO-GRAMMAIRIENNE: LES "LOIS PHONETIQUES": 34

L'APPROCHE STRUCTURALE: LES LARYNGALES: 37

EXCURSUS: TYPOLOGIE ET UNIVERSAUX: 41-43

IV. LE GROUPE DES LANGUES INDO-EUROPEENNES: LE TRIOMPHE DE LA GRAMMAIRE COMPAREE: 43

DIALECTES ET ISOGLOSSES DE L'INDO-EUROPEEN: 48

EXCURSUS: LA THEORIE DES ONDES: 49-50

LE PROTO-INDO-EUROPEEN: 50

LES "INDO-EUROPEENS": 53

EXCURSUS: UNE FABLE EN INDO-EUROPEEN: 54-55

EXCURSUS: LA PALEONTOLOGIE LINGUISTIQUE: 56-57

V. LES AUTRES FAMILLES DE LANGUES: DIFFICULTES ET LIMITES DE LA COMPARAISON: 59

LE CHAMITO-SEMITIQUE OU AFRO-ASIATIQUE: 60

EXCURSUS: LA TYPOLOGIE MORPHOLOGIQUE: 61-63

EXCURSUS: LA RACINE EN SEMITIQUE: 63-64

LE SINO-TIBETAIN: 64

L'OURALO-ALTAIQUE?: 65

DE L'ASIE A L'AMERIQUE: 67

EXCURSUS: LA GLOTTOCHRONOLOGIE: 67-69

EXCURSUS: LA COMPARAISON DE MASSE SELON JOSEPH H. GREENBERG: 70-72

CATALOGUE DES LANGUES DU MONDE D'APRES J. GREENBERG: 72-73

INDO-EUROPEEN, NOSTRATIQUE ET EURASIATIQUE: 73

EXCURSUS: LE NOSTRATIQUE: 74

EXCURSUS: UNE ETYMOLOGIE GLOBALE?: 76

DEUXIEME PARTIE: DE LA COMPARAISON A L'INTERCOMPREHENSION 77

I. OUVERTURE ET SENSIBILISATION A LA PERSPECTIVE HISTORIQUE: 77

II. ECHANGES ET HERITAGE DANS UN COUPLE DE LANGUES EUROPEENNES: 79

III. LA REDUCTION D'UNE LANGUE AU CONNU: 88

BIBLIOGRAPHIE GENERALE 101-106

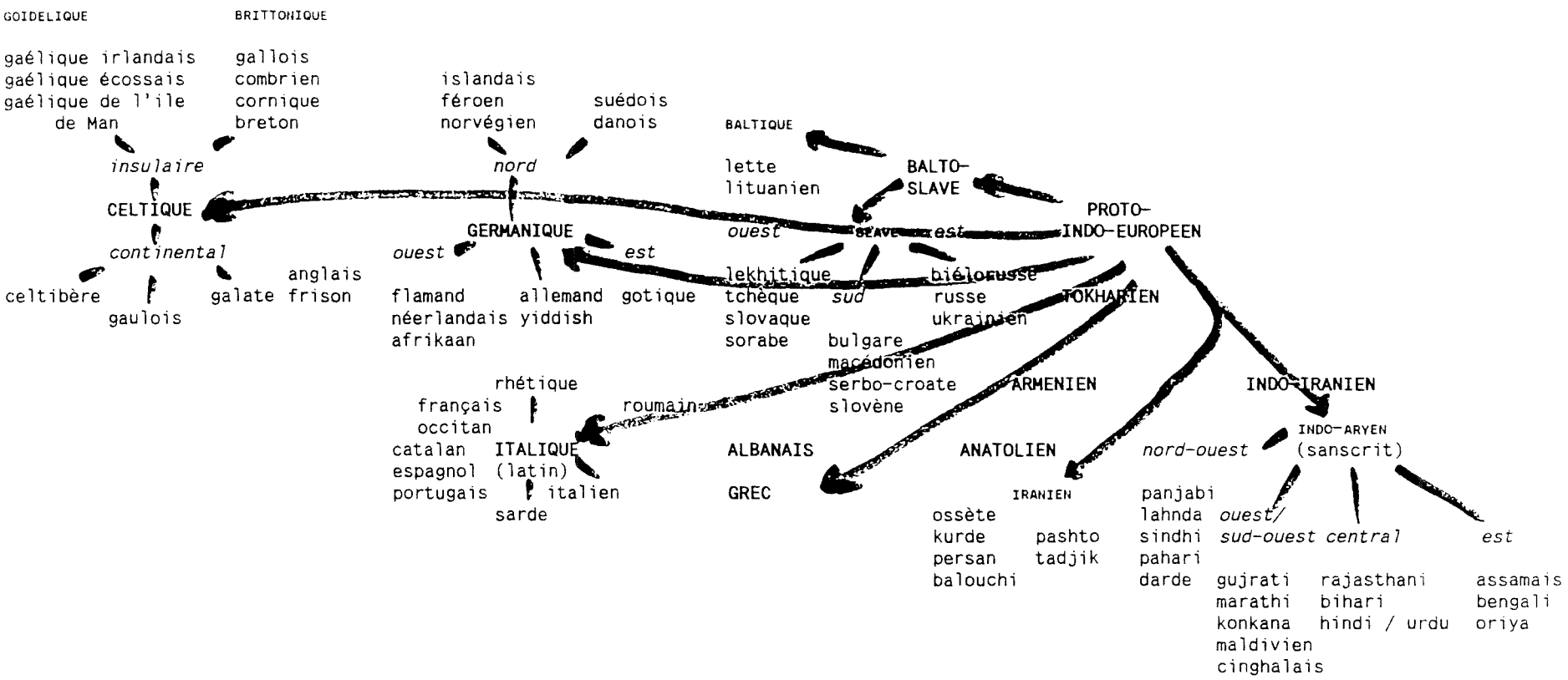
BIBLIOGRAPHIES SPECIALISEES:

