

Vocabularul reprezentativ al limbilor romanice, de Mihaela Bîrladeanu, M. Iliescu, Liliana Macarie, Ioana Nichita, Mariana Ploae-Hanganu, Marius Sala, Maria Theban, Ioana Vintilă-Radulescu; coordonator: Marius Sala; Universitatea București, Institutul de Lingvistică; Editura Științifică și Enciclopedică; București 1988, 629 pp.

1. L'étude statistique des langues, surtout celle du lexique, a une longue tradition et compte un nombre important d'ouvrages. Tout récemment, un groupe de linguistes de l'Institut de Linguistique de l'Université de Bucarest nous a donné un ouvrage consacré à cette discipline mais, d'après les mots mêmes de ses auteurs, conçu comme devant être différent de tous les précédents. Son but est de donner un lexique des langues romanes, défini comme représentatif (pour cette notion voir p. 578), basé sur des critères identiques pour tous les idiomes examinés (pp. 11—12) et strictement synchroniques (p. 12). Il s'agit d'une sélection d'un certain nombre de mots considérés représentatifs, et de critères destinés à réduire le plus possible le moment subjectif (p. 578). Les idiomes examinés sont le roumain, l'italien, le sarde, les dialectes rhéto-romans (appelés de ce terme dans tout l'ouvrage), le français, l'occitan, le catalan, l'espagnol et le portugais. Les unités lexicales sélectionnées vont de 2300 à 2600 (p. 15). Tout en étant bien entendu conscients des limites de la méthode statistique (p. 13), les auteurs déclarent en toute honnêteté préférer le risque de présenter une oeuvre peut-être imparfaite plutôt que renoncer au travail (p. 16).

2. Dans la préface on lit (aux pp. 16—17) que l'ouvrage consiste en deux parties. La première renferme les vocabulaires représentatifs (VR) de chacune des neuf langues, avec la discussion de certains problèmes, la liste des mots sélectionnés et les commentaires sur les critères et les classes morphologiques et étymologiques (pp. 19—482). La deuxième, «rezervată comentariul [sic] general romanic» (p. 17), est destinée à compléter le tableau des ressemblances et des différences entre les langues romanes (jusqu'à présent, au dire des auteurs, on a insisté surtout sur le second aspect) et à décrire ce que les auteurs appellent *dynamique* du VR roman (pp. 483—625). Un chapitre particulier de la 2^e partie (p. 538 et suiv.), qui à vrai dire pourrait être la troisième partie (d'autant plus qu'il porte le numéro 3), est consacré à l'examen de l'élément latin hérité, avec une liste de tous les mots latins inclus.

Les deux parties de l'ouvrage sont précédées de la préface, de la liste des abréviations et d'une courte bibliographie (19 titres) et suivies de la table des matières.

3. Comme déjà dit, la méthode est statistique et synchronique, et en plus de commentaires strictement quantitatifs les auteurs ont cherché à fournir aussi un tableau des implications linguistiques qui en résultent (p. 487). A plusieurs reprises les

auteurs constatent que les résultats du travail confirment leur méthode et même, dans une certaine mesure, les conclusions de certaines recherches antérieures de caractère impressionniste (p. ex., pp. 487, 498—499). Les calculs statistiques et les tableaux les résumant abondent dans tout l'ouvrage. Les critères et, dans la mesure du possible, les solutions aussi visent à être unitaires pour tous les idiomes traités (p. 12). A six des neuf langues (le roumain, l'italien, le français, le catalan, l'espagnol et le portugais) ont pu être appliqués tous les trois critères de sélection, à savoir: la richesse sémantique (S), la capacité dérivative (D) et la fréquence (l'usage) (F/U). Le lexique de chacun des six idiomes est étudié du point de vue des réponses aux critères, combinés ou pris isolément (SDF, SD, SF, DF, S, D, F); ensuite, sont étudiés le rapport entre les classes morphologiques et les critères et celui entre l'étymologie et les critères. Pour ce qui est de la dérivation, seuls les dérivés directs (formés à partir des bases respectives) ont été pris en considération (p. 14). Au sarde et à l'occitan, qui ne possèdent pas de norme standardisée, n'ont été appliqués que deux critères, S et D (p. 80 et suiv., p. 266 et suiv.). Le rhéto-roman est dans une position tout à fait particulière parce qu'il n'est ni unitaire ni ne possède de norme; aussi a-t-on renoncé dans ce cas aux calculs statistiques (avec les analyses ultérieures) et le corpus a été limité au seul élément latin hérité.

