

# *Le vocabulaire fondamental des langues romanes*

Marius SALA

1. Les études comparatives sur le lexique roman, à partir de celles de W. Meyer-Lübke jusqu'au récent travail de H. Lüdtke (*Geschichte des romanischen Wortschatzes*, 1968) ont eu comme principal but la présentation du vocabulaire dans une perspective historique; tout en mettant l'accent sur leur origine, on a procédé à une analyse des éléments constitutifs du vocabulaire des différents idiomes romans. Ces études ont mis sur le même plan des mots dont leur position dans l'ensemble du vocabulaire des langues respectives était fort différente: à côté des mots d'une large diffusion, caractérisés par une grande richesse sémantique et une famille nombreuse, il y en a d'autres, conservés uniquement dans certaines variantes diachroniques, diatopiques ou diastratiques de la langue.

On peut observer le même état de choses dans les études de géographie linguistique romane comme celles de G. Rohlfs figurant dans son célèbre ouvrage *Die lexicalische Differenzierung der romanischen Sprache. Versuch einer romanischen Sprachgeographie* (München, 1954). Ces études traitent d'un nombre réduit de faits auxquels on peut toujours leur opposer un nombre égal de faits contraires. Une fois encore, à côté des mots importants il y en a d'autres dont la valeur fonctionnelle est plus restreinte.

2. En tenant compte de toutes ces données, nous avons entrepris et réalisé dans le département des langues romanes de l'Institut de linguistique de l'Université de Bucarest un projet de recherche du vocabulaire roman qui élimine, dans la mesure du possible, le caractère aléatoire des résultats offerts par les recherches antérieures. Puisque, par de motifs évidents, la solution idéale d'une telle recherche — l'analyse du vocabulaire dans son entier — ne saurait être indiquée, nous

nous sommes proposés comme but de ce projet de faire distinguer entre les différents mots susceptibles d'une valeur fonctionnelle différente dans les langues romanes *actuelles*. Dans une deuxième étape, surces assises ainsi constituées, nous nous proposons d'élaborer une série d'études typologiques et comparatives de lexicologie romane. Ces études sont d'autant plus intéressantes que dans les recherches de typologie romane on a traité jusqu'à présent surtout des faits appartenant à la phonologie et à la phonétique (M. Pei, J. Greimas et F. B. Aggard, Bodo Müller) ou à la phonologie et à la morphologie (Ž. Muljačić, Maria Iliescu, G. B. Pellegrini, etc.). Pour ce faire, nous avons choisi un noyau de mots importants qui compose le vocabulaire fondamental de chaque idiome roman pris à part, en utilisant les mêmes principes de sélection pour chacun d'entre eux (le roumain, le sarde, l'italien, le rhétoroman, le français, l'occitan, le catalan, l'espagnol et le portugais). Ce noyau est représenté par environ 2.000-2.300 mots qui répondent, dans la variante actuelle de chaque idiome roman, aux critères de sélection utilisés dans des études similaires: la fréquence (les premiers 2.000 mots des dictionnaires de fréquence de A. Juilland), la puissance de dérivation (mots à 3 dérivés) et la richesse sémantique (mots à 5 sens). Vu le caractère synchronique de la recherche nous n'avons pas tenu compte du critère de l'ancienneté, de la stabilité du mot dans la langue. En même temps, nous n'avons pas pris en considération le critère de l'importance de la notion et, implicitement, celui de la disponibilité, parce que, à l'encontre des vocabulaires fondamentaux comme celui élaboré par G. Gougenheim pour le français, notre vocabulaire n'a pas de but pratique. Les auteurs du travail, intitulé *Le vocabulaire représentatif des langues romanes*, sont: Mihaela Bîrlădeanu, Mariana Hanganu, Maria Iliescu, Liliana Macarie, Ioana Ni-chita, Maria Theban, Ioana Vintila-Radulescu et l'auteur de la présente communication.

