



Cahiers d'Asie centrale

10 | 2002

Karakalpaks et autres gens de l'Aral : entre rivages et déserts

Les parlers karakalpak dans leur contexte

François Jacquesson



Édition électronique

URL : <http://journals.openedition.org/asiacentrale/658>
ISSN : 2075-5325

Éditeur

Éditions De Boccard

Édition imprimée

Date de publication : 1 mai 2002
Pagination : 93-137
ISBN : 2-7449-0191-1
ISSN : 1270-9247

Référence électronique

François Jacquesson, « Les parlers karakalpak dans leur contexte », *Cahiers d'Asie centrale* [En ligne], 10 | 2002, mis en ligne le 28 août 2009, consulté le 20 avril 2019. URL : <http://journals.openedition.org/asiacentrale/658>

Les parlers karakalpak dans leur contexte

François Jacquesson

I. Introduction¹

I.1. Les chemins du lac d'Aral

À l'est du lac d'Aral – ce mot *aral* signifie “archipel” – il existe deux de ces fleuves d'Asie centrale qui, pour une fois, ne se perdent pas dans les sables. Les fleuves d'Asie centrale, comme le Čüj ou le Zarafšan (le fleuve de Samarcande), généralement avortent. Ils tombent des hautes montagnes pour s'épuiser dans des déserts trop longs. C'est pourquoi les grands lacs de cette région, afin de porter au rang de fleuve ces rivières sans issue, sont promus au rang de mer : la mer Caspienne, la mer d'Aral. Le Syr Darya, l'Amou Darya, l'un au nord l'autre au sud, forment avec leur mer d'Aral un trio compliqué dont les niveaux capricieux au cours des siècles, les cours variables au gré des saisons, sont cependant le cœur pulsatile dont dépend la vie de toutes sortes de bêtes, et des hommes.

La fosse occidentale de la dépression de l'Aral une fois remplie, l'eau déborde vers l'est et vers le sud en marais et en lagunes où les réseaux des deux deltas viennent transvaser ce qu'il leur reste d'eau. Dans ces biotopes paradoxaux, poche aquatique en plein désert, sorte de Tchad du Nord, des îles plates proliféraient au milieu des roseaux géants et des herbages foncés marquant la zone transitoire entre barque et chariot. Il est probable que l'homme est ancien ici. Les Karakalpak, gens de pêche, de barques et de roseaux, sont pour une part les héritiers de cultures anciennes et fragiles accrochées comme on peut à ces rivages qu'on ne voit pas (ils sont trop plats, trop pleins d'herbes, et maintenant il n'y a plus d'eau) mais qui produisent incidemment des étangs, des bras morts, des “lacs résiduels” où la fourche et le filet ramènent encore de gros poissons gras.

Comme souvent les cultures fragiles, agrafées sur des milieux naturels encore plus fragiles qu'elles, ne vivent que par les échanges qu'elles font avec les éleveurs alentours, les chasseurs de gazelle de l'Ust-ürt autrefois, les vastes tribus organisées des steppes plus riches et dont certains clans abou-tissaient là. Ces groupes plus forts et plus rapides, gens de cheval, ont imposé depuis longtemps leur langue – le turk – aux rivages de l'Aral, et si autrefois les pêcheurs et cultivateurs des deltas ont eu une langue à eux (comme les Kets de l'Énisséi, ou les Youkaguirs de la Léna d'autrefois), il y a beau temps qu'ils ont appris le turk. Les Karakalpak ne sont, à tout prendre, qu'une forme de ce qu'on appelle pieusement un équilibre entre sédentaires et éleveurs. C'est en réalité un déséquilibre. Ils ne doivent, comme on verra, leur pérennité politique qu'à un accident de l'histoire récente où leur rôle est très modeste. Ils sont davantage un écho qu'une sonorité véritable, un détail dans le tableau plutôt qu'une thématique entière.

Pourtant, nous sommes devenus attentifs aux détails. Ces turkophones atypiques, appuyés sur une mer qui s'enfuit, tributaires des fleuves qui s'assèchent, ont conservé pourtant des parlers à peine différents des autres. Une sorte de singularité minimaliste à quoi ils se sont faits. C'est pourquoi, pour saisir ce que ces parlers représentent, il faudra longuement parcourir les parlers alentour, chercher patiemment ce qui leur reste.

1.2. “Langue”, “dialecte”, “parler”

Les gens qu'on appelait autrefois Karakalpak semblent avoir été assez nombreux, mais il n'est pas certain que ce nom ait toujours désigné les mêmes gens, d'autant qu'ils ont vécu dans des régions diverses. Et il est encore moins certain, sinon franchement douteux, que tous ces divers Karakalpak aient utilisé le parler dont provient celui des Karakalpak d'aujourd'hui². Les parlers karakalpak actuels sont au moins deux, ou l'étaient avant que l'école ne tende à les unifier. Tous deux appartiennent au groupe kipčak des langues turques, dont font aussi partie les parlers kazak et ceux de beaucoup des habitants de l'Ouzbékistan. D'une façon générale, tous ces gens se comprennent entre eux, et les différences dont il va être question dans cet article sont de l'ordre de la nuance.

Contrairement à l'idée que tente de répandre tout état, surtout s'il est nouveau et ambitieux, il n'y a pas nécessairement superposition entre “état” et “langue”. Les langues possèdent des trajectoires historiques complexes du fait qu'elles s'apprennent (ou s'oublient) et se modifient aisément. Tous les gens qui parlent aujourd'hui des langues turques ne sont pas nécessairement “Turks”. On ne peut pas induire d'une communauté linguistique actuelle la certitude d'une longue histoire commune. Le linguiste russe Berezin l'écrivait déjà en 1848, dans un livre écrit en français³ :

« Je crois que dans la définition du caractère distinctif des différents dialectes turcs-tatares, l'origine des peuplades ne doit pas être considérée comme une raison majeure ; dans le mélange des races, des langues, des religions, des

mœurs, des coutumes, qu'avait éprouvé la race turque, en passant à travers les siècles et les événements, il est bien facile de perdre non seulement la physiologie primitive, mais même de changer la langue maternelle. »

Nous allons étudier ici le cas de la "langue karakalpak". Comme toutes les langues officielles, elle est une fiction : il existe seulement des parlers karakalpak, qui sont les façons réelles qu'ont des gens de s'exprimer, et qui ne ressemblent pas nécessairement à ce qu'on écrit dans les grammaires et les dictionnaires. On dit souvent que la langue karakalpak n'est qu'un dialecte de la langue kazak. Mais "la langue kazak" elle aussi est une fiction : c'est un choix officiel parmi les différentes façons de parler qui existent dans les nombreuses communautés de ce qu'on appelle aujourd'hui le Kazakhstan. En réalité, c'est le kazak officiel qui est un dialecte du kazak, un dialecte soutenu par les grammaires et les dictionnaires officiels, certes, mais un dialecte pourtant. La réalité des parlers est celle des gens qui parlent, et non pas celle des frontières politiques.

Les différences d'un parler à l'autre tiennent à la prononciation des mots (c'est ce dont il sera question ici), ou au choix de mots ou d'expressions différents, ou à une façon différente de les arranger ensemble. Ces différences servent aux gens qui parlent comme à ceux qui écoutent à se repérer : en écoutant, on apprend que celui qui ne parle pas exactement comme soi vient de telle ou telle région, ou simplement de tel quartier de la ville, ou de telle partie du village. Dans la mesure où ces différences sont reconnues, elles peuvent être soumises à des pressions, conservatrices ou destructrices.

En théorie, plus on parle pareil, plus on habite près l'un de l'autre : c'est seulement quand on voyage loin de chez soi qu'on trouverait des gens qui parlent une langue qu'on ne comprend pas. Mais dans la réalité, ce n'est pas nécessairement le cas. Car il existe bien des cas de populations déplacées, soit en masse, lors des épisodes douloureux que sont les guerres ou les famines, soit par petits groupes. De sorte qu'on peut trouver près de chez soi des gens qui parlent une langue qu'on ne comprend pas, ou qu'on sent différente de la sienne malgré les ressemblances qu'on entend aussi. Une telle situation, qui est fréquente, n'est pas toujours traumatisante, et peut être au contraire fort bien vécue. En général, cette situation se modifie avec le temps : les moins nombreux, ou ceux qui se sentent plus faibles ou moins arrogants, apprennent à parler comme les autres, du moins en public. Ils peuvent naturellement conserver chez eux, ou entre eux, leur propre parler. Se développent ainsi des situations de bi- ou de pluri-linguisme. La distance n'est pas nécessairement le critère de la compréhension.

1.3. Questions d'orthographe

Les langues turkes ont été, et sont encore, écrites avec des écritures différentes. Elles en ont changé souvent au cours de l'histoire, notamment à l'époque soviétique. En Asie centrale, les premiers textes turks connus sont des inscriptions funéraires gravées sur les monuments de pierre, au moyen

d'une écriture dite "runique" (par comparaison avec les runes Scandinaves, qui étaient connues depuis longtemps). Cette écriture "runique" a été déchiffrée par le savant danois Vilhelm Thomsen en 1893⁴.

À partir des conquêtes musulmanes, l'écriture majeure des langues turques a été l'écriture arabe, tant au Proche et au Moyen-Orient qu'en Asie centrale. L'empire russe n'avait rien changé à cette situation (le savant russe Berezin, en 1848, écrit le turk avec l'alphabet arabe), mais le pouvoir soviétique, après une brève période de latence (jusqu'en 1928 pour le karakalpak⁵), a cherché à promouvoir dans les années 1930 l'écriture latine, jugée "internationale"⁶. Ensuite, devant l'échec de la révolution internationaliste, l'écriture du russe, l'alphabet cyrillique et "national", s'est substitué dans les années 1940 à ces tentatives en alphabet latin. La notation des langues turques en cyrillique, comme nous le verrons dans le cas du karakalpak (section IV.2.), a connu plusieurs phases différentes. Comme l'alphabet cyrillique ne permettait pas de noter des sons propres aux langues turques, on s'est résolu à y ajouter des lettres particulières, ce qui posait des problèmes concrets : la production de machines à écrire et de fontes typographiques.

En outre, les inventeurs des orthographes modernes ont proposé des conventions légèrement différentes selon les langues turques. En kazak et en ouzbek, on différencie dans la graphie la vélaire et l'uvulaire, tandis qu'on ne le fait pas en kirgiz et en türkmen. Les solutions actuelles sont celles-ci :

	sourde		sonore	
	vélaire	uvulaire	vélaire	uvulaire
kirgiz	к		Г	
türkmen	к		Г	
kazak	к	Қ	Г	Ғ
karakalpak	к	Қ	Г	Ғ
ouzbek (cyrill.)	к	Қ	Г	Ғ
ouzbek (latin)	k	q	g	g ^c

Il est vrai que cette question n'est pas simplifiée par les écritures successivement adoptées pour le karakalpak. Il y a eu au moins trois époques de graphies officielles⁷. La première a commencé avec le voyage de Malov (voir II.2. ci-dessous), mais a tourné court parce que Moscou a décidé que le cyrillique était beaucoup mieux que l'alphabet latin que Malov 1934 et Baskakov 1932 avaient d'abord utilisé (et avec lequel il existe quelques rares publications). La deuxième époque correspond à la première entreprise cyrillique, concrétisée dans le *Dictionnaire russe-karakalpak* de 1947, qui ne différencie pas /g/ et /k/ dans l'écriture des initiales, ni d'ailleurs les

voyelles d'avant et d'arrière comme /u/ et /ü/. Autrement dit, on a employé l'alphabet russe sans aucune modification. Le mot "jour" y est orthographié күн [kun]. La troisième époque commence avec la réforme orthographique de 1950 (le grand ouvrage de Baskakov, *Karakalpakskij âzyk*, va commencer à paraître en 1951, avec une phonologie, une grammaire et des textes dans plusieurs dialectes⁸) qui revoie la copie, et donne la graphie actuelle. Celle-ci est normalisée par exemple dans le *Karakalpaksko-russkij slovar'* [*Dictionnaire karakalpak-russe*] de 1958, toujours sous la direction de Baskakov, où l'on trouve күн [kün] pour "jour"⁹. Le grand dictionnaire russe-karakalpak de 1967, et toujours sous la responsabilité de Baskakov, a encore күн [kün]. Le parler de référence reste celui du NE, le plus proche du kazak (voir IV.2.).

Dans une publication sur les langues turkes, on est donc confronté à de nombreux systèmes d'écriture qui, s'ils ne sont pas reproduits tels quels, doivent être transcrits dans l'alphabet de la publication. Deux problèmes surtout se posent : celui des consonnes qu'on appelait autrefois "gutturales", et celui des voyelles. Phonétiquement, comme on vient de le voir, les langues turkes différencient deux positions de l'appareil phonatoire pour la réalisation des "gutturales" sourdes et sonore : une position au milieu de la bouche qui produit les consonnes classiques dites *vélaires* [k] et [g], et une position plus en arrière vers le fond de la bouche qui produit les consonnes *uvulaires* que l'API¹⁰ écrit [q] et [G] ; cette dernière, dans la tradition turkologique, est souvent écrite "ğ" ou "γ" (le gamma grec). En réalité, la prononciation de ces quatre positions dépend de la langue turke considérée. Toutefois, le problème se simplifie du fait que, dans les mots proprement turks ou bien turkisés (donc hors des mots récemment empruntés à une autre langue et / ou dont on essaie de copier la prononciation), les consonnes vélaires ne se trouvent que dans les mots contenant des "voyelles antérieures", tandis que les consonnes uvulaires ne se trouvent que dans les mots contenant des consonnes postérieures¹¹. Les langues turkes dont nous nous occuperons surtout ont quatre voyelles antérieures et quatre postérieures :

postérieures	[q] et [G]	a	ı	o	u
antérieures	[k] et [g]	e	i	ö	ü

Les graphies de ces huit voyelles sont traditionnelles en turkologie, et ce sont aussi les graphies qui sont employées pour écrire le turc de Turquie, ou du moins sa forme normée, orthographique. Pour le "ı", on trouve souvent la graphie "i" qui est moins logique.

Il existe en karakalpak, comme en kazak et d'autres langues turkes, une 9^e voyelle, notée "ə" (e à l'envers, dit aussi "chva") dans l'alphabet cyrillique modifié¹² et le plus souvent transcrite "ä". Cette voyelle se rapproche du "è"

(ou mieux “ê”) français, très ouvert, “à la parisienne”. Il s’agit d’une voyelle empruntée, et qui se trouve dans les mots d’origine arabe ou persane. Par exemple *kānt* “ville”, comme dans “Tachkent”; ce mot *kānt* est un emprunt ancien au sogdien, langue iranienne¹³.

Comme nous nous occuperons essentiellement de mots turks ou anciennement turkisés, il n’y a pas d’inconvénient à écrire nos consonnes vélares ou uvulaires avec seulement les lettres “k” et “g”. Il suffit de se rappeler qu’avec les voyelles postérieures elles ont en principe les prononciations [q] et [G]. On voit souvent les mots *kazak* et *karakalpak* écrits *qazaq* et *qaraqalpaq* : c’est pour cette raison. Nous, nous nous contenterons d’écrire *kazak* et *karakalpak*, et nous éviterons les graphies inutilement compliquées comme *kazakh*, qui sont souvent des transcriptions des mots russes¹⁴ adaptés des mots turks.

Signalons que les mots ouzbeks posent des problèmes spéciaux, parce que dans une partie des parlers ouzbeks – comme nous le verrons plus loin – et aussi dans la norme écrite ouzbèke, le système des voyelles est différent de celui qui vient d’être décrit. Cela est dû à la profonde influence des parlers iraniens sur ces parlers ouzbeks.