HISTORIQUE DE LA LINGUISTIQUE COMPAREE 9
LE CHANGEMENT LINGUISTIQUE 17
LA LINGUISTIQUE DES AIRES 20
LES PRINCIPES DE LA COMPARAISON 21
L'INDO-EUROPEEN 24
LA METHODE COMPARATIVE 32
LES DEVELOPPEMENTS DE LA COMPARAISON 41
TYPOLOGIE ET UNIVERSAUX 43
LES LANGUES INDO-EUROPEENNES 48
LA THEORIE DES ONDES 50
LE PROTO-INDO-EUROPEEN 53
UNE FABLE INDO-EUROPEENNE 55
LA PALEONTOLOGIE LINGUISTIQUE 57
LES "INDO-EUROPEENS" 59
LES LANGUES DU MONDE 59
LA TYPOLOGIE MORPHOLOGIQUE 63
LA RACINE EN SEMITIQUE 64

LE CHAMITO-SEMITIQUE 64
LE SINO-TIBETAIN 65
L'OURALO-ALTAIQUE 66
LA GLOTTOCHRONOLOGIE 69
LES LANGUES D'ASIE 70
LES LANGUES D'AMERIQUE 70
LA COMPARAISON DE MASSE 72
LES MACRO-FAMILLES DE LANGUES 73
LE NOSTRATIQUE 74
L'ETYMOLOGIE GLOBALE 76

ANNEXES:
LES LANGUES INDO-EUROPEENNES, SCHEMA SIMPLIFIE DE DISTRIBUTION GEOGRAPHIQUE
QUELQUES ETYMOLOGIES EURASIATIQUES SELON GREENBERG
LA RACINE INDO-EUROPEENNE *AK-
QUELQUES CORRESPONDANCES INDO-EUROPEENNES
TROIS ARTICLES D'UN DICTIONNAIRE DE CONCEPTS MULTILINGUE:
FIN
PLEURER
ESPOIR
TABLEAU SIMPLIFIE DE CORRESPONDANCES POUR LES LANGUES EUROPEENNES

ANNEXE : LES LANGUES INDO-EUROPENNES, SCHEMA SIMPLIFIE DE DISTRIBUTION GEOGRAPHIQUE.



ANNEXE : QUELQUES ETYMOLOGIES EURASIATIQUES SELON GREENBERG

	Ind-eur.	Oural.	Turc	Mongol.	Tongouse	Coréen	Japonais	Ainou	Gilyak	Tch.-K.	Esqu.-Aléoute
je[1]	*mê-[1]	*-m	men[2]	mini[3]	mini[4]	-ma[5]	mi[6]		me-[7]	-m	-ma
je[2]	-x[8]	*-k								-k	-k
qui?	*kʷi≈kʷo	ken[9]	*kim	ken	*xa-[10]	-ka[11]	ka[12]	-ka[13]	-ka[14]	k'e[15]	*kina
quoi?	*jo≈je[16]	*jo-[17]	je-[18]	javun	ja	ja[19]	i[20]≈ja[21]	ja[22]	ja-	jaq[23]	
duel	-k'[24]	*-k	iki[25]	ikire[26]				-ki	-gi	-k[27]	-k
pluriel		*-t	-t[28]	*-t	-te			*-ti	-t	-ti	-t
écorce		*kopa	*kâp[29]			këpcil[30]	kapa	sik-kap[31]	xip		
poisson	*(s)kʷalo-s	*kala			*xol-sa				q'ol	klxin[32]	iqaluk
loup	*wlkʷo-s	loqa[33]		noqa[34]	luka[35]			horokew	livs		
parler	*kel[36]	*kele[37]	*käla	kele(n)[38]					qlai[39]	kel[40]	kiliva[41]
arriver	*tek-[42]		teg-[43]			tah	tuku				tikippuq[44]
aiguiser	*(s)pik-[45]		*päkä		*päkä		piku[46]				*piv-