Tout d'abord le VR de chacun des idiomes a été établi indépendamment; ensuite, les étymologies ont été revues, pour obtenir des résultats comparables, en tenant compte toujours de la Romania tout entière (p. 511).

Les auteurs comparent plusieurs fois leur travail et ses résultats à certaines études antérieures (de A. Graur, P. Miclău, M. Mathy), en constatant soit des concordances soit d'intéressantes différences (mais sans négliger le fait que les données dans les dictionnaires antérieurs dépendent dans des mesures différentes des critères adoptés par leurs auteurs).

A la fin du livre on propose un choix de thèmes pour recherches ultérieures sur la base des VR établis (p. 579).

4. Les résultats du travail des auteurs sont nombreux et importants, bien que parfois (comme nous le verrons bientôt) ils ne fassent que présenter du point de vue statistique ce qui avait déjà été constaté à l'occasion de recherches précédentes. Pour ce qui est de chacun des idiomes, le roumain, p. ex., est caractérisé par un lexique relativement hétérogène (pp. 77—79, 536), une position moyenne du point de vue des critères adoptés (p. 491) et de celui des classes des mots principales (p. 510), ainsi que par une remarquable perméabilité aux emprunts (p. 536). Tout cela, selon le jugement des auteurs, ne compromet pourtant pas la romanité du roumain: il n'est pas moins roman que les langues soeurs mais uniquement, dans une certaine mesure, *diversement* roman d'elles (p. 579). Au sujet du sarde, on constate une charge fonctionnelle faible de la sémantique (p. 98), une capacité dérivative, elle aussi, faible (pp. 98, 120) et surtout une minceur de la couche latine savante [ce qui, du reste, était prévisible] (p. 110). Le nombre d'emprunts romans dans le sarde est considérable (pp. 119, 121), ce qui est une conséquence de son histoire (p. 527); enfin, dans le sarde (comme dans l'occitan, p. 307) est très nette — plus que dans les autres idiomes — la différence entre les mots sémantiquement «pleins» et les soi-disants «outils grammaticaux» (p. 115). Pour l'italien sont caractéristiques le nombre relativement

élevé d'emprunts germaniques anciens et français (p. 179), une distinction souvent difficile entre les mots populaires et les mots savants (latinismes) [cela aussi est prévisible, étant donné la conservativité de l'italien] (p. 524), une remarquable vitalité de la dérivation (p. 491) (tandis que celle-ci est relativement faible en français, *loco cit.*) et une distribution à peu près égale des classes étymologiques dans les catégories des mots (p. 181). L'occitan est «un limbaj preponderent rustic» (p. 265), sans norme; il possède une dérivation très riche (pp. 265, 296), un caractère fort créatif (p. 313) et un polysémantisme développé (p. 296), tandis qu'il est peu perméable aux emprunts (*ib.*). A certains égards l'occitan occupe une position unique dans la Romania (p. 266). Ce qui est un peu surprenant c'est le nombre relativement petit d'éléments celtiques en français par rapport à l'occitan, différence due au caractère rustique et traditionnel du lexique de ce dernier (p. 530). L'élément français est peu représenté en occitan; ceci ne reflète évidemment pas l'état réel des idiomes mais est la conséquence du purisme du vocabulaire occitan qui a servi de base (p. 532). Dans le catalan, la composante lexicale la plus nombreuse est celle des latinismes (*cultismes*) (p. 354), la productivité est remarquable (p. 364), mais surtout «*ponderea elementului latin este considerabilă*» dans cette langue (*ib.*). L'espagnol possède lui aussi un lexique productif (pp. 400, 420); ce qui surprend surtout c'est la faible proportion des éléments arabes (le 6^{ème} rang dans l'échelle des fréquences; 0,99%). Le portugais, comme l'occitan, est peu perméable aux emprunts (p. 477).