Les autres consonnes ne posent pas de problèmes compliqués. Le cas des consonnes dites *fricatives* et *affriquées* peut être résumé dans un petit tableau.

	la sourde		la sonore	
	ici	français	ici	français
fricatives	š	ch	ž	j
affriquées	tš ou č	tch	dž	dj

Comme souvent dans les publications de linguistique, nous utiliserons ici la lettre “j” non pas pour le son de la lettre française, mais pour le son français “y”. Un mot turk écrit *jol* doit donc être lu “yol”¹⁵.

Reste le problème classique des noms qui ont une orthographe plus ou moins accréditée dans les dictionnaires français. J’ai dit plus haut quelle solution j’adoptais pour *Kazak*, *Kirgiz*, *Karakalpak* et d’autres du même genre. Ces mots sont turks, et ne prennent donc pas les accords (féminin ou pluriel) du français. On ne peut pas être aussi rigoureux pour l’ouzbek ou l’ouïgour, qui ont des systèmes phonologiques assez différents. Pour éviter les complications inutiles dans le cadre de cet article, où il sera peu question des parlers *karluk* (les normes officielles ouzbèke et ouïgoure sont des parlers de type *karluk* – ce terme est expliqué plus loin), j’utilise pour ces deux mots les graphies françaises, et donc les accords français. Cela a aussi l’avantage de bien différencier les parlers *özbek* (certains parlers de l’Ouzbékistan, qui ne sont pas du type *karluk*) de la norme ouzbèke (qui est du type *karluk*).

Enfin : *turk* et *turc*. Dans les publications turkologiques françaises, il est d'usage de différencier *turk*, qui concerne l'ensemble des parlers *turks*, et *turc* qui concerne uniquement le parler de Turquie. Ces deux mots s'accordent.

II. Problèmes historiques

II.1. Généralités

Les langues turques se ressemblent beaucoup. À l'exception du tchouvache, on peut dire que leur degré de ressemblance est celui des dialectes italiens, avec cette différence que les langues turques, étendues sur un très vaste espace, ont emprunté beaucoup de mots à des langues étrangères : selon les cas et les époques, à l'arabe, au persan, au russe, au mongol etc. Quand on étudie les différences entre les langues turques afin d'essayer de saisir la singularité de l'une d'entre elles, on évite les mots empruntés, et on se concentre sur les mots plus proprement turks, dans la mesure où on peut s'en assurer¹⁶.

En Asie centrale, il existe trois grands types traditionnels de parlers turks qu'il faut signaler tout de suite, car ils reviendront constamment dans cet article. Ce sont les parlers de type *oguz*, de type *kipčak*, de type *karluk*. Le nom *kipčak* sera discuté plus loin. On trouve souvent, dans les ouvrages récents une nomenclature différente, plus géographique, par exemple chez Johanson¹⁷ :

oguz	Sud-Ouest
kipčak	Nord-Ouest
karluk	Sud-Est

La géographie de ces trois types ne correspond que très approximativement aux frontières modernes, surtout du fait qu'ils peuvent se trouver ensemble dans un même état. Toutefois, pour fixer les idées, nous pouvons proposer le tableau approximatif suivant :

	oguz	kipčak	karluk
Kazakstan		+	
Kirgizstan		+	
Ouzbékistan		+	+
Karakalpakistan	+	+	
Türkmenistan	+		

On peut déduire de cela que les parlers turks du Kirgizstan ou du Kazakstan, ou dans un autre registre ceux du Türkmenistan, sont relativement plus homogènes que ceux d'Ouzbékistan et de sa province le Karakalpakistan.

Il y a à cela des explications historiques connues. Pour ce qui concerne le Karakalpakistan, il faut tenir compte de la région toute proche, et historiquement associée, du Khorezm, dont les capitales traditionnelles sont Ourgentch et Khiva. Le Khorezm, qui est la région économiquement dominante du bas Amou Darya, était peuplé avant l'an mil surtout d'iranophones. Avant la conquête musulmane (VI^e-VII^e siècles), ces iranophones parlaient des langues iraniennes d'un type spécial, dont la principale est le *khorezmien*, sur quoi il existe des documents assez nombreux¹⁸. Cette langue était proche du célèbre *sogdien*, parlé plus à l'est dans la région de Samarcande. Dans les siècles qui ont suivi les incursions musulmanes, dont les langues de prestige étaient l'arabe et surtout le persan, c'est le persan qui a peu à peu remplacé les langues iraniennes plus anciennes, de sorte que les communautés iranophones dans l'Asie centrale actuelle parlent divers parlers persans connus sous le nom de *tadjik*. Les derniers témoignages qu'on ait du *khorezmien* datent du XIV^e siècle.

Quant aux parlers turks, ils se sont installés lentement à partir peut-être du X^e siècle. À cette époque en effet, plusieurs groupes turkophones fréquentent le bas Amou Darya, en particulier des gens de parlers *oguz*, dont une partie seront plus tard les *Türkmen*. Un peu plus tard, ou en même temps, s'installent plus à l'est (dans l'Ouzbékistan *stricto sensu* et au-delà vers le sud) des gens chez qui on constatera plus tard des parlers *karluk*. A cette époque-là, les parlers *oguz* et *karluk* devaient être très peu différents. Peter Golden, un des grands spécialistes de l'histoire des Turks, écrit, pour ce qui concerne les groupes repérables du X^e siècle¹⁹ :

« Most of these tribes constituted a linguistic community in which there was mutual intelligibility. Thus, al-Iṣṭaxrī comments, with regard to the Toquz Oğuz, Qırğız, Kimek, Oğuz and Qarluq, that "their languages are one, intelligible to one another." (ed. de Goege, p. 9). This is borne out by the more detailed analysis of Maḥmūd al-Kāšgarī. »

L'arrivée du troisième type de parlers, *kipčak*, date surtout de l'installation des *Özbek* historiques dans la région, au XVI^e siècle, ceux-là même qui délogèrent Babur et le poussèrent à conquérir l'Inde, fondant ainsi la dynastie des Mogols.

Il est donc essentiel de bien distinguer les *Ouzbékistanais*, citoyens de l'état actuel d'Ouzbékistan (que j'écris à dessein à la française), et les *Özbek* dont ils se réclament souvent. Les *Özbek* historiques ne sont qu'une des composantes de la population de l'Ouzbékistan. De même, les parlers des descendants de ces *Özbek* (si l'on ne tient pas compte des nombreux cas de changement de langue de telles régions ou de tels groupes), qui sont des parlers *kipčak*, ne sont qu'une des composantes de paysage linguistique de l'Ouzbékistan.

Au Karakalpakistan, où l'influence des parlers *karluk* est beaucoup moindre que dans le reste de l'Ouzbékistan (au point que la norme de la langue officielle ouzbèke est surtout tirée d'eux, beaucoup plus que des par-

lers *kıpčak* des Özbek !), on rencontre donc surtout des parlers de type *oguz*, notamment au Khorezm, et des parlers de type *kıpčak*.

II.2. Caractère politique des définitions actuelles

Lorsque Sergej Efimovič Malov (1880-1957) arrive en Karakalpakie en juillet 1930, il vient avec une mission bien précise : définir quel sera l'alphabet propre aux Karakalpak, et à cette époque il s'agit d'un alphabet en écriture latine. Nous savons exactement comment tout cela s'est passé, non seulement par le récit de Malov²⁰, mais aussi par le rapport de Baskakov, qui se distingue par son style militaire :

« En 1930, l'Académie des sciences d'URSS et le Comité central du nouvel alphabet kazak organisèrent le voyage du professeur S. E. Malov pour l'étude de la langue karakalpak et l'établissement d'un alphabet unifié. Pour l'arrivée de celui-ci, on avait réuni à Turtkul' un détachement de la Section de terminologie scientifique de la Région, qui comportait l'attaché scientifique N. A. Baskakov ; l'attaché du journal *Mijnetkeš karakalpak* T. Sapiev, le directeur du Musée régional N. V. Torčinskij, et l'enseignant de l'école de Čimbaj, Turdumurat Bekimbetov. Des cinq principaux districts karakalpak, ceux de Čimbaj, Kegejli et Taxta-Kupyr furent inspectés sous l'autorité directe du prof. S. E. Malov, et ceux de Kara-Uzâk et Kungrad furent commissionnés à l'attaché scientifique de la Région, N. A. Baskakov et à son aide, Turdumurat Bekimbetov ».

Le savant russe Aleksandr Nikolaevič Samojlovič (1880-1938), qui avait lui-même vécu dans la région vers 1910, avait produit pour le nouveau pouvoir soviétique en 1922 une classification des langues turkes, dans le cadre de l'importante Commission des nationalités, et en 1930 il avait une position académique bien plus importante que celle de S. E. Malov, qui s'était de son côté illustré par des voyages parmi les populations turkophones de Chine. Samojlovič était alors le secrétaire du département de Sciences humaines de l'Académie des sciences. Or, Samojlovič ne croyait pas à l'existence propre d'une langue karakalpak, où il voyait simplement un parler de type kazak. Malov, comme il l'a écrit lui-même, ne croit pas en arrivant sur les lieux que la langue d'un si petit peuple ait résisté à la concurrence de ses puissantes voisines : l'özbek, le türkmen, et bien sûr le kazak. Son enquête dans les mois qui suivent lui montre, ajoute-t-il, qu'il existe pourtant une langue karakalpak, en dépit des influences qu'il détaille de la façon suivante :

« L'influence de la langue türkmen à cet égard est très faible. Si l'on écoute une chanson karakalpak d'origine türkmen, alors, en effet, on peut y entendre plusieurs particularités du türkmen, comme la sonorisation des consonnes sourdes en début de mot, la désinence de datif en -a au lieu de -ya, et d'autres encore empruntées au lexique türkmen, mais c'est que, comme on dit, « dans les chansons, pas un mot ne changeons ». En dehors de ces chansons, des énigmes, et des vers, il y a peu de choses en karakalpak qui soit issu du türkmen. (...) On peut dire la même chose à propos de l'influence du dialecte ouz-

bek de Khiva sur la langue karakalpak. Bien sûr, il existe en karakalpak des locutions, des mots, qui viennent du khivien [le parler de la ville de Khiva], mais ils n'affectent en rien la structure générale et le caractère du karakalpak dans son ensemble. (...) Des singularités du dialecte ouzbek de Khiva, comme par exemple l'abrévement de la désinence de génitif (-ni), rien ici (...), etc.

La langue karakalpak a surtout subi l'influence du kazak. Ces deux tribus vivent depuis longtemps côte à côte (...). Mais il faut conserver à l'esprit que ces langues, le karakalpak et le kazak, sont apparentées et proches l'une de l'autre. Il peut paraître que le karakalpak a succombé sous l'influence du kazak, mais en réalité cette ressemblance des deux langues n'est pas le résultat de l'influence de l'une sur l'autre, elle tient à la longue et traditionnelle ressemblance de deux dialectes proches d'un groupe occidental, kipčak, des langues turques ».

Ainsi, Malov considère que le karakalpak, langue à part entière, est un dialecte kipčak au même titre que le kazak. Nous avons rapporté les raisons qui appuient ce constat, rédigé sur place, parce que deux questions se posent : une question politique d'abord, une question linguistique ensuite.

La question politique tient au fait que le statut du karakalpak comme langue était réglé d'avance. La Commission territoriale de 1924 avait déjà décidé de créer une Région (*oblast'*) autonome de Karakalpakie, et une sous-commission d'enquête pour en définir les frontières. En outre, il avait déjà été décidé de la rattacher au Kazakhstan. Les raisons qui ont poussé la Commission à créer une Région administrative karakalpak n'incluaient pas le souci de préserver l'intégrité de l'ethnie karakalpak, puisque Malov lui-même six ans plus tard pensa d'abord que ce peuple avait perdu sa langue, avant de découvrir avec enthousiasme qu'il n'en était rien. Il est plus probable qu'il s'agissait de contrevenir aux prétentions traditionnelles des Khiviens sur les Karakalpak, depuis que le khan de Khiva, en 1810, avait déporté une partie des Karakalpak dans les terres du delta afin qu'il les exploite pour son compte (voir l'article de Sv. Jacquesson). Et comme les Khiviens n'étaient pas bien vus, à cause du "nationalisme bourgeois"²¹ de la République qu'ils avaient eux-mêmes récemment instituée, le nouveau pouvoir de Moscou pensait qu'il convenait de réduire leurs prétentions, peut-être même de les affaiblir, en se faisant les garants de l'indépendance relative des Karakalpak. On ne saura jamais jusqu'à quel point Malov était sincère en 1930, mais il est certain qu'il était envoyé en Karakalpakie pour trouver une écriture aux Karakalpak, pour lesquels on avait donc déjà décidé en haut lieu qu'ils avaient une langue spécifique. La mission de Malov était de prouver que tel était bien le cas, et de sceller cette vérité par un alphabet non moins spécifique. Sa marge d'action était assez étroite.

II.3. Malov et Samojlovič

La question linguistique est en effet assez compliquée, comme le rapport de Malov le laisse transparaître. Ce rapport consacre plusieurs pages, intéressantes d'ailleurs, à détailler les auteurs qui ont jusque-là parlé des

Karakalpak, et à fustiger ceux qui les ont passés sous silence, comme Samojlovič l'a fait dans sa célèbre classification des langues turques, publiée d'abord en 1922, puis élaborée en 1926. C'est une des raisons pourquoi ce rapport n'a pas été publié *in-extenso* à l'époque. Mais au-delà des arguments d'autorité, Malov se trouvait confronté au fait qu'on trouve en karakalpak des traits oguz d'un côté et des traits kipčak de l'autre. Cela le mettait dans une situation embarrassante, parce que dans la théorie en vigueur les parlers turks étaient ou bien oguz, ou bien kipčak, et Malov était d'accord avec cette alternative, comme on le voit très bien dans l'extrait cité plus haut.

Parmi les traits oguz les plus nets, il y a la sonorisation des initiales à quoi il fait allusion : un certain nombre de mots qui normalement commencent par /k/ ou /t/ dans les autres langues turques, en karakalpak commencent par /g/ ou /d/ – comme si nous disions *golère* et *gapable* au lieu de *colère* et *capable*. Mais les traits kipčak lui semblent l'emporter.

On reste perplexe, pourtant, devant l'argument de Malov qui consiste à rejeter à peu près tous les éléments de l'orature traditionnelle d'une ethnie ("les chansons, les énigmes, et les poèmes", précise-t-il) en décidant qu'il s'agit d'emprunts aux Türkmen. Cette part du feu restera d'ailleurs vaine, puisque N. A. Baskakov, le grand linguiste russe qui s'imposera comme le spécialiste des Karakalpak (et plus généralement des parlers kipčak), montrera un peu plus tard (voir plus loin la section IV.2.) que la fréquence des initiales sonores n'est pas du tout limitée aux chansons, mais constitue l'un des traits les plus caractéristiques de certains parlers karakalpak. L'idée de Malov, de présenter ce trait comme une série d'emprunts au türkmen, n'avait d'autre but que de faciliter l'identification du karakalpak comme "dialecte kipčak" aux côtés du kazak – ce qui était aussi l'opinion de Samojlovič, et ce qui correspondait à la décision politique de rattacher le Karakalpakistan nouvellement formé au Kazakstan.