1. "moi".
2. En ouzbek.
3. "mon".
4. "mon".
5. Pour exprimer un souhait emphatique de la part du locuteur.
6. En japonais ancien.
7. "nous".
8. En hittite.
9. En finnois archaïque.
10. Base interogative.
11. Interrogatif enclitique.
12. Marque d'indéfinition.
13. Marque d'indéfinition.
14. Interrogatif enclitique.
15. En Kamtchadal.
16. Pronom relatif.
17. Pronom relatif.
18. Interrogatif enclitique.
19. Interrogatif de phrase.
20. Base interrogative.
21. Interrogatif de phrase.
22. Interrogatif de phrase.
23. En Kerek.
24. Pluriel arménien pour les noms, les pronoms et les verbes.
25. "deux.
26. "jumeaux".
27. Pluriel.
28. En turc ancien.
29. "couvrir".
30. En coréen moyen.
31. "paupière".
32. En Kamtchadal.
33. "renard".
34. "chien".
35. "lynx".
36. "appeler".
37. "langue, parole".
38. "langue, parole".
39. "converser".
40. "crier".
41. En Kuskokwim.
42. "atteindre".
43. En turc ancien.
44. En Groenlandais.
45. "point, ongle".
46. "moudre".

ANNEXE : LA RACINE INDO-EUROPEENNE *AK

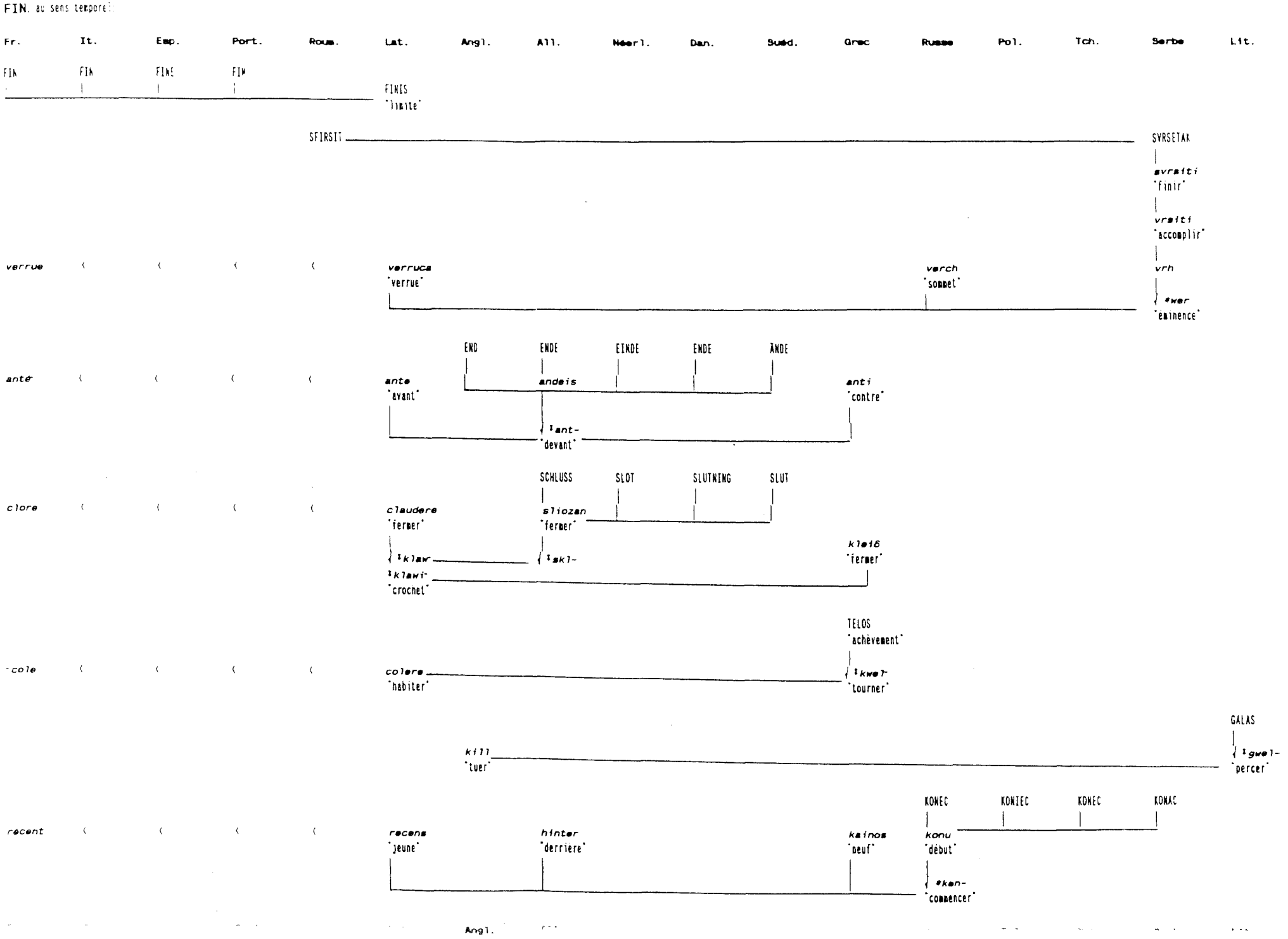
	ak	+i/j	+u	ôku-s*+mo/â	+men/mer	+n	+a/oker	akri/ro	okri/ro	akes/s	aksti	+t	oketâ	ok-tôu'8'kemen	kamer	ke/oi	ku
G	*akê* pointe *akôkê* pointe *akouô* pointer l'oreille	*akis*, *dos* pointe	*akhuron* balle	*ôkus* *akmê* pointe	*akmôn* enclume	*akaina* piquant *akonê* pierre à aig. *akôn*, *ntos* javelot *akontizô* *akanos* ± chardon *akanthos* chardon *akantha* épine *akathos* canot *akaté* chaussure	*akorna* chardon *akoros*	*akros* pointu *akron* pointe *akra* pointe *akris* pointe *akris*, *dos* *akremôn*	*okris* pointe	*akhnê* balle *akosté* orge *oxus* aigu *oxos* vinaigre *akakhmenos* appointé		*akté* côte	*oxiné* herse	*oktô*		*kônos* toupie	
