

Orthographe – Ecriture – Transcription Réflexions à partir du cas corse

Jean-Baptiste Marcellesi
Christiane Marcellesi
Université de Rouen

L'émergence du corse comme langue identifiée sur critères sociolinguistiques est relativement récente. En outre, dans les faits, malgré les tentatives unifiantes, aucun corsetage normatif, jusqu'ici, ne s'est imposé et n'a effacé la diversité linguistique interne. Il en va de même de son écriture fondée sur une adaptation transcriptive du système graphique italien. On peut ainsi analyser la démarcation de l'écriture pour reconnaître les différences linguistiques qu'elle s'attache à mettre en vedette (indicateurs scripturaux) et d'autres qu'elle tend à négliger et qui ne seront donc que des indicateurs oraux. En outre, s'agissant d'une écriture où le principe transcriptif domine, les mêmes processus se remarquent au niveau des variétés géographiques internes, les écarts étant à leur tour mis en exergue le plus possible (démarche différenciatrice) ou minimisés (écriture unifiante). On voit toutefois pointer à nouveau, depuis peu, notamment chez ceux qui n'ont pas eu le corse comme langue maternelle, une volonté de normalisation réductrice ou de création « langue de référence », avec, même, le désir de faire comme Richelieu : créer une « académie du corse ».

Nous avons prononcé cette communication au Colloque international de Tizi Ouzou en 2000. Les Actes de ce colloque n'ayant jamais paru, nous croyons que sa teneur doit être offerte à l'intelligence et à l'amour de la langue sous toutes ses formes que manifestait si bien notre Bernard qui n'a pas eu l'occasion d'exercer sa critique décapante en cette occasion.

Avant d'aborder le problème de l'écriture du corse et des réflexions qu'on peut en tirer sur le plan général, des rapports existant entre un système scriptural et l'auto-définition linguistique d'un groupe donné, il est nécessaire d'éclairer le phénomène d'émergence du corse comme langue identifiée sur critères sociolinguistiques. Comme l'a dit Marcel Cohen à propos du français, on n'a pas toujours, en Corse, parlé corse. D'abord, parce que, contrairement au berbère en Afrique du nord, dont on sait qu'il était « toujours déjà là » au début de l'histoire, l'ancêtre du corse, le latin, (si tant est, qu'on ne lui donne pas un ancêtre plus récent, le toscan, base de l'italien) ne peut pas avoir été introduit en corse avant la mainmise de Rome sur cette île, durant le III^e siècle avant notre ère.

Il est peu probable que les conquérants antérieurs, Etrusques, Grecs, Carthaginois aient imposé leur langue. On ne sait donc rien de ce que parlaient les populations de l'île : “Mégalithiques”, puis “Torréens”. En plus, on ne sait pas si la romanisation progressive, certainement avancée dans les derniers siècles de l'Empire Romain, a subsisté après les invasions vandale, byzantine et surtout sarrasine. Si bien qu'on peut opter pour l'idée que la romanisation a subsisté pendant cette période même si, par la suite, lors de la domination pisane, elle a été renforcée par un superstrat toscan, surtout

dans le nord de l'île ; ou au contraire que c'est la domination pisane qui a romanisé une seconde fois une île où la latinité ne subsistait plus que très partiellement. Nous n'avons pas les moyens de trancher sur ces problèmes de bouteille à moitié vide ou à moitié pleine.

En tout cas, au début du IIe millénaire de notre ère, commence pour la Corse une longue diglossie corso-toscane qui sera relayée progressivement par la diglossie et /ou le bilinguisme corso - français.