On a vu plus haut Malov parler du "dialecte ouzbek de Khiva" ou "khi-vien", et jusqu'à nos jours, en dépit des recherches dont les résultats sont parfaitement clairs, on continue de parler des "dialectes ouzbeks du Khorezm". Comme l'a montré F. A. Abdullaev dans son livre *Fonetika horezmskih govorov* [*Phonétique des parlers du Khorezm*], 1967, les parlers du Khorezm peuvent se grouper en deux dialectes : au nord-est, un petit tiers de la région utilise des parlers kipčak, et donc proches du kazak (ou du karakalpak, ou des parlers kipčak de l'Ouzbékistan) ; au sud, la majorité du terroir utilisait à cette époque des parlers oguz qui avaient, et ont encore parfois, toutes les caractéristiques des parlers türkmen, par exemple l'opposition de longueur chez les voyelles. Il n'y a aucune raison, du point de vue linguistique, d'appeler tous "ouzbeks" ces parlers de divers types. Les gens qui les parlent sont ouzbékistanais certes. Et bien entendu, la norme des écoles ouzbèkes contribue à modeler le paysage linguistique.

II.4. L'histoire des langues, l'économie, le milieu naturel

Une bonne classification linguistique est une classification qui ouvre une perspective sur l'histoire des langues, qui indique comment (et parfois quand) se sont faites les différences qui ont amené peu à peu, à partir de parlars communs, à des parlars différents. Cependant, les différences actuelles entre les parlars résultent souvent de la disparition de parlars transitionnels : de nos jours encore, entre le français du nord et le catalan, par exemple, il existe toute une série de parlars intermédiaires, et autrefois la frontière était plutôt vers le centre de la France que vers les Pyrénées. À mesure que la gamme de ces parlars disparaît pour ne laisser émerger que des entités franchement contrastées, notre impression que nous sommes en présence de langues distinctes s'affirme.

Dans le cas des langues turkes, non seulement les parlars intermédiaires sont très vivants, mais la facilité à se comprendre a favorisé les échanges que le mode de vie des éleveurs rendait par ailleurs naturels : ces gens-là avaient une idée des frontières, mais ce n'est pas l'idée qu'en ont les états modernes, c'était plutôt celle du partage des pâturages, qui variaient avec les saisons et les conditions climatiques. D'autre part, ce cordon des steppes²² est asymétrique : les prairies occidentales, vers l'Ukraine et la Hongrie, sont plus riches que les prairies ou les steppes orientales, qui sont soumises à des écarts climatiques plus intenses – et cela se sait. C'est ce qui explique dans une large mesure les “grandes invasions” successives qu'a connues l'Occident, et les Slaves subissaient en général le premier choc. Quand les conditions de pâture étaient trop rudes en Asie centrale, soit on mourrait, soit on se trouvait un chef assez déterminé pour aller vers l'ouest. Un groupe poussait l'autre, et l'extrémité occidentale subissait une partie de ces effets de vague, parfois amplifiés, parfois affaiblis. Enfin, outre le tropisme occidental qui visait d'abord les prairies plus riches, il y avait le tropisme des cités sédentaires.

On a d'abord pensé que l'élevage représentait le type archaïque de l'exploitation sociale de la nature. Les éleveurs représentaient donc alors le type archaïque de la société humaine, selon les critères hiérarchiques de cette époque, le plus proche de l'état “sauvage”. Nous savons aujourd'hui que c'est faux, et que l'agriculture est apparue avant l'élevage, du moins avant l'élevage spécialisé des steppes. Ce dernier, historiquement assez tardif puisqu'il est lié à la domestication du cheval, est toujours en relation économique avec les régions agricoles.

Il est indéniable que ce milieu économique particulier, spécialisé, de l'élevage à grande échelle, a été le contexte formateur d'un espace social où les populations turkophones et mongolophones se sont élaborées. Cela ne signifie pas que ces populations aient eu, du point de vue ethnique (et donc du point de vue de la génétique des populations), la même origine. Cela signifie au contraire que des groupes humains d'origines diverses se sont trouvés fusionner dans ce creuset particulier d'un mode de vie similaire et, à

certaines égards, uniformisé. C'est ce qui explique qu'on rencontre parmi les turkophones des gens "d'aspect européen ou méditerranéen" comme par exemple les Turcs de Turquie, et des gens "d'aspect oriental" comme certains Kazak ou Kirgiz.

C'est ce qui explique aussi l'uniformité de la plupart des langues turkes. Cette uniformité est due à deux causes. D'une part, au fait que ce sont des groupes dispersés sur un vaste espace mais en relations suivies ; or, ces groupes sont généralement plus conservateurs en fait de langue que les groupes denses réunis sur une zone réduite, parce que les innovations linguistiques doivent être avalisées par l'ensemble des co-locuteurs, sous peine de rupture des relations. Cette validation est évidemment plus lente quand elle doit faire le tour de populations répandues sur une espace vaste, alors qu'elle peut être l'affaire de quelques années, ou moins, dans une cité où les nouvelles circulent vite²³. D'autre part, les langues turkes sont des langues récentes : si vers le X^e siècle tous ces gens se comprenaient entre eux, cela signifie que le modèle linguistique turk était alors récent.

Nous ignorons, faute de documents écrits antérieurs aux inscriptions "runiques", comment s'est faite la singularisation des parlers turks, et c'est un problème très débattu que celui des relations du turk avec le mongol et le toungouse (qui ensemble sont souvent appelées les "langues altaïques"), et *a fortiori* avec le "modèle ouralien" (c'est-à-dire les langues finno-ougriennes et samoyèdes). Le problème est compliqué précisément parce que les interactions entre ces différentes populations ont dû impliquer autant les emprunts massifs dus à un mode de vie commun, que la supposée communauté d'origine. Les deux faits, et d'autres, sont inextricablement mêlés, comme il est naturel.

II.5. Les critères linguistiques

Quelques critères linguistiques simples permettent de distinguer, traditionnellement, trois groupes majeurs de langues turkes. Il existe deux autres groupes, minoritaires mais intéressants, que nous laisserons de côté ici parce qu'ils ne concernent pas les Karakalpak. Ces trois groupes majeurs sont le groupe *oguz*, le groupe *kipčak*, et le groupe *karluk*. Si l'on se borne aux parlers que représentent les langues officielles, certaines différences apparaissent :

		jour	montagne
<i>Clauson</i> ²⁴		*gün	*da:ğ
oguz	turc	gün	<i>dag</i>
	azeri	gün	<i>dag</i>
	türkmen	gün	<i>daag</i>

		jour	montagne
<i>Clauson</i> ²⁴		*gün	*da:ğ
kipčak	baškort ²⁵	kön	taw
	tatar	kön	taw
	nogaï	kün	taw
	kazak	kün	taw
	kırgız	kün	too
	<i>Codex cumanicus</i>	kün	taw, tag
	tatar de Crimée	kün	dag
	karačaj	kün	taw
	karaïm	kün	tag, dag, taw
karluk	ouzbek	kün	tâg
	ouïgour	kün	tag

1/ Dans de nombreux mots (mais non dans tous les cas), la consonne initiale sonore en oguz correspond ailleurs à une sourde. Dans l'exemple "jour", le /g/ oguz correspond au /k/ kipčak ou karluk. Dans l'exemple "montagne", le /d/ oguz correspond au /t/ kipčak ou karluk. Si cette règle était absolue, elle impliquerait que l'opposition entre sourde et sonore à l'initiale, vivante en oguz, puisqu'on y trouve aussi bien des mots en k- qu'en g-, serait inexistante ailleurs où l'on n'aurait de mots qu'en k-. Mais nous verrons que cette règle n'est pas absolue, et que l'opposition de sonorité à l'initiale est problématique en dehors de l'oguz (voir plus loin l'exemple de "quatre"). Et spécialement en karakalpak.

2/ Exemple "montagne". L'uvulaire sonore finale ("g", mais prononcée /ɣ/, proche du "r" du français du nord), est réduite à l'approximante [w] dans la majorité des parlers kipčak, et cette réduction va même en kırgız jusqu'à créer des voyelles longues nouvelles (tay > taw > too), distinctes des anciennes voyelles longues qui apparaissent encore régulièrement en türkmen (daag). En territoire ouzbékistanais, il existe des parlers kipčak où "montagne" se prononce /taw/ ou /tau/, et des parlers karluk où ce mot se prononce /tag/.

3/ Un /a/ très postérieur (noté ici "â", qui ressemble à la prononciation champenoise du "â" de *pâte*), est typique de la norme ouzbèke, et parfaitement populaire au Ferghana. La graphie ouzbèke officielle l'écrit "o" tant en cyrillique qu'en latin. Ce passage de /a/ à /ɔ/ (un "o ouvert" comme dans *port*) et en réalité souvent à /ɑ/ semble bien s'être fait à l'imitation du passage de "a long" à /ɔ/ dans les parlers iraniens occidentaux comme le persan/tadjik.

Dans ce cadre, les parlers karakalpak se comportent exactement comme les parlers kazak. Il faut donc regarder de plus près les parlers kipčak.

II.6. Le nom “kipčak” et Klaproth

Ces trois noms “oguz”, “karluk”, et “kipčak” ont une histoire compliquée, mais passionnante, et sont importants dans l’histoire des populations comme des langues turkes. Ils sont importants aussi dans l’histoire de la science turkologique. On peut consulter là-dessus Golden 1992, qui écrit surtout en historien, ou bien les intéressants articles de *l’Encyclopédie de l’Islam*, plus anciens. Je me contenterai de commenter le terme *kipčak*.

Ces noms sont d’abord des noms de groupements ethniques, et non pas de groupes de langues. C’est pourquoi les linguistes modernes essaient parfois de leur substituer des noms géographiques (voir plus haut), qui ont d’autres inconvénients puisque l’opinion générale est que l’efficacité de l’article de Samojlovič (1922) sur la classification des langues turkes tenait au fait qu’il s’émancipait des regroupements géographiques qui avaient dominé jusqu’alors.

Pourtant, les savants du XIX^e siècle avaient déjà largement utilisé les critères linguistiques. Il a Nikolaevič Berezin, dans sa classification de 1848 en français, donne des critères phonétiques (et d’autres morphologiques et lexicaux) dont la plupart seront repris par Samojlovič. Et il utilise, pour les langues turkes dont il ne sait rien d’autre, les vocabulaires fournis par l’érudit Jules Klaproth, un des grands pionniers de la linguistique de l’Asie.

Jules Klaproth, ou Julius von Klaproth (1783-1835), fut un des grands orientalistes du début du XIX^e siècle. Il était allemand d’origine, mais russo-phonie et francophone aussi bien, et a longtemps vécu en France, où il a été un des fondateurs de la Société asiatique, en 1822. Il a voyagé dans le Caucase, en Russie, en Chine, et a publié en 1823 la 1^{re} édition de son *Asia polyglotta* (la 2^e est de 1831). Il est comme l’articulation entre l’époque des grandes compilations linguistiques et la science ultérieure, qui tendait à distinguer soigneusement les différentes disciplines. Son œuvre se distingue donc des entreprises universalistes comme le *Mithridates, oder allgemeine Sprachkunde mit dem “Vater unser” als Sprachprobe in bey nahe fünfhundert Sprachen und Mundarten* de Johann Christoph Adelung (1817), ou *l’Atlas ethnographique du Globe, ou classification des Peuples anciens et modernes d’après leurs langues* d’Adrien Balbi (1826). C’est Klaproth, semble-t-il, qui utilisa le premier le terme *kipčak* dans le domaine linguistique. C’est aussi lui qui redécouvrit le *Codex cumanicus*. Ce document exceptionnel des parlers *kipčak* avait été acquis au début du XIV^e siècle par le poète et érudit italien Francesco Petrarca. Pétrarque avait légué ce manuscrit à la Bibliothèque de la basilique Saint-Marc de Venise. Klaproth en avait obtenu une copie, qu’il a publiée (sans jamais voir l’original !) dans le 3^e volume de ses *Mémoires relatifs à l’Asie, contenant des recherches historiques, géographiques et philologiques sur les peuples de l’Orient* (1828). Klaproth est donc le premier savant à avoir eu des langues turkes une vision étendue, personnelle grâce à ses voyages, et historique. En outre, il a composé et publié plusieurs cartes intéressantes de nos jours encore.

Je me permets de citer un assez long passage de Klapproth, qui est le premier exposé sur les Kıpçak et les Comans. On y trouve ce style panoramique si caractéristique de l'érudition orientaliste qu'illustrera encore Paul Pelliot plus d'un siècle plus tard²⁶. Une partie des ancêtres de nos Karakalpak, encore indistincts, se trouvent quelque part dans ce tourbillonnement²⁷ :

« Les Comans et les Petcheneghes formaient le peuple appelé Kiptchak ou Kaptchak. Ruysbroeck²⁸ dit que les premiers se donnaient eux-mêmes le nom de *Capchat* ; dans un autre endroit il les appelle *Coman-Capchat*. Selon les historiens byzantins, les Petcheneghes (Πατζινακοι, Πετζινακοι, Πετζινακισται) avaient habité originairement près des fleuves *Atil* (Volga) et *Gheikh* (Iaïk) ; ils en furent chassés vers l'année 894 ou 899, par les Ouzes et les Khazares réunis. Quelques Petcheneghes retournèrent volontiers chez les Ouzes, et se confondirent avec les vainqueurs ; cependant ils en étaient distingués par un habillement qui leur était propre. Le reste des Petcheneghes passa le Don, et fonda sur les Hongrois, dont une grande partie s'enfuit vers *Atelkouzou*, aujourd'hui la Moldavie et la Transylvanie. Les Petcheneghes les poursuivirent encore et les poussèrent plus à l'ouest. Depuis ce temps (900 ans après J.-C), ils furent maîtres de toutes les côtes de la mer Noire, depuis le Don jusqu'au Danube. Les Comans habitaient alors plus à l'est, au delà du Don. D'après le témoignage de Constantin Porphyrogénète, les Patsinakites ou Petcheneghes portaient anciennement le nom de *Kangar*, qui, à ce que cet auteur assure, signifiait dans leur langue *générosité* et *courage*. Ce nom se trouve aussi chez Ruysbroeck, sous la forme de *Kangle*. Ce voyageur partant, en 1253, du camp de Batou-khan, fut conduit à l'est du Volga, et se dirigea à l'orient en traversant le pays des *Kangle*, qui descendaient des Comans. Il avait alors à sa gauche et au nord la *Grande Bulgarie*, et à sa droite ou au sud la mer Caspienne. Ces *Kangle* étaient vraisemblablement les Petcheneghes qui étaient restés parmi les Ouzes, et dont j'ai parlé plus haut. Jean du Plan-Carpin qui, en 1245, ainsi peu de temps avant Ruysbroeck, fut envoyé en Tatarie par le pape Innocent IV, s'explique encore plus clairement : « Le pays de *Comanie*, dit-il, a immédiatement au nord, après la Russie, les Morduins, les Bilères, c'est-à-dire la Grande Bulgarie ; puis les Parasites et les Samoïèdes, qu'on dit avoir la face de chien, qui sont sur les rivages déserts de l'océan. Au midi, il a les Alains, les Circasses, les Gazares (la *Crimée*), la Grèce et Constantinople (...) ».

Les Comans et les Petcheneghes (*Kangar*, *Kangle* ou *Cangites*) ne formaient donc qu'un seul peuple, avec la différence que ces derniers se montrèrent en Europe vers 894, un siècle avant les autres. C'étaient les قیچاق *Kibchâk* dont le nom de tribu était قنقلی *Kankly* ou قانقلی *Kânkly* qui, selon Abû al-Ghâzî, dérive du bruit que font les roues des voitures, قانق *kânak*, dont ils prétendaient être les inventeurs. Les restes des Comans et des *Kankly* se sont confondus avec les Turcs Nogaj, dont plusieurs hordes portent encore le nom de *Kankly*.

