

LES BALKANS LINGUISTIQUES

JACQUES ALLIÈRES

La Péninsule des Balkans, que l'on ne sait trop comment délimiter au Nord mais à laquelle le linguiste aimerait bien voir annexée la Roumanie, est à la fois un lieu de passage précisément dans sa partie septentrionale, et dans sa partie méridionale un cul-de-sac peut-être si l'on s'en tient aux voies terrestres, mais un cul-de-sac « ouvert » sur l'une des mers les plus fécondes qui soit dans l'histoire de l'humanité, la Méditerranée. Il constitue depuis des millénaires un lieu de rencontres et de mélanges, un creuset culturel et linguistique, où se sont succédé des populations pré-indo-européennes mal connues, puis des ethnies indo-européennes appartenant à diverses branches, qui y ont coexisté de façon plus ou moins pacifique à travers les siècles. Dresser un bilan est pour le linguiste un réflexe tout naturel devant une situation si complexe.

Il constate d'abord que ce ne sont pas moins de sept entités linguistiques qu'on y dénombre : s'y rencontrent en effet le *roumain*, l'*albanais*, le *serbo-croate*, le *macédonien*, le *bulgare* — ces trois dernières appartenant à une même famille slave —, le *turc* et le *grec*. Avec des interpénétrations : îlots « aroumains » en Grèce, macédo-roumains en Bulgarie et en Macédoine, istro-roumains en

Istrie (NO de l'ex-Yougoslavie), albanais en Grèce et en Macédoine (le Kossovo), etc. En outre, les appartenances linguistiques sont diverses : le roumain est une langue romane, donc indo-européenne, comme l'albanais et le grec qui pour leur part constituent des langues isolées ; sont également indo-européennes les langues slaves susnommées, parmi lesquelles le bulgare offre la singularité de reposer sur un substrat peut-être ouralo-altaïque (les "Proto-bulgares") ; reste le turc, authentiquement altaïque quant à lui. Peut-être pourrait-on ajouter à ce tableau le *tsigane*, disséminé dans la Péninsule, et qui est lui aussi indo-européen.

Or, en dépit de cette diversité d'origine, ces idiomes parlés dans les Balkans offrent des traits communs, diversement partagés, qui ont intrigué les linguistes. Ils ont fait l'objet de deux ouvrages principaux, à savoir *Linguistique balkanique, problèmes et résultats*, publié à Paris en 1930 chez Champion par Kristian Sandfeld, et *Die Balkansprachen*, publié à Heidelberg en 1975 chez Carl Winter par Helmut Wilhelm Schaller.

Le premier traite en premier des mots d'emprunt, puis des concordances inter-balkaniques en dehors du lexique, confrontant d'abord les langues deux par deux et traitant ensuite des concordances générales en dehors du lexique en énumérant 13 items essentiellement morphosyntaxiques. Quant au second, après un développement général sur les langues balkaniques, les principaux traits de leur structure actuelle et de leur développement historique, il énumère les « balkanismes » touchant successivement la phonétique, la morphosyntaxe, la syntaxe *stricto sensu* et le lexique.

I. Nous aborderons en premier le volet *phonético-phonologique*. On constate que le *vocalisme* offre généralement trois degrés d'aperture répartis en deux ou trois séries : palatale/vélaire ou palatale étirée/vélaire étirée/vélaire arrondie. D'où les trois schémas suivants :

A. Type gréco-serbo-croate	B. Type bulgare	C. Type roumain
[i] [u]	[i] [ɨ] [u]	[i] [ɨ] [u]
[e] [o]	[e] [o]	[o] [e] [e]
[a]	[a]	[a]