Il est évident que la frontière qui sépare le noyau ainsi établi du reste du vocabulaire ne saurait être déterminée avec beaucoup d'exactitude puisque, tel que le montrait I. Coteanu (*SCL*, IX, 1958, p. 400) «le vocabulaire ressemble à une nébuleuse dont la matière se raréfie constamment du centre vers la périphérie. La densité de plus en plus réduite ne présentant pas de barrières ne diminue pas brusquement». Il nous semble inutile d'insister sur d'autres difficultés comme celles déterminées par le fait que bon nombre de langues romanes n'ont pas les mêmes instruments de travail (des dictionnaires) qui permettent la mise en pratique des critères de sélection énoncés ci-dessus. A côté des langues qui disposent de dictionnaires de fréquence ainsi que de riches dictionnaires explicatifs, essentiels pour appliquer nos critères de sélection, il existe d'autres langues qui n'ont pas de tels instruments (dans ces cas-là, les auteurs qui ont élaboré les vocabulaires re-

présentatifs ont dû utiliser des listes de fréquence, avec des indications succinctes, ce qui est le cas pour le catalan et le portugais, ou même des dictionnaires bilingues, ce qui est arrivé pour le rhétoroman).

3. Le commentaire que nous avons fait, obtenu par les données offertes par les inventaires ainsi établis, a mis en évidence toute une série de faits intéressants à la fois chaque langue romane et la comparaison entre ces langues.

3.1. Le premier commentaire au niveau de chaque langue romane a été déterminé par la constatation que l'on peut établir une hiérarchie entre les divers mots constituant les vocabulaires représentatifs, en fonction de leur manière de répondre aux critères de sélection: il y a des lexèmes qui répondent à tous les trois critères de sélection (il est à noter que pour le rhétoroman, l'occitan et le sarde il n'existe pas d'informations sur le critère de la fréquence), d'autres qui ne répondent qu'à deux ou même à un seul critère de sélection (il s'agit surtout du critère de la fréquence). Selon ces faits, on a établi pour chaque langue romane (à l'exception des trois langues mentionnées plus haut) trois groupes de mots:

a) mots qui répondent à tous les trois critères (pour le roumain leur nombre est de 207)

b) mots qui répondent à deux critères. Ce groupe, formé par 742 mots, se divise à son tour en trois sous-groupes:

b') mots qui répondent aux critères US - 468

b'') mots qui répondent aux critères UD - 162

b''') mots qui répondent aux critères SD - 112

c) mots qui répondent à un seul critère (1.642) et sont répartis en trois sous-groupes:

c') mots qui répondent au critère de l'U - 1.074

c'') mots qui répondent au critère de S - 258

c''') mots qui répondent au critère de D - 310

On établit de la sorte différents groupes de mots en fonction de leur importance fonctionnelle.

Un autre commentaire, commun à chaque langue, se rapporte à la composition étimologique et à la structure morphologique des vocabulaires représentatifs.

A part cela, il existe d'autres commentaires qui pourraient être mis à contribution dans un proche avenir (les auteurs du travail n'ont pas voulu trop augmenter son volume). On peut comparer par exemple la manière dont chaque langue romane prise à part continue le voca-

bulaire représentatif du latin. Pour la comparaison on peut utiliser un ouvrage de Maurice Mathy, *Vocabulaire de base du latin*, 2.<sup>e</sup> éd., Paris, 1956, et le vocabulaire établi pour le latin par Al. Graur, dans son livre consacré au vocabulaire fondamental du roumain, *Incercare asupra fondului principal lexical al limbii române*, Bucuresti, 1954, pp. 94-99. Le premier auteur s'est servi de la fréquence en tant que critère de sélection des mots pour le vocabulaire de base du latin; il a établi une statistique des mots existants dans les textes latins utilisés dans les écoles françaises et, en s'appuyant sur 5.500 mots, il a obtenu une liste de 2.212 mots à une fréquence d'au moins 4 occurrences. Le deuxième auteur, Al. Graur, a utilisé d'autres principes dans la sélection des mots en vue de former le vocabulaire de base du latin, à savoir l'ancienneté, la richesse sémantique et la puissance de dérivation. La liste de Al. Graur comprend 787 mots. Le nombre sensiblement réduit de l'inventaire de Al. Graur s'explique par le fait que les familles de mots y sont représentées par un seul mot. Ces deux vocabulaires de base latins se révèlent être complémentaires et ils offrent l'image d'un vocabulaire latin fondé sur les critères qui ont présidé à l'élaboration des vocabulaires fondamentaux des idiomes romans.

3.2. Autrement intéressants sont les commentaires qui visent la comparaison romane. Les différents commentaires au niveau de la Romania pourraient être groupés à leur tour en deux catégories: les uns touchent explicitement tout l'ensemble de la Romania, les autres comprennent implicitement cette comparaison.