Le mot *kipčak* est le nom d'une fédération assez lâche de divers groupes dont la majorité étaient turkophones. À cause de l'histoire de ce groupement, le mot *kipčak* a aussi désigné une région, qui correspond à peu près aux steppes centrales, non seulement l'actuel Kazakhstan mais aussi les régions plus à l'ouest jusqu'à la Crimée. Historiquement, les groupes que les Russes médiévaux appelaient *Polovec* ("les Pâles") et les documents byzantins aussi les *Comans* (d'où le nom du *Codex comanicus*) faisaient partie de cet ensemble. À l'intérieur de ce vaste ensemble, les petits groupes se déplaçaient en fonction des besoins du moment. Et c'est une des raisons des déplacements probables de nos Karakalpak.

III. À l'intérieur des parlers *kipčak*

III.1. Les sous-groupes du *kipčak*

Cette section est consacrée aux différences entre les divers parlers *kipčak*. J'utilise ici les normes officielles des dictionnaires²⁹, sauf pour le tatar de Crimée pour lequel je remercie Alije Akimova³⁰. Pour le karakalpak, où la norme a changé dans les années 1950, j'utilise le dictionnaire russe-karakalpak compilé par N. A. Baskakov et quatre savants karakalpak, publié en 1967.

III.1.1. "inversion" vocalique en tatar et baškort

Les langues *kipčak* que nous étudions ici pour nous guider dans l'appréciation du karakalpak se divisent d'abord aisément en deux groupes sur le critère du vocalisme³¹. J'ajoute les formes que donne le *Codex cumanicus* d'après le lexique de Grønbech 1942³² :

	veuf	jour	feu	quatre
<i>Clauson</i>	*du:l	*gün	*o:d	? *do:rd
baškort	töl	kön	ut	dürt
tatar	töl	kön	ut	dürt
nogaj	tul	kün	ot	tört
kazak	tul	kün	ot	tört
karakalpak	tullıq	kün	ot	tört
kırız	tul	kün	ot	tört
<i>Codex cumanicus</i>	tul	kün	ot	tört
karačaj	tul	kün	ot	tört
tatar de Crimée	tul	kün	(ateš)	dört
karaım	tul	kün	ot	dört

Il s'agit d'une "inversion" des voyelles arrondies en baškort et en tatar, par rapport à la situation ordinaire en turk : /o/ y devient /u/ et inversement, et de même avec les arrondies antérieures : /ø/ devient /ü/ et inversement.

Un phénomène similaire a lieu dans certains cas avec le /e/ ordinaire, qui devient /i/ (et inversement) en baškort et tatar :

	venir	viande
<i>Clauson</i>	*gel	*et
baškort	kil-	it
tatar	kil-	it
nogaj	kel-	et
kazak	kel-	et
karakalpak	kel-	et
kirgiz	kel-	et
<i>Codex cumanicus</i>	kel-	et
tatar de Crimée	kel-	et
karaïm	kel-	et

Il faut prendre garde ici à certaines distinctions. Dans le domaine vocalique de [e-i], il existe une légère différence de traitement, qui concerne le türkmen, selon qu'il s'agit étymologiquement d'un *e bref ou d'un *ee long, d'un *i bref, ou d'un *ii long³³. Les deux exemples du tableau précédent concernaient le *e bref. Voici des exemples pour les autres cas. Les langues kipčak sont encadrées.

	*ee		*ii		*i	
	rein	tardif	entrer	dent	un	savoir
<i>Clauson</i>	*bé:l	*gé:č	*gir-	*dɪ:š	*bi:r	*bil
turc	bel	geč	gir-	diš	bir	bil-
azeri	bel	gedž	gir-	diš	bir	bil-
türkmen	biil	gič	giir-	diiš	bir	bil-
baškort	bil	kis	ker-	teš	ber	bel-
tatar	bił	kič	ker-	teš	ber	bel-
nogaj	bel	keš	kir-	tis	bir	bil-
kazak	bel	keš	kir-	tis	bir	bil-
karakalpak	bel	keš	kir-	tis	bir	bil-
kirgiz	bel	keč	kir-	tiš	bir	bil-
<i>Codex cuman.</i>	bel	keč	kir-	tiš	bir	bil-
tatar de Crimée	bel	keč	kir-	tiš	bir	bil-
karaïm	bel	keč	kir-	tiš	bir	bil-
ouzbek	bel	keč	kir-	tiš	bir	bil-
ouïgour	bäl	keč	kir-	tiš, čiš	bir	bil-

En baškort et tatar (ou du moins dans les dialectes de ces deux langues qui ont servi de base à leur norme actuelle), on observe donc une “inversion” caractéristique de /o/ et /u/, de /ö/ et /ü/, de /e/ et /i/.

III. 1. 2. résistance relative des “gutturales”

Comme nous l’avons vu en II.4., les langues kipčak ont tendance à affaiblir les “g” en fin de mot (le mot “montagne” *tag* devient *taw*), et aussi en position intervocalique. C’est là un dossier difficile qu’il n’est pas question de traiter ici de façon détaillée. Parmi les langues kipčak, cet amuissement a été mené à des degrés différents selon les parlers.

L’examen du sort des monosyllabes, comme c’est le cas de nombreuses racines verbales, donne des résultats à peu près homogènes : après une voyelle postérieure le *g étymologique (prononcé comme une spirante /ɣ/, souvent noté “ğ”) s’amuit en /w/ dans la plupart des parlers kazak-karakalpak et tatar-baškort. Après une voyelle antérieure, ce même *g s’amuit en /j/ dans presque tout le kipčak. Je donne quelques exemples dans le tableau ci-dessous, où l’on verra que certaines langues ajoutent des suffixes. En outre, j’ai ajouté les formes en “vieux turk”, c’est-à-dire celui des inscriptions sur pierre (VIII^e et IX^e siècles).

	voyelle postérieure		voyelle antérieure	
	lien, cordon	naissance	humide, cru	toucher
<i>Clauson</i>	*ba:ğ	*duğ-	*čig	*deg-
turc	bag	dog-um	čig	(dokun-)
azeri	bag	dog-um	čij	(toxun-)
türkmen	baag	dog-luš	čiiğ	deg-mek
baškort	baw	tuw-ıw	sej	tej-ew
tatar	baw	tuu	či	tij-ü
nogaj	baw	tuw-uw	šij	tij-üw
kazak	baw	tuw-uw	šij-ki	tij-üw
karakalpak	baw	tuw-ıw	šij-ki	tij-iw
kirgiz	boo	tuu	čij-ki	tij-üü
<i>Codex cuman.</i>	baw-, bag-	tow-, tog-	čijik	tej-, teg-
tatar de Crimée	bag	dog-ma	čij	tij-
karačaj	baw	tuw-uw	čij	tij-erge
karaïm	baw, bag	tog-ırma	čij	tij-
ouzbek	bâg	tug-liš	(xâm)	teg-mâk
ouïgour	bag	tug-uš	(xam)	täg-mäk
vieux türk	baag	tog-um	či	teg

L’introduction des suffixes dans certaines langues montre que la situation est analogue pour les bisyllabes qui sont formés sur des racines monosyllabiques.

Mais dans d'autres bisyllabes, de formation plus ancienne sans doute, ou en tout cas qui ne paraissent pas décomposables au moment où l'amuissement de cette consonne finale s'est fait sentir, la consonne a souvent mieux résisté. La conséquence est que dans les bisyllabes où *g est intervocalique, les résultats sont plus variés :

	arbre	fil	courir
<i>Clauson</i>	*iġa(:)č	*oġul	*jüġür
turc	agač	ogul	(koš-)
azeri	agač	ogul	jüjür-
türkmen	agač	ogul	jüvür-
baškort	agas	ogul, ul	jöger-
tatar	agač	ul	jöger-
nogaj	agaš	ul	jör-
kazak	agaš	ugul, uul	žügür-
karakalpak	agaš	ul	žuwır-ıw
<i>Codex cuman.</i>	agač	ogul	jüġür-
tatar de Crimée	agač	ogul, owul	jövür-
karaïm	agač	ogıl, ogul, owul	juwur-
karačaj	agač	ul	džür-
kırğız	džıgač	uul	džügür-
ouzbek	(däräxt)	ogıl	jüġür-
ouïgour	jagač	ogul	jügri-

Si l'on s'en tient au -g final après voyelle postérieure, on constate chez les langues du sous-groupe "coman", dit aussi "kıpčak-polovec" une résistance relativement plus grande à l'amuissement du -g.

C'est des raisons pour quoi on peut diviser le domaine kıpčak en trois sous-groupes, en fonction d'une part de "l'inversion vocalique", de l'autre de la résistance à l'amuissement du -g final. Nous verrons bientôt qu'il existe d'autres raisons à cette sub-division.

		o	-ig	-ag
kıpčak-bulgar	tatar	u	-ij	-aw
	baškort	u	-ej	-aw
kıpčak-nogaj	nogaj	o	-ij	-aw
	karakalpak	o	-ij	-aw
	kazak	o	-ij	-aw
kıpčak-polovec	<i>Codex cumanicus</i>	o	-ij	-ag/w
	tatar de Crimée	o	-ij	-ag
	karaïm	o	-ij	-ag/w
	karačaj	o	-ij	-aw
kırğız		o	-ij	-oo

III.2. Réduction des affriquées sourdes en *kipčak*

Une importante caractéristique des parlers *kipčak* est le traitement des fricatives (/ʒ/ et /ʃ/) et affriquées (/dʒ/ et /tʃ/, ce dernier aussi écrit /č/). Nous avons parlé de ces sons dans l'introduction. Nous allons d'abord considérer le sort de l'affriquée sourde/č/.

Regardons le tableau suivant, où les mots entre parenthèses sont empruntés à d'autres langues non turques (c'est pourquoi nous donnons en bas les mots iraniens tadjiks) :

	humide, cru	fleur	cheveu
<i>Clauson</i>	*čig	*čéčeg	*sač
turc	čig	çiçek	sač
azeri	čij	çiçäk	sač
türkmen	čiiġ	(gül)	sač
baškort	sej	säskä	säs
tatar	či	čäčäk, čäčkä	čač
nogaj	šij	šešekej	šaš
<i>kazak</i>	šijki	šešek	šaš
karakalpak	šijki	(gül)	šaš
kirgiz	čijki	čeček	čač
<i>Codex cuman.</i>	čijik	čiček	sač
tatar de Crimée	čij	čeček	sač
karačaj	čij	čiček	čač
karaïm	čij	čiček, cecek	sač, cac, čač
ouzbek	(xâm)	čečäk	sâč
ouïgour	(xam)	čečäk	čač, sač
(tadjik)	xom	gul	

Les langues *kipčak* sont encadrées. On voit avec “humide, cru” et “fleur” que certaines langues ont simplifié l'affriquée (č > š). Dans le sous-groupe baškort-tatar, le baškort a /s/. Dans le sous-groupe du nogaj, on a /č/. Dans le troisième sous-groupe, on a /š/. C'est une autre des raisons pour lesquelles N. A. Baskakov, que j'ai évoqué plus haut, divisait les langues *kipčak* en trois sous-groupes (Baskakov 1962), mais il met le kirgiz à part :

kipčak-bulgar	tatar baškort
kipčak-nogaj	nogaj karakalpak kazak

kipčak-polovec	<i>Codex cumanicus</i> tatar de Crimée karaïm kumık karačaj et balkar
----------------	-----------------------------------------------------------------------------------

L'exemple "cheveu" est intéressant, parce qu'une dissimilation s'est produite en oguz, dans la plupart des parlers karluk, et même dans quelques parlers kipčak : l'origine est bien *čač, mais dans ces langues le premier des deux /č/ a été simplifié en /š/ puis en /s/ pour le différencier du second. Dans les parlers kipčak, cette simplification en général n'a pas eu lieu (on la voit dans le *Codex cumanicus*, en tatar de Crimée, et dans certains parlers karaïm). Cela signifie peut-être que les parlers oguz et karluk n'ont pas toujours été séparés les uns des autres, comme ils le sont aujourd'hui, par les parlers kipčak.

Nous sommes donc en mesure de caractériser nos sous-groupes kipčak :

		*e	*č
kipčak-bulgar	baškort	i	s
	tatar	i	č
kipčak-nogaj	nogaj	e	š
	kazak	e	š
	karakalpak	e	š
kipčak-polovec	tatar de Crimée	e	č
	karaïm	e	č
	karačaj	e	č
	<i>Cod. Cuman.</i>	e	č
	kırğız	e	č

Les changements qui affectent l'affriquée dans certains parlers kipčak (en particulier ceux qui ont servi à définir la norme écrite en baškort et en kazak) posent un problème intéressant. Il y avait en turk ancien, selon toute vraisemblance, un *č, un *š, et un *s. Le passage de *č à /š/ en kazak pouvait donc aboutir à une confusion dans cette langue de *č et de *š. Mais il se trouve que ce *š est devenu /s/ dans le sous-groupe nogaj et dans une partie du karaïm, et c'est donc une convergence entre *š et *s qui s'est produite en réalité. Voici ce qui est arrivé au *š :

	trouer	trou	âne	écouter
<i>Clauson</i>	*del-/teš-	*delük/tešük	*ešgek	*ėšid-
turc	de/-	de/ik	(merkep)	išit-
azeri	däl-	dä/ik	eššäk	ešit-
türkmen	deš-	dešik	ešek	ešit-
baškort	?	tišek	išäk	išet-
tatar	tiš-	tišek	išäk	išet-
nogaj	tes-	tesik	ešek	esit-
kazak	tes-	tesik	esek	esit-
karakalpak	tes-	tesik	ešek	esit-

	trouer	trou	âne	écouter
<i>Clauson</i>	*del-/teš-	*delük/tešük	*ešgek	*éšid-
<i>Codex cuman.</i>	teš-	tešik	ešek	ešit-
tatar de Crimée	teš-	tešik	ešek	ešit-
karačaj	teš-	tešik	ešek	ešit-
karaïm	teš-, tes-	tešik, tesik	ešek, esek	ešit-, esit-
kırgız	teš-	tešik	ešek	ešit-
ouzbek	teš-	tešik	ešäk	ešit-
ouïgour	täš-	töšük	ešäk	išit-

Ce tableau montre en outre que le *š en kıpčak (et en karluk) a sans doute deux origines différentes, puisqu'il a des correspondances distinctes (/š/, mais aussi /l/) dans certains parlers oguz. Il est cependant homogène en kıpčak, où il passe à /s/ dans tous les cas dans le sous-groupe du nogaj. Cette régularité montre que les formes /ešek/ pour "âne" en karakalpak et nogaj sont des emprunts. En karakalpak, la forme /ešek/ est un emprunt à l'ouzbek.

