Le système linguistique des langues slaves de l'Ouest (Nord et Sud)

P. Pognan

« L'unité linguistique du slave est évidente. Aujourd'hui encore elle apparaît partout au premier coup d'œil ; et, si l'on examine les formes anciennes des langues slaves, les ressemblances sont si grandes qu'elles approchent souvent de l'identité. »

Meillet, *Le slave commun*, introduction, page 2

1 Situer l'apprentissage d'une langue dans le système du groupe de langues

L'enseignement des langues, notamment en France, se fait très généralement de manière « tubulaire » pour reprendre l'expression de Pierre Escudé, spécialiste de l'intercompréhension romane. Trop souvent, l'information donnée est « hors contexte », située dans les limites strictes de la langue étudiée sans même de renvois à sa propre histoire. Ce type d'approche prive d'une vue d'ensemble, ce qui est préjudiciable à une compréhension intime de la langue.

C'est pourquoi, dans le cadre d'une recherche commune dotée d'un PHC Stefanik franco-slovaque « Conception d'une nouvelle grammaire slovaque pour les francophones », nous avons tenté de définir le système linguistique des langues slaves de l'Ouest, d'y situer le slovaque et de donner à ce système linguistique slave une existence pratique dans nos expériences d'enseignement.

D'autres expériences nous y ont aidés, notamment la recherche menée en parallèle sur le slovène dans le cadre d'un PHC Proteus franco-slovène, « Grammaire du slovène à l'usage des francophones – Morphologie ».

Les langues slaves constituent un excellent laboratoire pour l'étude du multilinguisme [1]. Les études en intercompréhension des langues romanes sont nombreuses, débutées il y a déjà quelques décennies avec des pionniers tels que Paul Teyssier, Louise Dabène et Claire Blanche-Benvéniste. Il est étonnant que l'intérêt pour l'inter-

1. Elles ne sont pas les seules. Les langues germaniques, les langues turques, les langues fenniques, les différentes langues berbères se prêtent bien à une modélisation du système linguistique.

compréhension slave ne prenne de l'ampleur que maintenant. Mais cela peut constituer une nouvelle donne pour l'étude du système linguistique.

En fait, l'intercompréhension slave existe naturellement, au moins jusqu'à un certain point, qui en certaines périodes a pu être améliorée par une vie fédérative (Tchécoslovaquie, Yougoslavie).

« *Dobra kniha druhdy ryzy złoto přesaha* » est un énoncé qui, si vous êtes Slave ou slavisant, ne sera en aucun cas hermétique. Vous en comprendrez le sens global ou un sens plus précis, voire tout à fait exact suivant la proximité géo-linguistique de vos connaissances d'une ou plusieurs langues slaves. Malgré cette compréhension à l'écrit (la compréhension orale d'une langue n'est pas identique à celle de l'écrit, cette dernière étant généralement meilleure), il n'est pas sûr que vous puissiez déterminer, au moins de manière immédiate, de quelle langue il s'agit. Simplement, vous savez au minimum que cette langue est « slave [2] ».

À une certaine époque, les ménagères tchécoslovaques avaient dans leur cuisine de nombreux produits venant de Yougoslavie. Ces produits portaient des inscriptions en caractères latins. Les recettes étaient préparées à l'aide d'un mode d'emploi compréhensible rédigé dans une langue non spécifiquement déterminée.

Cette compréhension peut être variable suivant les langues considérées, mais même si la distance est grande, même s'il n'y a pas (ou plus) l'habitude du contact de l'autre langue, il y a un fond commun qui devrait permettre d'éviter entre Slaves le recours à l'anglais pour communiquer ou encore mieux pour accéder à une information. Malheureusement, ce n'est pas ce qui est fait en général.

Une vision globale des langues slaves est utile à leur traitement automatique, à leur enseignement (en tant que langue maternelle ou étrangère) et à leur comparaison. La connaissance du système linguistique nous semble nécessaire pour la compréhension de ces langues. En effet, ceux qui pratiquent le mieux une langue, de manière parfaitement intuitive, à savoir les locuteurs natifs, ne connaissent généralement pas ou peu ou mal le système et les règles qui sous-tendent cette langue. De manière surprenante, l'école – quelle qu'elle soit et où qu'elle soit – n'arrive pas à ancrer ces connaissances dans la (sub-)conscience des locuteurs. La connaissance du système linguistique du groupe, connaissance de ce qui est commun et connaissance de ce qui est divergent, permet une bien meilleure compréhension de chaque langue du groupe, des phénomènes ténus dans une langue pouvant être éclairés par la présence massive de ce même phénomène dans une autre langue. Par exemple, la présence de la double mouillure en haut-sorabe fait prendre conscience que celle-ci est impossible en tchèque, que la réalisation fluctuante de la métathèse en haut-sorabe met l'accent sur le fait que la métathèse est constante en tchèque.

Nous reviendrons sur ce genre de phénomènes dans le cadre général du groupe de langues considérées, les langues slaves de l'Ouest Nord et Sud.

2. La solution sera donnée à la fin de ce chapitre.

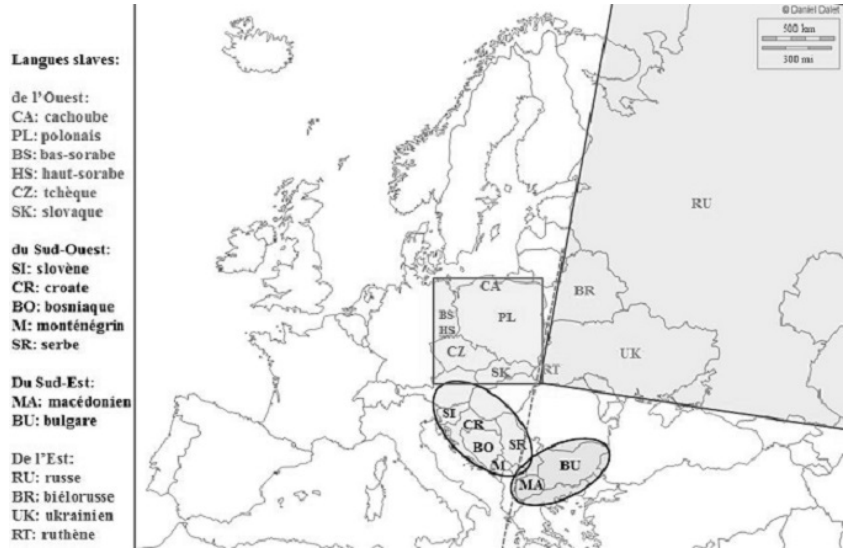


Figure 1 – Répartition des langues slaves

2 Les langues concernées

2.1 Répartition des langues slaves

De manière traditionnelle, les langues slaves sont divisées en trois groupes : langues slaves de l'Est (russe, biélorusse, ukrainien et ruthène – souvent oublié), langues slaves du Sud (bulgare, macédonien, serbe – monténégrin – bosniaque – croate, slovène) parfois divisées en Sud-Est (les deux premières) et Sud-Ouest (du serbe au slovène), langues slaves de l'Ouest (bas-sorabe, haut-sorabe, tchèque, slovaque, polonais et cachoube). Ce dernier groupe peut être subdivisé en trois sous-groupes : langues sorabes, tchèque et slovaque, polonais et cachoube (sous-groupe léchitique).

Starostin (2004) (cité par Blažek) émet l'hypothèse sur la base d'études de glottochronologie d'une part de la scission première du bulgare et du macédonien du reste des langues slaves et d'autre part, l'appartenance des langues slaves du Sud-Ouest (slovène, BCMS [3]) au groupe des langues slaves de l'Ouest. Nous ne sommes pas en mesure de nous prononcer sur les datations, par contre une proximité forte des langues slaves du Nord-Ouest et du Sud-Ouest semble être confirmée par les travaux de Pognan et Panevová sur la génération automatique de lexèmes slaves à partir des racines vieux slaves présentes dans le dictionnaire de Derksen.

2.2 Langues slaves de l'Ouest

Les langues de référence seront les langues slaves de l'Ouest au sens traditionnel du terme, c'est-à-dire le polonais, le bas-sorabe, le haut-sorabe, le tchèque et le slovaque. Le cachoube n'est pas pris en compte faute de documentation adéquate et d'expérience

3. BCMS = bosniaque, croate, monténégrin, serbe.

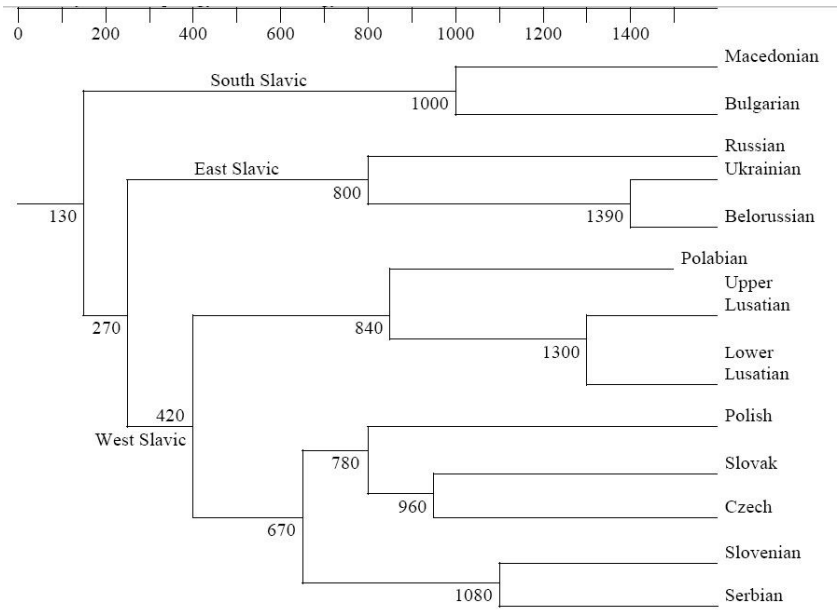


Figure 2 – La répartition de Starostin : slovène et BCMS – langues slaves de l'Ouest

de cette langue. Bien qu'elles soient classées traditionnellement comme langues slaves du Sud et bien que quelques traits leur soient spécifiques (groupe consonantique « cv » là où les langues du Nord-Ouest ont « kv », chute de la dentale dans les groupes « dl » et « tl », caractéristique d'ailleurs partagée avec les langues slaves de l'Est), le slovène et le croate (ainsi que d'autres variantes du BCMS) entrent parfaitement dans une description cohérente et unifiée du système linguistique.

Remarquons que l'ensemble de langues ainsi obtenu possède un code écrit en caractères latins et représente aussi l'ensemble des langues slaves écrites en caractères latins [4].

En raison de l'importance du russe et de sa connaissance prioritaire chez les Slavisants nous aurons aussi des références à cette langue prise en compte dans nos travaux.

3 À la recherche du système linguistique

Les travaux présentés ci-après sont le résultat d'un long cheminement passant par l'analyse automatique d'une langue slave de l'Ouest, le tchèque, les questions de grammaire et de formalisation associées, puis l'approfondissement de la grammaire historique avant l'élargissement aux langues slaves de l'Ouest (groupe traditionnel), ce qui appelait la comparaison avec un représentant de l'Est, le russe et un représentant du Sud, le slovène, bientôt intégré à l'étude et rapidement accompagné du croate. C'est donc un groupe de sept à huit langues slaves qui est soumis à notre étude.

4. Il existe cependant une tradition d'écriture latine même pour le serbe.

Nous comprenons le système linguistique comme la conjonction d'un axe spatial (études synchroniques comparées) et d'un axe temporel (études diachroniques comparées). C'est l'ensemble des faits ainsi rassemblés qui constitue le système d'une famille ou d'un groupe de langues.