3.2.1. Les commentaires se rapportant explicitement à tout l'ensemble de la Romania mettent dans une nouvelle lumière ce qui unit et ce qui sépare les langues romanes au niveau du vocabulaire. Dans ce qui suit nous allons insister sur ce qui unit les langues romanes, fait ignoré jusqu'à ce jour dans les études sur le vocabulaire roman où, à notre connaissance, la catégorie des mots panromans n'a pas été analysée.

Il existe bon nombre de mots latins transmis à toutes les langues romanes. Une liste de ces mots, dressée par I. Fischer (*Istoria limbii române*, vol. II, Bucuresti, 1969, pp. 111-112) conformément au dictionnaire étymologique de la langue latine de A. Ernout et A. Meillet, comprend un nombre de 488 mots (d'après nos recherches, ce nombre atteint le chiffre de 502). Les mots panromans n'ont pas la même position dans l'ensemble du vocabulaire de chaque langue romane. Si l'on tient compte de la manière dont chaque mot a été sélectionné dans les divers vocabulaires représentatifs romans ainsi que de leur présence ou absence dans le vocabulaire de base du latin, on peut distinguer quatre grandes catégories de mots: des mots qui appartiennent à la fois aux vocabulaires représentatifs romans et à celui latin, des mots qui appartiennent uniquement aux vocabulaires représentatifs romans,

d'autres appartenant seulement au vocabulaire de base du latin et enfin des mots qui ne figurent dans aucun des vocabulaires mis en discussion.

VR Roman	VB Latin
+	+
+	—
—	+
—	—

3.2.1.1. Parmi les mots sélectionnés dans le VR roman et le vocabulaire de base latin on peut reconnaître deux groupes:

3.2.1.1.1. Les mots sélectionnés dans tous les VR romans et dans le VB latin sont au nombre de 81. De cette catégorie, nous donnons à titre d'exemple les mots commençant par les lettres A, B, et C: *ad, altus, aqua, arbor, arcus, bene, bonus, campus, can(t)o, caput, caro, clavis, cognosco, cohors, contra, cornu, curro*.

3.2.1.1.2. Les mots appartenant au vocabulaire représentatif de certaines langues romanes sont au nombre de 187: *acer* (5) \*, *albus* (2), *anima* (7), *annus* (8), *ante* (4), *area* (2), *arma* (7), *ars* (2), *asper* (5), *audio* (7), *bibo* (8), *bos* (6), *bracchium* (8). Il est à remarquer que presque deux tiers de ces mots sont sélectionnés dans les vocabulaires représentatifs d'un grand nombre de langues romanes (6, 7 ou 8). Sont rares les cas où ces mots sont sélectionnés dans un ou deux vocabulaires représentatifs: *invito, neque* (dans un seul VR), *albus, area, aro, mane, mollio, rado, servus, socer, vado* (dans deux VR).

3.2.1.2. Les mots panromans appartenant au vocabulaire représentatif des langues romanes et ne figurant pas dans le vocabulaire de base du latin peuvent être répartis en deux groupes:

3.2.1.2.1. Un nombre de 36 mots sont sélectionnés dans le vocabulaire représentatif de toutes les langues romanes, mais sont absents du vocabulaire de base du latin: *aer, barba, battuo, camisia, capra, carrico, chorda, cooperio, costa* (les exemples sont extraits des lettres A, B et C).

3.2.1.2.2. Nombreux sont les mots (173) qui appartiennent au vocabulaire représentatif de certaines langues romanes, mais qui sont absents du vocabulaire de base du latin: *adiuto* (8), *allium* (2), *altare* (2), *annelus* (5), *angelus* (8), *aprilis* (4), *arunca* (4), *argentum* (5), *asinus* (3), *augustus* (6), *aurus* (6), *ausculto* (8). Il faut remarquer que, à la différence de la catégorie précédente, cette fois-ci le nombre des

\* Le chiffre mis entre parenthèses indique le nombre de ces langues romanes dans les vocabulaires représentatifs actuels desquels on a sélectionné le mot respectif.

mots sélectionnés dans le vocabulaire représentatif de plusieurs langues romanes (6, 7 ou 8) est moins élevé (il ne touche même pas à la motié). Il faut aussi souligner le nombre élevé de substantifs qui appartiennent à cette catégorie par rapport au nombre de verbes, très peu élevé.

