

APERÇU DE LA STRUCTURE ÉTYMOLOGIQUE DU ROUMAIN

Il est bien connu que le roumain a une position particulière dans le monde néo-latin. C'est évidemment une langue néo-latine contenant beaucoup de particularités qui ne permettent pas de réduire les problèmes à l'héritage latin seulement. La plupart, pour ne pas dire tous les ouvrages dédiés à l'étude du roumain, notent ses relations avec les autres langues romanes (par exemple Contenu et Sala 1987; Sala et al. 1988); on peut ajouter, bien sûr, les influences des substrats, en premier lieu slave mais aussi des autres langues. C'est le cas du dernier livre de N. D. Raevskij (1988). Pour faire justice à l'auteur, il faut préciser brièvement que c'est un livre utile à l'étude de la romanité orientale. L'auteur utilise souvent une méthode originale pour délimiter les périodes historiques de la cristallisation d'un spécifique roman "danubien".¹ Il introduit (à ma connaissance) la formule "perioadă etnică" (période ethnique). Le but de cette discussion n'est pas d'analyser cet ouvrage mais de signaler quelques aspects d'ordre plus général liés à un problème qui me semble fondamental: le rôle du substrat thrace (ou thraco-dace, selon une autre formule²) dans la formation de la langue roumaine.

Il faut dire que N. D. Raevskij, le co-auteur d'un dictionnaire étymologique de la langue "moldave" (1979) réduit - comme d'ailleurs la plupart des auteurs - le rôle du substrat thrace dans la formation de la romanité orientale à quelques dizaines de mots. C'est vrai que cette méthode (adoptée par une autorité comme Al. Rosetti, 1986) a ses raisons bien connues. Les linguistes aiment utiliser des données écrites qui leur permettent de tracer une évolution plus sûre d'un phénomène phonétique, en réfutant - en général - les reconstructions, étant donné le risque d'erreur. Ce principe général doit avoir des exceptions, car il est absurde d'affirmer que le

¹ Beaucoup d'auteurs utilisent le terme "latin danubien" (par exemple Fischer 1985 qui l'analyse): Raevskij ne fait pas exception. Je le maintiens si bien que je le désavoue. Il serait plus adéquat de choisir un terme comme "latin carpathique" ou "carpatho-balcanique".

² "Thrace" et "thraco-dace" ne sont pas des formules opposées, c-à-d. deux langues distinctes. Les auteurs antiques utilisent l'ethnikon "dace" (*Daci, Dacisci, Dakoi* etc.) et "gète" (*Getes, Getai* etc.) pour désigner les Thraces nord-danubiens. Le Danube n'a jamais été une barrière linguistique dans le monde thrace. Les hypothèses (que je ne peux pas analyser ici) qui voient deux ou bien trois (!) langues de type thrace parlées dans l'antiquité sont parfaitement erronées. Mais c'est un sujet pour amples discussions.

roumain ne peut avoir que (disons) 180 mots d'origine thrace parce qu'on ne possède pas de textes anciens thraces ou proto-roumains. Il est évident que la structure étymologique d'une langue n'a presque rien à faire avec ses attestations écrites. Je dis "presque" parce que l'apparition de l'écriture est liée aux contacts culturels qui peuvent influencer le vocabulaire. Donc, le roumain doit avoir un certain nombre d'éléments indigènes non parce qu'il y a ou il n'y a pas de textes anciens mais purement et simplement parce que la civilisation thrace si originelle, celle qui avait toujours fasciné les grecs, fut trop puissante et trop bien représentée par les autochtones conquis. Les données archéologiques (il y a des centaines d'études qu'on ne peut pas citer ici) prouvent que la romanisation n'a pas été un phénomène de prépondérance numérique de la part des colonistes romains, bien au contraire: les indigènes eurent la supériorité. La romanisation a été un phénomène de prestige culturel comme résultat d'un prestige militaire. Enfin, un autre détail: il y a des différences nettes entre le substrat du roumain (thrace, c'est-à-dire un idiome satem) et les autres langues romanes où le substrat a été celto-italique ou ibéro-celte (idiome centum, aussi non-indo-européen, voire le basque).