Le système linguistique (et tout particulièrement sa composante synchronique) est présenté suivant une dichotomie système nominal / système verbal en raison de particularités très différentes. Cette distinction, qui nous semble valide pour la description et la compréhension des langues slaves, n'est pas sans rappeler la distinction qui est faite dans les visions grammaticales des langues sémitiques entre nominal et verbonominal.

La composante diachronique est particulièrement concernée par la phonologie et les questions de graphie antérieures aux questions d'orthographe. Dans le regard synchronique porté sur un groupe de langues slaves, cette composante diachronique a le pouvoir d'éclairer les variations lexicales entre langues parentes et d'apporter ainsi un support important au multilinguisme et à l'intercompréhension.

Les langues slaves sont des langues indo-européennes marquées par une flexion externe [5], c'est-à-dire qui possèdent, en fin de mot, une désinence qui donne la fonction du mot. Il en résulte un ordre relativement libre des mots.

De manière surprenante, le sous-groupe tchèque et slovaque, grâce au fait que la longueur soit marquée dans l'écriture, apporte par la prise en compte de la préfixation des informations morphologiques complémentaires permettant de lever certaines ambiguïtés qui, sinon, ne seraient pas résolues.

3.1 Le système nominal

3.1.1 Le genre

Les langues slaves possèdent trois genres : le masculin, le féminin et le neutre. En dehors des humains et des animaux supérieurs (mammifères, oiseaux) où l'opposition masculin – féminin a un sens, le genre d'un mot n'est pas donné par sa sémantique, mais par sa forme. Nous pensons qu'en slave, celle-ci prime et doit être le fondement des études de morphologie. La morphologie doit être interprétée au seul niveau de la morphologie par l'intermédiaire de la forme [6]. Ce n'est qu'après que l'on peut différencier des interprétations en faveur de la syntaxe ou de la sémantique. Aller directement aux interprétations, comme cela est souvent fait dans les exposés didactiques des pays slaves, c'est se priver de l'appui des régularités qui organisent la langue et, au-delà, de la compréhension qui en résulte.

Ce qui est plus particulier, ce sont les sous-genres du masculin. Il s'agit de caractéristiques acquises au cours de l'évolution des langues slaves. C'est d'abord la distinction animé / inanimé qui est introduite. Plus tardivement, certaines langues (sorabes, polonais et slovaque, même s'il n'est pas décrit en ces termes) subdivisent le caractère

5. Il convient de rappeler que les langues slaves du Sud-Est, le bulgare et le macédonien, ont perdu leur système flexionnel.

6. Il peut s'agir aussi bien d'une analyse faite par l'homme que par une machine. Surtout dans le second cas, le code graphique est d'une très grande importance, car il est l'un des principaux vecteurs de la calculabilité.

« animé » en « humain » et « animal ». Ce sont donc ces trois sous-genres masculins qui définissent la dimension maximale du système que nous retenons.

3.1.2 Le nombre

Les langues slaves possèdent trois nombres : le singulier, le duel et le pluriel. Cela veut vraisemblablement dire que mentalement pour les Slaves « deux » n'est pas (ou n'était pas) « plusieurs ». Pour l'immense majorité des Français, « deux » est « plusieurs ». Dans l'expression du nombre, le duel est le « maillon faible » et il a effectivement disparu de la majorité des langues slaves où il ne se conserve que par la présence d'ensembles très réduits de termes, notamment ce qui est lié aux parties du corps ou aux organes externes présents par paires ou à la numération. La prise en compte du duel permet d'analyser correctement ces formes résiduelles et de conserver le pluriel pour des usages dérivés. Par exemple, en tchèque, *ucho*, l'oreille donne « *uši* » (duel) s'il s'agit d'oreilles et « *ucha* » (pluriel) s'il s'agit d'un sens dérivé par analogie de forme, les anses (de panier).

Le slovène, le bas-sorabe et le haut-sorabe conservent un duel actif. Cela veut dire que le verbe et toutes les parties du syntagme nominal susceptibles d'être fléchies seront au duel s'il s'agit de deux (nous deux, vous deux, eux deux). Cependant, la question de la pérennité du duel en slovène peut se poser. On peut entendre « *gremo* » au lieu de « *greva* » (allons, nous deux) sans que cela ne choque.

3.1.3 Les cas

Les langues slaves étant flexionnelles, elles expriment la fonction des mots à l'aide de cas. Ils sont au nombre de sept : nominatif (expression du sujet), génitif (complément de nom), datif (complément d'attribution), accusatif (complément d'objet ; expression du lieu avec mouvement), vocatif (cas de l'appel), locatif (expression du lieu sans mouvement), instrumental (instrument, moyen). Les valeurs qui sont données ici sont celles qui sont les plus fréquentes, mais elles sont en fait très variées, notamment en combinaison avec les prépositions.

En ce qui concerne les cas, c'est le vocatif qui est le « maillon faible », déjà disparu dans certaines langues et en régression dans d'autres comme nous allons le voir dans le tableau ci-dessous.

La tradition tchèque exprime ces cas par un numéro d'ordre allant de 1 à 7 dans l'ordre qui est donné ici, ce qui est très utile pour une utilisation informatique. Cependant l'ordre des cas peut être différent non seulement en fonction des traditions de chacune des nations slaves, mais même dans ce cadre varier suivant les auteurs.

En raison des outils de traitement automatique et des bases de données que nous avons réalisés, nous conservons l'ordre tchèque qui est appliqué à l'ensemble des langues que nous comparons. Pour les langues sans vocatif, ce cas est marqué par une ligne vide.

Les démonstratifs, les adjectifs et les substantifs sont concernés par l'ensemble des valeurs genre / nombre / cas.

Tableau 1 – Les cas

1	**2**	**3**	**4**	**5**	**6**	**7**	← cas
nominatif	génitif	datif	accusatif	vocatif	locatif	instrumental	↓ langues
N	**G**	**D**	**A**	**V**	**L**	**I**	tchèque
N	**G**	**D**	**A**	**L**	**I**		slovaque (pas de vocatif)
N	**G**	**D**	**A**	**I**	**L**	**V**	haut-sorabe
N	**V**	**G**	**D**	**A**	**I**	**L**	
N	**G**	**D**	**A**	**I**	**L**		bas-sorabe (pas de vocatif)
N	**G**	**D**	**A**	**I**	**L**	**V**	polonais
N	**V**	**A**	**G**	**D**	**I**	**L**	
N	**G**	**D**	**A**	**L**	**I**		slovène (pas de vocatif)
N	**G**	**D**	**A**	**V**	**I**	**L**	BCMS
N	**G**	**D**	**A**	**I**	**L**		russe (pas de vocatif)
N	**A**	**G**	**D**	**I**	**L**		

3.1.4 Les démonstratifs

Nous souhaitons faire une place aux démonstratifs pour une raison simple. La linguistique générale fait parfois appel à des langues lointaines pour montrer un système de démonstratifs à trois niveaux de distance, ce que, fondamentalement, possède le système slave. Ainsi, en croate, mais aussi en polonais, les trois niveaux de distance sont afférents au sujet parlant (1re personne) pour ce qui est tout proche, ici, dans ma proximité, à l'interlocuteur (2e personne), ce qui est là, près de toi et à celui qui n'intervient pas dans le dialogue, ce qui est près de lui là-bas (3e personne). Les formes de deuxième et troisième personnes sont particulièrement parlantes par leur forme rattachée à la forme du pronom personnel de 2e ou 3e personne. Voici un tableau pour le croate :

Tableau 2 – Les démonstratifs

	croate	tchèque
près de moi (1ère pers.)	ovaj, ova, ovo	(tento, tato, toto)
près de toi (2e pers.)	taj, ta, to	ten, ta, to
près de lui (3e pers.)	onaj, ona, ono	onen, ona,ono

La comparaison avec le tchèque montre que dans cette langue le démonstratif afférent à la 1re personne n'existe pas et qu'il est remplacé par un renforcement du démonstratif de 2e personne pour exprimer une plus grande proximité (que l'on peut montrer du doigt).

3.1.5 Les adjectifs

Les adjectifs peuvent être présentés par une double dualité. En premier lieu, nous avons une opposition entre forme longue et forme courte de l'adjectif. Cette opposition sépare les langues slaves du Nord (formes longues dominantes) et les langues slaves du Sud (formes courtes dominantes). Les formes courtes existent dans les langues slaves du Nord en nombre réduit et avec des fonctions divergentes de celles de l'adjectif qualificatif standard. Une forme courte peut être employée, lorsqu'elle existe, en fonction d'adjectif attribut (jamais en tant qu'épithète), mais sous sa forme neutre singulier, cette forme courte peut prendre progressivement une valeur d'adverbe, puis de substantif.

Les langues slaves du Sud possèdent aussi une forme longue au nominatif et accusatif inanimé du masculin. L'opposition entre forme longue et courte a été ici spécialisée sémantiquement en déterminé vs indéterminé : en slovène : « *lep človek* » (un bel homme) face à « *lepi človek* » (le bel homme).

Rappelons ici que la construction de la forme courte au masculin peut être concernée par la disparition et la vocalisation des jers, ce qui sera abordé dans les apports de la grammaire historique à la compréhension du système.

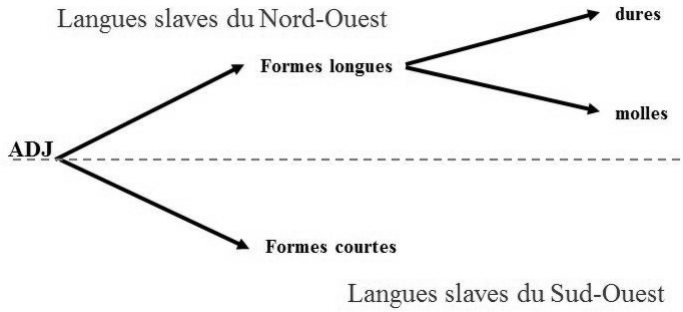


Figure 3 – L'organisation des adjectifs slaves – opposition entre formes courtes et longues

La deuxième dichotomie concerne uniquement les formes longues qui peuvent se diviser entre formes longues dures (le support vocalique de la désinence est une voyelle dure) et formes longues molles (le support vocalique de la désinence est une voyelle molle). Le classement des adjectifs correspondants se fait ici aussi par la forme. La répartition quantitative et qualitative peut être variable suivant les langues considérées. Le slovaque n'a que très peu d'adjectifs « mous » alors qu'en tchèque la répartition entre formes dures et formes courtes est équilibrée. Les adjectifs d'origine étrangère sont empruntés en tchèque principalement en tant qu'adjectifs mous dont ils représentent environ 60 %. Les adjectifs durs sont marqués par une proportion très

importante d'adjectifs autochtones et d'adjectifs d'origine germanique empruntés de longue date, généralement non reconnus comme étant d'origine étrangère.

3.1.6 Les substantifs

Le classement des paradigmes nominaux, réalisé à partir de la forme, peut être présenté à l'aide d'une série d'oppositions.

La première est l'opposition entre terminaison consonantique (modèles entourés de rouge) et terminaison vocalique (modèles entourés de vert). Sont, en principe, des masculins les substantifs terminés par une consonne (c'est toujours la forme de nominatif singulier qui est considérée). Ces substantifs sont terminés en consonne du fait de la disparition du « jer » vieux slave qui était à la fin du mot [7]. Les terminaisons vocaliques englobent les modèles féminins (essentiellement en « -a ») et les modèles neutres (en « o » ou en « e »). Malheureusement, cette répartition n'est pas totalement exacte du fait de quelques modèles féminins consonantiques et de quelques modèles masculins vocaliques. Dans cette dernière catégorie, certains mots peuvent être simultanément masculin et féminin.