Et en baškort, où le *č devient /s/, y a-t-il confusion avec *s ? Non plus, car dans cette langue, le *s est passé à /h/ !³⁴ :

	os	vache	mot
<i>Clauson</i>	? *süyük	*sıgır	*söz
turc	(kemik)	(inek)	(kelime)
azéri	sümük	(inäk)	(käl mâ)
türkmen	süñk	sıgır	söz
baškort	höjäk	hıjır	hüz
tatar	söjäk	sıjır	süz
nogaj	süjek	sıjır	söz
kazak	süjek	sıjır	söz
karakalpak	süjek	sıjır	söz
<i>Codex cuman.</i>	?	sıgır	söz
tatar de Crimée	süjek	sıgır	söz
karačaj	süjek	sıjır	söz
karaïm	süjek, süwek	sıgır, sıjır	söz
kırgız	söök	(inek)	söz
ouzbek	süjäk	sıgır	söz
ouïgour	süjäk, söñäk	sıjır	söz

On peut résumer l'essentiel de la façon suivante pour le domaine kipčak (les langues où *č a évolué sont encadrées) :

	*e	*č	*š	*s
baškort	i	s	š	h
tatar	i	č	š	s
nogaj	e	š	s	s
kazak	e	š	s	s
karakalpak	e	š	s	s
<i>Codex cuman.</i>	e	č	š	s
tatar de Crimée	e	č	š	s
karaïm	e	č	š	s
karačaj	e	č	š	s
kirgiz	e	č	š	s

III.3. Création d'affriquées sonores

Le sous-groupe kazak-karakalpak (pas la norme nogaj), certains parlers ouïgours, et le kirgiz sont bien connus pour une innovation importante³⁵, au regard des autres langues que nous étudions. Là où les autres langues ont un /j/ (le son “y” du français, comme dans *yatagan*) en début de mot, ces langues ont développé une affriquée sonore /dž/ (le “dj” français comme dans *djinn*). Puis le kazak et le karakalpak, de même qu'ils réduisaient /č/ (= /tš/) à /š/ (*tch-* à *ch-*), ont réduit /dž/ à /ž/ (*dj-* à *j-*) :

	serpent	année	visage	cœur
<i>Clauson</i>	*djila:n	*jil	*ñu:z	*jürek
turc	jılan	(sene)	(bet)	jürek
azeri	ilan	il	üz	üräk
türkmen	jılaan	jıl	jüz	jürek
baškort	jılan	jıl	jöz	jöräk
tatar	jılan	jıl	jöz	jöräk
nogaj	jılan	jıl	jöz	jörek
kazak	žılan	žıl	žüz	žürek
karakalpak	žılan	žıl	žüz	žürek
<i>Codex cuman.</i>	jılan	jıl	jüz	jürek
tatar de Crimée	jılan	jıl	jüz	jürek
karaïm	jılan, ilan	jıl, il	jüz, jiz, iz	jürek, jirek
karačaj	?	džıl	džüz	džürek
kirgiz	džılan	džıl	džüz	džürök
ouzbek	ilân	jıl	jüz	jüräk
ouïgour	žlan	žıl / jıl	jüz	j / žüräk

C'est le trait le plus facile pour identifier un parler kazak ou kirgiz (et pour les différencier entre eux), à condition de se souvenir par exemple des dialectes ouïgours homologues. Par analogie avec les parlers slaves qui présentent le même phénomène, on parle souvent à ce propos de "dialectes yé-kaviens" (avec /j/ initial), "djé-kaviens" (avec /dž/), ou "jé-kaviens" (avec /ž/).

Ces nouvelles affriquées sont sonores, à la différence des précédentes qui sont sourdes. Il n'y a pas d'affriquées ni de fricatives sonores étymologiques en turk, du moins au niveau envisagé ici, de sorte que la création de ces sonores n'amenait aucune confusion avec d'autres sons préexistants.

Il existe cependant des fricatives sonores dans des emprunts arabo-persans, sur quoi il est intéressant de jeter un coup d'œil :

	âme	djinn
turc	džan	džin
azeri	džan	džin
türkmen	džaan	džin
baškort	jän	jın
tatar	džan	džin
nogaj	jan	?
kazak	žan	žin
karakalpak	žan	žin
tatar de Crimée	džan	džin
karaïm	džan, dzan	?
karačaj	džan	džin
kirgiz	džan	džin
ouzbek	džân	džin
ouïgour	džan	džin
(tadjik)	džân	džin

On voit que kazak et baškort ont transformé l'affriquée du mot d'origine, parce qu'on ne peut pas la prononcer dans ces langues. La réduction est différente dans les deux langues, et s'apparente à celle qui a eu lieu pour la sourde correspondante :

	*dž	*č
baškort	j	s
nogaj	j	š
kazak	ž	š
karakalpak	ž	š
tatar	dž	tš

	*dž	*č
tatar de Crimée	dž	tš
karaïm	dž	tš
karačaj	dž	tš
kirgiz	dž	tš

III.4. Synthèse des critères de classement

Au terme de cette exploration des critères phonétiques pour le classement des langues turques – ou du moins d’une partie importante d’entre elles –, il est possible de faire un essai de synthèse. Nous avons d’abord (II.4.) esquissé les grandes différences entre parlers oguz, kypčak et karluk. Ensuite, tout en conservant un œil sur les autres, nous nous sommes concentré sur les sub-divisions des parlers kypčak. Nous avons considéré la question des voyelles (III.1.1.) qui permettent d’identifier les parlers “kypčak-bulgar” ; puis celle des -g en fin de mot étymologique (III.1.2.) qui permettent de mettre à part le cas des langues “kypčak-polovec”, vérifiant ainsi le bien-fondé des classifications courantes (Benzing, Menges, Baskakov³⁶). Enfin, nous avons abordé le problème du traitement des fricatives et affriquées sourdes (III.2.), puis sonores (III.3.), qui concernent tout spécialement le sous-groupe du “kypčak-nogaj”, où il apparaît que justement le nogaj d’une part, et kazak et karakalpak de l’autre ont des traitements différents.

	d/t-	*o	-ag	*c	*j-
turc	dag	o	-ag	č	j-
azeri	dag	o	-ag	č	ø-
türkmen	daag	o	-ag	č	j-
baškort	taw	u	-aw	s	j-
tatar	taw	u	-aw	č	j-
nogaj	taw	o	-aw	š	j-
kazak	taw	o	-aw	š	ž-
karakalpak	taw	o	-aw	š	ž-
<i>Codex cuman.</i>	taw, tag	o	-ag	č	j-
tatar de Crimée	dag	o	-ag	č	j-
karaïm	dag, taw	o	-ag/w	č	j-
karačaj	taw	o	-aw	č	dž-
kırgız	too	o	-oo	č	dž-
ouzbek	tâg	o	-ag/w	č	j-
ouïgour	tag	o	-ag/w	č	j-

La conclusion en est que, sur tous ces critères, le karakalpak ne se distingue en rien du kazak. N’y a-t-il donc aucune bonne raison, au plan linguistique, pour identifier des parlers comme spécifiquement karakalpak ?

IV. L'énigme karakalpak

IV.1. Le problème du "papillon"

La singularité la plus frappante de la langue des Karakalpak (du moins dans certains parlers) est l'abondance des occlusives sonores initiales³⁷. Ž. Aralbaev, dans son étude de 1988, *Qazaq fonetikasy bojnša êtjudter* [Études sur la phonétique kazak], donne de longues listes de mots où le karakalpak a une sonore initiale, et le kazak une sourde. Donnons seulement quelques-uns de ses exemples³⁸ :

karakalpak	kazak	
gedej	kedej	pauvre
gereŋ	kereŋ	sourd
gewde	kewde	corps
göže	köže	brouet fermenté
göbelek	köbelek	papillon
göne	köne	vieux
gübi	kübi	baratte en bois
gümis	kümis	argent
gün	kün	jour
gürek	kürek	pelle en bois
güres	küres	lutte (sport)
güriš	küriš	riz
güz	küz	automne

Ce trait est connu des Karakalpak eux-mêmes, qui savent que dans ces cas-là les Kazak prononcent la sourde correspondante. Par exemple dans le cas remarquable du mot "papillon", que les Karakalpak prononcent *göbelek*. Lors d'une enquête en novembre 2000, faite dans le cadre du Projet NSH du CNRS³⁹, nous avons demandé à une vingtaine de personnes, interrogées dans diverses localités du delta de l'Amou Darya, de prononcer toute une série de mots, parmi lesquels "papillon". Le résultat est le suivant :

	köbelek	göbelek
Kazak	6	0
Karakalpak	1	14

Ce résultat a été confirmé lors d'une enquête plus extensive en avril 2001, menée aussi dans le cadre du Projet NSH, mais l'enquête était cette fois menée auprès de plusieurs dizaines de Karakalpak de la région de Mojnak

(tous disent *göbelek*) et de Kazak et d'Ouzbeks de la région de Koñırat, dont voici les résultats par village :

KAZAK	kepelek	köbelek	göbelek
Kök Darya	7	8	4
Kızılküş	0	3	0
Rawşan	3	20	1
TOTAL (sur 46)	10	31	5

OUZBEK	kepelek	köbelek	göbelek
Adebijät	0	0	3
Kıtaj	0	0	7
Kök Darya	0	0	8
Kızılküş	0	0	9
Örnek	1	0	5
Sajat	0	0	3
TOTAL (sur 36)	1	0	35

On s'aperçoit que le fameux trait de sonorité, dans ce cas du moins, n'est donc pas caractéristique des Karakalpak, puisque les Özbek du Karakalpakistan l'ont aussi.

IV.2. Contrastes de sonorité

IV.2.1. deux hypothèses extrêmes pour les sonores initiales

Le problème paraît donc être le suivant : les langues oguz, le turc par exemple, et le karakalpak ont des initiales sonores et sourdes, tandis que le kazak n'a que des sourdes. Dans cette perspective deux hypothèses classiques se présentent.

Ou bien l'opposition de sonorité est ancienne, "proto-turk", et les sonores initiales du karakalpak sont héritées de cette opposition ancienne, de même que celles des langues oguz comme le turc. On doit alors retrouver en karakalpak les mêmes sonores qu'en turc, à peu près pour les mêmes mots : c'est l'hypothèse "conservatrice". Un argument dans ce sens est la prononciation *gün* "jour", en karakalpak comme en turc : dans ce cas, le karakalpak est "conservateur".

Ou bien l'opposition de sonorité est une innovation récente, qui s'est produite indépendamment en oguz et en karakalpak. Le kazak et les langues kıpçak "ordinaires" témoignent alors d'un stade plus ancien. Dans cette hypothèse, il est probable qu'on ne trouvera pas les initiales sonores dans les mêmes mots qu'en turc : c'est l'hypothèse "innovatrice". Un argument est la prononciation *göbelek*, que le karakalpak partage avec l'ouzbek local, mais pas avec les langues oguz : on ne peut pas expliquer cette prononciation par la conservation d'une opposition ancienne.

IV 2. 2. corrélation des voyelles longues et de la sonore qui suit

La complication de la question des sonores turkes est en partie due à la “sous-question” particulière des sonores de fin de syllabe, et à l’histoire de ce thème dans la discipline turkologique. La première langue turke à être connue en Occident était évidemment le turc de Turquie, longtemps dit “ottoman”. En 1851, l’essai de O. Böhlingk, *Über die Sprache der Yakuten*, a commencé d’attirer l’attention sur des langues turkes nettement différentes. Puis sont venus les débats autour du *Codex cumanicus* et, à la fin du XIX^e siècle, les premières publications sur la langue tchouvache. Cela explique que le premier ouvrage véritablement rigoureux de comparaison des langues turkes ne date que de 1902. Il s’agit du *Forstudier til tyrkisk lydhistorie* [*Études préliminaires sur l’histoire des sons turks*], de Vilhelm Grønbech. Malheureusement, cet important ouvrage était écrit en danois, et n’a eu que très progressivement l’impact qu’il méritait (il n’a été traduit en anglais qu’en 1979). De sorte que les idées de V. Grønbech sur la corrélation entre les voyelles longues perceptibles en yakoute et la sonorisation des consonnes *qui suivent la voyelle longue* (j’insiste sur le fait qu’il ne s’agit donc pas des consonnes initiales) a été longtemps ignorée. Il a fallu attendre l’article du turkologue hongrois Louis Ligeti, paru en français dans le *Journal asiatique* de 1938⁴⁰, pour que cette corrélation apparaisse sur le devant de la scène ; en même temps en Russie, le turkologue E. D. Polivanov faisait des observations similaires. À vrai dire, ce renouveau d’intérêt était largement provoqué par la récente publication, en 1929 à Achkhabad, de la première étude systématique de la langue türkmen, *Rukovodstvo dlâ izučeniâ turkmenskogo âzyka* [Guide pour l’étude de la langue türkmen] de A. M. Poceluevskij.

En türkmen en effet, comme nous l’avons remarqué dès notre section II.4., il existe une opposition très claire entre des voyelles longues et des voyelles brèves. Or, ces voyelles longues du türkmen correspondent à celles dont témoigne le yakoute et, dans certaines autres langues comme le turc “ottoman” où ces voyelles longues ne sont plus sensibles, correspondent cependant à la sonorisation de la consonne qui suit. V. Grønbech avait ainsi remarqué les corrélations suivantes :

	turc	yakoute	yakoute	turc	
	longue		brève		
nom	ad	aat	at	at	cheval
feu	od	uot	ot	ot	herbe
affamé	adž	aas	as	atš	ouvert
cuisse	bud	buut			
lait	süd	üüt			
garder	güd	küüt			

On allait ensuite retrouver sporadiquement des traces diverses de ces anciennes voyelles longues, y compris dans les “dialectes ouzbeks du Khorezm”, qui sont en fait des parlers oguz très proches du türkmen.

Toutefois, si la réalité de la corrélation était prouvée, le débat s’installait en même temps pour savoir si c’étaient les voyelles longues, proto-turkes, qui avaient provoqué la sonorisation des consonnes suivantes, au point que bien souvent cette sonorisation restait la seule trace des anciennes voyelles longues, ou bien à l’inverse si les sonores étaient plus anciennes et avaient été la cause de l’allongement des voyelles.

IV. 2. 3. les deux dialectes karakalpak selon Baskakov et selon Nasyrov

Dans son ouvrage fondamental, *Karakalpakskij âzyk*, N. A. Baskakov a décrit les critères qui différencient selon lui les dialectes karakalpak⁴¹. Ces dialectes sont deux, et la ligne de partage suit une oblique descendante à travers le delta : il existe donc un dialecte “nord-est” (cantons de Mojnak, Taxtaköl, Kara-özek) et un dialecte sud-ouest (les autres cantons). Deux des critères de Baskakov recourent les nôtres :

/ž/ au NE, contre /dž/ au SW (kazak /ž/)

/t/ et /k/ au NE, contre les sonores correspondantes au SW, en début de syllabe

Ce qu’on peut résumer de la façon suivante, en rapprochant les faits NE de ceux du kazak :

	karakalpak		kazak
	SW	NE	(norme)
affriquées	dž	ž	ž
initiales	d - g	t - k	t - k
ex. “sel”	duz	tuz	tuz
ex. “jour”	gün	kün	kün

Pour ce qui concerne la sonorité, les parlers du NE sont donc proches du kazak (officiel), tandis que ceux du sud-ouest sont plus originaux. C’est parce que les parlers du NE ont d’abord été choisis comme norme karakalpak (par exemple dans le dictionnaire de 1967 dont nous avons parlé) que la singularité du karakalpak a souvent été occultée.

Cette question dialectale a connu un rebondissement passionnant avec un article de D. S. Nasyrov, “O rudimentah âzyka rodoplemennyh ob"edinenij v dialektah i govorah karakalpakskogo âzyka” [*Les traces de la langue des groupes claniques dans les dialectes et parlers de la langue karakalpak*], (Nasyrov 1974). Cet article montre, en reprenant et affinant les critères de Baskakov, que le partage dialectal correspond bien aux deux grands ensembles en quoi sont divisés les Karakalpak, les *Koŋırat* d’une part, les *On tört uruw* de l’autre.

Les Karakalpak se divisent en effet en deux ensembles nommés *aris*, ce qui signifie “brancard (de la célèbre charrette karakalpak)”, et évoque leur association indéfectible. L’*aris* des *On tört uruw* (voir l’article de S. Jacquesson) est divisé en 4 *ülken uruw* “grands clans”. L’*aris* des *Koñırat* est divisé en deux sections, *Žawıngır* et *Šüllik*, dont la seconde est divisée en 8 *ülken uruw*. Et naturellement chaque *ülken uruw* est divisé en *uruw* “clans”, qui sont souvent à leur tour sub-divisés. Les Karakalpak sont de nos jours l’un des rares peuples turks où l’exogamie clanique est très vivante. Les petits enfants, dès leur plus jeune âge, savent leur clan (celui de leur père), et l’appartenance clanique est toujours indiquée sur la tombe des défunts.