5. Très intéressants sont les résultats qui concernent l'ensemble des langues romanes. Pour la plupart d'entre elles les auteurs constatent «*profunda latinitate*» ou «*caracterul romanic*» (mais pourtant pas pour le roumain, malgré ce qui est dit aux pages 578—579; v. *supra*). Quant aux classes morphologiques, le substantif, le verbe, l'adjectif et l'adverbe occupent les premières places dans toutes les langues (p. 494). L'importance des classes morphologiques correspond aux prévisions, mais avec des écarts considérables (p. 498). La structure morphologique confirme pourtant, jusqu'à un certain degré, les constatations impressionnistes des travaux antérieurs (p. 510). L'élément latin hérité est en tête dans tous les idiomes (p. 516), tandis qu'il y a des différences p. ex. dans les rapports entre l'élément hérité et les formations internes ultérieures (p. 511). Ce sont surtout les différences entre le roumain et les langues romanes occidentales qui sont remarquables (pp. 512, aussi pp. 525, 527). Du point de vue des emprunts latins et romans mis ensemble l'italien occupe la première place et le roumain la dernière (pp. 528—529). Il y a également des différences importantes dans l'élément latin savant (p. 525). En général, les langues romanes se distinguent les unes des autres plus dans leur structure étymologique que morphologique (p. 535) [c'est là une autre confirmation de ce que l'on savait déjà, c'est-à-dire l'hétérogénéité du lexique en face de la stabilité de la morphosyntaxe, avec l'implication bien connue pour la classification typologique]. L'élément panroman est l'élément d'unité romane au niveau lexical (p. 545); il est caractérisé par l'unité mais présente aussi des différences (pp. 546—547). La catégorie des éléments panromans «*formează elementul de stabilitate din lexicul romanic*» (p. 577), c'est le secteur le plus important de la composante latine héritée, quoique les mots qui reflètent les divers changements romans soient beaucoup plus nombreux (*ib.*). Tous les idiomes romans ne conservent pas le lexique de base latin dans la même mesure

[constatation tout à fait prévisible, v. § 6.5] (pp. 553, 556, 560). Enfin, et c'est là un résultat extrêmement important pour les études diachroniques et surprenant en même temps, le substrat est dans tous les idiomes romans une composante lexicale de faible importance (p. 530). Ce résultat est en contraste net avec la linguistique traditionnelle qui admettait l'influence du substrat — si tant est qu'il y en avait — justement dans le lexique.

6. Passons maintenant à la discussion de certaines formulations et à quelques objections critiques.

6.1 Le fait qu'il n'a pas été possible d'appliquer tous les trois critères à toutes les neuf langues soulève le problème de l'hétérogénéité, donc de la comparabilité des résultats. Dans plusieurs cas les résultats des calculs statistiques ont été ensuite modifiés et/ou réinterprétés (§§ 1.2.0.2.1, 1.3.0.2.1, 1.6.0.2.1, 1.7.0.2.1, 1.9.0.2.1). A la différence de la structure étymologique, ce procédé n'a pas été appliqué dans la structure morphologique, pour des motifs qui dépendent des domaines spécifiques respectifs (p. 492) [quelques explications de ces motifs seraient très utiles]. Ces circonstances imposent une certaine réserve devant les résultats et les conclusions qu'en tirent les auteurs.

6.2 Presque pour toutes les langues, les auteurs constatent eux-mêmes que les résultats des statistiques surprennent, car certaines classes de mots y sont peu représentées, ce qui ne correspond pas à leur position réelle (pp. 51, 104, 157, 239, 296, 346, 401, 456). L'inclusion des mots à fonction morphosyntaxique multiple modifie le tableau, mais seulement en partie. Voilà qu'une fois de plus la même question se pose: s'il en est ainsi, quelle valeur peuvent avoir les calculs et les déductions qui se basent sur eux? Dans certains cas les pourcentages sont vraiment très faibles (voir p. ex. les chiffres qui représentent certaines couches lexicales du catalan, p. 368), et cela aussi rend les résultats sujets à caution.

6.3 L'objection suivante nous semble grave. Pour l'italien et pour l'occitan les auteurs ont renoncé à inclure les formations avec préfixes parce que, vu leur fréquence dans les deux langues, leur inclusion aurait augmenté démesurément le corpus (pp. 124, 265, 485). Or, si l'importante catégorie des formations avec préfixes est négligée, le tableau résultant ne peut être que déformé: il ne reflète pas la situation réelle et les résultats ne sont pas comparables à ceux qui ont été obtenus par l'analyse de la formation des mots dans les autres langues. La richesse des formations avec préfixes, loin d'être un argument pour les omettre, en est au contraire un, très important, pour les inclure dans l'analyse!