Dans le type B, [i] correspond à « ѣ » du bulgare et à « ë » de l'albanais (tosque) ; ce modèle vaut également pour le macédonien et certains dialectes serbes. Le modèle (daco-) roumain est aussi celui du macédo-roumain. La présence de ce type vocalique vélaire étiré, très caractéristique des langues balkaniques comme on le voit, est également propre à la plupart des langues du domaine slave (russe « ы », polonais et vieux-tchèque « y ») et rappelle d'assez près le « ı » (« i sans point ») du turc. Tout comme la présence d'une série palatale arrondie, partielle ou complète, dans des idiomes européens occidentaux ([y] en français et en occitan, dans des parlers rhéto-romans (Engadine) et « gallo-italiens », en haut-et bas-allemand, en néerlandais, en breton, jadis en gallois et en vieil-anglais, etc. ; [ø] et [œ] dans certaines de ces langues, dont le français), on peut se demander s'il s'agit d'un phénomène de substrat, de propagation, de « génération spontanée » ou de pur hasard... La coïncidence géographique, quelque imparfaite qu'elle soit dans les deux cas, ne laisse pas d'être à tout le moins surprenante.

Du point de vue du *consonantisme*, dont l'originalité est moins marquée, on enregistrera néanmoins deux faits notoires :

1. En premier lieu, des phénomènes de *palatalisation* qui, s'ils affectent en roumain proprement dit non seulement les consonnes vélares ou dentales suivies de [j] ou les vélares suivies d'une voyelle palatale, comme dans les autres langues romanes, mais aussi les occlusives dentales, la latérale [l] et la sifflante [s], touchent en outre dans les dialectes daco-roumains du NE — qui palatalisent les dentales également devant [e] — la vibrante [r] ainsi que... les labiales [p], [b], [f], [v] et [m], auxquelles viennent se substituer [s] et [z] ! Ces processus prolongent en fait ceux qui, d'une façon plus générale, ont affecté l'ensemble des langues slaves : on se rappelle d'abord que celles-ci appartiennent au groupe « satem », où les vélares indo-européennes se sont palatalisées à date ancienne devant voyelle palatale, par opposition à celles du groupe « centum » où ce processus n'a pas eu lieu ; mais à partir de l'époque du slave commun se sont développées trois autres « crises » de palatalisation qui ont culminé en russe et en polonais par exemple, marquant également, dans une moindre mesure, certains dialectes bulgares et macédoniens — mais allant jusqu'à affecter quelque peu — par « contagion » ? — des parlers grecs de Macédoine

2. En second lieu, la présence des *fricatives interdentes* [θ] et [ð] dans les systèmes respectifs du grec moderne (« θ », « δ »), de l'albanais (« th », « dh ») et de l'aroumain (enclavé dans l'aire grecque : le phénomène concerne essentiellement des emprunts lexicaux au grec) qui forment de la sorte une aire continue.

II. En matière de *morphosyntaxe*, sept items peuvent être énumérés, où les correspondances sont tantôt générales, tantôt limitées :

1. La *coalescence formelle du génitif et du datif* en bulgare, macédonien, albanais, roumain et grec, contrairement au serbo-croate où cette distinction se maintient. Dans un cadre plus large, on enregistre une très nette tendance balkanique à passer du synthétisme à l'analytisme, contrastant avec le maintien d'une riche flexion substantivo-adjectivale dans les langues du pourtour, serbo-croate, hongrois, russe ou turc ; cette tendance aboutit soit à la polyvalence des formes, soit au passage à l'analytisme avec emploi de prépositions ; ainsi, parmi les langues slaves, tantôt, comme en serbo-croate, l'idiome est resté foncièrement de type synthétique, vieux modèle indo-européen, mais avec une nette tendance à l'analytisme, tantôt, comme en bulgare et en macédonien, il est devenu essentiellement analytique, mais présente d'importants vestiges de synthétisme. Quant aux causes du phénomène, étant donné qu'il est peu vraisemblable qu'une seule langue ait déclenché le mouvement, et que d'autre part on ne voit pas quel substrat aurait pu servir de modèle (le proto-bulgare était plus synthétique, et quel substrat "balkanique commun" pourrait être invoqué ?), on ne peut qu'en conclure que c'est sans doute le grec, nettement orienté vers l'analytisme au cours d'une évolution historique parfaitement connue, et surtout le roumain, devenu analytique relativement à la source latine comme les autres langues romanes (encore que dans une moindre mesure), qui ont servi d'exemple aux langues voisines.