Aussi longtemps que la langue italienne ne sera pas relativement normalisée, on ne peut pas dire si les textes insulaires que nous avons (ceux des chroniqueurs ou des actes officiels) sont en corse toscanisé ou en toscan avec des corsicismes. Tout ce qu'on peut constater, c'est que la référence de ceux qui font des études est le toscan, comme du reste, dans une grande partie de la péninsule (faut-il rappeler que l'Etat italien n'existe que depuis le milieu du XIXe siècle, sous la férule du roi de Piémont-Sardaigne) et que, vers 1950, selon Tullio de Mauro, on constatait que seuls 5 ou 10% des Italiens avaient comme langue maternelle, la langue de cet Etat ?). La diglossie corso-italienne est manifeste pendant la brève indépendance de l'île entre 1729 et 1768. Les textes et les débats nombreux que nous avons de cette époque, rédigés par les chefs corses, sont tous en italien, et le concurrent de l'italien ne semble pas être le corse, mais le latin qui occupait une place comme langue savante et théologique à l'Université de Corse, instaurée par Pascal Paoli.

Le corse ne commence à apparaître, dans les écrits, que sous la forme de variété B, réservée notamment aux genres et aux personnages du comique ou du burlesque. C'est la manifestation habituelle de la diglossie. Ainsi l'identité linguistique corse repose sur un phénomène de distanciation, d'individuation qui s'affirme à partir du milieu du XIXe siècle, c'est-à-dire à partir de la naissance d'un Etat italien, par distanciation avec la langue de ce nouvel Etat.

D'où l'utilisation du corse dans divers genres littéraires, et surtout, l'affirmation de l'existence d'une « langue corse » qui n'est pas l'italien, même si les uns consentent quand même à reconnaître la proximité avec ce dernier, alors que d'autres prétendent que le corse se rapproche plus d'autres langues (parfois même non-romanes) à partir de tel ou tel rapprochement ponctuel.

Il doit être bien clair que s'il existe une langue corse, c'est à l'intégration de l'île dans la France que cela est dû. Si la Corse avait été intégrée à l'ensemble politique italien, l'identité linguistique ne serait jamais apparue et le corse aurait été assimilé, avalé par la « Botte ».

C'est avec l'émergence d'une conscience d'identité linguistique que s'est développée, à partir du système graphique italien, une écriture identifiant des différences qui font que tel texte est corse et non italien. Il y a donc eu toute une activité scripturale différenciatrice par rapport à l'italien. Il faut noter que cette individuation se fait de langue à langue. Le corse est proclamé langue différente de l'italien et, par conséquent, peu importe qu'il y ait des dialectes de la péninsule qui aient les mêmes variations que le corse. La distanciation se fait par rapport à la langue de l'Etat italien. Certains Corses du nord de l'île ont essayé pendant un certain temps d'appliquer la même exigence de distanciation vis-à-vis du sarde, en faisant la guerre aux points communs entre variétés sudistes et sardes. Jusqu'au début des années 70, les Sudistes qui envoyaient des textes avec -dd- (notant le d cacuminal) les voyaient publiés selon les mots avec -ll- ou -gli-. En même temps le rêve d'une unification au moins de l'écrit corse, d'un corsetage normatif, de création d'une orthographe identique pour toute l'île s'est écroulé. Dans le dernier quart du XXe siècle, la pluralité du corse a été admise et théorisée sous le terme de « langue polynomique ». Chacun, donc, prend des règles d'écriture commune ou plutôt de transcription et transcrit son oral : c'est le principe dont nous verrons qu'il a eu le

mérite d'éviter les conflits campanilistes et a permis de revendiquer, pour tout de suite, l'enseignement de la langue, dans sa diversité, sans attendre que les Corses fassent, en quelques années, ce que la France a mis un millénaire pour le faire. L'enseignement d'abord associatif, puis scolaire, à partir du moment où, en 1974, le corse a été admis au bénéfice de la loi Deixonne sur l'enseignement des langues régionales (qui datait de 24 ans plus tôt), a progressé sur ces principes. Quand, au début des années 90, le CAPES (concours de recrutement des professeurs de collèges et de lycées) a été créé pour le corse, le jury a posé « qu'aucune variété ne sera favorisée ». La proposition de création d'une « Académie du corse » battue en brèche alors, vient certes de réapparaître, en même temps que la revendication de l'obligation du corse, aussi bien dans la prise de position, en mars 2000, de la majorité territoriale décentralisatrice, que dans celle de la minorité autonomiste ou indépendantiste. Il appartiendra à l'histoire de mesurer les efforts collatéraux positifs ou négatifs de ces points s'ils sont adoptés définitivement.