3.2.1.3. Les mots panromans appartenant au vocabulaire fondamental latin qui n'ont été sélectionnés dans aucun vocabulaire représentatif roman sont au nombre de 3: *ager*, *furtum*, *oro*. Parmi ceux-ci, *ager* seulement est sélectionné tant par Mathy que par Graur, tandis que *furtum* et *oro* par Mathy seulement. L'absence de *ager* et de *oro* dans tous les vocabulaires romans actuels s'explique par la concurrence avec leurs synonymes *campus* et *rogo*. En ce qui concerne le mot *furtum*, il faut observer que certaines langues romanes actuelles ont sélectionné le verbe appartenant à la même famille (*furare*), tandis que d'autres ont remplacé ce même verbe par d'autres mots.

3.2.1.4. Les mots panromans qui n'ont été sélectionnés ni dans le vocabulaire représentatif d'une langue romane quelconque, ni dans le vocabulaire de base du latin, sont au nombre de 18: *aratrum*, *balteus*, *carpinus*, *coma*. La plupart sont des noms de plantes.

3.2.1.5. L'examen des quatre catégories mentionnées ci-dessus montre que plus de la moitié (268) des mots panromans représentant la première catégorie ont été sélectionnés dans le vocabulaire représentatif d'une ou de plusieurs langues romanes actuelles, ainsi que dans le vocabulaire de base du latin. Ils sont, évidemment, les plus importants du point de vue de la stabilité du vocabulaire, attestant la continuité au même niveau du latin aux langues romanes. Ils doivent figurer dans toute description typologique de ce que l'ensemble des langues romanes a conservé du latin.

Une continuité également, mais dans le sens négatif, présente aussi bien la dernière catégorie, celle qui réunit les mots panromans non-sélectionnés dans le vocabulaire de base du latin, de même que dans nul vocabulaire représentatif roman. Ces mots —très peu nombreux (19)— ont aujourd'hui le même statut dans les langues romanes qu'autrefois dans le latin: ce sont des mots connus, mais négligés. Ils appartiennent à la section habituellement nommée vocabulaire disponible où se trouvent également les termes désignant diverses plantes moins importantes (*carpinus*, *fraxinus*, *hedera*, *ienuperus*, *lambrusca*, *laurus*, *salvia*, *sambucus*, *tilia*), mais connues (à l'appui de cette dernière affirmation étant le fait même qu'ils ont été conservés dans toutes les langues romanes).

Les deux catégories restantes réunissent les situations qui ne présentent pas une continuité, au niveau du vocabulaire de base, du latin aux langues romanes actuelles. Les plus nombreux (202) sont les mots non-sélectionnés dans le vocabulaire de base du latin mais présents

dans les vocabulaires fondamentaux romans. On est arrivé à cette situation grâce d'abord aux importantes modifications de vocabulaire produites dans le latin tardif, menant à l'élimination d'une série de synonymes en faveur de leurs concurrents plus expressifs (tel est le cas typique de *caballus*). La plupart des mots de cette catégorie appartenait dans le latin à ce que nous avons appelé plus haut vocabulaire disponible: noms de plantes, d'animaux, etc. Dans les langues romanes ces mots ont passé du vocabulaire disponible dans le vocabulaire fondamental, ce qui constitue un processus complexe qui n'a pas été jusqu'à présent signalé. A la différence de leur situation dans le latin, ces mots ont été employés toujours plus fréquemment, acquérant des sens nouveaux et devenant base pour des mots dérivés. Par exemple, *capra* a été sélectionné par les langues romanes pour des sens et des dérivés (le roumain, le rhétoroman, l'occitan), pour des dérivés (l'italien, le sarde, le français), pour des sens et fréquence (le catalan) et pour des dérivés et fréquence (l'espagnol, le portugais). L'absence d'une continuité dans le cas des mots panromans sélectionnés dans le vocabulaire de base du latin mais non-sélectionnés dans les vocabulaires représentatifs romans dont 3 mots seulement mentionnés plus haut font preuve, peut être considérée comme négligeable.

3.2.1.6. Des conclusions pareilles à celles présentées plus haut peuvent être formulées aussi bien si l'on examine tout le vocabulaire hérité du latin par les langues romanes. Nous avons largement présenté des conclusions concernant la modification de l'ensemble du vocabulaire latin dans son passage aux langues romanes dans notre article *Del latín a las lenguas romances. Dinámica del vocabulario fundamental*, dans «Anuario de Letras» XIX (1981), pp. 31-46.