La seconde opposition, toujours en considérant le dernier graphème du mot au nominatif singulier, est l'opposition « dur / mou ». C'est dans toutes les langues slaves, l'opposition majeure [8]. Elle s'applique aussi bien aux consonnes qu'aux voyelles [9]. On peut en voir l'étendue en tchèque où pratiquement tous les paradigmes du masculin, du féminin et du neutre sont concernés. La totalité des modèles durs possède un équivalent mou, ce qui ne laisse de côté que le modèle à augment « *kuře* » poulet.

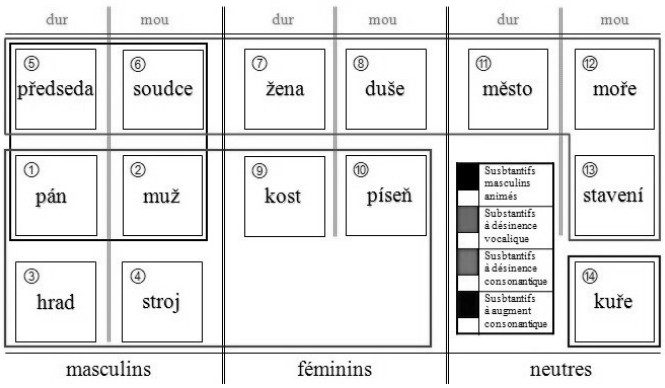


Figure 4 – Les paradigmes substantivaux tchèques

7. Nous examinerons ce phénomène lors de la présentation des éléments historiques qui influent sur le fonctionnement actuel des langues slaves.

8. Certains linguistes, en particulier spécialistes de la langue russe, considèrent les termes de cette opposition comme dépassés. Nous les conservons parce qu'ils sont révélateurs des phénomènes concernés, constituent en même temps une pré-formalisation des processus informatiques qui peuvent être utilisés, par exemple pour l'analyse et qu'ils ont une validité générale par rapport à notre ensemble de langues.

9. N'oublions pas que dans une suite consonne – voyelle, c'est la non identité du caractère de dureté qui provoque des alternances vocaliques ou consonantiques (palatalisation) sachant que c'est toujours l'élément « dur » qui est transformé en « mou ».

Les tableaux qui sont présentés ici offrent la modélisation des paradigmes repris tels quels de la tradition scolaire correspondante ou légèrement modifiée en fonction de notre vision du système linguistique. Ils sont utilisés dans nos bases de données, dans un onglet spécialisé, comme aide-mémoire des classifications adoptées.

L'opposition suivante, binaire pour les trois tableaux présentés dans l'ordre tchèque, slovaque et slovène, concerne l'opposition masculin animé / masculin inanimé. Elle devrait être ternaire pour le polonais et les sorabes en masculin humain / masculin animal / masculin inanimé. Nous rendons cette différence visible en entourant en noir le masculin animé.

Le dernier élément à noter correspond également à une opposition très ancienne : face à la flexion vocalique (pratiquement seule aujourd'hui), nous trouvons les restes d'une flexion consonantique avec cinq paradigmes marqués par la présence d'un augment et dont il semble qu'ils aient déjà été en régression dans la période la plus ancienne que nous connaissons (IX^{e}-X^{e} siècles). Ces paradigmes sont entourés de bleu.

Il convient de remarquer qu'il existe peu de ces paradigmes encore actifs aujourd'hui et ce ne sont généralement pas les mêmes dans les différentes langues slaves. Le modèle tchèque et le modèle slovaque représente le paradigme en « voyelle – t » issu de l'ancien augment -nt-. Le slovène est vraisemblablement la langue ayant le mieux conservé la flexion avec augment puisqu'on y trouve, comme pour le tchèque et le slovaque, le paradigme en « voyelle – t » (*tele* – le veau), mais aussi le paradigme en « voyelle – n » (*vreme* – le temps) que l'on trouve comme modèle vivant en russe, le paradigme en « voyelle – s » (*telo* – le corps), tous modèles de neutre, mais aussi un modèle féminin en « voyelle – r » (*mati* – mère).

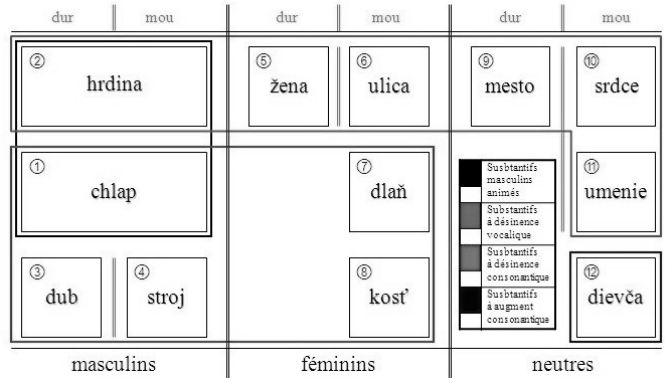


Figure 5 – Les paradigmes substantivaux slovaques

D'autres commentaires peuvent être faits à propos de ces tableaux. Le slovaque n'a pas dépalatalisé les labiales et les dentales comme l'a fait le tchèque et n'oppose donc pas un modèle dur face à un modèle mou pour les féminins consonantiques. De même, il n'a pas besoin pour les masculins animés consonantique et vocalique de correspondants mous, puisqu'il n'a pas changé le « a » en « e » derrière une consonne molle comme l'a fait le tchèque.

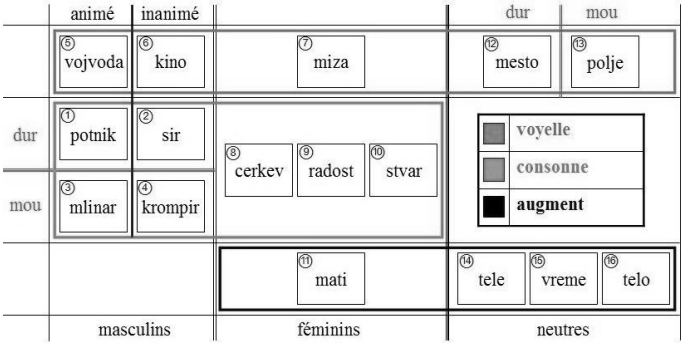


Figure 6 – Les paradigmes substantivaux slovènes

Le slovène ayant presque entièrement perdu les mouillures, les oppositions « dur » / « mou » y sont minimales, existant notamment pour le neutre entre « *mesto* » (ville) et « *polje* » (champ). Au masculin, les modèles mous terminés en « -r » sont calculables ajoutant un « j » (un moyen possible d'exprimer la mouillure dans les langues slaves écrites en caractères latins) aux cas obliques à quelques exceptions près.

Le fait que le slovène classe les mots étrangers terminés en voyelle, notamment « o », dans les masculins et non parmi les neutres, l'écarte un peu plus du principe de classification en fonction de la forme. C'est une difficulté ou au moins une surprise pour un apprenant étranger qui possède déjà une autre langue slave.

Remarquons aussi le nombre élevé de modèles consonantiques dont celui des féminins à terminaison en « -ev » (*cerkev* – église) qui en slovène a conservé une flexion autonome. Notons également le paradigme « *radost* » avec le suffixe « -ost » dur à cause de la forte dépalatalisation en slovène. Cette forme est à comparer avec celle du slovaque en « *kosť* ». D'une manière générale, le suffixe « -ost » montre bien la situation des différentes langues slaves vis-à-vis de la mouillure :

- Russe (mouillure) : « -ость. »
- Slave de l'Ouest (mouillure) : slovaque « -osť », sorabes « -ość », polonais « -ość », mais le tchèque a une situation différente (dure) « -ost » qui le rapproche du slovène également dur en « -ost ».

3.2 Le système verbal

Bien évidemment, ce qui marque le plus le système verbal des langues slaves est la présence de l'aspect. Nous distinguons quatre valeurs aspectuelles : perfectif, imperfectif, itératif (qui est par définition un imperfectif) et bi-aspectuel (ayant simultanément les valeurs perfective et imperfective).

Il convient de remarquer que dans le domaine du verbe, nous aurons besoin aussi des trois genres, des trois sous-genres et des trois nombres auxquels nous associerons les trois personnes (1re, 2e et 3e).

Dans un premier temps, nous donnons un tableau des formes verbales slaves que nous diviserons en formes personnelles et formes impersonnelles et que nous présenterons ensuite dans le détail :

Tableau 3 – Formes verbales slaves

imperfectif	**formes communes**	**perfectif**	**adjectifs autres liés**	**type**
présent		présent (futur sémantique)		
impératif négatif		impératif positif		
imparfait		aoriste		
	participe en -l			
	adj. de part. en -l			dur
	passé composé			
	plus-que-parfait			
	conditionnel présent			
	conditionnel passé			
	futur avec participe en -l			
	futur avec infinitif			
gérondif présent		gérondif passé		
adj. issu de ger. prés.		adj. issu de ger. pass.	adj. verbal de but	mou
	part. prés. passif (-m)			
	adj. de p.prés.p			dur
	part. passé passif (-n/t)			
	adj. de part.passé.pass		adj. prédicatif	dur
	substantif verbal			
	infinitif			
	supin			

Le tableau [tab3] est organisé en trois colonnes principales (à gauche) et deux colonnes annexes (à droite). Nous avons dans la deuxième colonne de gauche tous les temps qui peuvent être utilisés soit avec le perfectif, soit avec l'imperfectif, donc communs aux deux. Dans la première colonne de gauche, nous avons les temps qui relèvent spécifiquement de l'imperfectif à l'exception du présent sur lequel nous reviendrons. L'imparfait et le gérondif présent sont liés à l'imperfectif.

Dans la troisième colonne de gauche se trouvent les temps qui relèvent du perfectif, en particulier l'aoriste et le gérondif passé qui sont le « pendant » de l'imparfait et du gérondif présent.

Il est nécessaire de revenir sur le présent. Les langues slaves, dans de très nombreux cas, ont perdu les formes simples, notamment du passé (imparfait et aoriste) au profit des formes composées. En dehors de formes de futur composé, il n'existe pas morphologiquement de forme spécifique du futur. C'est la forme morphologique du présent du perfectif qui donne une valeur sémantique de futur. Ceci est l'une des très nombreuses situations où il convient de faire d'abord une analyse au niveau morphologique, puis de l'interpréter au niveau du sens. Cette interprétation est aussi un mode de détermination de l'aspect : si le présent morphologique est aussi un présent sémantique, alors le verbe est imperfectif. Cependant, si le présent morphologique a un sens de futur, le verbe est perfectif.

Horizontalement, ce tableau est organisé en deux parties. La partie supérieure est afférente aux formes dites personnelles, c'est-à-dire dans lesquelles la référence à la personne est marquée dans la morphologie. Cette partie est elle-même subdivisée en temps simples (quel que soit le mode) : présent, impératif, imparfait / aoriste et en temps composés : passé composé, plus que parfait, conditionnels présent et passé, futurs composés. La partie inférieure concerne les formes impersonnelles que nous avons divisées en formes gérondives, formes participiales (dont le substantif verbal formé sur le participe passé passif) et formes afférentes à l'infinitif (infinitif, supin). Toutes ces formes impersonnelles sont simples.

Ce tableau a pour but de représenter l'intégralité des formes verbales des langues slaves de l'Ouest. C'est pourquoi cette matrice s'appuie essentiellement sur le bas-sorabe pour les formes personnelles et le supin et sur le tchèque pour les formes impersonnelles gérondives et participiales, parce que ces deux langues sont les plus complètes pour les points concernés. Aucune description des langues de l'étude n'est égale à la description du système. Certaines s'en rapprochent plus que d'autres, mais toutes sont un sous-ensemble du système.