2. L'existence d'un *vocatif* dans la flexion nominale : c'est le cas en bulgare, en macédonien, en serbo-croate et en roumain pour les deux genres, pour les seuls masculins en -o- en grec, l'albanais ignorant ce trait. Il peut s'agir d'une influence slave, mais en tout état de cause le fait est indo-européen (cf. le grec, où le vocatif φίλε de φίλος est en usage depuis l'Antiquité). Quant au vocatif

en *-o* des substantifs féminins roumains en *-ă* il s'agit de toute évidence d'une influence relativement récente exercée par le slave.

3. La *postposition du déterminant-article au substantif* caractérise le roumain (*om-omul, casă-casa*, le bulgare (мъж-мъжът, жена-жената, перо-перото), le macédonien et l'albanais (*shok-shoku, vajzë-vajza, mish-misht*). Cette disposition — que partagent du reste également des langues scandinaves, l'arménien et le basque ! —, caractérise ici une aire compacte d'idiomes aux appartenances linguistiques différentes, et ne saurait trouver d'explication que dans quelque action d'un substrat de type « thrace », car on voit mal à partir d'où une propagation serait imaginable. Ce substrat, qui aurait favorisé en proto-roumain la tournure latine *homo ille* au détriment de *ille homo* « occidental », serait en albanais d'essence daco-mysienne (?), d'essence dace en roumain.

4. Le roumain, le bulgare, l'albanais, le grec — et même le turc — ont un *comparatif* et un *superlatif analytiques* du type fr. (*le*) *plus grand* face à lat. *maior maximus*, allem. *größer - der größte*, etc. : bulgare по-добър, най-добър, albanais *më bukur, shumë bukur*, roumain (*cel*) *mai bun*, grec (ὁ) *πὸ καλός* « meilleur, le meilleur ». Phénomène « classique » pour un romaniste, la substitution de l'analytique au synthétique, qui rappelle notre premier thème abordé, est un processus banal. Ici, le caractère compact de l'aire concernée suggère clairement que son point de départ peut être cherché précisément dans le roumain, qui à l'instar des autres langues romanes a délaissé le système des comparatifs et superlatifs synthétiques du latin classique au profit de tournures construites au moyen d'un adverbe — cf. déjà en latin *magis arduus* face à *doctior*, etc. —, et sans doute dans le grec où les anciennes formes, encore usitées — *καλύτερα* « mieux » — tendent à se figer.

5. Un autre schéma apparaît dans la *formation des numéraux de 11 à 19* : ici, l'addition des unités à dix s'effectue selon le modèle slave par le biais de la préposition *na* (на) « sur » : ainsi, à bulg. *единадесет, дванадесет, тринадесет*, etc., macéd. *единаесет, дванаесет, тринаесет*, etc., serbo-cr. *jedanaest, dvanaest, trinaest*, etc., correspondant roum. *unsprezece, doisprezece, treisprezece* etc. et alban. *njëmbëdhjetë, dymbëdhjetë, trembëdhjetë*, etc., contre le grec *ένδεκα, δώδεκα, δεκατρείς-δεκατρία*, etc. Le point de départ slave, remontant au vieux-slave, est ici manifeste, d'autant

plus que plus au nord le hongrois en a également hérité : *tízenégy*, *tízenkettő*, *tízenhárom*, etc. (*tíz* = « dix », *en* = « sur », mais l'ordre des composants est inversé).