La répartition dans le delta des *On tört uruw* et des *Koñırat* est différente. Les *Koñırat* sont dispersés sur presque tout le delta, même si l’on en trouve des concentrations dans et autour de la ville qui porte leur nom. Les *On tört uruw* sont au contraire bien groupés dans la partie orientale du delta.

Nasyrov s’est demandé si ces groupements recoupaient les critères linguistiques qu’avait autrefois découvert Baskakov. Nasyrov avait parmi ses critères d’enquête un critère n° 8 qui concerne l’opposition g / k à l’initiale. Ses résultats sont les suivants (nous les donnons presque au complet, afin de susciter la curiosité de nos lecteurs), le n° 8 est bien net :

- 3- /a/ ou /ə/ au lieu de /e/ : *kasa* “bol, tasse”, *səkkiz* et non *segiz* “huit”
- 4- harmonie vocalique labiale plus dense : *tülkünün balası mırşaj*
- 5- n + l > nn : *katınnar* “femmes” ; *künnör* et non *künler* “jours”
- 6- s et non š : *jesek* “âne”, *samal* “vent, tempête”, *mısık* “chat”
- 7- suffixe *-mıs* (et non *-mız*) : *balamıs* “notre fils”, *üjüjüs* “votre maison”.
- 8- g- et non k- : *göpür* “pont”, *geme* “barque”, *gındik* “nombriil”, *garga* “corneille”, *garbız* “arbouse”.
- 9- (clan kırık des Kitaj) *-p-* au lieu de *-w-* au passé : *kar žaptı* “il a neigé” et non *kar žawdı*
- 10- présent en *-ıj-* et non en *-a-* : *tisim awrıjdı* « mes dents (me) font souffrir »
- 11- chute de *-l-* devant le suff. négatif : *kımajdı* et non *kılmajdı* « il n’écoute pas, il n’obéit pas »
- 12- cas de /č/ au lieu de /š/ : *čakkı* (ou *čappa*) “couteau pliant”, *čeker* “seau”

aris		ülken uruw	3	4	5	6	7	8	9	10	11	12
<i>On tört uruw</i>		Kıtaj	+					+	+	+	+	
		Kıpčak	+					+		+	+	
		Keneges						+			+	
		Mañıt						+			+	
<i>Koñırat</i>	Zawıngır	Kıjat		+		+	+					
	Šüllik	Ašamajlı		+		+	+					
		Koldawlı		+	+	+	+					

aris		ülken uruw	3	4	5	6	7	8	9	10	11	12
		Kos-tamgalı		+			+					
		Balgalı		+								
		Kandekli		+			+					+
		Karamojın		+								
		Müjten		+	+	+	+					

Ce qui signifie que les sonores initiales /g/ sont caractéristiques des *On tört uruw*, beaucoup plus que des *Koñırat*.

Ce type d'analyse est intéressant parce qu'il est plus fin que celui de Baskakov. Si Nasyrov a raison, il est possible de rapprocher ce que nous savons de l'histoire des groupes karakalpak et les tendances de leur prononciation.

IV.2.4. tendance aux sonores initiales dans l'ouest

Les spécialistes distinguent trois grands types de parlers kazak : le dialecte du NE de l'actuel Kazakstan, sur quoi la norme officielle a été établie, celui du sud qui est frontalier du Kirgızstan et aussi de l'Ouzbékistan, enfin celui de l'ouest qui occupait autrefois la région à l'occident d'une ligne entre l'Oural et la mer Caspienne, et donc géographiquement plus proche de langues kipčak occidentales comme le nogaj et le tatar. Depuis les travaux, maintenant un peu vieilliss, de Särsen Amanžolov, notamment son livre *Questions de dialectologie et d'histoire de la langue kazak* (1958, rééd. 1997), on sait que les parlers occidentaux du Kazakstan présentent une tendance à la sonorisation des /k/ et /t/ initiaux. On peut aussi étudier le détail dans Nurmagambetov 1978, *Kazak tili govorlarının batis tobi*, qui décrit les parlers kazak occidentaux.

Ce penchant du kazak occidental pour les sonores initiales doit être rapproché d'une remarque de Baskakov, selon qui le karakalpak et le nogaj se distinguent ensemble du kazak (la norme kazak) par les traits suivants :

a/ absence de l'alternance conditionnée l/d/t, notamment pour le suffixe *-lar* du pluriel, caractéristique du kazak :

	nogaj & karak.	kazak
pierres	tas-lar	tas-tar
filles	kız-lar	kız-dar

b/ des traits morphologiques, comme l'existence du nom verbal *-adžak / -edžek*, absent en kazak.

Il existe donc un ensemble de traits, parmi lesquels le goût des réalisations sonores à l'initiale tient une place centrale, qui est caractéristique des parlers kipčak occidentaux : le kazak occidental, le nogaj et dans une

moindre mesure le tatar – et le karakalpak. L'existence de réalisations sonores des initiales apparaît donc comme un trait sporadique en kıpčak, plus fréquent dans l'ouest, mais attesté jusqu'en kirgiz. Dans cette dernière langue en effet, un mot comme “bleu, ciel”, orthographié *kök* avec une sourde initiale, est le plus souvent prononcé *gök*⁴². L'atlas dialectologique kirgiz n'a jamais été publié, pas plus d'ailleurs que l'atlas dialectologique karakalpak, pourtant évoqué dans plusieurs articles⁴³, de sorte que nous ne pouvons pas connaître le détail des faits.

La plupart des langues kıpčak révèlent à l'examen un petit nombre, variable selon les parlers, de mots commençant par un *d-* ou un *g-*.

Pour le kumik, Johannes Benzing citait (in Deny 1959 : 396) une petite série de mots à initiales *d-* : *dala* ‘Ebene, Steppe’, *de-* ‘sagen’, *deŋ(ŋ)iz* ‘Meer’, *doŋuz* ‘Schwein’, *dögürek* ‘rund’, *dön-* ‘sich umwenden’, *dört* ‘vier’, *dügü* ‘Reis’ ; et à initiale *g-* : *gibek* ‘Franse’, *gibin* ‘Tliege’, *gij-* ‘sammeln’, *guw-* ‘waschen’.

Pour le baškort, Maksūtova (1976 : 36) cite dans le dialecte oriental quelques mots en *d-* qui sont en *t-* dans la norme : *dulkin* ‘vague’, *dugaj* ‘prairie’, *dulaw* ‘se battre’, etc. Et inversement cite des exemples de mots qui ont *d-* dans la norme et *t-* dans le dialecte méridional : *daga* ‘fer à cheval’, *doŋja* ‘monde’, *därt* ‘vieillesse’, *došman* ‘ennemi’, *döröč* ‘vérité’, *dürt* ‘quatre’.

Certains dialectes kirgiz présentent de tels faits, y compris celui de la vallée du Čüj⁴⁴ ou de la région d'Oš⁴⁵. J'ai déjà évoqué le cas de *gök*, écrit *kök* dans les dictionnaires.

IV.2.5. les sonores initiales karakalpak et les autres

Prenons pour base le dictionnaire karakalpak-russe de Baskakov (1958), et cherchons la proportion des sonores et des sourdes. Cela signifie que nous comptons les entrées de dictionnaire commençant par *d-*, *g-*, *ğ-*, *k-*, *q-*, *t-*. Nous excluons du décompte les mots copiés du russe (*gastronom* ou *gigant*, *traktor* ou *turist*, etc.). Nous comptons toutes les entrées comportant une traduction en russe, c'est-à-dire que nous incluons non seulement les racines mais tous les composés. Dans les grilles ci-dessous, on trouve les initiales en abscisse, les voyelles qui suivent en ordonnée. Suivent à chaque fois cinq colonnes numérotées de (a) à (e) :

(a) donne le total des occurrences de la ligne. Par exemple il existe 1 mot en *ga-*, 135 en *ğa-*, 3 en *ka-*, 1 052 en *qa-*, soit au total pour les mots en “gutturale” + voyelle /a/ : 1 191.

(b) donne le pourcentage du total de la ligne par rapport au total général du tableau. Par exemple, les 1 191 cas de gutturale + /a/ représentent 30,53 % de tous les mots commençant par une gutturale.

(c) fait la somme de ces pourcentages pour les voyelles postérieures, puis antérieures, puis /ä/. Ici, les gutturales + voyelle postérieures représentent

donc 61,42 % des cas de mots commençant par une gutturale, tandis qu'il n'y a que 33,04 % de mots en gutturale + voyelle antérieure.

(d) donne le pourcentage des initiales sonores sur l'ensemble de la ligne. Par exemple, sur les 1191 mots commençant par gutturale+/a/, ceux qui commencent par une sonore (g- ou ğ-) sont 1+135 : 11,4 %

(e) fait la moyenne (et non pas la somme) des pourcentages de mots en sonores pour les voyelles postérieures, puis antérieures, puis /ä/. On constate qu'il existe une grande différence entre la relative rareté des sonores+voyelle postérieure (5 %), et la relative abondance (presque 30 %, 6 fois plus) des sonores+antérieure.

Bien entendu, il convient de prendre avec prudence ce relevé, d'abord parce qu'il dépend du dictionnaire choisi ; ensuite parce que, malgré l'exclusion des emprunts récents au russe, les emprunts de diverses origines demeurent très nombreux, comme on peut le constater avec les mots en /ä/ ; enfin parce que le décompte comprend les composés, et gonfle donc abusivement l'importance des formes verbales. En effet, ce type de dictionnaire donne dans des entrées distinctes les divers composés verbaux.

Un dernier point préliminaire important est que ce dictionnaire donne parfois plusieurs orthographes d'un mot. Il donne par exemple pour "papillon" : *göbelek* et *gübelek* ; mais l'entrée *göbelek* renvoie à l'entrée *gübelek* qui seule comporte une traduction, et donc est seule prise en compte dans les tableaux. Ces entrées multiples reflètent des différences dans les divers parlers karakalpak.

	g-	ğ-	k-	q-	TOT	%	%		
					(a)	(b)	(c)	(d)	(e)
-a-	1	135	3	1 052	1 191	30,53	61,42	11,4	5,0
-ı-	0	19	0	479	498	12,77		3,8	
-o-	0	9	0	309	318	8,15		2,8	
-u-	0	8	0	381	389	9,97		2,1	
-e-	97	0	389	0	486	12,46	33,04	20,0	29,6
-i-	29	0	108	0	137	3,51		21,2	
-ö-	25	0	279	0	304	7,79		8,2	
-ü-	249	0	113	0	362	9,28		68,8	
-ä-	17	32	40	126	215	5,51	5,51	22,8	22,8
TOT	418	203	932	2 347	3 900	100	100,00		
%	11	5	24	60	100				

	d-	t-	TOT	%	%	%	%
			(a)	(b)	(c)	(d)	(e)
-a-	197	527	724	29,1	55,9	27	24
-i-	34	156	190	7,6		22	
-o-	43	217	260	10,5		20	
-u-	61	155	216	8,7		28	
-e-	48	177	225	9,0	34,0	21	30
-i-	80	154	234	9,4		34	
-ö-	31	72	103	4,1		30	
-ü-	100	186	286	11,5		35	
-ä-	154	97	251	10,1	10,1	61	61
TOT	748	1 741	2 489	100,0	100,0		
%	30	70	100				

On peut extraire de ces tableaux les remarques suivantes :

1/ les formes en sonore initiale sont plus fréquentes chez les dentales (*d-* 30 %) que chez les gutturales (*g-* + *ğ-* 16 %)

2/ elles sont plus fréquentes devant voyelle antérieure, en général.

Est-il possible de cerner le statut historique de ces réalisations sonores à l'initiale en karakalpak ? Il suffirait en principe de comparer des mots karakalpak à sonore initiale avec les mots analogues des autres langues turques. Voici des exemples :

	corps	lutte	oie	pelle	aveugle
<i>Clauson</i>	*gövdöŋ ?	*güreš	*ka:z ⁴⁶	*kürge:k	*gör- ⁴⁷
	G1	G2	G3	G4	G5
turc	<i>gövde</i>	<i>müdzaadele</i>	<i>kaz</i>	<i>kürek</i>	<i>kör</i>
azeri	<i>gövdä</i>	<i>müdzaadilä</i>	<i>gaz</i>	<i>küräk</i>	<i>kör</i>
türkmen	<i>gövre</i>	<i>göreš</i>	<i>gaaz</i>	<i>kürek</i>	<i>köör</i>
başkort	<i>gävdä ?</i>	<i>köräš</i>	<i>kaz</i>	<i>köräk</i>	(<i>hukır</i>)
tatar	<i>gäüdä</i>	<i>köräš</i>	<i>kaz</i>	<i>köräk</i>	(<i>sukır</i>)
nogaj	<i>kevde</i>	<i>küres</i>	<i>kaz</i>	?	(<i>sokır</i>)
kazak	<i>kevde</i>	<i>küres</i>	<i>kaz</i>	<i>kürek</i>	<i>kör</i>
karakalpak	<i>gewde</i>	<i>güres</i>	<i>gaz</i>	<i>gürek</i>	<i>gör</i>
kırgız	(<i>tulku</i>)	<i>küröš</i>	<i>kaz</i>	<i>kürök</i>	<i>kör</i>
ouzbek	<i>gävdä</i>	<i>küräš</i>	<i>gâz</i>	<i>küräk</i>	<i>kör</i>
ouïgour	<i>gävdä</i>	<i>küräš</i>	<i>gaz</i>	<i>küräk</i>	<i>kör</i>

	porc	genou	enfiler	piège	sel
<i>Clauson</i>	*doɲuz	*di:z	*diz-	? *duzak	? *du:z
	D1	D2	D3	D4	D5
turc	<i>domuz</i>	<i>diz</i>	<i>diz-</i>	<i>tuzak</i>	<i>tuz</i>
azeri	<i>donuz</i>	<i>diz</i>	<i>düz-</i>	<i>duzag</i>	<i>duz</i>
türkmen	<i>doɲuz</i>	<i>duz</i>	<i>düz-</i>	<i>duzak</i>	<i>duz</i>
baškort	<i>duɲgız</i>	<i>tez</i>	<i>tiz- ?</i>	<i>tozak</i>	<i>toz</i>
tatar	<i>duɲgız</i>	<i>tez</i>	<i>tez-</i>	<i>tozak</i>	<i>toz</i>
nogaj	<i>doɲız</i>	<i>tiz</i>		<i>tuzak</i>	<i>tuz</i>
kazak	<i>doɲız</i>	<i>tize</i>	<i>tiz-</i>	<i>tuzak</i>	<i>tuz</i>
karakalpak	<i>doɲuz</i> ⁴⁸	<i>dize</i>	<i>diz-iw</i>	<i>duzak</i>	<i>duz</i>
kirgız	<i>tongız</i>	<i>tize</i>	<i>tiz-</i>	<i>tuzak</i>	<i>tuz</i>
ouzbek	<i>toɲguz</i>	<i>tizzä</i>	<i>tiz-</i>	<i>tuzâk</i>	<i>tuz</i>
ouïgour		<i>tiz</i>	<i>tiz-</i>	<i>tozak</i>	<i>tuz</i>

À première vue, de telles listes seraient favorables à l’hypothèse “conservatrice” : la plupart des mots cités correspondent à des réalisations sonores en oguz. Toutefois G4 et G5 font exception, et d’autres aussi, comme “papillon” *göbelek*. En outre, en G1 et D1 presque toutes les langues ont une sonore initiale, ce qui révèle une situation transitionnelle. Cela signifie que dans presque toutes les langues turques, à des degrés très divers, les locuteurs sentent une différence de prononciation entre sourde et sonore initiale. Quand les cas sont très minoritaires, il ne peut pas s’agir d’une réelle opposition phonologique, où par exemple un mot ***das* s’opposerait par ce seul trait de sonorité initiale à un mot ***tas*. Nos **** marquent que ces “mots” sont fictifs. Il est même très difficile de trouver de tels exemples, et toujours quand on en trouve l’un des deux est un emprunt à une langue non-turque qui possède beaucoup d’initiales sonores, comme le persan ou l’arabe ou le russe. Un exemple classique est en karakalpak (ou en kazak) l’opposition de *kül* “cendre” et *gül* “fleur”, ce dernier étant un emprunt persan.