6.4 Dans les calculs statistiques la catégorie «étymologie incertaine» figure de pair avec les étymologies «certaines», c'est-à-dire sûres, connues. Toutefois, à notre avis, l'étymologie incertaine (donc, inconnue) ne devrait pas avoir la même position dans les calculs que les étymologies connues. Du reste, la statistique des classes étymologiques pour l'occitan ne prend pas en considération les étymologies incertaines (p. 313, note 13), ce qui est tout à fait justifié, d'autant plus que les auteurs eux-mêmes se rendent compte que la différence des proportions des étymologies incertaines reflète avant tout le stade des recherches étymologiques (p. 530). S'il en est ainsi, on peut bien se demander pourquoi la même réserve n'a pas été appliquée à tous les VR, c'est-à-dire pourquoi les étymologies incertaines n'ont pas été écartées, ou pour

le moins présentées séparément. La même objection vaut aussi pour les étymologies dites «multiples», dont les auteurs disent qu'elles sont une catégorie plus subjective que les autres, parce qu'elles dépendent dans une large mesure des auteurs des dictionnaires utilisés comme base des calculs (p. 532).

6.5 Il y a des formulations qui se répètent presque littéralement pour chacune des langues. On en a vu un cas dans le § 6.2; un autre exemple est la conclusion sur le caractère «profondément roman» ou la «profonde latinité» de la plupart des langues, elle aussi déjà citée. Parfois notre *Vocabularul reprezentativ* annonce comme résultats «intéressants» [donc nouveaux?!] des faits connus depuis longtemps, vraiment banals, p. ex. p. 553: «este interesant de observat de asemenea că nu toate VRRom păstrează în aceeași măsură vocabularul de bază al latinei», affirmation répétée presque mot à mot aux pages 556 et 560. Si les VR de chacun des idiomes doivent refléter la situation réelle de l'idiome respectif, de telles différences sont tout à fait normales et données d'avance; donc, il n'y a rien là de particulièrement intéressant. On a un peu l'impression que les parties, élaborées chacune par leur auteur, ont été mises ensemble sans retouches ultérieures, d'où quelquefois des répétitions inutiles et des formulations quelque peu naïves.

6.6 A la fin de la section consacrée aux rangs des classes étymologiques dans les VR on constate que «rangurile ocupate de diversele clase etimologice corespund în general așteptărilor» (p. 517). Sur plusieurs points il n'en est pourtant pas tout à fait ainsi: les celtismes, p. ex., manquent en français, l'élément arabe en espagnol est à peine au 6^{ème} rang, le substrat est absent du français et du sarde (quoiqu'on lise à la p. 515 que dans le sarde le substrat se cache entièrement sous l'étiquette des étymologies incertaines, ce qui implique que dans le sarde il y a des restes lexicaux du substrat).

7. On peut faire aussi plusieurs autres objections, de moindre importance.

7.1 Page 105: il nous semble que la différence entre 187 et 47 termes (resp. 38,79% et 9,75%) ne soit pas 'très petite' («foarte mică») mais plutôt considérable.

7.2 Page 120: selon nous, si le sarde est resté différent de l'italien, ce n'est pas tant parce que la forte influence des deux idiomes ibéro-romans a empêché les italianismes d'y pénétrer, mais pour bien d'autres raisons encore: la romanisation particulière de l'île (voir, tout récemment, E. Blasco Ferrer, *Le parlate dell'Alta Ogliastra*, Cagliari, 1988), la genèse du sarde et toute son évolution ultérieure.

7.3 Page 125: pour l'italien, seul le dictionnaire étymologique de C. Battisti et G. Alessio (DEI) a été consulté, alors qu'on a également le vocabulaire étymologique de A. Prati (VEI) et en plus, de nos jours, celui de M. Cortelazzo e P. Zolli.

7.4 Page 300: on a de la peine à voir comment la catégorie de l'article (en occitan) peut participer à la catégorie [c'est-à-dire: le critère] D. Doit-on comprendre que l'article sert de base aux dérivés?

7.5 Page 427: que signifie *vorbitor negativ* ('parlant négatif')? Est-ce une coquille pour *vorbitor nativ*?

7.6 Pages 427—428: les listes de fréquence du portugais ont été élaborées selon les «centres d'intérêt» (d'après Gougenheim), tandis que dans l'élaboration des autres VR ce critère n'a pas été appliqué (ou, du moins, ce n'est pas dit expressément).

Cette incohérence est, elle aussi, une source d'hétérogénéité et d'incommensurabilité, d'autant plus que ces «centres d'intérêt» ne sont pas commentés.

7.7 Page 490: l'affirmation que le VR portugais est basé sur une «apreciere ne-cantitativă a frecvenței cuvintelor» ('évaluation non quantitative de la fréquence des mots') nous est incompréhensible, car elle comporte une vraie *contradictio in adiecto*: comment est-il jamais possible d'établir la fréquence d'un élément quelconque de façon non quantitative?