Beaucoup de linguistes d'autorité (parmi eux Al. Rosetti) invoquent le fait qu'on ne peut pas analyser un mot roumain d'origine obscure en se rapportant directement à une racine indo-européenne (j'ajoute, *horrible dictu*, pré-indo-européenne aussi). Autrement dit, la langue roumaine sera perpétuellement non-analysable aussi complètement que possible parce qu'on ne peut pas invoquer une racine primitive reconstruite en utilisant les lois de la reconstruction linguistique. Il est difficile de comprendre pourquoi le lithuanien et le letton, langues très archaïques et avec des textes écrits plus tardifs que ceux en roumain, sont entrées dans tous les manuels de grammaire indo-européenne comparée tandis que les mots roumains n'en ont pas ce droit.

Les similitudes du roumain et des langues baltiques ne sont pas seulement "typologiques" mais elles sont plus profondes. Roumain *iar* "et" (en quelques constructions typiques) est évidemment apparenté au lithuanien *ir* "et". Aussi roum. *daina* (forme dialectale en Transylvanie), *doina* (forme littéraire courante) - lith. *dainà*, let. *daina* "chant populaire". Et les exemples peuvent continuer.

Les relations roumaines-hongroises sont, à mon avis, plus complexes que certains linguistes ne le veulent (non seulement hongrois). Il est courant de considérer que roum. *hotr* "limite, frontière" reflète le mot hongrois *hatr* (lire *hatar*); mais il y a un mot presque identique en albanais: *hatër*, *htër*. L'exemple le plus sensationnel est représenté par le rapport entre roum. *oraş* (dial. aussi *uraş*) et hong. *város*. Il est habituel de considérer le mot roumain comme une influence hongroise. Mais les Thraces avaient un terme presque identique pour désigner la

ville: *ora, oros, oron*. Un simple hasard? Une analyse attentive nous permet de conclure qu'on ne peut pas invoquer ici le hasard (Paliga 1987 a; 1989 a).³ Le mot roumain fait partie d'un héritage très ancien: le substrat pré-indo-européen d'où proviennent aussi le latin *urbs*, basque *uri*, aussi *iri* (cf. le canton helvétique *Uri*) et, plus ou moins épouvantable, le sumérien *Ur, Uruk*. Toutes ces formes reflètent un vieux terme pour désigner la structure urbaine plus ou moins primitive. J'ai signalé tout à l'heure l'importance capitale du fond pré-indo-européen pour expliquer des aspects importants de l'héritage culturel du sud-est européen (Paliga 1986; 1987 a; 1988 a; 1989 a, c). Je n'y insisterai plus.

Bien sûr, le substrat pré-indo-européen est seulement un aspect mais il est bien fondamental. Il y a beaucoup de problèmes presque aussi intéressants que importants. On dit, par exemple, que le roumain a 4000 mots environ d'origine inconnue (Vraciu 1984). Dans un livre récent (Coteanu et Sala 1987: 112) le lecteur est averti qu'"on ne peut pas prétendre, comme on faisait à l'époque romantique, que tous les mots roumains à l'étymologie inconnue sont d'origine thrace". Une telle précaution me semble inutile. A ma connaissance, personne n'a affirmé une telle absurdité (et, moi, je ne comprends pas ce que les auteurs veulent dire par "époque romantique" non précisée dans le texte). Mais ces auteurs-ci, comme la majorité d'ailleurs, ne proposent aucune méthode plus ou moins radicale pour réduire le nombre impressionnant de "black holes" du vocabulaire roumain. En effet, le nombre de 4000 mots dont l'origine reste inconnue semble trop grand. Malheureusement, le nombre exact de tels mots ne peut pas être établi purement et simplement parce que le dictionnaire explicatif (DEX) qui a probablement été la source d'inspiration de cette estimation, contient beaucoup d'incertitudes pour ne pas dire de graves erreurs. En premier lieu, ce dictionnaire ne fait jamais distinction entre: (1) mots d'origine thrace (certaine ou probable), (2) mots d'origine discutable (pour lesquels on a suggéré quelques explications), (3) mots non expliqués plausiblement. Toutes ces catégories sont analysées en bloc comme "d'origine inconnue". Il n'est pas étonnant donc qu'on enregistre 4000 mots d'origine inconnue. Avec de telles manières commodes de résoudre les problèmes complexes de l'étymologie roumaine, on peut citer 5000 ou bien 10000 mots obscurs si nous ajoutons les formes dialectales. Deux exemples y seront utiles.