Il est indéniable que beaucoup de mots turcs à initiale sonore sont des emprunts, et que dans les langues où ces emprunts ont été massifs, il a pu se créer un clivage nouveau, dont on peut parfaitement concevoir qu’il ait gagné même des mots d’origine turque. Voici des exemples de mots turcs (de Turquie) qui sont certainement des emprunts :

turc	karakalpak	tadjik	
<i>gaga</i>	(<i>tumsık</i>)	(<i>nuul</i>)	bec (d’oiseau)
<i>gam</i>	(<i>kajgı</i>)	<i>γam</i>	chagrin, regret
<i>gıdżık</i>	<i>kıtıkla-w</i>	<i>qıtiq</i>	irritation de la gorge
<i>gıdżır</i>	(<i>kajraw</i>)	<i>γidżirrosi</i>	gomme à mastiquer

turc	karakalpak	tadjik	
gıda	(ajkat)	γizo	aliment, portion
gına	(kurgın)	(xušholii)	chant, contentement
gır < arabe			jaserie, mensonge
gırtlak	kegirdek ?	(rohi nafas, gardan)	gosier, gorge
gıjap	(žoklık)	γajp	absence, défaut
godžuk	(ton, postin)	(puustin)	pelisse grossière
gurup	(batırw)	γurub	coucher du soleil
gurur	ar (šillıq)	γurur	orgueil, présomption

On remarquera que ceux qui ne sont pas directement d'origine arabe ressemblent de près à des mots persans (ici la forme tadjike – ces mots iraniens pouvant eux-mêmes être empruntés à l'arabe), et d'ailleurs qu'ils sont souvent absents du lexique kıpčak, ici représenté par le karakalpak.

Cette sensibilité aux réalisations sonores, pour être “secondaire” et créée ou développée par des emprunts, explique certainement que dans des conditionnements phonétiques particuliers certaines consonnes aient pris l'allure des sonores d'abord étrangères. Car de tels phénomènes se produisent même dans la norme kazak, qui est certainement une des moins riches en initiales sonores. Dans les trois exemples ci-dessous, la sonorisation de l'initiale est probablement liée à la consonne labiale (w, b, p) qui suit :

	voix	sac	balle, ballon
<i>Clauson</i>	*tawiš ⁴⁹	?	*to:b
turc	sedaa	torba	top
azeri	sādaa	torba	top
türkmen	sedaa	toorba	top
baškort	tawiš	torba tok	tup
tatar	tawiš	(kapčık)	tup
kazak	dıbıs	dorba	dop
karakalpak	dawıs	?	top
kırgız	dobuš	torbo	top
ouzbek	tovuš	torvā	top
ouïgour	tavuš	torva	top
tadjik	sado	tuurba	tuub
vieux türk			top

V. L'énigme turke et la contribution karakalpak

Selon Aleksandr Mihajlovič Šerbak, le grand spécialiste russe de la linguistique turke comparée⁵⁰, il n'existe tout simplement pas de sonore initiale étymologique en proto-turk. Celles qui existent en oguz sont la conséquence d'un développement plus intense d'une tendance répandue partout, et pour laquelle il ne donne pas d'explication.

Je ne sais pas si l'on peut être aussi radical. Dire que les initiales sonores sont "secondaires", c'est-à-dire un résultat ultérieur, et ne pas proposer de conditionnement historique qui les ait produites (ce qu'on a fait au contraire pour les sonores qui suivent les voyelles longues), aboutit à une impasse. Il faut donc se résoudre à une solution mixte, et envisager à la fois des sonores empruntées et des sonores originales.

Il ne fait aucun doute qu'une partie des sonores initiales soit le résultat d'emprunts, parfois massifs, ni que ces emprunts massifs aient entraîné une prononciation "sonore", dans quelques cas, de mots d'origine turke.

Mais il reste beaucoup de cas, en oguz et ailleurs, et surtout quand il existe une corrélation entre oguz et karakalpak (de même qu'autrefois Ligeti mettait en évidence celle qui unissait le türkmen et le yakoute pour les voyelles longues), où les initiales sonores paraissent plus anciennes que les emprunts d'époque musulmane. Clauson admet la présence de *d- et *g- en "pre-eighth century turkish"⁵¹.

Enfin, à mi-chemin entre ces sonores étymologiques qu'il faut admettre, puisqu'on ne saurait les expliquer autrement, et des sonores empruntées avec les mots qui les portent, il existe aussi, diversement selon les parles, des initiales sonores dont l'apparition repose sur le croisement d'un "terrain" général favorable aux sonores initiales (à cause des emprunts par exemple) et des conditionnements phonétiques particuliers (des sonores intervocaliques par exemple). C'est sans doute dans ce cadre qu'il faut comprendre notre "papillon" karakalpak, *göbelek*. En se souvenant que de tels mots sont souvent aussi ceux des Özbek du delta.

Un des faits les plus étonnants qui apparaît, à mesure qu'on connaît moins mal la diversité des langues turkes, qui sont si proches les unes des autres, c'est que cette proximité est elle-même énigmatique⁵². Certes, comme nous l'avons vu plus haut, la proximité entre elles s'explique en partie par leur "conservatisme" relatif, qui tient lui-même à la dispersion des locuteurs et donc à leur mode de vie. Mais même si la démographie des populations turkophones d'autrefois rend compte de cette proximité continuée, il reste qu'il y a eu une époque, sans doute vers le début de l'ère chrétienne, où les langues turkes formaient un réseau intercompréhensible. Pour dire les choses d'une façon plus frappante : les langues turkes sont récentes.

Les historiens romantiques voyaient naguère les tribus sauvages qui déferlaient sur l'Europe comme l'émergence d'êtres venus de la nuit des temps. Mais pour le linguiste, c'est le contraire qui saute aux yeux. Sans qu'il

puisse préjuger des mouvements complexes de peuples, leurs unions et leurs ruptures sur un si vaste espace, il est clair que la formation des langues turkes – quant à elles – est un phénomène relativement récent. La dispersion, le mode de vie, les épisodes historiques dramatiques que leur histoire connaît rendent largement compte de l’extension et des recoupements qu’on observe d’un bout à l’autre de la zone actuelle de leurs parlers.

On a parfois suggéré que le proto-turk avait pu être un sabir des steppes, une *lingua franco* qui a eu du succès, ce qui expliquerait sa présence parmi des populations par ailleurs assez différentes, sa réceptivité aux emprunts ; et aussi sa formation récente. Naturellement, un créole a aussi des sources, ce qui repousse la discussion dans d’autres régions (les Mongols, les Toungouses, les parlers ouraliens orientaux). Et une mixité est aussi une histoire. Mais il est certain que lorsqu’on remonte par la comparaison des langues l’histoire des parlers turks, on arrive assez vite à un horizon sur quoi l’on butte et l’histoire des sonores initiales en est un exemple : avec elle, et avec les sonores des Karakalpak du sud-ouest, nous touchons à cet horizon. Si le long débat toujours ouvert sur le statut historique du lien entre langues turkes, mongoles et toungouses (les langues dites ensemble “altaïques”) et aussi avec certaines au moins des langues ouraliennes ne parvient pas à se conclure, c’est pour la même raison.

Il est probable que l’histoire des sonores initiales du turk, pour difficile qu’elle soit, et encore largement mystérieuse, repose sur le chevauchement de plusieurs époques. Celles où les sonores progressaient, celles où elles régressaient et se confinaient dans certains mots et chez certaines ethnies. Le conservatisme ethnique des Karakalpak est des plus remarquables. J’ai déjà souligné qu’ils sont, avec les Türkmén sans doute, parmi les populations turkes d’aujourd’hui qui ont conservé de façon très vivante leur structure clanique. Cela n’est pas décisif, mais c’est certainement un élément du dossier. Car même si le linguiste souligne toujours que l’histoire des langues, parce qu’elles s’apprennent et s’oublient, suit des voies différentes de celle des groupes humains qui les parlent à un moment de leur histoire, il reste vrai bien sûr que l’histoire des langues et celle des populations sont intimement mêlées.

Ainsi, isolés aujourd’hui dans une région que tous proclament sinistrée, derniers tenants des anciennes cultures riveraines de la mer d’Aral, les Karakalpak témoignent avec leur faible voix de la sonorité maintenue de certaines initiales. On pourrait sourire de ce détail fragile, de ce “papillon” presque perdu, comme d’une note légèrement décalée dans un concert orchestré par d’autres puissances. Mais la persistance de ce détail n’est pas seulement émouvante, car elle fait écho à d’autres voix, chez les Oguz notamment, et cette résonance acquiert par les rapprochements une valeur unique dans cette enquête sur l’horizon historique des peuples de l’Asie centrale.

François Jacquesson
LACITO – CNRS

NOTES

1. Une partie des données que cet article contient (section IV. 1.) a été obtenue dans le cadre du “Projet NSH” du CNRS. Voir la note consacrée à ce Projet sous IV.1. Une première version a été rédigée à Tachkent, grâce à la bibliothèque de l’Institut Français d’Études sur l’Asie centrale (IFÉAC), et à celle de Svetlana Jacquesson, que je remercie chaleureusement pour ses conseils.
2. Voir là-dessus la position prudente de Peter Golden (Golden 1992 : 403-4). Cet historien est revenu plus récemment sur la question des équivalences supposées, mais certainement fausses, entre Karakalpak et “Capuchons Noirs”, *Černii Klobouci* (Golden 1996).
3. Berezin 1848 : 14-15.
4. Thomsen (1842-1927) en a raconté les circonstances dans sa première publication (Thomsen 1893), dans laquelle il publiait sa lecture de l’alphabet. Son rival russe, Friedrich Wilhelm Radloff (1837-1918), s’est empressé de publier des traductions grâce à ce déchiffrement, ce dont Thomsen se plaint dans ses publications plus complètes de 1894-96. Sur les écritures du turk ancien, voir Gabain 1974.
5. Baskakov 1932 : 11.
6. Dans Baskakov 1932, le karakalpak est dans son nouvel alphabet latin.
7. On les trouve reproduites dans Baskakov 1952 : 128-9.
8. Sous le titre *Karakalpakskij âzyk*, Baskakov lui-même n’a publié que deux volumes : Baskakov 1951 et 1952. Les savants de Noukous publieront bien plus tard la troisième partie : Baskakov 1995.
9. Ce dictionnaire comporte d’ailleurs une table de correspondances entre les orthographes anciennes et nouvelles.
10. API : Alphabet Phonétique International, utilisé par les linguistes du monde entier lorsqu’il s’agit de décrire la prononciation exacte des parlars.
11. “Postérieur” signifie ici : prononcé plus en arrière dans la bouche, plus vers le fond de la bouche. Il est donc logique que les “gutturales” postérieures accompagnent les voyelles postérieures.
12. Dans l’ordre alphabétique, elle vient en deuxième position, après le A.
13. Clauson 1972 : 728a.
14. Le mot “kazak” est écrit kazax en russe, ce qui se transcrit *kazakh* ou *kazax*.
15. Cette notation par “j” du son français “y” est une des caractéristique de l’API. Les savants allemands et slaves ont joué un rôle important dans la création de l’API.
16. Cette entreprise consistant à repérer les emprunts est associée surtout aux noms de Gerhard Doerfer et de Sir Gerard Clauson. L’œuvre de ce dernier savant culmine dans son *Etymological Dictionary of Pre-Thirteenth Century Turkish* (Clauson 1972), qui rassemble tous les mots turks connus avant la grande période des emprunts. La méthode de cet ouvrage est expliquée dans Clauson 1962.
17. Johanson 1998 : 87.
18. Pour le khorezmien, voir Humbach, dans Schmitt 1989. Pour le sogdien, consulter Sims-Williams dans le même ouvrage. Ces deux articles contiennent des bibliographies. Les inscriptions khorezmiennes les plus anciennes, en alphabet araméen, sont du II^e siècle de notre ère. Elles ont été trouvées dans le palais de la ville (aujourd’hui en ruines) de Toprak-kala. Les plus récentes, en alphabet arabe, sont du XIII^e ou XIV^e siècle.
19. Golden 1992 : 195.
20. S. E. Malov, *Zametki o karakalpakskom âzyke*, ed. Baskakov & Nasyrov, Noukous, 1966. Cette trentaine de pages n’ont été publiées qu’en 1966. Les deux éditeurs expli-

- quent que, malgré l'intention de Malov de les publier sur place (le manuscrit dactylographié est à Noukous), les circonstances ont contraint l'auteur à n'en publier qu'un résumé : Malov 1934.
21. Tolstov S. P. ; Ždanko T. A. et als. (Eds.), *Očerki istorii Karakalpakskoj ASSR* [Études sur l'histoire de la RSSA du Karakalpakistan], Tachkent : Fan, 1964 ; vol. II, pp. 134-135.
 22. Voir l'article "The Steppe", de William H. McNeill, dans *Encyclopedia Britannica 2001*.
 23. Voir sur ce thème Jacquesson 1999, 2001.
 24. J'inclus dans ces tableaux les mots tels que G. Clauson les reconstitue dans *An Etymological Dictionary of Pre-Thirteenth-Century Turkish*. Les voyelles suivies de "·" sont des voyelles longues.
 25. Les Russes disent "baškir", et cette prononciation s'est ensuite glissée dans les encyclopédies et les publications européennes. Mais personne n'a jamais utilisé ce mot chez les "Baškir", où l'on a toujours dit baškort. C'est donc cette prononciation que j'utilise ici.
 26. En particulier Pelliot 1920.
 27. Klaproth 1828 : 115-6.
 28. Il s'agit de Guillaume de Rubrouck. Traduction en français : Kappler 1993.
 29. Plusieurs d'entre eux sont indiqués à la fin de cet article. Le but de ces normes officielles étant précisément de "dépasser" la diversité dialectale, elles n'indiquent qu'une partie de la vérité, mais elles permettent quand on les rassemble de donner une idée de la gamme des faits. La part de diversité dissimulée est plus ou moins grande selon les cas. Un exemple de grande diversité dialectale est celui du baškort (Bajšev 1955), où l'on a choisi pour la norme une série de traits originaux par rapport au kupčak courant.
 30. Le tatar de Crimée, longtemps, n'a pas eu de dictionnaire parce que cette population a été proscrite par Staline lors de la Deuxième Guerre mondiale. Encore de nos jours, cet ostracisme n'est pas complètement levé, et la plupart des "Tatars de Crimée" vivent en exil. Alije Akimova travaille à l'Institut Français d'Études sur l'Asie centrale (IFÉAC), à Tachkent.
 31. La première description systématique de ce fait semble être dans Vilhelm Grønbech (1902). John Krueger a donné une traduction anglaise de cet important ouvrage du savant danois : Grønbech 1979.
 32. Kaare Grønbech (fils de Vilhelm), qui a publié par ailleurs un fac-simile du mss., reconstruit dans son *Wörterbuch* des normes turkes. Il donne également les graphies du mss. La première édition complète du *Codex* (celle de Klaproth ne l'était pas, et comportait des fautes) est celle de Kuun 1880 : *Codex cumanicus bibliothecae ad templi divi Marci Venetiarum*. Voir aussi Gabain, in Deny 1959.
 33. Clauson et le türkmen ne sont pas toujours d'accord pour les voyelles longues.
 34. Du moins dans la norme, qui correspond ici aux parlers du sud-est. Il existe plusieurs parlers baškort où l'on prononce /s/, /θ/, /ð/. Voir la carte dans Bajšev 1955, entre pages 64-65.
 35. Certains linguistes, dont Šerbak, pensent qu'il ne s'agit pas exactement d'une innovation, mais qu'au contraire ce sont les *j*- initiaux qui sont une innovation. Ce problème implique les autres groupes de langues turkes que nous n'abordons pas ici (les langues de l'Altaï, le tchouvache). Voir la position de Šerbak dans Šerbak 1970 :158 sqq.
 36. Benzing et Menges dans Deny 1959, Baskakov dans Baskakov 1962.
 37. Il existe deux descriptions du karakalpak en anglais. La première est de Menges 1947. Karl Menges avait fait deux voyages en Russie. Au retour de son deuxième voyage,