Le corse a donc comme base d'écriture le système graphique italien. En principe, les lettres ou les groupements de lettres correspondent aux mêmes sons. La spécification se fait par la création d'autres combinaisons pour transcrire des sons que l'italien n'a pas ou ne transcrit pas (les considérant ainsi comme des déviances). Des enquêtes systématiques ont permis d'établir une liste de 12 traits pan-corses qui distancient le corse de l'italien. Cela paraît peu, mais, dans la réalité, leur fréquence est telle qu'on ne peut guère écrire un mot sans que la corsité apparaisse. Parmi ces traits, 6 sont morpho-syntaxiques et donc ne posent pas de problèmes particuliers d'écriture. Nous les citons pour mémoire et pour permettre aux berbérophones de rêver (comme du reste pour les 6 autres) à des coïncidences intéressantes.

Donc tous les signes et rien que les signes d'une machine à écrire. Si l'accent tonique qu'on est souvent obligé d'indiquer, vu son importance pour le sandhi, est facile à noter pour è et à ; c'est beaucoup plus difficile pour e, o, u pour lesquels on peut utiliser l'accent circonflexe. Aucun moyen non plus de noter directement les différences de timbre entre o fermé et o ouvert, entre e fermé et e ouvert. On peut jouer là sur le redoublement des consonnes, mais cela a des inconvénients (il y a une opposition phonologique entre ce qu'on note par des consonnes doubles et ce qu'on note par des consonnes simple = opposition fortes /douces). Les accents du clavier français, non utilisés en italien, sont utilisés en corse, mais pour l'accent tonique.

Les points qui sont sentis comme identificateurs pan-corse sont les suivants :

- l'adverbe affirmatif *iè*, ce qui a permis à Mathée Giacomo de désigner le corse comme « langue d'*iè* » ; à noter la simple approbation hé, comme en kabyle ;
- la forme vocative o + nom, le nom étant débarrassé de la partie post-tonique ;
- le système pronominal très simplifié du point de vue morphologique ;
- l'utilisation – dans les subordinées – du pronom sujet comme simple morphème de conjugaison ;
- la construction syntaxique de “ l'accusatif ” des noms propres avec la préposition à ;
- la constitution d'un « futur proche » composé de la suite « auxiliaire autonome avè (« avoir ») + à + infinitif », à côté d'un futur ordinaire, de même composition que dans les autres langues romanes (infinitif + forme non-autonome dérivée de habere, en corse *avè*) ;
- la distinction morphologique des interrogatifs et des conjonctifs.

Pour les six traits phonétiques :

- la finale des mots en –u en face de –o italien ; sans qu'on puisse dire que le –u n'est pas le maintien de la finale latine ;

- la phonématisation d'une série de consonnes médio-palatales notamment [tj], écrit chj et [dj], écrit ghj ;
- l'apophonie des voyelles intérieures non toniques qui fait que dans cette position, tendanciellement, les timbres vocaliques sont réduits à 3 ([a], [i], [u]), (sauf restitution étymologique de la voyelle radicale) ;
- le sandhi des consonnes initiales suivant une finale de mot atone (sandhi généralisé dans le Nord et limité à une partie seulement des consonnes dans le Sud) ;
- l'article défini sans consonne initiale (réduit donc à la partie vocalique), un [l] apparaissant à l'initiale dans le style élevé ou comme consonne euphonique ;
- l'infinitif débarrassé du –re final étymologique.

D'abord il faut noter que si l'on ne pose pas l'existence d'une volonté distanciatrice, serait-elle inconsciente, on pourrait écrire toutes ces formes « à l'italienne » et les prononcer “ à la corse ” (comme la finale –o- prononcée –u- en portugais, la suite –ou- notant un o ; ou la lettre latine –u- finissant par noter en français une voyelle centrale, distincte de la voyelle d'arrière notée –o-, l'attribution, en français, aux lettres notant des post-palatales latines c et g, devant i et e, la prononciation chuintante etc.).