6. En matière de morphologie verbale, à la notion de *futur*, très abstraite, se substituent fréquemment dans l'histoire des langues des formules plus « vécues », recourant aux concepts de *volonté*, *devoir* ou *aptitude*, à moins que son expression se transpose dans le jeu aspectuel. Or, tandis que ce dernier caractérise des langues slaves telles que le russe ou le polonais, c'est la formule précédente qu'ont adoptée celles du groupe balkanique : « je travaillerai » se dit ainsi en grec θὰ δουλεύω (duratif) δουλέψω (ponctuel) — θὰ provenant de θέλω (ou θέλει, « il veut », figé) νὰ « je veux que... », donc « je veux que je travaille » —, en bulgare ще работя, en macédonien ќе работам, formules qui recourent au même auxiliaire « vouloir » (bulgare хотѣти) sous des formes réduites comme en grec, tandis que l'albanais se partage ici entre son expression tosque (dialecte méridional, officiel) équivalente *do të punoj* (*dua* « vouloir ») et son expression guègue (nord) *kam më punoj*, litt. « j'ai à travailler », que le serbo-croate peut soit employer le même auxiliaire que le bulgare, soit recourir à l'aspect, et que le roumain emploie couramment la formule *o să lucrez*, *o* représentant un héritier « suractivé » de *VOLEŌ (cl. VOLŌ), mais connaît aussi la variante *am să lucrez* « j'ai à travailler ». On voit de la sorte comment, ignorant les solutions aspectuelles propres au slave plus septentrional, les langues balkaniques ont toutes suivi ici deux modèles essentiels, le modèle grec avec l'auxiliaire « vouloir », formule majoritaire (E-SE), et secondairement le modèle latino-roman avec « avoir à » + IF (N-NE).

7. Enfin, un dernier phénomène mérite mention, la *perte de l'infinitif* dans la flexion verbale. Il a disparu en tant que tel en grec, en macédonien et en bulgare, et n'a gardé que quelques emplois en albanais, roumain et serbo-croate. En grec, certaines des anciennes formes sont restées figées comme substantifs neutres ; on citera ainsi τὸ φαγί « la nourriture, le repas », de l'ancien φαγεῖν, infinitif aoriste de ἐσθίω ἔφαγον, ou encore τὸ φιλί « le baiser », de φιλεῖν, « aimer », infinitif prés. ; de son côté, par exemple, le roumain distingue deux infinitifs, l'un « court » qui sert à former le futur, ainsi *cânta* dans *voiu cânta* « je chanterai », l'autre « long » qui joue le rôle de substantif : *cântarea* « le chant ». A se placer au

niveau dialectal, on s'aperçoit du reste que l'effacement de ce mode s'accroît à mesure que l'on se rapproche du foyer grec : le serbe l'ignore — *хочу да радим* —, non le croate — *hoću raditi* ; en roumain, le NE de l'aire se montre beaucoup plus fidèle à son emploi que le reste. La tournure qui se substitue à lui se fonde soit sur l'emploi de l'indicatif précédé des conjonctions *че* ou *мо* en bulgare, *që* ou *se* en albanais, *πὼς*, *πὸὺ* ou *ὅτι* en grec, *că* en roumain pour des constatations ou affirmations — « je le vois venir » —, ou bien *да* en bulgare, *të* en albanais, *và* en grec et *să* en roumain pour des tournures exprimant une volonté, un souhait — « je veux le voir », soit sur la parataxe des deux éléments, comme dans la tournure bulgare *той започва и работи* « il commence à travailler », litt. « il commence et il travaille ». L'histoire du grec nous indique que dès les premiers textes l'on voit la construction finale avec *ἵνα* remplacer l'infinitif, phénomène qui s'accroît avec le temps et caractérise plus particulièrement la langue populaire, surtout à partir du XI^e siècle. L'origine grecque du phénomène est rendue encore plus vraisemblable par le fait qu'en Italie du sud, au contact des îlots hellénophones, il commence à affecter les parlers romans locaux — néanmoins, certains verraient plutôt une source latino-romane à la base.

III. Peut-être peut-on réserver à la *syntaxe stricto sensu* trois autres faits de concordance :

1. L'*anticipation de l'objet direct* par un pronom anaphorique, d'où un *redoublement* de cet objet, comme dans le grec (τὸ) βλέπω τὸ βιβλίον « je (le) vois le livre », cf. le bulg. а Герго хич муха не (го) ухапала « la mouche ne (l')a pas du tout piqué Gergo », alban. (e) urdhëroi të mbesën « il (l')ordonna qu'ils restent », roum. (il) duce pe fiul său la școală « il (le) conduit son fils à l'école », macéd. му (го) подаде жакет « il m(e) l'a tendue la veste ». Quant à l'origine du phénomène, certains y verraient un moyen de compenser la confusion formelle des cas évoquée en II-1 — encore que l'on puisse noter un certain parallèle en occitan et en ibéro-roman, puisque dans ces domaines le phénomène se produit lorsque l'objet représente un être humain : castillan *lo he visto a Pedro/a tu padre*, occitan *l'ai vist a Pèire/a ton paire* (mais dans ce cas l'« anticipation » est obligatoire).