L	*acere* être acide *acidus* acide *acetum* vinaigre	*acies* tranchant	*acus* aiguille *acuere* aiguiser *acumen* pointe *acia* fil *aculeus* piquant *accipiter* faucon	*ocior* (c.)		*agna* épi	? *acer* érable	*âcer* âcre *acerbus* âpre-hauteur aigre	*medi-ocris* à mi	*acus*, *eris* balle *acervus* tas			? *occa* herse	*octô*		*catus* rusé *cos*, *cotis* p. à aiguiser	*cuneus* coin *culex* cousin
G/g						*ahana* balle								*ahtau*	*himins* ciel		
a		*ails* barbe													*heaven* ciel	*hammer* marteau	*hone* p. à aiguiser
al		*ecke* coin arête	*achel* barbe			*ahne*	*acher* érable			*âhre* épi							
d							*ar* érable										
s	*ag* herbe (marais) *agg* rancune																
S/r		*osëlok* p. à aiguiser									*oste* pointe barbe						
t		*osina* barbe									*osten* piquant						
B/l	*akuotas* barbe *as'aka* son		*a's'utai* crin		*a's'mens* lame *akmuo'* pierre	*as'nis* pointe	*asêry's* perche	*as'rus* tranchant			*akstis* *akstinas* piquant		*akec'ios* herse	*as'tuo-ni*			
Ar		*aseln* aiguille												*ut'*		*sur* aigu *sair* lame *sadrem* il aiguise	*slak* pique
Al	*athêtê* aigre													*tetê*			
I/s				*âs'u*	*as'man* ciel - pierre *as'mara* pierreux			*as'rih.* coin arête				*aba-as'-tha* p. de flèche		*as.t.â*		*s'ila* pierre *s'i-s'a-ti* il aiguise	*s'ula* pique
p	*âs* meule					*as'anih.* pointe - de flèche										*san* p. à aiguiser	*sôzan* aiguille *sôx* épi

Abréviations: G = grec; L = latin; G/g = Germanique / gotique; a = anglais; al = allemand; d = danois; s = suédois; S/r = Slave / russe; t = tchèque; B/l = Balte / lituanien; Ar = arménien; Al = albanais; I/s = Indo-iranien / sanscrit; p = persan