Nous allons montrer la complétude du bas-sorabe en ce qui concerne les formes personnelles, d'abord présent, imparfait et aoriste :

Infinitif:	kupować	*Supin*:	kupowat

	Présent [ipf]: singulier	*ou* duel	*futur [pf]*: pluriel
1.	kupuju / -jom	kupujomej	kupujomy
2.	kupujoš	kupujotej	kupujośo
3.	kupujo	kupujotej	kupuju

	Imparfait [ipf]: singulier	duel	pluriel
1.	kupowach	kupowachmej	kupowachmy
2.	kupowašo	kupowaštej	kupowaśćo
3.	kupowašo	kupowaštej	kupowachu

	singulier	duel	*Aoriste [pf]*: pluriel
1.	nakupowach	nakupowachmej	nakupowachmy
2.	nakupowa	nakupowaštej	nakupowašćo
3.	nakupowa	nakupowaštej	nakupowachu

Figure 7 – Bas-sorabe – Présent, imparfait, aoriste

Puis de formes composées du passé :

Passé composé:

Singulier

	masculin hum.	masculin animal	masculin inan.	féminin	neutre
1		som kupował		1 som kupowała	1 som kupowało
2		sy kupował		2 sy kupowała	2 sy kupowało
3		jo kupował		3 jo kupowała	3 jo kupowało

Duel

	masculin hum.	masculin animal	masculin inan.	féminin	neutre
1			smej kupowałej		
2			stej kupowałej		
3			stej kupowałej		

Pluriel

	masculin hum.	masculin animal	masculin inan.	féminin	neutre
1			smy kupowali		
2			śćo kupowali		
3			su kupowali		

Figure 8 – Bas-sorabe – Formes composées du passé

Plus-que-parfait:

Singulier

	masculin hum.	masculin animal	masculin inan.	féminin	neutre
1		běch kupował		1 běch kupowała	1 běch kupowało
2		běšo kupował		2 běšo kupowała	2 běšo kupowało
3		běšo kupował		3 běšo kupowała	3 běšo kupowało

Duel

	masculin hum.	masculin animal	masculin inan.	féminin	neutre
1			běchmej kupowałej		
2			běštej kupowałej		
3			běštej kupowałej		

Pluriel

	masculin hum.	masculin animal	masculin inan.	féminin	neutre
1			běchmy kupowali		
2			běšćo kupowali		
3			běchu kupowali		

Figure 9 – Bas-sorabe – Plus-que-parfait

En ce qui concerne les formes composées de futur, nous présenterons le futur slovène, futur qui nous vient d'un passé lointain, qui s'efface actuellement en polonais devant le futur composé en auxiliaire et infinitif. Il a disparu depuis longtemps des formes tchèques.

Futur avec le participe passé actif en –l:

Singulier

	masculin hum.	masculin animal	masculin inan.	féminin	neutre
1		bom potoval		1 bom potovala	1 bom potovalo
2		boš potoval		2 boš potovala	2 boš potovalo
3		bo potoval		3 bo potovala	3 bo potovalo

Duel

	masculin hum.	masculin animal	masculin inan.	féminin	neutre
1		bova potovala		1 bova potovali	
2		bosta potovala		2 bosta potovali	
3		bosta potovala		3 bosta potovali	

Pluriel

	masculin hum.	masculin animal	masculin inan.	féminin	neutre
1		bomo potovali		1 bomo potovale	1 bomo potovala
2		boste potovali		2 boste potovale	2 boste potovala
3		bodo potovali		3 bodo potovale	3 bodo potovala

Figure 10 – Slovène – Formes composées de futur

L'étendue des formes impersonnelles peut être montrée sur la base du tchèque, mais pour des raisons de volume, nous nous concentrerons sur le gérondif passé (le gérondif présent et l'adjectif en dérivant étant relativement fréquents) et le participe passé passif et ses dérivations :

Gérondif passé [pf]:

Singulier

masculin hum.	masculin animal	masculin inan.	féminin	neutre
	nakupovav		nakupovavši	

Duel

masculin hum.	masculin animal	masculin inan.	féminin	neutre

Pluriel

masculin hum.	masculin animal	masculin inan.	féminin	neutre
		nakupovavše		

Figure 11 – Tchèque – Gérondif passé

Déclinaison au singulier de l'adjectif issu du gérondif passé :

Adjectif issu du gérondif passé [pf]:

Singulier

	masculin hum. masculin animal	masculin inan.	féminin	neutre
1	nakupovavší	1 nakupovavší	1 nakupovavší	1 nakupovavší
2	nakupovavšího	2 nakupovavšího	2 nakupovavší	2 nakupovavšího
3	nakupovavšímu	3 nakupovavšímu	3 nakupovavší	3 nakupovavšímu
4	nakupovavšího	4 nakupovavší	4 nakupovavší	4 nakupovavší
5	nakupovavší	5 nakupovavší	5 nakupovavší	5 nakupovavší
6	nakupovavším	6 nakupovavším	6 nakupovavší	6 nakupovavším
7	nakupovavším	7 nakupovavším	7 nakupovavší	7 nakupovavším

Figure 12 – Tchèque – Adjectif issu du gérondif passé

Nous donnons enfin ci-dessous les formes du participe passé passif, l'adjectif qui en dérive et le substantif verbal :

Participe passé passif (en –n ou t):

Singulier

masculin hum.	masculin animal	masculin inan.	féminin	neutre
	kupován		kupována	kupováno

Duel

masculin hum.	masculin animal	masculin inan.	féminin	neutre

Pluriel

masculin hum.	masculin animal	masculin inan.	féminin	neutre
kupováni		kupovány		kupována

Figure 13 – Tchèque – Participe passé passif

Adjectif issu du participe passé passif

Singulier

	masculin hum.	masculin animal	masculin inan.	féminin	neutre
1	kupovaný		1 kupovaný	1 kupovaná	1 kupované
2	kupovaného		2 kupovaného	2 kupované	2 kupovaného
3	kupovanému		3 kupovanému	3 kupované	3 kupovanému
4	kupovaného		4 kupovaný	4 kupovanou	4 kupované
5	kupovaný		5 kupovaný	5 kupovaná	5 kupované
6	kupovaném		6 kupovaném	6 kupované	6 kupovaném
7	kupovaným		7 kupovaným	7 kupovanou	7 kupovaným

Duel

	masculin hum.	masculin animal	masculin inan.	féminin	neutre
1	1		1	1	1
2	2		2	2	2
3	3		3	3	3
4	4		4	4	4
5	5		5	5	5
6	6		6	6	6
7	7		7	7	7

Pluriel

	masculin hum.	masculin animal	masculin inan.	féminin	neutre
1	kupovaní		1 kupované		1 kupovaná
2	kupovaných		2 kupovaných		2 kupovaných
3	kupovaným		3 kupovaným		3 kupovaným
4	kupované		4 kupované		4 kupovaná
5	kupovaní		5 kupované		5 kupovaná
6	kupovaných		6 kupovaných		6 kupovaných
7	kupovanými		7 kupovanými		7 kupovanými

Figure 14 – Tchèque – Adjectif issu du participe passé passif

Substantif verbal:

		singulier	duel	pluriel
nominatif	1.	kupování		kupování
génitif	2.	kupování		kupováních
datif	3.	kupování		kupováním
accusatif	4.	kupování		kupování
vocatif	5.	kupování		kupování
locatif	6.	kupování		kupováních
instrumental	7.	kupováním		kupováními

Figure 15 – Tchèque – Substantif verbal

Dans le système que nous utilisons depuis longtemps, notamment pour l'analyse automatique, nous considérons, comme nous l'avons annoncé plus haut, une dérivation verbale et une dérivation nominale. Cela a l'avantage de pouvoir décrire par exemple le substantif verbal avec les valeurs de dérivation verbale et de substantif en ce qui concerne la catégorie lexicale. De même, les adjectifs dérivés des gérondifs et des participes ont une dérivation verbale et une catégorie lexicale d'adjectif. Une sous-catégorie de l'adjectif désigne l'origine gérondive ou participiale de cet adjectif.

Dans le tableau des formes verbales, les adjectifs dérivés des gérondifs sont sur fond rose pour indiquer qu'il s'agit d'adjectifs mous (en tchèque), ceux qui sont dérivés des participes sont sur fond bleu pour indiquer qu'ils sont durs (toujours en tchèque). Cette valeur est portée dans la cinquième colonne du tableau.

Ce tableau prend également en compte, en quatrième colonne, des spécificités du tchèque : l'adjectif verbal de but, formé sur le lexème du passé (qui sert à faire ce qu'exprime le sens de la racine) et qui est aussi un adjectif mou et l'adjectif prédicatif (qui correspond aux adjectifs français en « -able » et « -ible ») formé également sur le verbe et qui est un adjectif dur.

Ce tableau peut être utilisé comme une véritable matrice qui permet de montrer que les difficultés à interpréter certains points de grammaire proviennent d'un manque systémique (d'un « trou » dans la matrice) ou du « brouillage » d'un élément pourtant existant. Certains de nos outils, comme l'interface homme-machine d'une base de données, en permettent un usage de type didactique.

Parmi les « manques systémiques » peut figurer l'absence de forme longue de l'adjectif dans les langues slaves du Sud qui rend non perceptible la différence entre le participe passé passif et l'adjectif (tous deux ayant une forme courte, *cf.* tableau ci-dessous). En polonais, les formes du participe ne sont plus usitées, sauf dans la langue religieuse. Il ne demeure que la forme du neutre singulier.

La situation du slovaque est plus singulière, puisque l'adjectif vient aussi à la place du participe alors que les formes afférentes du participe existent potentiellement. On perd ici aussi la distinction entre participe et adjectif malgré la possibilité du système de les distinguer.

Tableau 4 – Participe passé passif

	tchèque *acheter*	slovaque *travailler*	polonais *aimer*	slovène *voyager*
participe passé passif (ppp)	kupován	(od)pracovaný	(—) miłowano	potovan
adjectif du ppp	kupovaný	(od)pracovaný [10]	miłowany	potovan
substantif verbal	kupování	pracovanie	miłowanie	potovanje

10. Malgré des manques éventuels dus à la sémantique notamment en ce qui concerne l'imperfectif, la dérivation « participe passé passif, adjectif dérivé et substantif verbal » appartient aux deux aspects.

Tableau 5 – Gérondif

	tchèque	slovaque	polonais	slovène
gérondif présent	kupuje/íc/íce	pracujúc	miłując	potujoč
adjectif issu du gérondif présent	kupující	pracujúci	miłujący	potujoč
gérondif passé	nakupovav/avši/avše	—	umiłowawszy	(potovavši)
adjectif issu du gérondif passé	nakupovavší	odpracovavší	—	—

En tchèque, il y a présence du gérondif et de l'adjectif qui en découle, et le gérondif a trois formes propres : masculin singulier, féminin et neutre du singulier, pluriel.

4 De la diachronie cachée pour la compréhension multilingue du lexique slave

Dans l'apprentissage d'une langue, il est possible de fournir de manière discrète, surtout aux débutants, c'est-à-dire sans faire un cours explicite de grammaire historique, des éléments de la diachronie, qui viennent en plus de la présentation du système linguistique et qui sont essentiels pour la compréhension du fonctionnement du groupe de langues, non seulement de la langue étudiée prioritairement, mais aussi des langues voisines. L'ouvrage de Teyssier, sur les langues romanes, est un bon exemple de ce qui peut être fait. Une telle pratique permet un élargissement notable de l'intercompréhension naturelle.

Cette manière de faire nous semble justifiée pour la compréhension de phénomènes synchroniques qui ont besoin d'un éclairage diachronique alors que les apprenants n'ont pas encore le niveau de langue suffisant pour aborder un vrai cours de grammaire historique. Ces informations pourront être approfondies plus tard dans un cours spécialisé.