2. L'emploi du génitif du pronom personnel pour exprimer la possession : il s'agit d'une forme spécifique des rapports entretenus par le pronom personnel et le possessif correspondant ; de toute façon, l'histoire des langues indique que $\mu\acute{\epsilon}$ et $\acute{\epsilon}\mu\acute{o}\varsigma$, $m\bar{e}$ et *meus*, *mich* et *mein*, *me* et *my*, $\mu\eta\epsilon$ et $\mu\omicron\upsilon\iota$ sont étroitement liés du point de vue étymologique. L'expression de la possession par le pronom personnel au génitif n'est donc qu'une façon de réactiver, renouveler en quelque sorte ce rapport. Tandis que le grec ancien pouvait déjà dire \acute{o} $\pi\alpha\tau\acute{\eta}\rho$ $\mu\omicron\upsilon$ $\acute{\epsilon}\rho\chi\epsilon\tau\alpha\iota$ comme \acute{o} $\acute{\epsilon}\mu\acute{o}\varsigma$ $\pi\alpha\tau\acute{\eta}\rho$, annonçant donc le moderne $\acute{\epsilon}\rho\chi\epsilon\tau\alpha\iota$ \acute{o} $\pi\alpha\tau\acute{\epsilon}\rho\alpha\varsigma$ $\mu\omicron\upsilon$, le bulgare, le macédonien et, dans une certaine mesure, le roumain suivent ici le modèle grec : bulg. $\kappa\acute{\upsilon}\sigma\eta\tau\alpha$ $\mu\upsilon$ « sa maison », macédon. $\mu\omicron\upsilon$ « sa voix », et même roum. $\acute{\imath}$ *ara-mi* « mon pays », *ochi- $\acute{\imath}$ i* « tes yeux » (des dat. *mihim \bar{i} tibi*). En bulgare, le fait est ancien puisque l'on rencontre en vieux-bulgare (vieux slave) l'emploi de formes de datif des pronoms personnels pour exprimer la possession (cf. en français « une cousine à moi... » évitant « une mienne cousine » ?) ; voir la confusion précitée du génitif et du datif... En tout état de cause, s'il est impossible d'identifier sûrement la source du phénomène, au fond banal — cf. le basque *nire etxea* « ma maison », litt. « de-moi la-maison » —, on se contentera de souligner que, comme dans les cas précédents, l'aire balkanique qu'il couvre est compacte, ce qui semble exclure le hasard.

IV. En ce qui concerne la *lexique*, on distinguerait souvent fort mal entre faits de commun substrat, choix identiques et emprunts. Aussi nous contenterons-nous de mettre ces derniers en relief, car nous en connaissons la source et le cheminement. Dans un tel monde, les contacts sont permanents et quasi-général le plurilinguisme. Or le lexique est le domaine le plus perméable de la langue.

Nous tâcherons de préciser ici les sources, les domaines lexicaux, les périodes et les points d'impact — mais sans donner des exemples qui grossiraient inutilement notre contribution tout en restant en deçà d'un choix assez éclairant. Le grec a servi de source au bulgare d'abord par la voie religieuse — traductions de textes en vieux slave aux XII^e et XIII^e siècles —, plus tard, mais dans son prolongement, en matière scientifique ou « savante », parfois par l'intermédiaire du turc, puis au roumain par l'intermédiaire du slave, du VII^e au X^e siècle d'abord, ensuite à l'époque