Deux batailles ont marqué ces questions, la première concerne l'utilisation et la mise en pratique de la notation des nasales, au début du XXe siècle.

Les nationalistes italiens développent, dès cette époque, le mot d'ordre des *terre irredente*, des régions qui n'auraient pas été « rachetées », c'est-à-dire libérées... et annexées à l'Italie. C'est le véritable irrédentisme bien que le mot soit aujourd'hui très souvent employé pour désigner toute revendication linguistique intransigeante ou irréductible. Les implications sont considérables, qu'on en juge : après avoir établi l'*Atlas linguistique de la France*, Jules Gillieron et son enquêteur Edmond Edmont s'attaquent, dans la première décennie du XXe siècle à l'*Atlas Linguistique de la Corse*. Edmond Edmont a cru (à raison) déceler dans l'articulation corse des voyelles devant consonne nasale en syllabe fermée une résonance nasale sur la voyelle. C'est un fait articulatoire assez banal, même ailleurs. Mais les cartes commencent à être publiées alors que la guerre de 1914 a éclaté et que la France et l'Angleterre essaient de détacher, des Empires Centraux, l'Italie ; la cheville ouvrière de cette opération, du côté italien, étant le futur dictateur Mussolini. Une violente campagne à laquelle les linguistes italiens ne sont pas les derniers à participer se déclenche : dans la péninsule, on accuse les auteurs de l'*Atlas linguistique de la Corse* d'être les instruments politiques de la « désitalianisation » de l'île. Gillieron sacrifiera l'Atlas sur l'autel de la patrie et les cartes suivantes resteront dans les cartons. Elles sont encore à la Bibliothèque Nationale.

La deuxième bataille est une bataille intra-corse : c'est celle de la phonématisation de la série de médio-palatales « mouillées », issues soit de dentales, soit de post-palatales latines suivies de –i- ou de –e-. Il s'agit de sons dont le premier élément est une implosion consonantique médio-palatale et l'explosion est celle d'un –j-. Par commodité, nous les appelons médio-palatales. Il s'agit d'un trait fortement identitaire bien qu'il y ait un grand flottement, certains tendant à réintroduire, dans certains mots, les sons italiens correspondants, les autres tendant à faire disparaître entièrement ces derniers. Quand il s'est agi de marquer la distanciation par la graphie, certains ont proposé d'écrire les médio-palatales en utilisant les dentales correspondantes suivies de i + voyelle. On avait donc, dans les années 70, deux écritures pour les médio-palatales : d'un côté t+i ou d+i devant voyelle, de l'autre ch+i et g+i devant voyelle. L'écriture avec t ou d a été qualifiée de « française » : on ne voit pas pourquoi. Dans ce genre de débat, tous les arguments sont bons. L'autre, utilisant ch ou g, supprimant dans l'écriture toute marque identitaire (les mots corses, différents à l'oral, ressemblent à l'italien) était prônée au nom de la tradition insulaire (et pour cause !). Le mouvement vers la démarcation et la spécification

s'est fait quand la graphie par les dentales a utilisé -j- pour noter l'explosion ; par la suite l'utilisation du j dans le système utilisant ch et g et la généralisation de h après g, a donné la notation -chj- pour la sourde et ghj pour la sonore. C'est ce système qui s'est imposé dans l'écriture transcriptive actuelle, quasiment hégémonique. On remarque que, dans ce cas, les lettres comme -j- ou -h- non utilisées (ou très peu utilisées) dans le système italien sont employées en combinaison pour noter des phénomènes spécifiques. Ainsi, pour distinguer les trois *g* du corse, article féminin pluriel transcrit par -e- tout seul, è avec accent grave pour la conjonction signifiant "et" et l'adverbe d'approbation è, enfin hè pour la 3^e personne du singulier du verbe être, ce qui fait que dans ce cas on ajoute la lettre h non étymologique.