Certains phénomènes phonologiques, les plus anciens en particulier, marquent la séparation progressive des langues slaves les unes des autres et sont susceptibles d'expliquer des formes apparemment distantes en donnant les régularités de transformation.

Ce que nous présentons ci-dessous sont des transformations phonologiques générales dont la connaissance aide à la compréhension des langues du groupe. Ces connaissances peuvent aussi constituer une base pour une étude plus précise et plus conventionnelle de la grammaire historique du groupe de langues.

Nous allons, à titre d'exemples, examiner quelques phénomènes importants.

4.1 La métathèse

Il s'agit de l'un des phénomènes les plus anciens que nous connaissons, vraisemblablement de la période autour des VIII^{e} et IX^{e} siècles. La métathèse peut être présentée à des Francophones à l'aide du français (« fromage » avec métathèse) face à l'italien (« *formaggio* » sans métathèse).

La métathèse en slave est en fait double, puisqu'elle peut être réalisée avec une voyelle « e » ou une voyelle « o ». Un tableau valant souvent mieux qu'un long discours, nous allons expliquer la métathèse en « o » d'après le schéma suivant :

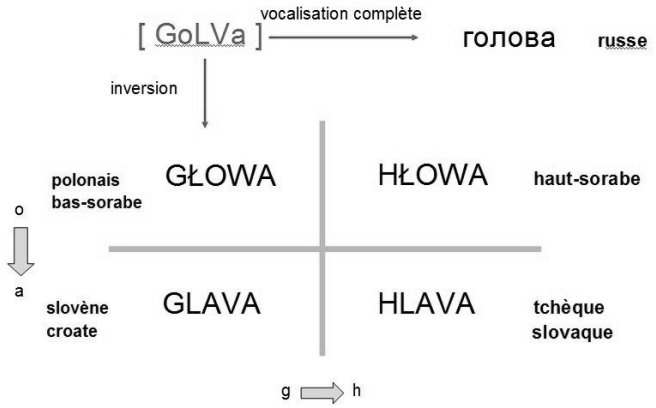


Figure 16 – La mécanique de la métathèse en « o »

Nous devons, en premier lieu, exposer les conditions qui provoquent la métathèse. La suite de caractères « (C) {e | o} {l | r} C » permet la réalisation de la métathèse. « C » désigne une consonne quelconque. La présence, dans cet ordre, d'une voyelle « e » ou « o » suivie d'une consonne liquide « l » ou « r » devant une consonne permet la réalisation de la métathèse si la langue le permet. La présence ou non d'une consonne avant ce patron importe peu, d'où la notation d'une consonne facultative entre parenthèses.

Dans le cas de « o », le russe a une vocalisation complète en « (C) o {l | r} o C », le bas-sorabe, le haut-sorabe et le polonais ont une forme en « (C) {l | r} o C », le tchèque et le slovaque remplacent le « o » par « a » « (C) {l | r} a C » ce que font aussi les langues slaves du Sud. Remarquons que cette métathèse présente une régularité élevée. Le problème le plus important est que le haut-sorabe ne forme pas toujours la métathèse (qu'elle soit en « e » ou « o »).

Dans la pratique, les mots transformés par cette métathèse avec la voyelle « o » peuvent être modifiés par d'autres phénomènes. C'est le cas ici où le « g » est transformé en « h » en haut-sorabe, tchèque et slovaque (phénomène de la période allant de la fin du X^{e} siècle à la fin du XIV^{e} siècle). Le schéma présenté permet de combiner les deux transformations « o » vers « a » et « g » vers « h ».

Un dernier phénomène concerne la notation du « l » dur en l barré « ł » pour les sorabes et le polonais.

Sur cette base, la correspondance entre les huit langues concernées semble claire et facilement utilisable pour faire des rapprochements entre mots concernés, voire de prédire la forme d'un mot dans une langue que l'on ne connaît pas ou que l'on connaît moins.

Un tableau, moins parlant quant à la « mécanique », comme celui ci-dessous, a aussi le mérite d'apporter des éléments de classement. On y voit ce qui semble être la position du balte (lituanien, letton) : pas de métathèse, voyelle en « a ».

Tableau 6 – La métathèse

domaine du « o »				domaine du « a »					
vocalisation complète	métathèse								sans métathèse
russe	polonais	bas sorabe	haut sorabe	tchèque	slovaque	slovène	croate	serbe	balte (lit/let)
голова	głowa	głowa	hłowa	hlava	hlava	glava	glava	глава	galva
			changement de « g » à « h »						

La métathèse en « e » est sans alternance vocalique (« e » reste « e »), mais peut présenter des irrégularités.

À partir d'une suite de caractères « (C) e {l | r} C », on obtient :

- en russe : une vocalisation complète en « (C) e {l | r} e C » ;
- pour toutes les autres langues : « (C) {l | r} e C ».

Ainsi, sur la base de la racine BeRGЪ, on obtiendra en russe une forme BeReG (берег, la BeRGe, la rive) qui donnera « BŘeH » en tchèque, par conséquent « BRZeG » en polonais, « BReH » en slovaque et BReG en slovène, « BRijeG » en croate.

La forme irrégulière est ici celle du haut-sorabe qui fait « BRjóH » dont le « ó » (prononcé entre « u » et « o » – la « *smuška* » ne représente pas la longueur, mais le timbre) alterne dans la flexion avec « o », par exemple la forme « *brjoha* » au génitif. J. E. Smoler[11] (Schmaler) dans son dictionnaire *Njemsko-Serski słownik*[12] publié en 1843 à Budyšin / Bautzen note encore, p. 132, « *brjeh* » à côté de « *brjóh* ».

Les transformations « g » vers « h » sont les mêmes que précédemment.

Les différentes formes en présence attestent d'une mouillure derrière le « r » visible en particulier en croate (segment « ije »). La présence de la mouillure est visible en tchèque par la palatalisation de « r » en « ř ». La forme du polonais en découle

11. La librairie de la *Domowina* porte son nom, « *Smolerjec kniharnja* ».
12. Numérisé par Google, il est accessible en ligne sur le site de Lexilogos.

avec le remplacement de « ř » par « rz ». Du groupe de langues considérées, le slovène est certainement celle qui a perdu le maximum de mouillures et ceci très tôt (vraisemblablement vers le XI^{e} siècle). Le tchèque, ayant abondamment perdu la mouillure des labiales et des dentales [13], occupe une position intermédiaire entre les autres langues du Nord-Ouest et le slovène.

La situation du slovaque mérite un commentaire :

Tableau 7 – L'évolution du « r » en slovaque

	mou	dur	dur long
tchèque	ř	r	
slovaque	r		ŕ

L'évolution du « r » mouillé en slovaque est marquée par la perte de la mouillure, ce qui fait du « r » slovaque une consonne mixte.

Le tchèque et le slovaque marquent la longueur vocalique dans le système graphique : « áéíúý » + « ů » pour le tchèque et « ô » pour le slovaque. Pour « l » et « r » en position de support vocalique, le slovaque maintient le parallélisme avec les voyelles en gardant une opposition de longueur : « l » face à « Ĺ » et « r » face à « ŕ ».

4.2 La contraction

Le phénomène de contraction le plus ancien consiste, dans une suite « voyelle – j – voyelle », en la chute du « yod » (« j ») ce qui donne une suite « voyelle – voyelle » qui devient dans les langues ayant la longueur (tchèque, slovaque) une voyelle longue et une voyelle standard dans les autres. Ce phénomène a été particulièrement bien décrit par Marvan et présenté sous forme d'un tableau que nous reproduisons ci-dessous [tab. 8].

Par exemple, « *pojasЪ* » (phénomène 4 dans le tableau précédent) donnera, après les transformations adéquates (oja > oa > á – c'est généralement la seconde voyelle qui donne la valeur finale), une forme « *pás* » en tchèque et en slovaque, « *pas* » en polonais et en slovène (pour les langues qui ont contracté), mais « *pojas* » en croate, bosniaque et serbe qui n'ont pas contracté pour cette unité lexicale. Le tableau est organisé en « I » lorsque la langue concernée contracte toujours, « — » quand elle ne contracte jamais et « + » quand elle contracte de manière intermittente. Dans le cas de « *pojasЪ* », le slovène et le serbo-croate ont « + », mais le slovène réalise la contraction (« *pas* ») et le serbo-croate non (« *pojas* »).

13. En tchèque, la série des consonnes mixtes est issue de cette « dépalatalisation » massive (« b[f]lmpsvz »), ce qui pose d'ailleurs au niveau du système graphique le problème de l'absence des caractères « mous » correspondants. Il est des cas où cette absence se fait sentir, par exemple pour la déclinaison au pluriel de mots tels que « *doupě* » (terrier), « *poupě* » (bourgeon) ou pour les formes de passé et du participe passé passif d'un verbe tel que « *napnout* » (tendre). Dans le premier cas, les Tchèques renoncent à la mouillure (« *doupata* » et « *poupata* »), dans le second ils se hasardent à ajouter un « j » (« *napjal* » et « *napjat* » qui donnera ensuite de manière régulière et juste le substantif verbal « *napětí* » (tension).

Tableau 8 – La contraction

		pol	b-s	h-s	tch	slk	slo	cr/s	bul	ukr	rus		
1	dobrajego	I	I	I	I	I	I	I	I	I	I	20	1
2	dobroje	I	I	I	I	I	I	I	—	I	—	16	2
3	dělaje	+	I	I	I	I	I	+	+	+	—	14	3
4	pojasъ	I	I	I	I	I	+	+	—	—	—	12	4
5	sъměje	I	—	I	I	I	+	I	—	—	—	11	5
6	ženojǫ	I	I	I	I	+	I	+	—	—	—	10	6
7	vějati	+	+	I	I	I	+	—	—	—	—	9	7
8	mojego	+	—	—	I	+	I	+	—	—	—	7	8
9	stojati	+	—	—	I	I	+	—	—	—	—	6	9
10	moja	—	—	—	+	+	+	—	—	—	—	4	10
11	mojimi	+	—	—	I	—	—	—	—	—	—	3	11
12	vojevoda	—	—	—	I	—	+	—	—	—	—	3	12
13	zajęcь	—	—	—	—	—	+	+	—	—	—	2	13
14	moji	—	—	—	—	—	—	—	—	—	—	0	14
15	směje sę	—	—	—	—	—	—	—	—	—	—	0	15
		14	11	14	19	17	15	10	3	5	2		
		15	11	14	23	17	17	11	3	5	2		

Ce phénomène de contraction peut être comparé au pavé lancé dans la mare : là où le pavé tombe les ondes sont les plus fortes, puis diminuent en intensité au fur et à mesure que l'on s'éloigne du point de chute.

C'est exactement ce que l'on a avec la contraction : le point de chute est l'aire tchèque et, plus on s'en éloigne, moins les contractions se réalisent. C'est ainsi que les langues slaves de l'Est et du Sud-Est ont très peu contracté. À l'inverse, il y a l'exception de « *zajęcь* » qui donne « *zec* » en BCMS alors qu'il n'y a pas contraction ailleurs : « *zając* » en polonais, « *zajac* » en slovaque, « *zajíc* » en tchèque et « *zajec* » en slovène et, bien sûr, par définition pour les langues slaves de l'Est : « *заяц* » en russe, « *заєць* » en ukrainien.

Marvan présente l'extension de la contraction de la manière suivante :

1. Foyer : empire de Grande Moravie – tchèque et slovaque de l'Ouest.
2. Territoire central : haut sorabe, polonais (sud), slovaque de l'Est, čakavien (nord), slovène (nord) (avec le kajkavien).
3. Périphérie : bas sorabe, polonais, quelques dialectes slovaques de l'Est, slovène, dialecte čakavien du Sud.
4. Périphérie extrême : polabe, cachoube et polonais (nord), slovène (sud), slovaque central, štokavien.