« phanariote »¹, au *bulgare* et au *serbo-croate* dès l'époque byzantine (XIV^e siècle), plus récemment l'*albanais* dans tous les domaines du lexique, du plus quotidien aux abstractions ; le *turc* a fourni de multiples éléments partout, du fait de l'extension de l'Empire ottoman, pendant des siècles, de l'Asie mineure à l'Europe centrale : surtout dans le *bulgare* voisin, de l'époque où se rencontrèrent et se fusionnèrent, aux VII^e-VIII^e siècles, les « Proto-Bulgares », d'origine altaïque comme les Turcs, et les Slaves qui parlaient la forme première de leur langue, le *vieux slave*, destinée à l'emporter sur celles des susnommés avec la christianisation du royaume ; ensuite en *albanais*, où le vocabulaire contient environ 1 700 termes d'emprunt, moins en *roumain* (vocabulaire quotidien), et moins encore en *grec* (vocabulaire militaire, de la vie urbaine, du costume) ; le latin, langue importée et longtemps présente (comme langue officielle) dans l'Empire byzantin, alimenta largement le lexique *albanais*, dans les registres les plus divers, jusqu'à y introduire certaines prépositions — on a pu évoquer l'image d'une « langue mixte » romano-albanaise ! —, et aussi le lexique *grec* jusqu'au VI^e siècle (invasions barbares en Occident), sans oublier celui des langues voisines, *macédonien* et *serbo-croate* ; enfin, bien évidemment, le *vieux slave* servit de source pour l'*albanais* et le *roumain* du VI^e au IX^e siècle (on peut imaginer alors chez les peuples concernés un certain statut de bilinguisme), surtout dans les registres de la vie urbaine et de l'agriculture, son influence étant encore plus nette, dans le second cas, en *macédo-roumain* et en *istiro-roumain* — voir la géographie —, moins en *aroumain*, tandis que le *grec* lui-même compte environ 300 emprunts slaves, surtout dans le domaine agricole, et particulièrement dans les parlers septentrionaux.

À titre de bilan, on en conclura que c'est dans un « noyau central » — *bulgare*, *macédonien*, *albanais* et *roumain* — que s'observent les correspondances les plus nombreuses. Un second niveau inclurait le *grec* et le *serbo-croate* (vocalisme, pas d'article postposé ni de numéraux du type « un sur dix »). Enfin, tandis que le *turc* reste à part et n'a servi que de source lexicale (avec les faits

1. Le « Phanar » — τὸ φανάρι « le phare » — était un quartier de Constantinople habité par une bourgeoisie grecque riche et libérale qui joua aux XVII^e-XVIII^e siècles un grand rôle dans les affaires, la culture et la politique turques.

culturels correspondants), *slovène* et *hongrois* appartiennent à l'Europe centrale, non aux Balkans. Historiquement se distinguent trois périodes : 1. Au cours du 1^{er} millénaire, après que grec au sud, latin au nord se sont partagé le rôle de langues de communication et de commerce, la scission en deux de l'Empire romain en 395, puis l'invasion slave des V^e-VI^e siècles fondent la « balkanité » linguistique ; 2. Du XI^e au XVII^e siècle se développent les « balkanismes », sensibles en bulgare dès le XI^e siècle, en albanais et en roumain aux XVI^e et XVII^e siècles ; 3. Depuis le XVII^e siècle, cette « balkanité » stabilisée s'effrite, tandis que va s'effaçant l'influence turque.

Ainsi, le « fait balkanique » est ici particulièrement patent, même si les populations péninsulaires partagent bien d'autres traits de la vie sociale et culturelle. Quant à l'origine du phénomène, s'il apparaît désormais comme oiseux de la faire remonter à quelque « substrat » dont nous ne savons rien, ou si l'on ne sait s'il faut attribuer cette spécificité aux Grecs qui ont joué depuis si longtemps et si largement dans la région un rôle civilisateur et dominateur, ou les exclure au contraire radicalement et situer plus au nord, dans les Balkans proprement dits, son foyer d'irradiation, la clé de l'énigme appartiendra sûrement aux chercheurs qui, les conflits locaux une fois apaisés, mèneront une étude pluridisciplinaire dans cet étrange, difficile et passionnant domaine, où chaque ethnie a pu, à travers les siècles, apporter sa contribution.

*Université de Toulouse-Le Mirail,
UFR de Lettres modernes - CLID
(Centre de linguistique et de dialectologie)*