ANNEXE : QUELQUES CORRESPONDANCES INDO-EUROPEENNES

	Sanscrit	Grec	Latin	v. Irl.	Gotique	v. Slave	Lituanien	Arménien	Albanais	Hittite	Tokh. A	Tokharien B
un	éka-	oinos[1]	ûnus	oin, ôen	ains	inù	vienas		një			
deux	dvâ', dva'u	du'(w)o	duo	dâu, dô	twai	dùva	dù	erku	dü	dâ-	wu	we
trois	tra'ya-	treis	três	tri	*Preis	trije	trys	erek'	tre, tri	tri-	tre	trai
quatre	catvâ'ra-	téttares	quattuor	ceth(a)ir	fidwôr	c'etyre	keturi	c''ork'	katër		s'twar	s'(t)war
cinq	pañca-	pénte, pémpe	quinque	côic	fimf	peti	penki	hing	pesë	panta[2]	pëñ	pis'
six	s.a't.-	héks	sex	sê	saihs	s'esti	s'es'i	vec'	gjashtë		säk	s.kas
sept	sapta'-	hepta'	septem	secht	sibun	sedmi	septyni	evt'n	shtatë	s'ipta-	s.pät	s.uk(t)
huit	as.t.a'(u)	oktô'	octo	ocht	ahtau	osmi	as'tuoni	ut'	tetë		okät	okt
neuf	na'va-	enné(w)a	novem	noi	niun	devet	devyni	inn	nëntë		ñu	ñu
souris	mû's.-	mûs	mûs		mûs	mys'i		mukn	mi			
loup	vr.'ka-	lu'kos	lupus	olc[3]	wulfs	vluku	vilkas		ulk			
vache	go'-	boûs	bôs	bô	ky'r[4]	gumuno[5]	gùovs[6]	kov			ko	kau
mouton	a'vi-	o'(w)is	ovis	ôi	ouwi[7]	ovica	avis	hoviw[8]				
porc	sûkara'-	hûs	sûs		swein	svinija	suvens[9]		thi			suwo
chien	s'va'n-	ku'ôn	canis	cû	hunds		s'uô, s'uñs	s'un			ku	ku
cheval	a's'va-	hippos	equus	ech	eoh[10]		es'và[11]				yuk	yakwe
pied	pa'd-	pou's,podos	pes, pedis	is[12]	fôtus	pes'i[13]	pâdas[14]	otn	(për)posh[15]	pat-	pe	pai
coeur		kardi'â	cordis	cride	hairtô	sridice	s'irdis	sirt		kard-		
oeil	a'ks.i-	o'psomai[16]	oculus	enech	augô	oko	akis	akn	sü		ak	ek
langue	jihvâ'		lingua	teng	tuggô	jezyku	liezùvis	lezu			käntu	kantwo
mère	mâta'r-	mâ'têr	mâter	mâthir	môôir[17]	mati	mote[18]	mayr	motrë		mâcar	mâcer
père	pita'r-	patêr	pater	athir	fadar			hayr			pâcar	pâcer
soeur	sva'sar-	éor[19]	soror	siur	swistar	sestra	sesuo	k'oyr				
frère	bhrâ'tar-	phrâ'têr[20]	fràter	bràth(a)ir	brôPar	bratru	brolis	elbayr			pracar	procer
plein	pûrn.a-	plêrês	plênus	làn	fulls	plunu	pilnas	li	plot			
race	jana-	génos	genus	gein[21]	kuni			cin[22]				
mois	mâ's-	mên	mênsis	mi	mêna, mênô	Pmeseci	menuo	amis	muaj		mañ	meñe
mort/mourir	mr.ta'-	ambrotos[23]	mortuus	marb	maurPr	miro, mreti	mirti			merta		
vieux	sana-	hénos[24]	senex	sen	sineigs[25]		senas	hin				
vomi	vamiti	eméô	vomo		vâma[26]			vémti				