La connaissance de ces faits permet de comprendre la correspondance entre la suite « voyelle – j – voyelle » et « voyelle » dans les langues slaves de l'Ouest (Nord et Sud) sachant que le problème ne se pose pratiquement pas pour les langues slaves de l'Est et du Sud-Est. Il s'agit donc d'un phénomène essentiellement afférent aux langues que nous avons présentées comme « slaves de l'Ouest » ce qui pourrait soutenir davantage l'hypothèse de Starostin.

4.3 L'évolution des nasales

Les nasales du vieux-slave sont « ǫ » et « ę ». Elles semblent avoir disparu dès le milieu du X^{e} siècle pour le russe, vers la fin du X^{e} siècle pour les langues slaves de l'Ouest à l'exception du polonais où subsistent deux nasales « ą » et « ę ».

Si le polonais n'était pas à considérer, nous aurions dans les langues slaves une tendance générale de la vocalisation de « ǫ » en « u ». Mais la présence du polonais, extrêmement importante pour la prise de conscience de l'emplacement des nasales et de leur nature, nécessite la prise en compte de l'accentuation pour offrir une typologie de la vocalisation des nasales. Il n'y a pas, en effet, de correspondance biunivoque entre les deux séries de nasales comme nous le verrons ci-dessous.

Nous utilisons le projet néerlandais « Indo-European Etymological Dictionary », base de données compilant de nombreux dictionnaires étymologiques. Pour les langues slaves, cette base de données, mise en place par Rick Derksen sous l'égide de l'Université de Leiden et de l'éditeur Brill, a donné lieu à l'édition d'un dictionnaire, *Etymological Dictionary of the Slavic Inherited Lexicon*, qui nous sert de référence première et de moyen d'orientation. Par rapport à nos besoins et à nos orientations de recherche, le seul défaut de ce dictionnaire est le nombre d'entrées dont nous souhaiterions qu'il soit plus élevé.

Nous avons repris à notre compte le système accentuel qu'il utilise (ton et quantité) en six éléments : ton ascendant long (aigu), ton ascendant bref, ton descendant long (circonflexe), ton descendant bref, voyelle inaccentuée longue, voyelle inaccentuée brève. Ceci nous permet de trouver un certain nombre de critères de calculabilité qui, sinon, ne serait pas assurée.

Voici deux exemples montrant les processus de calcul [14] avec la nasale « ǫ » avec un ton ascendant long (1er exemple) ou en tant que nasale non accentuée longue (2e exemple) :

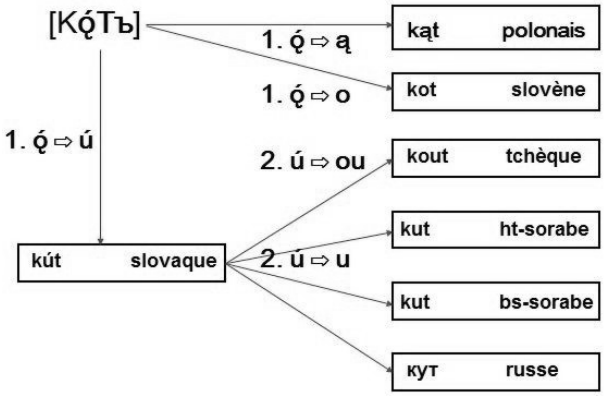


Figure 17 – Transformation de la nasale « ǫ » avec un ton ascendant long

Nous avons la transformation systématique de la nasale « ǫ » avec un ton ascendant long en « o » en slovène. En polonais, la nasale « ą » est typique pour cette transformation.

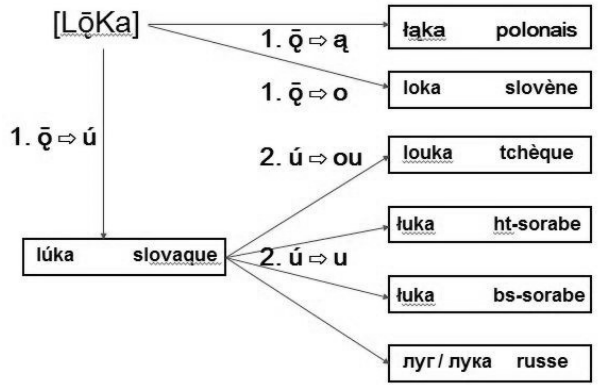


Figure 18 – Transformation de la nasale « ǫ » non accentuée longue

Dans ce second cas, il convient de remarquer que la forme « *лука* » (*luka*) en russe est dialectale. En croate, le mot « *luka* » n'apparaît que dans le syntagme « *zračna luka* » (*aéroport*).

Dans ces deux exemples, le système est le même et ne connaît pas d'exception. À partir de la nasale « ǫ », nous obtenons la transformation en « ą » pour le polonais et en « ú » pour le slovaque.

14. Qui ne correspondent pas nécessairement à une réalité historique attestée.

C'est la forme slovaque qui peut ensuite servir de base pour la génération automatique dans les autres langues slaves (forme en « u ») avec la forme du tchèque, qui par une transformation plus tardive, fait passer « ú » à « ou ». Le slovène transforme la nasale « ǫ » en « o ». Cela amène en slovène des homographies issues d'un véritable « o » et d'un « o » qui est originellement un « u » : « *kos* » peut correspondre au « *kos* » du tchèque (merle) ou au « *kus* » du tchèque (morceau) :

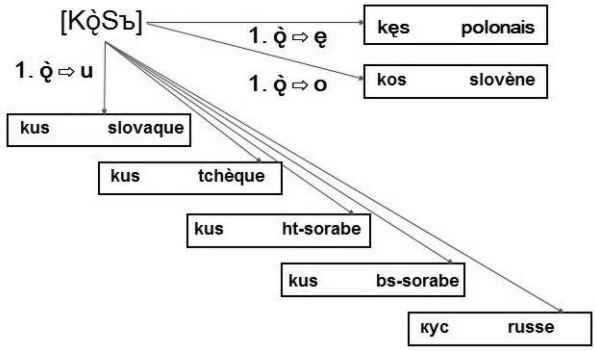


Figure 19 – Transformation de la nasale « ǫ » avec un ton descendant long

Dans ce troisième exemple, celui de la nasale « ǫ » avec un ton descendant long, nous avons un « u » en slovaque (u bref) qui donne « u » en tchèque, en sorabe (haut et bas), en russe, en croate où nous avons également « *kus* » et « o » en slovène : « *kos* ». Dans ce cas, nous avons généralement « ę » en polonais (mais il peut y avoir aussi des alternances avec « ą » ou « ą » / « ę »). Nous observons de nombreuses séries régulières, par exemple : « *гусmoй* » en russe, « *hustý* » en tchèque et slovaque, « *gęsty* » en polonais, « *gust* » en croate et « *gost* » en slovène (dense).

De manière générale, la vocalisation de la nasale « ę » a été accompagnée d'une mouillure derrière labiale. Elle semble se faire non pas d'après le type d'accentuation, mais en fonction du contexte qui l'entoure et qui produit deux séries différentes.

La première série peut être représentée par « *męso* » (viande) qui donnera d'abord une forme mouillée « *m'ęso* », puis la série « *мясо* » en russe (forme mouillée), « *mięso* » en polonais (forme mouillée), « *mäso* » en slovaque (le « ä » a été une forme intermédiaire en tchèque, transformée ensuite en « a »), « *maso* » en tchèque, « *meso* » en slovène, bosniaque, croate et serbe.

Un exemple de la seconde série sera « *pętь* » (cinq) qui donne un ensemble avec « p » mouillé « *пять* » en russe, « *pięć* » en polonais, « *pět* » en tchèque et non mouillé avec « *päť* » en slovaque et « *pet* » en slovène et BCMS.

4.4 La disparition et la vocalisation des jers

Il s'agit d'un phénomène essentiel qui concerne toutes les langues slaves. Dûment expliqué, il ouvre une vaste compréhension du multilinguisme slave.

Les jers sont de nature vocalique, prononcés de manière très brève. Ils se trouvent dans un mot – supposé être majoritairement de structure « consonne – voyelle – consonne – voyelle… » – en position de voyelle. Ils sont soumis, comme les voyelles de l'époque, aux environs des années 800, à une opposition de mouillure (jer dur, représenté par « ъ », opposé à un jer mou, représenté par « ь »).

La loi de Havlík rend compte de la disparition et de la vocalisation des jers. Nous allons montrer ces processus, calculés à partir de la fin du mot, à l'aide du schéma ci-dessous :

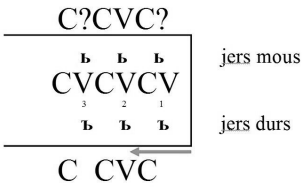


Figure 20 – Visualisation de la loi de Havlík – vocalisation des jers pairs, « disparition » des jers impairs

4.4.1 Série impaire

Les jers de rang impair disparaissent, mais de manière différenciée, ce qui apparemment n'est pas abordé par la loi de Havlík : un jer dur de rang impair disparaît sans laisser de trace. C'est justement cela qui fait qu'un substantif masculin est terminé par une consonne, parce que le jer, vocalique, qui terminait la chaîne est tombé et a laissé la consonne « à nu ». À notre avis, dans un cas de ce type et seulement dans un cas de ce type, on peut utiliser un « morphème zéro » qui rend compte de cette disparition / absence. Par contre, le jer mou impair disparaît en tant que tel, mais demeure sous une forme combinée avec la consonne précédente où il représente la mouillure de cette consonne. Nous donnons ci-dessous l'exemple qui nous semble le plus intéressant, la transformation de la suite « t – ь » :

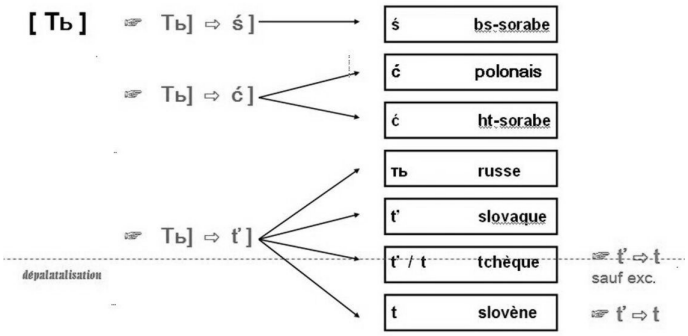


Figure 21 – Transformations lors de la disparition du jer mou impair

Le résultat premier est un « t mouillé » qui demeure en russe (t + signe mou), en slovaque (« t' »), ce qui est aussi le cas du tchèque, mais qui a considérablement

supprimé la mouillure des labiales et des dentales si bien que les formes mouillées sont minoritaires (à l'exception de « ň » qui reste abondant). Le slovène a supprimé presque toutes les mouillures, si bien qu'ici la forme est « t », donc la seule langue où le jer mou impair a entièrement disparu.

En polonais et en haut-sorabe, c'est le graphème « ć » qui en résulte et représente la forme palatalisé de « t », ce qui explique que jusqu'à une malheureuse réforme d'il y a quelques années, l'ordre alphabétique en haut-sorabe était « t, ć » pour montrer cette relation de mouillure entre les deux graphèmes. En bas-sorabe, le caractère résultant est « ś ».

La série de transformation obtenue est d'autant plus intéressante qu'elle représente le caractère final du morphème d'infinitif pour toutes les langues représentées, sauf les langues du Sud-Est qui possèdent un infinitif terminé en « -ti ».

D'autres séries de transformation sont données par les suites « Dь », « Nь », « Rь » ou « Sь ».