³ Il est utile à observer que l'hypothèse d'un vieux mot hongrois pour désigner la ville ou la forteresse n'est pas plausible étant donné que les populations ougriennes et fenniques n'avaient pas des établissements urbains comme le prouve la situation en finlandais où les mots pour "ville" (*kaupunki*) et pour "marché" (*kauppala*) sont germaniques. Pour accepter l'idée que les hongrois auraient un très vieux mot pour désigner la ville ou la cité il faudrait invoquer des arguments solides qui, à ma connaissance, n'existent pas, mais bien au contraire.

Naiba, mot populaire pour "diable", est considéré d'origine inconnue. Mais il est évident que son étymologie est très simple: *n-aiba* (*parte*) "qu'il n'ait pas (part)". C'est donc une création euphémique parce que le nom du diable était un tabou (cf. Paliga 1989 b où l'on analyse la situation similaire de *zîna*, "fée").

Un autre exemple. *Vatrai* "tisonnier, attisoir" est considéré un emprunt au bulg. *vatral*, s.-cr. *vàtrali* "id. ". Il est évident que *vatrai* doit être apparenté à *vatra* "foyer, âtre" qui se trouve aussi en bulgare, serbo-croate ainsi que dans d'autres langues balkaniques ou est-européennes (alb. *vatrë* etc.) où a le sens de "feu". Le mot est certainement d'origine thrace (si bien que l'étymologie n'est pas claire). Les auteurs du dictionnaire ont pensé probablement que le roumain a emprunté comme telle une forme dérivée du bulgare et/ou serbo-croate de *vatra* qui est, à son tour, un élément proto-roumain d'origine thrace dans ces langues-ci. Mais peut-on accepter un tel point de vue? Ne serait-il pas plus normal de juger que *vatrai* soit dérivé purement et simplement de *vatra*? Il est vrai, une telle explication doit résoudre un autre point difficile, c'est-à-dire la dérivation de type *vatra* + *-i* qui n'est pas fréquente mais bien présente en quelques mots obscurs, par exemple *malai* "farine de mas", sens primitif "farine de millet" de **mala*, sens inconnu. Il est inutile de dire que ce suffixe *-i* doit être indigène (c'est-à-dire thrace) et ne peut avoir rien à faire avec l'*i* hongrois, sens locatif (*budapesti* "de Budapest").

Il est significatif que le roumain n'a pas encore de dictionnaire étymologique-historique complet de haut niveau scientifique: celui de Cihac est inutilisable aujourd'hui et ne peut constituer éventuellement qu'un exemple de comment on ne doit pas faire un dictionnaire étymologique; celui de Gabinskij et Raevskij (1978) est un dictionnaire scolaire pour ne pas dire pire; celui de Candrea et Densusianu est admirable comme méthode scientifique mais malheureusement incomplet (1914, *a-putea*, seulement pour les éléments latins).

Un beau jour, le roumain aura son dictionnaire étymologique. Mais pour achever une telle tâche et pour ne pas répéter ou bien amplifier les erreurs habituelles, je suggère une réorganisation (aussi radicale que possible) des méthodes utilisées et du matériel linguistique. Je le résume en quelques points.

1^e Une analyse profonde et complète des mots présents dans toutes les langues dites "balkaniques" et la précision s' il s'agit vraiment de mots slaves ou des emprunts au substrat thrace, illyrique ou à d'autres langues. La simple présence d'un mot en roumain et dans une ou plusieurs autres langues slaves ne peut signifier automatiquement que ce mot roumain doit être slave mais il peut être aussi bien un mot roumain en une langue slave ou thrace en roumain et en slave.

2^e Les relations entre le roumain et le hongrois doivent être réanalysées en profondeur. Il y a des dizaines de mots communs aux deux langues considérés en roumain comme éléments hongrois et en hongrois comme des mots d'origine obscure; de tels mots semblent parfois être des mots roumains en hongrois (le cas de *gond*, *határ*, *talp*, *város*, *oriás* etc.).