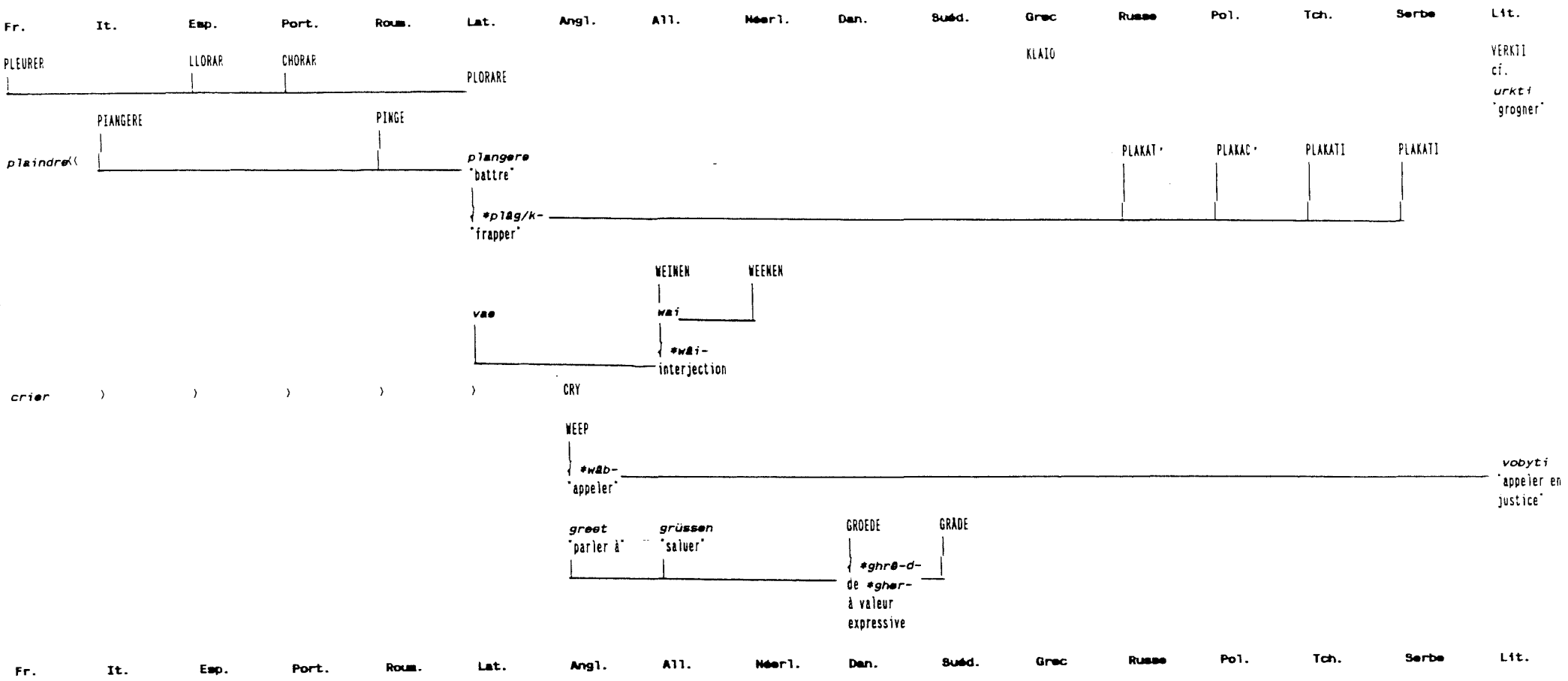
1. "la une" aux dés.
2. En hittite hiéroglyphique.
3. "mal".
4. En vieil islandais.
5. "aire d'une grange".
6. En lette.
7. En haut-allemand.
8. "berger".
9. En lette, "porcelet".
10. En vieil anglais.
11. En lituanien ancien, "jument".
12. "en dessous".
13. "à pied".
14. "semelle".
15. "sous".
16. Forme verbale, "je verrai".
17. En vieil islandais.
18. "femme".
19. Au vocatif.
20. "membre d'une fratrie".
21. "naissance".
22. "naissance".
23. Avec préfixe négatif, au sens de "immortel".
24. "de l'année précédente".
25. En vieil islandais.
26. "maladie".



FIN, au sens temporel:

Fr.	It.	Esp.	Port.	Roum.	Lat.	Angl.	All.	Néerl.	Dan.	Suéd.	Grec	Russe	Pol.	Tch.	Serbe	Lit.
FIN	FIN	FINE	FIM		FINIS 'limite'											
				SFIRSIT											SVRSETAK	
															svrsiti 'finir'	
															vrsiti 'accomplir'	
verrue	(	(	(	(	*verruca* 'verrue'							*verch* 'sommet'			*vrh*	
															**wer* 'éminence'	
						END	ENDE	EINDE	ENDE	ÄNDE						
anté	(	(	(	(	*ante* 'avant'		*andeis*				*anti* 'contre'					
							**ant-* 'devant'									
							SCHLUSS	SLOT	SLUTNING	SLUT						
clore	(	(	(	(	*claudere* 'fermer'		*sliozan* 'fermer'									
					**klaw-*		**skl-*				*kleiô* 'fermer'					
					**klawi-* 'crochet'											
											TELOS 'achèvement'					
-cole	(	(	(	(	*colere* 'habiter'						**kwel-* 'tourner'					
																GALAS
						kill 'tuer'										**gwel-* 'percer'
												KONEC	KONIEC	KONEC	KONAC	
récent	(	(	(	(	*recens* 'jeune'		*hinter* 'derrière'				*kainos* 'neuf'	*konu* 'début'				
												**ken-* 'commencer'				
						Angl.										