- en 1934, il avait écrit une description en allemand. Mais en 1936 il dut fuir l'Allemagne en abandonnant la plupart de ses notes et toute sa bibliothèque, de sorte qu'il ne put publier qu'une phonologie, bien plus tard, grâce au Cercle de linguistique de New York. Stefan Wurm, qui avait aussi étudié la langue, sachant que Menges avait perdu ses archives, publia aussitôt son propre travail, Wurm 1951. Aucune des deux études ne porte une attention particulière au problème des sonores initiales.
38. *Op. cit.*, pp. 58-59.
 39. Le Projet NSH, de son nom complet "Echanges linguistiques et génétiques entre Orient et Occident au Nord et au Sud de l'Himalaya" est un projet pluridisciplinaire du CNRS, dans le cadre de l'action "Origine de l'Homme, du Langage, des Langues", sur quoi on peut consulter le site web du CNRS. Le Projet NSH regroupe des linguistes, dont l'auteur de cet article, des ethnologues, dont la responsable de ce dossier des *Cahiers d'Asie centrale*, et des généticiens du Musée de l'Homme et de l'Institut Pasteur. Il concerne d'une part l'Asie centrale, et d'autre part certaines régions de l'Inde. Pour l'Asie centrale et en particulier l'Ouzbékistan, il travaille avec l'appui de l'Institut Français d'Études sur l'Asie centrale (IFÉAC), et en liaison avec deux Instituts de l'Académie des sciences d'Ouzbékistan : l'Institut d'histoire, d'archéologie, et d'ethnographie de la filiale de Noukous (Karakalpakistan) dirigé par le professeur Vadim Nikolaevič Āgodin, et l'Institut d'immunologie de Tachkent dirigé par le professeur Ruslan Ruzibakiev, assisté par Tatiana Hegaj.
 40. Ligeti explique dans son article qu'il a signalé sa découverte dès 1936, juste avant de partir pour un voyage d'études en Afghanistan. Son article n'a été publié qu'après son retour.
 41. Baskakov 1951 : 4.
 42. Je remercie Amantur Džaparov, secrétaire scientifique de la Section des sciences humaines de l'Académie des sciences du Kirgizstan, à qui je dois cette information.
 43. Nasyrov 1975 et 1983.
 44. Bakinova 1959 : 34-36.
 45. Bakinova 1956 : 36-38.
 46. « Almost certainly an early loan-word from some Indo-European language, probably Tokharian (...) », Clauson 1972 : 679b.
 47. « Obviously a very old etymological connection with *kō:z* "eye" », Clauson 1972: 736a.
 48. La graphie est *t-*, mais la prononciation commune est *d-*.
 49. Clauson 1972 : 446a signale « a wide range of phonetic changes », à commencer par *t-/d-*.
 50. Šerbak 1970.
 51. Clauson 1962 : 170-1.
 52. Johanson 1992.

OUVRAGES CITÉS

- Abdullaev 1967 : Abdullaev F. A., *Fonetika horezmskih govorov* [Phonétique des parlers du Khorezm], Tachkent, 1967.
- Amanžolov 1997 [1958] : Amanžolov S., *Voprosy dialektologii i istorii kazahskogo âzyka* [Questions de dialectologie et d'histoire du kazak], Almaty, 1997 ; 607 p.
- Aralbaev 1988 : Aralbaev Ž., *Qazaq fonetikasi bojnša etjudter* [Études sur la phonétique kazak], Almaty, 1988 ; 142 p.
- Bajšev 1955 : Bajšev T., *Baškirske dialekty v otnošenii k literaturnomu âzyku* [Les dialectes baškort (baškirs) en relation avec la langue littéraire], Moscou, 1955 ; 112 p. et cartes.
- Baskakov 1932 : Baskakov N. A., *Kratkaâ grammatika karakalpakskogo âzyka* [Petite grammaire de la langue karakalpak], Turtkul', 1932 ; 150 p.
- Baskakov 1951 : Baskakov N. A., *Karakalpakskij âzyk I : materialy po dialektologii* [La langue karakalpak I : matériaux de dialectologie], Moscou, 1951 ; 411 p.
- Baskakov 1952 : Baskakov N. A., *Karakalpakskij âzyk. II : fonetika i morfologiâ* [La langue karakalpak II : phonétique et morphologie], Moscou, 1952 ; 543 p.
- Baskakov 1962 : Baskakov N. A., *Vvedenie v izučenie tûrkskih âzykov* [Introduction à l'étude des langues turkes], Moscou, 1962 ; 332 p.
- Baskakov 1995 : Baskakov N. A., *Karakalpakskij âzyk III : morfologiâ, slovoizmenenie, sintaksis, slovosocetanie i predloženie* [La langue karakalpak III : morphologie, morpho-phonologie, syntaxe, composition nominale et énoncé], Noukous, 1995 ; 291 p.
- Berezine 1848 : Berezine E., *Recherches sur les dialectes musulmans I : Système des dialectes turcs*, Casan [Kazan], 1848 ; 95 p.
- Clauson 1962 : Clauson G., *Türkish and Mongolian Studies*, London : Royal Asiatic Society, 1962 ; 261 p.
- Deny 1959 : Deny J. ; Grønbech K. ; Scheel H. ; Togan Z. V. (Eds.), *Philologiae Turcicae Fundamenta I*, Aquis Mattiacis : Steiner, 1959 ; 813 p.
- Gabain 1959 : Gabain Annemarie von, "Die Sprache des Codex cumanicus", dans : Deny 1959 ; pp. 46-73.
- Gabain 1974 : Gabain Annemarie von, *Alttürkische Grammatik*, 3 Aufl., Berlin, 1974 ; 398 p.
- Golden 1992 : Golden P. B., *An Introduction to the History of the Turkic Peoples*, Wiesbaden : Harrassowitz, 1992 ; 483 p.
- Golden 1996 : Golden P. B., "The Černii Klobouci", dans : Berta A.; Brendemoen B. ; Schönig C. (Eds.), *Symbolae Turcologicae : Studies in Honour of Lars Johanson*, Stockholm : Swedish Research Institute in Istanbul, 1996 ; 246 p. (*Transactions*, vol. VI)
- Grønbech 1902 : Grønbech V., *Forstudier til tyrkisk lydhistorie*, Copenhagen, 1902 ; 122 p. Traduction anglaise par John R. Krueger, *Preliminary Studies in Turkic Historical Phonology*, Bloomington, 1979, 162 p. (*Uralic and Altaic Series*, vol. 135)
- Humbach 1989 : Humbach H., "Choresmian", dans : Schmitt 1989 ; pp. 193-203.

- Jacquesson 1999 : Jacquesson F., "L'évolution des langues dépend-elle de la densité des locuteurs ?", *Études Finno-Ougriennes*, n° 31 ; pp. 27-34 (Actes du Colloque *Gyarmathi*).
- Jacquesson 2001 : Jacquesson F., "Pour une linguistique des quasi-déserts", dans A. M. Loffler-Laurian (Ed.), *Études de linguistique générale et contrastive : hommage à Jean Perrot*, pp. 199-216.
- Johanson 1992 : Johanson L., *Strukturelle Faktoren in türkische Sprachkontakten*, 1992.
- Johanson 1998 : Johanson L. & Csató É.Á. (Eds), *The Turkic Languages*, London ; New York : Routledge, 1998 ; 474 p. (*Routledge Language Family Descriptions*)
- Kappler 1993 : Kappler Claude-Claire & René, *Guillaume de Rubrouck : voyage dans l'empire mongol, 1253-1255*, Paris : Imprimerie nationale, 1993 ; 301 p.
- Klaproth 1828 : Klaproth J.-von, *Mémoires relatifs à l'Asie, contenant des recherches historiques, géographiques et philologiques sur les peuples de l'Orient III*, Paris, 1828 ; 520 p.
- Kuun 1880 : Kuun G., *Codex cumanicus bibliothecae ad templi divi marci Venetiarum*, Budapest, 1880 ; 395 p.
- Lebedynsky 2001 : Lebedynsky I., *Les Scythes : la civilisation des steppes (VII^e-III^e siècles av. J.-C.)*, Paris : Errance, 2001 ; 267 p.
- Malov 1934 : Malov S. E., "Karakalpaksčij âzyk i ego izučenie" [La langue karakalpak et son étude], dans : *Karakalpakiâ : trudy pervoj konferencii po izučeniû proizvoditel'nyh sil Karakalpakskoj ASSR* [Karakalpakie : travaux de la première conférence sur l'étude des forces productrices de la RSSA du Karakalpakistan], Leningrad : AN SSSR, 1934 ; vol. II, pp. 200-207.
- Malov 1966 : Malov S. E., *Zametki o karakalpakskom âzyke* [Notes sur le karakalpak], Noukous, 1966.
- Menges 1947 : Menges Karl, *Qaraqalpaq Grammar I : Phonology*, New York, 1947, 110 p. et carte.
- Nasyrov 1974 : Nasyrov D. S., "O rudimentah âzyka rodoplemennyh ob"edinenij v dialektah i govorah karakalpakskogo âzyka" [Les traces de la langue des groupes claniques dans les dialectes et parlers de la langue karakalpak], *Vestnik KKFAN UzSSR*, n° 1, 1974, pp. 64-67.
- Nasyrov 1975 : Nasyrov D. S., "O rabote na dialektologičeskim atlasom karakalpakskogo âzyka" [Les travaux de l'atlas dialectologique du karakalpak], *Vestnik KKF AN UzSSR*, n° 1, 1975, pp. 70-74.
- Nasyrov 1983 : Nasyrov D. S., "Izučenie dialektov i govorov karakalpakskogo âzyka metodom lingvističeskoj geografii" [Étude des dialectes et des parlers du karakalpak avec la méthode de la géographie linguistique], *Vestnik KKF AN UzSSR*, n° 3, 1983, pp. 68-72.
- Nurmagambetov 1978 : Nurmagambetov Ê., *Kazak tili govorlarınıń batis tobi* [Le groupe occidental des parlers kazak], Almaty, 1978 ; 216 p.
- Pelliot 1920 : Pelliot P., "À propos des Comans", *Journal asiatique*, XV, 1920 ; pp. 129-195.
- Samojlovič 1922 : Samojlovič A., *Nekotorye dopolneniâ k klassifikacii tureckih âzykov* [Quelques compléments à la classification des langues turkes], Petrograd, 1922 ; 15 p.

- Ŝerbak 1970 : Ŝerbak A. M., *Sravnitel'naâ fonetika tûrkskih âzykov* [Phonétique comparée des langues turkes], Leningrad : Nauka, 1970 ; 204 p.
- Schmitt 1989 : Schmitt R. (Ed.), *Compendium linguarum iranicarum*, Wiesbaden : Reichert, 1989 ; 529 p.
- Sims-Williams 1989 : Sims-Williams N., "Sogdian", dans : Schmitt 1989, pp. 173-192.
- Thomsen 1893 : Thomsen V., "Déchiffrement des inscriptions de l'Orkhon et de l'Iénisséi : notice préliminaire", *Bulletin de l'Académie royale des sciences et des lettres de Danemark*, 1893, pp. 285-299.
- Wurm 1951 : Wurm Stefan, "The Karakalpak Language", *Anthropos*, 46, 1951, pp. 487-610.

DICTIONNAIRES CITÉS

- Abdullaev 1994 : Abdullaev Ė. ; Umerov M., *Russko-krymskotatarskij učebnij slovar'* [Dictionnaire scolaire russe-tatar de Crimée], Simferopol', 1994 ; 384 p.
- Bakinova 1956 : Bakinova G., *Kirgiz tilinin Oš govorloru* [Les parlers kirgiz d'Oš], Frunze, 1956 ; 126 p.
- Bakinova 1959 : Bakinova G. ; Kondučalova S. ; Sıdıkov S., *Kirgiz tilinin Čüj gororu* [Le parler kirgiz du Čüj], Frunze, 1959 ; 128 p.
- Baskakov 1956 : Baskakov N. A. & al., *Russko-nogajskij slovar'* : 20 000 slov [Dictionnaire russe-nogaj : 20 000 mots], Moscou, 1956 ; 735 p.
- Baskakov 1958 : Baskakov N. A., *Karakalpaksko-russkij slovar'* : 30 000 slov [Dictionnaire karakalpak-russe : 30 000 mots], Moscou, 1958, 892p., suivi d'un *Očerk grammatiki karakalpakskogo âzyka* [Essai de grammaire karakalpak], pp. 789-866, et d'une *Ključ k novoj karakalpakskoj orfografii utverždennoj v 1957* [Clef pour la nouvelle orthographe du karakalpak promulguée en 1957], pp. 867-888.
- Baskakov 1967 : Baskakov N. A. (Ed.), *Russko-karakalpakskij slovar'* : 47 000 slov [Dictionnaire russe-karakalpak : 47 000 mots], Moscou, 1967 ; 1124 p.
- Baskakov 1974 : Baskakov N. A., Zajončkovskij A., Žapšal S. M., *Karajmsko-russko-pol'skij slovar'* : 17400 slov [Dictionnaire karaïm-russe-polonais : 17400 mots], Moscou, 1974 ; 688 p.
- Clauson 1972 : Clauson G., *An Etymological Dictionary of Pre-Thirteenth-Century Turkish*, Oxford : Clarendon Press, 1972 ; 989 p.
- Gazizov 1971 : Gazizov R. S., *Russko-tatarskij slovar'* : 50 000 slov [Dictionnaire russe-tatar : 50 000 mots], Kazan, 1971 ; 803 p.
- Grønbech 1942 : Grønbech K., *Komanisches Wörterbuch*, København, 1942 ; 315 p.
- Ūdahin 1965 : Ūdahin K. K., *Kirgizsko-russkij slovar'* : 40 000 slov [Dictionnaire kirgiz-russe : 40 000 mots], Moscou, 1965 ; 973 p.
- Karimova 1954 : Karimova G. R., *Russko-baškirkij slovar'* : 14 000 slov [Dictionnaire russe-baškort : 14 000 mots], Moscou, 1954 ; 600 p.
- Musabaev 1978 : Musabaev G. G. (Ed.), *Russko-kazahskij slovar'* : 65 000 slov [Dictionnaire russe-kazak : 65 000 mots], Almaty, 1978 et 1981 ; 2 vols.
- Tenišev 1989 : Tenišev E. R. ; Süjünč X. I., *Karačaj-malkar-orus sözlük / Karačævo-balkarsko-russkij slovar'* : 30 000 slov [Dictionnaire karačaj-balkar-russe : 30 000 mots], Moscou, 1989 ; 832 p.