4.4.2 Série paire

Les jers de rang pair (à partir de la fin du mot) sont transformés en voyelle qu'ils soient durs ou mous. Nous donnons ci-dessous un tableau qui montre que la vocalisation est en « a » pour les deux jers en croate, en « e » pour le tchèque, le polonais, le sorabe et une partie du slovaque, le slovaque central présentant une situation plus complexe. Le slovène pose problème dans la mesure où chacun des jers peut être vocalisé soit en « a », soit en « e ».

Tableau 9 – Transformation des jers

	jer mou (Ь)	**jer dur (Ъ)**
tchèque	e	e
polonais	e	e
sorabe	e	e
slovaque (est et ouest)	e	e
slovaque central	e (o) a	o / e a
slovène	e / a	e / a
serbo-croate	a	a
russe	**e**	**o**
macédonien	**e**	**o**
bulgare	**e**	**Ъ**

Cette situation ne permet pas de déduire la nature du jer sous-jacent. Seuls le russe, le macédonien et le bulgare différencient dans leur vocalisation le type du jer d'origine. La connaissance du même mot dans l'une de ces langues permet de lever le mystère :



Figure 22 – Vocalisation du jer 1

Ainsi, nous avons « e » dans le mot russe « *день* », ce qui indique la présence d'un jer mou dans cette position vocalique. Nous pourrons, dans les bases de données sur les langues slaves, reporter cette valeur pour les langues où la nature du jer sous-jacent n'est pas visible (figures ci-dessus et ci-dessous).

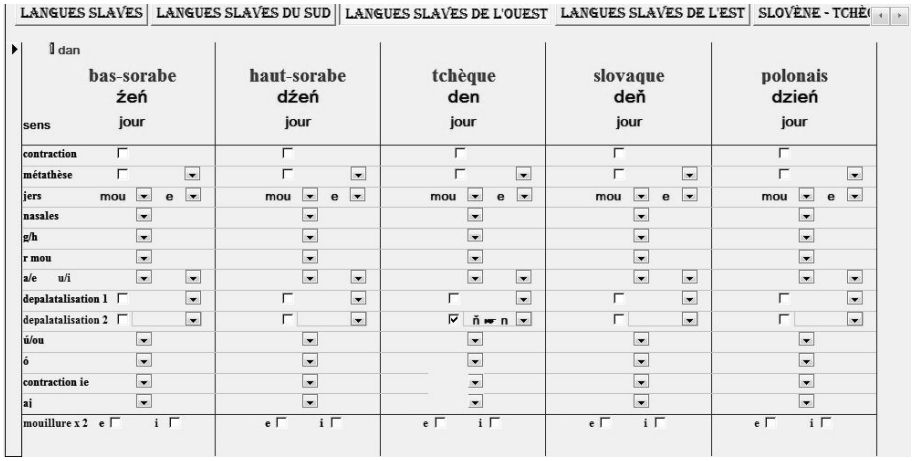


Figure 23 – Vocalisation du jer 2

Les quatre phénomènes que nous venons de présenter (métathèse, contraction, évolution des nasales, disparition et vocalisation des jers) sont ceux qui marquent le plus la différenciation entre les langues slaves.

Les phénomènes plus tardifs que nous allons examiner maintenant (période des XI^{e}-XIV^{e} siècles et période des XV^{e}-XVI^{e} siècles) concernent un nombre de langues plus restreint, concentré sur les langues slaves de l'Ouest ou certaines d'entre elles. Nous examinerons ces phénomènes parce qu'ils permettent de montrer en quoi le tchèque et le slovaque se distinguent l'un de l'autre.

4.5 Le passage de « g » à « h »

Cette transformation a eu lieu en haut-sorabe, tchèque et slovaque et oppose ces trois langues au reste du groupe slave de l'Ouest. Une conséquence favorable est la calculabilité induite par le phénomène. En effet, dans ces langues, le « g » slave étant devenu « h », si l'on trouve dans un texte la présence d'un « g », le mot qui le contient est d'origine étrangère. C'est l'un des critères de reconnaissance automatique des emprunts (d'origine gréco-latine) utilisé dans l'analyse automatique du tchèque.

4.6 L'évolution du r mouillé

Le « r » demeure en situation mouillée en russe, par exemple dans un mot comme « *море* » (mer). On le retrouve en sorabe sous forme d'un digraphe « rj » : « *zwěrjata* » (animaux), « *brjóh* » (berge), mais aussi sous la forme d'un graphème spécifique en bas-sorabe, par exemple dans des mots tels que « *farář* » (prêtre) ou « *rybář* » (pêcheur). Le « r » mouillé a donné « ř » en tchèque (« r » avec *háček*) et son correspondant « rz » en polonais. Ces formes mouillées s'opposent ainsi au « r » du slovaque et du slovène qui ont perdu la mouillure. C'est la raison pour laquelle « r » en slovaque est une consonne mixte (*cf.* le tableau donné en fin de a) la métathèse).

4.7 La transformation des voyelles dures (« a » et « u ») en voyelles molles (« ě » et « i »)

Vers le XII^{e} siècle, la suite consonne molle – voyelle dure « a » se transforme en une suite consonne molle – voyelle molle « ě ». Les consonnes molles sont les molles d'aujourd'hui (ď, ť, ň, č, ř, š, ž, c) et les molles qui sont devenues les mixtes d'aujourd'hui par perte de la mouillure (b', l', m', p', s', v', z'). On assiste donc dans ce contexte à la transformation de « a » bref en « ě » et de a long en « ie ».

Nous pouvons donner des exemples tels que :

« *ovca* » – « *ovcě* » (mouton), « *zem'a* » – « *země* » (terre) ou « *duša* » – « *dušě* » (âme)

Cette transformation est très importante pour au moins deux raisons :

- Elle ne concerne que le tchèque et le sépare ainsi de toutes les autres langues slaves.
- À partir d'une certaine époque, la suite issue de cette transformation s'est également transformée et le « ě » est devenu « e », ce que nous interprétons comme une impossibilité de « double mouillure », « š » étant une consonne mouillée et « ě » une voyelle mouillée.

Cette impossibilité de double mouillure est mise en valeur par la comparaison avec le haut sorabe (tchèque « *jed* » (poison) face à « *jěd* » en haut-sorabe) [fig. 24].

Dans le jeu de la double mouillure, le *háček* s'efface devant le « j ». C'est toujours lui le maillon faible.

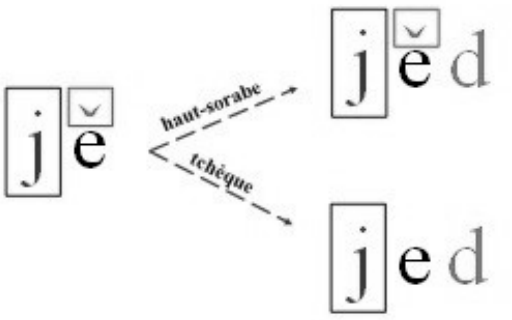


Figure 24 – Double mouillure en haut-sorabe

Au milieu du XIV[e] siècle, mais cette fois-ci uniquement en Bohême, intervient la transformation consonne molle – voyelle dure « u ». La voyelle dure « u », brève (u) ou longue (ú) se transforme en voyelle molle « i » brève (i) ou longue (í). Ainsi « *jug* » (sud) devient « *jih* » en tchèque, « *ľud* » (*peuple*) devient « *lid* ».

Par l'application successive des deux transformations, « *pľúca* » (poumons) devient « *plíce* ».

Ces deux transformations singularisent le tchèque face aux autres langues slaves et le sépare en particulier du slovaque.

Une autre conséquence est la multiplication des paradigmes de flexion nominale par rapport au slovaque (voir ci-dessus, « À la recherche du système linguistique – Les substantifs »).

4.8 La dépalatalisation des labiales et des dentales en tchèque

La dépalatalisation des labiales et des dentales en tchèque (« *peču* » [je fais cuire] face à « *pjeku* » en haut-sorabe, « *kost* » [*os*] face à « *kosť* » en slovaque) place le tchèque en position intermédiaire entre les langues slaves de l'Ouest, mouillées et le slovène, fortement dépalatalisé.

4.9 La diphtongaison du « u » long (ú) tchèque en « ou »

La diphtongaison du « u » long (ú) tchèque en « ou » sépare cette langue du slovaque avec une correspondance régulière (ú/ou) : « *súd* » / « *soud* » (tribunal).

4.10 La « monophtongaison » de « uo » et de « ie »

La « monophtongaison » de « uo » et de « ie » est une transformation importante.

Le « o » long (ó) devenu « uo » se transforme ensuite en « ů » en tchèque et en « ô » en slovaque. Les deux formes se correspondent de manière régulière : « *původ* » (origine) face à « *pôvod* » en slovaque.

Une première conséquence de cette transformation est que tout o long apparaissant dans un texte tchèque indique une origine étrangère (« *cirhóza* », « *mykóza* », « *skle-*

róza »). Il devrait en être de même en slovaque pour les mêmes raisons : le o long slave a été transformé.

La transformation de « ie » en « i mou long » (í) isole le tchèque des autres langues slaves et notamment du slovaque. Mais, en même temps et avec la transformation de « uo » en « ů », cela lui apporte une calculabilité plus grande. En effet, la seule diphtongue autochtone en tchèque étant « ou », toute contiguïté de deux voyelles indique une origine étrangère. Pour rechercher des emprunts, il est nécessaire dans les autres langues de faire la liste des paires de voyelles autochtones.

4.11 La transformation de la suite « aj » en « ej » en tchèque

Enfin, par souci d'exhaustivité, signalons qu'en tchèque la suite « aj » au sein d'une syllabe se transforme en « ej ». Nous avons donc le préfixe de superlatif en « nej- » et « *vejce* » (œuf) au nominatif singulier, mais « *vajec* » au génitif pluriel. Les autres langues slaves conservent le « aj ».

En conclusion, le tableau ci-dessous a le mérite de montrer, sur la base de faits historiques, que le tchèque pendant six siècles et par huit phénomènes différents s'est constamment éloigné de la plateforme commune dont le slovaque ne s'est écarté que par deux phénomènes (la perte du r mouillé et la transformation du o long).

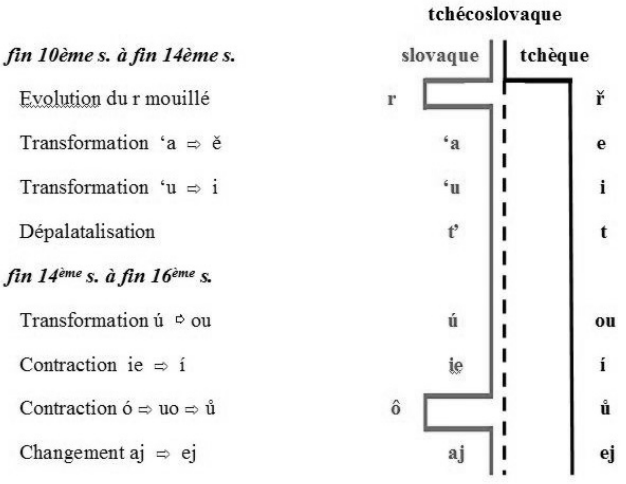


Figure 25 – La plateforme tchéco-slovaque commune et les divergences tchèques et slovaques

5 Systèmes graphiques et calculabilité

Les éléments, synchroniques et diachroniques présentés ci-dessus, qui nous servent à définir le système linguistique des langues slaves de l'Ouest doivent être complétés par une étude importante, celle des systèmes graphiques mis en présence. Si plusieurs de

ces systèmes ont été l'objet de notre attention [15], l'étude systématique des différents systèmes de manière globale n'en est encore qu'à ses débuts. Cette étude est nécessaire pour définir un cadre général en vue de l'analyse automatique de ces langues qui puisse donner une place importante à la calculabilité. Nous définissons la calculabilité comme la possibilité de faire correspondre une forme et une valeur, forme graphique dans un texte d'une part et p. ex. valeur morphosyntaxique d'autre part [16].