3.2.1.7. Il est possible de comparer toutes les langues romanes du point de vue des différents éléments constitutifs au niveau du vocabulaire fondamental. L'élément latin hérité se présente sous des pourcentages qui vont de 30,45 % et 35,37 %, enregistrés pour le roumain, respectivement pour le français, jusqu'à 48,80 % et 49,50 % (le sarde, respectivement le catalan). Une comparaison faite au niveau des vocabulaires fondamentaux des langues romanes et du latin prouve d'une manière évidente que le roumain (12,42 %) et l'italien (14,70%) sont les langues les plus éloignées du latin, tandis que le sarde (19,14%) et le catalan (18,8%) sont les plus proches.

3.2.1.8. Il est possible de comparer seulement deux langues romanes au niveau du vocabulaire fondamental. Des résultats intéressants dans ce domaine fournit la thèse de doctorat de Víctor Manuel Sánchez Corrales (C. Rica), *L'élément lexical latin hérité en espagnol et en roumain*.

3.2.2. Les études comprenant implicitement une comparaison sont ceux qui examinent une langue romane dans l'ensemble du contexte

roman. Nous avons analysé, en employant les données du *Vocabulaire représentatif des langues romanes*, deux catégories de mots amplement discutés dans la linguistique roumaine (les mots panromans sauf roumain et les mots conservés exclusivement en roumain) et nous sommes arrivés à des conclusions différentes de celles communiquées dans les études antérieures.

3.2.2.1. Les mots panromans, à l'exception du roumain, ont été analysés par I. Fischer dans un article intitulé *Panroman sauf roumain* (*RRL IX*, 1964, pp. 595-602), après avoir été signalés pour la première fois par S. Puscariu dans son très connu discours de réception à l'Académie Roumaine, consacré à la place occupée par la langue roumaine parmi les langues romanes. On a établi un inventaire de 214 mots.

En employant les données fournies par l'ouvrage *Le vocabulaire des langues romanes actuelles*, un aspect complémentaire peut être pris en considération, à savoir la recherche de la position dans une langue des éléments latins conservés dans toute la Romania à l'exception du roumain. Une analyse sommaire de cet inventaire de 214 mots nous permet d'observer que les mots panromans sauf roumain n'ont pas tous une même position dans l'ensemble de la Romania. Certains d'entre eux n'ont été sélectionnés par aucun vocabulaire représentatif roman, d'autres ont été sélectionnés seulement par 1-2 vocabulaires représentatifs (21, respectivement 32); à l'opposé, il y a les mots latins sélectionnés par les vocabulaires représentatifs de toutes les langues romanes (9 termes) ou bien par un grand nombre de langues (5 à 7). Il s'ensuit que les mots de cette catégorie n'ont été pour la plupart sélectionnés par nul vocabulaire représentatif roman, ou bien ont été sélectionnés par un ou deux vocabulaires représentatifs seulement. Ils sont donc des mots sans importance dans l'ensemble de la Romania, leur absence en roumain ne prouvant rien d'autre que cette langue avait éliminé plus nettement les mots «faibles» du vocabulaire latin. Aux mots panromans sauf roumain sélectionnés par plusieurs (au moins 5) vocabulaires représentatifs romans correspondent le plus souvent des synonymes latins ou des dérivés à partir des mots latins; le remplacement de ces mots latins s'est passé depuis longtemps dans le latin carpatho-danubien, suite à un processus de différenciation stylistique.

3.2.2.2. On arrive à une conclusion semblable si l'on étudie les mots conservés exclusivement en roumain et disparus dans les autres langues romanes. Sans se perdre dans le détail, il convient de préciser que, parmi les 107 mots conservés exclusivement en roumain, 17 seulement ont été sélectionnés dans son vocabulaire représentatif et que, de même que dans le cas des manques du roumain (mots absents du roumain mais hérités par d'autres langues), aux mots latins con-

servés en roumain seulement correspondent dans les autres langues romanes d'autres mots latins (*lingula-cochlearium*, *ovis-ovicula*, *vervex*, *(ovis)feta*) ou bien des emprunts savants (*imperator*).

3.2.2.3. On peut conclure qu'en observant les données offertes par les vocabulaires représentatifs romans, les deux catégories de mots en question ne présentent pas l'importance qu'on leur attachait jusqu'à présent.