PLEURER


Fr.
It.
Esp.
Port.
Roum.
Lat.
Angl.
All.
Néerl.
Dan.
Suéd.
Grec
Russe
Pol.
Tch.
Serbe
Lit.
PLEURER
LLORAR
CHORAR
PLORARE
KLAIO
VERKTI
cf.
urkti
"grogner"
PIANGERE
PINGE
plaindre
plangere
"battre"
*plāg/k-
"frapper"
PLAKAT'
PLAKAC'
PLAKATI
PLAKATI
WEINEN
WEENEN
vae
wai
*wāi-
interjection
crier
CRY
WEEP
*wāb-
"appeler"
vobyti
"appeler en justice"
greet
"parler à"
grüssen
"saluer"
GROEDE
GRADE
*ghrē-d-
de *gher-
à valeur
expressive

ESPOIR

Fr.	It.	Esp.	Port.	Roum.	Lat.	Angl.	All.	Néerl.	Dan.	Suéd.	Grec	Russe	Pol.	Tch.	Serbe	Lit.
ESPOIR ESPERANCE	SPERANZA	ESPERANZA	ESPERANÇA													
espérer	*sperare*	*esperar*	*esperar*		*sperare* "espérer"											
					spes, "espoir"											
					{ *sp(h)ê- "prospérer, grossir, s'étendre"		*spät* / *Speck* "tard" / "lard"					*speti* "prospérer"				
				NADEJDE								NADEZ~DA	NADZIEJA	NADE~JE	NADA	
												na-de~jati "espérer, = se placer sur"				
					facere "faire"						*thêkê* "réceptacle"		*deti* (v. sl.) "placer"			
					*dhê-k						*dhê-k		{ *dhê- "placer, poser"			
					volup-tas "plaisir"						ELPIDA					
					*wol-p-						*elpis*					
					volo "vouloir"		*Wille* / *wählen* "volonté/choisir"				*wel-p- { *wel- "vouloir, choisir"					VILTIS *pa-velti* "vouloir, permettre"
					cupio ? "vouloir"	HOPE étymologie obscure	HOFFNUNG	HOOP	HAAB	HOPP						
Fr.	It.	Esp.	Port.	Roum.	Lat.	Angl.	All.	Néerl.	Dan.	Suéd.	Grec	Russe	Pol.	Tch.	Serbe	Lit.