7.8 Page 497, lignes 4—12: les chiffres entre parenthèses ne se rapportent pas aux mots à fonction morphologique multiple mais à ceux qui sont ajoutés dans les listes supplémentaires: en effet, les chiffres qui résultent de l'adjonction des mots à fonction multiple sont entre crochets dans tout le livre.

7.9 Page 515: si dans le VR sarde le substrat se cache entièrement dans les mots à étymologie incertaine, c'est une preuve que la statistique ne reflète pas toujours la réalité linguistique; car il est possible qu'il y ait aussi des mots à étymologie incertaine dans les couches lexicales postérieures, ainsi qu'il existe sans doute des étymologies prélatines certaines. Le substrat et l'étymologie incertaine ne se recouvrent donc qu'en partie.

7.10 A la même page, dans le tableau qui représente les rangs des classes étymologiques dans chacun des VR, le substrat est présenté de façon assez incohérente et mal précisée. Tout d'abord on cite le «substrat neprecizat» (dénomination en soi douteuse, et encore est-elle présente seulement en italien — pourquoi? comment?; justement en italien, quoique les substrats d'Italie nous soient mieux connus que ceux du reste de l'Empire). Ensuite nous trouvons le substrat préroman, suivi par les substrats ibérique, celtique et thraco-dace. Puisque la catégorie 'préroman' inclut aussi les autres substrats spécifiés, l'insuffisance logique est évidente. En plus: comment savons-nous qu'il s'agit de substrat, si celui-ci n'est pas précisé?

7.11 Page 521: l'abréviation VRP ne figure pas dans le tableau à la p. 520, et l'abréviation VRF devrait, elle aussi, être une faute, car il est absurde de calculer les emprunts français dans un vocabulaire français. Comment faut-il corriger?

7.12 Page 531: dans le tableau un des deux VRF doit être remplacé par VRP: lequel?

7.13 Ib.: sauf erreur de notre part, apparaît ici pour la première fois la catégorie de l'adstrat (turc, arabe), de sorte qu'on ne sait pas ce qui lui correspond dans les nombreuses analyses précédentes. D'après les auteurs (loco cit.) l'adstrat est une catégorie «mai laxă decât substratul sau superstratul».

8. Outre les tout à fait banales coquilles, faciles à corriger (ainsi que, bien entendu, la vérification des innombrables calculs, pourcentages etc.), nous relevons les erreurs suivantes: 1) Bibliographie: corriger deux fois *Vocabulario* en *Vocabolario* et *Dicionarie* en *Dicionário*; remplacer pour DEI l'année 1957 par les années 1950—1957 (de même à la page 125). — 2) Page 81: lire *paffuto* au lieu de *paffutto*. — 3) Page 103: lire *vapore*, non *vopore*. — 4) Page 124: corriger *maleta* en *malato* (ou *malata*?). — 5) Pages 152—155: corriger *eccelenzza*, *suporre*, *abboccare*, *appicare*, *sciale*, *tabbaro* resp. en *eccellenza*, *supporre*, *abboccare*, *appicare*, *scialle*, *tabbarro*. — 6) Page 186: dans le titre du dictionnaire de Pizzinini e Plangg corriger *Paroles* en *Parores*. — 7) Pages 187—204: corriger p. ex. *altetzga*, *avrer*, *bader*, *intosr-*

he, minți etc. resp. en *altezia, arver, badar, intorshe, minti*; le mot sursilvain autochtone pour 'quand' est *cura*, non *cont*; COXA signifie en roumain 'coapsă' ('cuisse'), non 'coasă' ('faucille'); la forme engadinoise *guavid* devrait être une faute, car le REW 9321 donne *vaidg*; et il y a aussi d'autres fautes mineures. — 8) Page 257, ligne 5: corriger 697 en 693 (ce chiffre figure aussi au tableau de la même page). — 9) Page 296: corriger deux fois VRC en VRO. — 10) Page 357, ligne 7: corriger 399 en 1399, et dans le tableau, auprès de Numeralul, remplacer [VII] par [VIII]. — 11) Page 523, ligne 6: remplacer Elementele de substrat par Elementele de superstrat. — 12) Page 529: remplacer le renvoi au § 5.7.1.3 par le renvoi au § 2.5.7.6. — 13) Page 541: lire *texere* au lieu de *toxere*. — 14) A la page 546 le chiffre des éléments latins hérités sélectionnés dans le VRRom est de 2740, à la page suivante 2470: quel est le chiffre exact?

Pavao Tekavčić