3^e Une analyse approfondie sans idées préconues (qui sont toujours le fruit de la commodité et de l'ignorance) en ce qui concerne les relations entre le roumain et les autres langues balkaniques d'une part, et les autres langues indo-européennes ou non-indo-européennes (aussi pré-indo-européennes) d'autre part, pour noter les correspondances lexicales évidentes ou possibles. Cela implique une bonne préparation dans le domaine comparatif. Il n'est pas étonnant d'observer que les meilleures histoires ou analyses de la langue roumaine sont dues aussi aux linguistes spécialisés en langue romanes qu'à ceux spécialisés en langues indo-européennes (B. P. Hasdeu, Gh. Ivanescu, A. Vraciu). J'ajouterais les ouvrages remarquables de M. M. Radulescu (par exemple 1981, 1984, 1987) qui, selon notre avis, ont rendu clairs beaucoup d'aspects de l'héritage dace en roumain.

4^e On ne peut pas s'imaginer une analyse profonde, ample, sérieuse et aussi complète que possible du vocabulaire roumain sans coopération entre les linguistes et les archéologues d'une part, et entre diverses branches de la linguistique comparée (les langues slaves, romanes, germaniques, turciques, balkaniques - ce qui implique, encore une fois, la thracologie) d'autre part. Il est évident que les problèmes complexes liés au substrat thrace dans les Balkans ne sont pas une "affaire" roumaine ou bulgare mais qu'ils impliquent et réclament une coopération internationale. Et je pense qu'en premier lieu, il n'y a que des linguistes yougoslaves et bulgares qui puissent offrir des sujets de discussion intéressants.

Evidemment, cette note ne peut pas épuiser tous les problèmes concernant le substrat thrace. En ce qui concerne le problème de l'écriture thrace, je soutiens l'hypothèse avancée il y a trois ans (Paliga 1988 c) fondée sur des données archéologiques et historiques, que la société thrace en général ainsi que la société thraco-dace en particulier, a été de *type initiatique et aniconique* et, de plus, avec *l'interdiction sacrée des représentations visuelles y compris l'écriture*. Les représentations visuelles dans le monde thrace sont le résultat des influences étrangères ou bien une évolution tardive comme résultat du syncrétisme religieux. Le fait que la société thrace a eu un caractère aniconique peut avoir des conséquences essentielles pour la compréhension du caractère spécifique de cette civilisation qui a fasciné les Grecs et qui a influencé l'ethnogénèse du sud-est européen. Comme j'ai souligné dans l'étude citée, on peut observer l'aura magique qui entoure les Thraces et les Etrusques, une aura amplifiée par notre ignorance

doublée par l'ignorance des Grecs et des Romans vis-à-vis de la civilisation des Thraces. C'est la tâche des recherches à venir de préciser si l'apparition tardive de l'écriture chez les Roumains ne constitue pas une survivance de l'interdiction sacrale des représentations visuelles dans le monde thraco-dace.⁴ Cela peut prouver - indirectement - que l'influence subie par la civilisation daco-romane/proto-roumaine de la part de la civilisation thraco-dace a été considérable, beaucoup plus ample et profonde de ce qu'on admet d'habitude. Si c'est le cas (que je suis incliné à soutenir par des données linguistiques et mythologiques), il est évident qu'on ne doit pas utiliser des principes simplistes pour expliquer le rôle du substrat thraco-dace dans l'éthnogénèse roumaine ou, en général, le rôle du substrat dace ou thrace dans les Balkans. Un tel problème réclame de profondes et sérieuses recherches absentes pour le moment dans la plupart des ouvrages consacrés à la langue roumaine et à d'autres langues balkaniques.