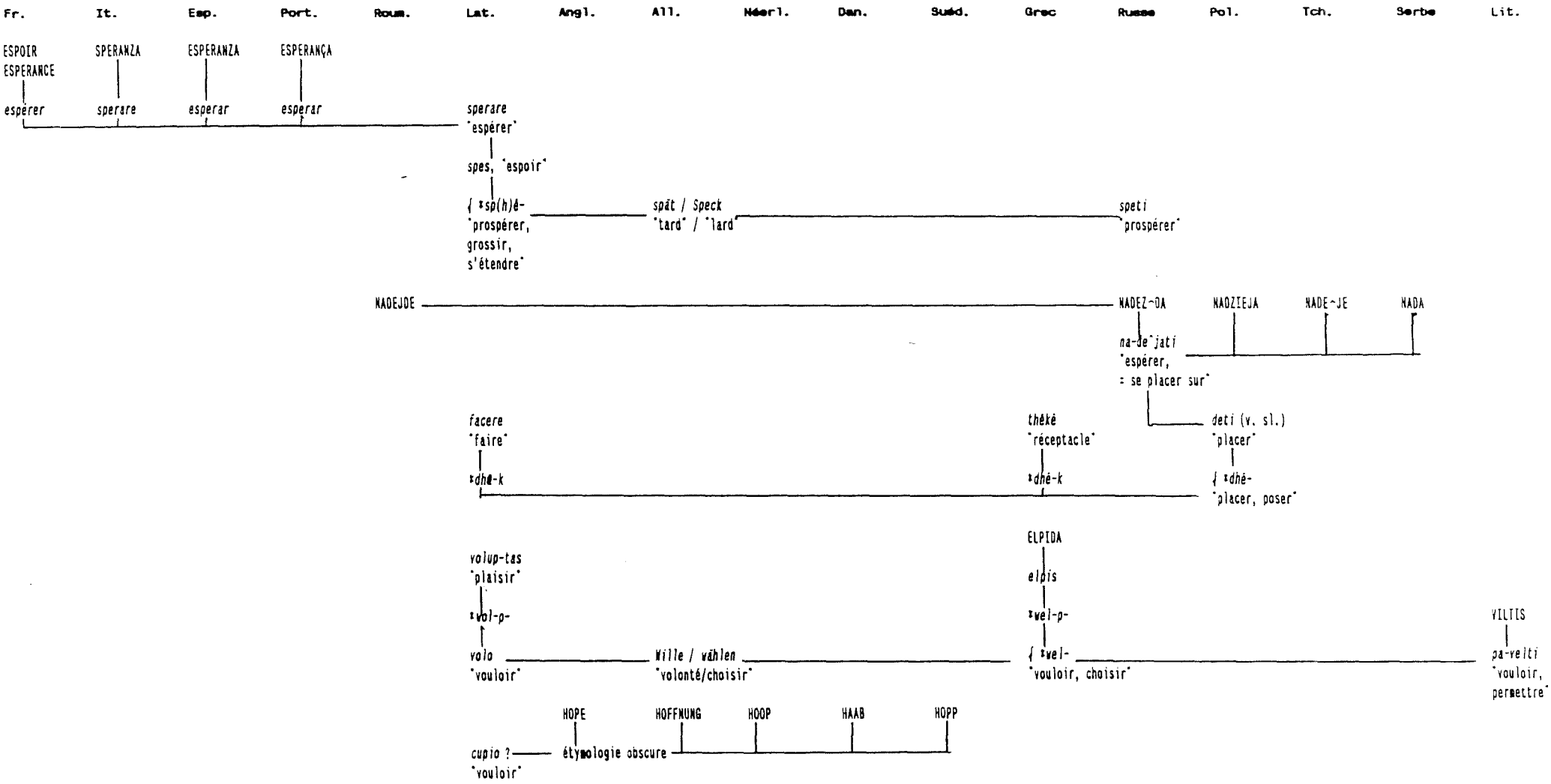


Tableau de correspondance:

F.	It.	E.	Por.	R.	A.	All.	N.	D.	Su.	G.	R.	Pol.	T.	S.	L.	Irl.
marteau	martello	martillo	martelo	ciocan	hanner	hanner	hamer	hammer	hammar	sphuri	molot	młotek	kladiso	c'ekik	plaktukas	orcc
du Lat. martellus	:	:		du Turc çakan	"outil de pierre"					cf. sphuron				par le Turc	cf lat	?
cf malleus ⌊ *mel	:	:			⌊ *ak	:	:	:	:		⌊ *mel	:	#	du Pers.	plangere	

Fr.	It.	Esp.	Port.	Roum.	Angl.	All.	Néerl.	Dan.	Suéd.	Grec	Russe	Pol.	Tch.	Serbe	Lit.	Irl.
oeil	occhio	ojo	olho	ochiu	eye	auge	oog	øje	ögа	mati	glas	oko	oko	oko	akis	súil
oculus					ae. eage	ouga				de omnation	sens 1 "pierre"					Lat sol
⌊ *ok"	:	:	:	:	:	:	:	:	:	de omma	#	:	:	:	:	

Fr.	It.	Esp.	Port.	Roum.	Angl.	All.	Néerl.	Dan.	Suéd.	Grec	Russe	Pol.	Tch.	Serbe	Lit.	Irl.
suivre	seguire	seguir	seguir	urma	follow	folgen	volgen	følge	följa	akolouthò	sledit'	is'c'	na sledovati	sljedeti	sekti	sechur
Lsequi › sequere					ae folgian											
⌊ *sek"																

Fr.	It.	Esp.	Port.	Roum.	Angl.	All.	Néerl.	Dan.	Suéd.	Grec	Russe	Pol.	Tch.	Serbe	Lit.	Irl.
nouveau	nuovo	nuevo	novo	nou	new	neu	nieuwe	ny	ny	kainourgès	novyj	nowy	novy	nov	naujas	nüe
Lnovellus ‹ novus		:		:	nùve	niuwi				kainos	novu					
⌊ *newos	:				:	:				cf L. recens						

Fr.	It.	Esp.	Port.	Roum.	Angl.	All.	Néerl.	Dan.	Suéd.	Grec	Russe	Pol.	Tch.	Serbe	Lit.	Irl.
hiver	inverno	invierno	inverno	iarnà	winter	winter	winter	vinter	vinter	kheimónas	zima	zima	zima	zima	ziema	gan
L hiems › hibernum (tempus)						G. wintrus				kheimón						gem
⌊ *ghejm ‹ *ghei						cf wato "eau"					:				:	:

Fr.	It.	Esp.	Port.	Roum.	Angl.	All.	Néerl.	Dan.	Suéd.	Grec	Russe	Pol.	Tch.	Serbe	Lit.	Irl.
homme	uomo	hombre	homem	om	man	mensch	mensch	menneske	människa	anthrôpos	c^elovek	człowiek	c^love^k	c^ovjek	z^mogus	duine
homo cf humus "terre"						G mannisks				⌊ *okw	*c^elo+ve^ku^ membre de maisonnée				cf germ +gutturale	
⌊ *						cf ⌊ *men ou l manus										

Fr.	It.	Esp.	Port.	Roum.	Angl.	All.	Néerl.	Dan.	Suéd.	Grec	Russe	Pol.	Tch.	Serbe	Lit.	Irl.
sommeil	sonno	sueñ^o	sono	somn	sleep	schlaf	zlaap	søvn	svefn	hupnos	span'e	spanie	spa'nek	spavanje	miegas	suan
somniculus ‹ somnus					mais oe swefn	⌊ *être nou :										
⌊ *swep swep-no-s swop-no-s sup-no-s																