En revanche, certains autres traits spécifiques, comme l'affaiblissement des consonnes ou de certains groupes consonantiques initiaux notamment derrière l'article, ce que nous appellerons le sandhi consonantique, ne sont pas notés par l'écriture. Or, là encore, c'est un trait individuateur important : on reconnaît l'italien à ce qu'il ne fait aucun sandhi et le corse de telle ou telle région à ce qu'il en fait certains. Prenons par exemple le mot désignant « le vent » : en initiale absolue, après une pause ou un mot oxyton, la consonne initiale est -v- (comme en français et en italien). Autrement (notamment après l'article "u") la consonne initiale est un -w- écrit -v-.

La raison invoquée pour ne pas noter le sandhi est que les mots à l'initiale, dans le dictionnaire notamment, auraient deux formes différentes. Mais on aurait pu identifier le mot par son initiale faible et poser que les formes à initiale forte s'expliquaient par la position. Il semble que dans ce cas, l'étymologie (se référant soit au latin, soit au toscan, selon les écoles) ait prévalu. Mais il ne s'agit que d'étymologie par l'écriture puisque, en latin, "v" notait un "w".

Parfois, les principes permettent d'adopter une écriture plus ou moins identitaire. Là encore, on va trouver des ressemblances avec le Maghreb, mais nous ne nous étendrons pas là-dessus. A l'intérieur des mots, les voyelles non toniques intermédiaires, -e- et -o- ont tendance à se confondre avec -i- et -u-, si bien que, dans cette position, on se trouve devant un système vocalique réduit à trois sons. Dans les mots composés et dans les paradigmes verbaux, deux tendances contraires s'opposent : d'un côté la prononciation identitaire, de l'autre la restitution étymologique par laquelle les radicaux verbaux gardent la même forme. On aura donc dans ce cas une écriture plus identitaire avec fermeture en i et u des e et o non toniques à l'intérieur des mots et une écriture plus étymologique ne notant pas cette fermeture. Les mêmes tendances opposées se trouvant également dans la prononciation ; on a donc un système relativement libre.

On pourrait aussi parler de l'article : l'alternance d'une forme avec ou sans "l" initial crée une alternance stylistique faisant de la première une variante haute.

Si on passe aux variétés régionales, on rencontre les mêmes problèmes : mise en évidence de certaines différences ou prise en compte de certaines autres. Par exemple, les parlars sudistes seront écrits de manière plus ou moins identitaire selon qu'on marquera par l'écriture les particularités : importance accordée à l'opposition phonématisée consonnes fortes / consonnes douces, système spécifique du pluriel des noms avec la substitution à l'alternance u/a masculin vs féminin, propre au singulier, d'une alternance i/a, masculin et féminin vs neutre, maintien de -i- et -a- brefs latins, passés ailleurs à -e- et -o-, correspondance d'un d noté -dd- avec les formes d'autres régions ayant un l double ou un l mouillé. De même, dans le Nord de l'île où souvent les consonnes intervocaliques se sont affaiblies, on choisira ou non de montrer ce phénomène par l'écriture.

On pourrait de même noter la non prise en compte du renforcement de s derrière consonne en affriquée, par exemple *corzica* au lieu de *corsica* (qui n'est pas seulement phonétique puisque dans les mêmes séquences les corsophones qui ont ce trait en corse ne l'ont pas en français). Non notation non plus des j de transition entre voyelle d'avant et

voyelle d'arrière (phénomène au contraire que les sudistes ont en corse et en français).

On comprend pourquoi la polynomie a été jusqu'ici une planche de salut pour le corse. Selon le degré d'affirmation locale ou de volonté d'intercompréhension, on peut ou tendre vers une écriture localiste ou vers une écriture plus commune. C'est cette souplesse qui permet une grande variété d'expression et un dépassement d'une normativisation qui aurait coupé la langue de ses racines populaires et de ses locuteurs authentiques. Contrairement à ce qu'on imagine, quand on n'envisage comme mode de fonctionnement d'une langue que le « corsetage » de la norme, le corse serait une langue extraterrestre si elle s'imposait une règle académique. Un large consensus existe dans l'île sur ce point. Pourvu que ça dure !