Je voudrais accentuer l'idée que la cause principale de l'absence d'un dictionnaire étymologique du roumain à haut niveau scientifique est due en premier lieu à une optique dépassée et déformée pour ne pas dire erronée et dépourvue de fondement (malgré les affirmations contraires de ceux qui l'adoptent). Si on changeait radicalement une telle optique, on pourrait espérer qu'un dictionnaire étymologique du roumain pourrait être écrit au moins dans une première forme provisoire. Celui-ci devrait être un dictionnaire qui comprendrait une analyse sérieuse de toutes les correspondances possibles entre le roumain et les autres langues antiques et modernes (y compris les formes thraces enregistrées par Dačev et d'autres). Il devrait aussi faire allusion au fond pré-indo-européen sans lequel on ne peut pas espérer une analyse étymologique sérieuse. Cette tâche une fois accomplie, on pourrait faire un autre pas plus difficile: le dictionnaire étymologique thesaurus des langues balkaniques qui offrirait une vision d'ensemble sur le rôle du substrat thrace et encore pré-thrace (pré-indo-européen) dans l'évolution ethno-linguistique du sud-est européen. Pour le moment, c'est encore un rêve, mais il y a beaucoup d'exemples qui montrent que parfois les rêves deviennent des réalités vivantes.

⁴ Un exemple typique de la survivance d'une société initiatique de type orale aniconique et le *ius valachicum*, la loi coutumière (jamais écrite!) des Roumains, appliquée pendant le moyen âge partout où ils vivaient.

REFERENCES

- Candrea, I.-A., Densusianu, O. 1914. *Dictionarul etimologic al limbii române. Elementele latine (a - putea)*. București: Socec
- Coteanu, I., Sala, M. 1987. *Etimologia și limba română*. București : Editura Academiei
- Fischer, I. 1985. *Latina dundăreană*. București : Editura Științifică și Enciclopedică
- Paliga, S. 1987 a. Thracian terms for 'township' and 'fortress' and related place-names. *World Archeology* 19, 1: 23-29
- 1987 b. The social structure of the southeast European societies in the Middle Ages. A linguistic view. *Linguistica* 27: 111-126
- 1988 a. A pre-Indo-European place-name: Dalmatia. *Linguistica* 28: 105-108
- 1988 b. Slovansko *s to - izzivalen problem? (avec un résumé en anglais: Slavic *s to - a challenging problem?). *Slavistična revija* 36,4: 349-358
- 1988 c. Dialogul interdisciplinar si tracologia. *Contemporanul* no. 42/21 oct., p. 11
- 1989 a. Oras. Fascinatia unei etimologii. *Noi Tracii* 18 (172): 16-21
- 1989 b. Zeitati feminine ale basmelor romanesti: zîna, Sînziene. Originea cuvîntului si a cultului profan. *Limba romana* 38, 2: 141-149
- 1989 c. Types of mazes. *Linguistica* 29: 57-70.
- Rădulescu, M. M. 1981. Daco-Roumanian-Baltic common lexical elements. *Ponto-Baltica* 1: 15-113
1984. Illyrian, Thracian, Daco-Mysian. The substratum of Romanian and Albanian. *The Journal of Indo-European Studies* 12: 77-131
1987. The Indo-European position of Illyrian, Daco-Mysian and Thracian *The Journal of Indo-European Studies* 15: 239-271
- Raevskij, N. D. 1988. *Contactele romanicilor răsăriteni cu slavii*. Chișinău: Știința
- Raevskij, N. D., Gabinskij, M. 1978. *Scurt dictionar etimologic al limbii moldovenesti*. Chișinău: Redactia principala a Enciclopediei Sovietice Moldovenesti
- Rosetti, Al. 1986. *Istoria limbii române*. Editie definitivă. București: Editura Științifică și Enciclopedică
- Sala, M. (coord.) 1988. *Vocabularul reprezentativ al limbilor romanice*. București: Editura Științifică si Enciclopedică
- Vraciu, A. 1984. Préface de A. Berinde, S. Lugojan, *Contribuții la cunoașterea limbii dacilor*. Timișoara: Facla

Rezumat

VEDERE DE ANSAMBLU DESPRE STRUCTURA ETIMOLOGICĂ LIMBII ROMÂNE

Pornind de la lucrarea lui Raevskij (1988), autorul subliniază principalele probleme cu care se confruntă analiza etimologică a limbii române, subliniind erorile (unele grave) și stângăcia cu care s-a abordat adeseori problema substratului traco-dac, propunând și aici (vezi referințele) soluții pentru depășirea "blocajului" din demersul etimologic, ce face ca limba română să fie singura limbă romanică (și printre foarte puținele limbi europene) fără un dicționar etimologic adecvat poziției sale cheie, atât ca idiom neolatin cât și continuator al substratului traco-dac.