Une écriture alphabétique, système graphique, est en fait un code formel destiné à consigner à l'écrit la chaîne éventuellement parlée d'une langue. « Éventuellement », étant donné que l'homme moderne communique autant, sinon plus, à l'écrit qu'à l'oral. L'histoire des langues nous a montré qu'un même type d'écriture peut servir à la communication de plusieurs langues, parentes ou non. En premier lieu, on peut se référer à l'écriture cunéiforme qui a servi à des langues sémitiques, indo-européennes ou même autres, la première d'entre elles étant le sumérien. L'alphabet cyrillique permet de consigner les langues slaves de l'Est et du Sud-Est. Mais, à l'inverse, et cela est encore plus probant, une même langue peut avoir reçu des codes écrits différents, par exemple arabe, puis latin, cyrillique et à nouveau en caractères latins différents des premiers pour les langues turques d'Asie centrale. Plus près de nous, le serbe écrit en cyrillique possède aussi une tradition ancienne d'écriture en caractères latins et a été, plusieurs siècles durant, écrit en caractères arabes [17].

Les langues slaves de l'Ouest (Nord et Sud), écrites en caractères latins, sont souvent regroupées sous l'appellation « Slavia romana ». Elles ont en commun d'avoir subi des périodes d'évolution naturelle, notamment en ce qui concerne la phonologie, mais aussi des périodes de réformateurs qui ont imprimé leur marque avec un plus ou moins grand bonheur. La période des renouveaux nationaux (depuis la fin du XVIII^{e} siècle) qui a permis d'ériger des nations slaves après la première guerre mondiale a été particulièrement active en ce sens.

Deux autres grandes périodes ont marqué l'évolution des écritures slaves de l'Ouest. C'est en premier lieu la création de l'écriture glagolitique qui est liée à deux aires géographiques des langues slaves de l'Ouest : la Grande Moravie, lieu de son essor et la Croatie, lieu de sa conservation et de son évolution. L'aire bulgaro-macédonienne a été le lieu de sa transformation en cyrillique.

C'est le prince morave, Rostislav qui, dans le but de s'affranchir de la tutelle des Francs, obtint de l'empereur byzantin Michel III l'envoi de missionnaires aptes à former un clergé morave. Constantin et Méthode, originaires de Salonique, parlant le slave représenté en Macédoine et en Grèce, arrivèrent en Moravie en 863 ou 864. Constantin forgea une nouvelle écriture, l'écriture glagolitique, à partir de l'onciale grecque en la complétant, pour les phonèmes propres au slave, par des signes inspirés des alphabets orientaux (il connaissait l'hébreu et l'araméen). Le trait de génie de Constantin est d'avoir créé une écriture dans laquelle un graphème unique renvoyait à un phonème, ce que conserve le cyrillique.

15. *Cf.* Pognan, P. (1999) pour le tchèque.
16. La calculabilité fait apparaître des régularités également utiles à l'enseignement de la langue.
17. Cette thématique a été l'un des nombreux sujets de recherche du Professeur Werner Lehfeldt de l'Université de Göttingen.

L'alphabet glagolitique se compose d'un ensemble de 38 caractères (les 24 caractères du grec et 14 nouveaux caractères « slaves »). F. Mareš a publié le tableau suivant [18] [fig. 26].

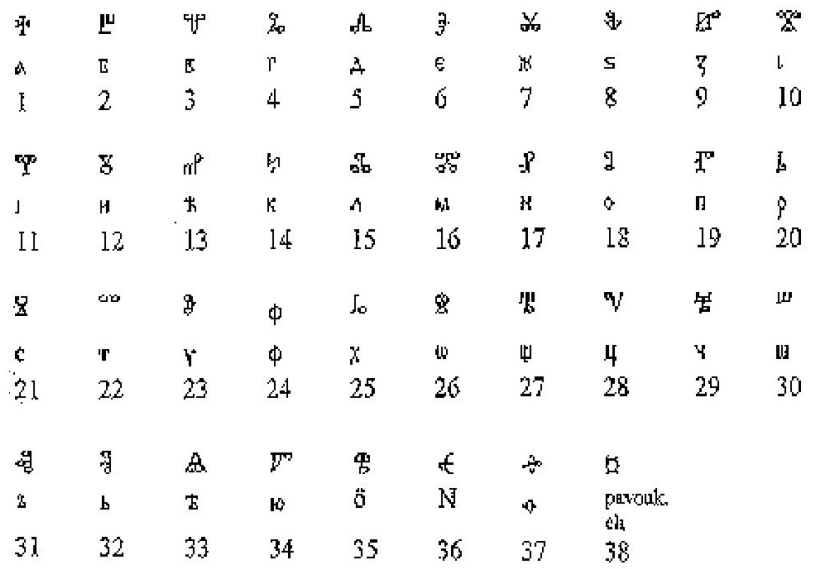


Figure 26 – Alphabet glagolitique

Les successeurs de Méthode, bannis de Moravie, emportèrent avec eux le rite slave et l'écriture glagolitique en Bulgarie, en Croatie et en Bohême. En Bohême, elle se maintient avec la liturgie slave jusqu'à la fermeture du couvent des Bénédictins slaves à Sázava en 1097. L'empereur Charles IV fait appel à des moines croates pour restaurer la liturgie slave au monastère d'Emmaüs (Prague) en 1347.

La Croatie conserve l'usage de la liturgie slave, privilège accordé officiellement par le pape Innocent IV en 1248, privilège renouvelé en 1935. Cet usage fait que l'écriture glagolitique de Croatie suit sa propre évolution. Elle devient plus carrée que l'écriture primitive, morave, ronde et les textes sont marqués par l'influence du croate.

La seconde période importante est certainement, 550 ans après la Grande Moravie, la réforme apportée par Jan Hus dans le désordre graphique qui régnait dans les Pays tchèques (*Ortographia bohemica* paru en 1412, republié à Wiesbaden en 1968). Elle consiste en le maintien essentiel de la correspondance graphie – phonie (à l'exception du digraphe « ch » près) notamment en remplaçant les digraphes par des caractères diacrités.

L'esprit de cette réforme, acquise seulement au XVI^{e} siècle grâce à l'action des Frères moraves, a vraisemblablement présidé à d'autres actions en faveur de langues slaves, mais aussi, plus tardivement, en faveur de l'écriture du lituanien qui s'en est inspiré

18. La première ligne est l'alphabet glagolitique, la seconde l'alphabet cyrillique (ancien) correspondant.

explicitement. C'est vraisemblablement le polonais qui est le moins concerné, ayant conservé de nombreux digraphes dans son système graphique qui nécessite l'élaboration d'une graphotaxe pour un traitement automatique efficace. Cependant, « Stanisław Zaborowski (v. 1470-1530) fut le premier à appliquer cette idée, dans un opuscule publié en 1515, plaçant des points au-dessus des consonnes chuintantes [19]. »

Même fondés sur un alphabet, les systèmes graphiques peuvent avoir des caractéristiques différentes.

Ces différences peuvent résulter de décisions d'aménagement. Ainsi le tchèque conserve-t-il dans l'écriture la distinction entre les deux « i » bien que leur prononciation se soit unifiée aux environs du XVII^{e} siècle, ce qui assure une certaine vision étymologique de la langue et conserve souvent le parallélisme p. ex. avec le russe.

Un avantage fondamental du tchèque (plus réduit en slovaque) est la conservation dans l'écriture de la longueur qui assure à la langue un degré plus élevé de calculabilité. En ce qui concerne le « i », nous avons donc quatre formes : « i », « í », « y », et « ý ». En tchèque, un mot terminé par « -ý » est toujours, à deux exceptions près, un adjectif dur au masculin singulier nominatif ou accusatif inanimé face au « -y » qui peut être une terminaison d'adverbe, une forme fléchie de substantif ou même de verbe (forme du passé, participes). En polonais, la distinction n'existant pas, la calculabilité est nettement inférieure.

Que dire alors du slovène qui ne possède que « i » pour les quatre formes tchèques ? Le slovène marque dans les dictionnaires et certains exposés didactiques les différentes formes d'accentuation, mais pas dans les textes courants. Il en résulte que la manipulation automatique du « e » intercalaire qui dépend en grande partie de sa prononciation pose problème dans des processus de flexion automatique.

Une distinction essentielle, bien visible dans les langues slaves de l'Ouest, est la différence entre une conception phonétique et une conception phonologique de l'écriture. Le lieu où s'applique l'assimilation régressive de sonorité en est une excellente illustration.

Prenons l'exemple de l'adjectif tchèque « *blízký, blízká, blízké* » (proche) avec la forme d'adverbe en « *blízko* ». Dans la prononciation, le « z » sera assourdi sous l'influence du « k », lui-même sourd. La forme « *blíže* » (de manière plus proche) est parfaitement compréhensible, étant dans le système qui assure par palatalisation le rapport entre « z » et « ž ». L'écriture de l'adjectif tient compte de la loi phonétique d'assimilation qui s'exerce à l'oral, mais conserve à l'écrit la vision du système. Nous parlerons ici d'alphabet phonologique. Si nous considérons le croate, la situation sera « *blizak, bliska, blisko* ». Nous avons ici typiquement une conception phonétique de l'écriture. Grâce à la forme courte de l'adjectif, sur la racine « BLiZK », il y a vocalisation du jer pair en « a », ce qui nous donne la forme « *blizak* » où la sonore « z » se fait entendre. La forme « *bliže* » (plus proche) peut se comprendre par rapport à la forme masculine de l'adjectif. Il n'en reste pas moins que l'exposé didactique se complique sérieusement. Mais le polonais, fonctionnant avec les formes longues de l'adjectif, donnera la série « *bliski, bliska, bliskie* » et l'adverbe « *blisko* ». L'ensemble

19. Zaremba, C. (2009), p. 41.

répond à un critère phonétique. La forme de comparatif de l'adverbe sera « *bliżej* ». L'adoption d'une vision phonétique de l'écriture a pour conséquence que le système graphique n'est pas en accord avec le système linguistique, « ż » n'étant pas dans le domaine de la palatalisation de « s ». Il complique l'exposé didactique (comment expliquer que « s » se transforme en « ż ») et la compréhension de la langue par l'apprenant. Si pour le traitement automatique on est obligé de considérer toute suite de sourdes comme étant l'assourdissement éventuel d'une sonore suivie d'une sourde, on perd de la calculabilité, car un certain pourcentage de ces suites sera réellement un assemblage de sourdes.

Un des points essentiels à élucider est le fonctionnement et la représentation des phénomènes de mouillure dans l'écriture latine de ces langues slaves.

Nous avons vu une première approche de ce problème lors de la perte du jer mou impair fusionnant avec la consonne le précédant et une seconde avec la « double mouillure » liée au « e ».

Les questions de mouillure associée à la présence du « i mou » doivent être aussi étudiées attentivement. Les grammaires tchèques présentent les transformations « ď + i = di », « ť + i = ti » et « ň + i = ni » comme de simples réécritures. Nous pensons qu'il s'agit d'un phénomène beaucoup plus profond : l'expression de la mouillure dans les systèmes graphiques des langues slaves de l'Ouest. Ainsi, le polonais fait clairement des équivalences régulières entre « ość / ości », « d – dź / dzi », « zd – źdź / ździ » ... où le « i » est nettement, par sa nature, une marque de mouillure. Dans les cas où il est présent, la « *kreska* » du polonais, maillon faible comme le *háček* dans la mouillure de « e » en tchèque, s'efface.

L'importance de la manière dont la mouillure est traitée peut être démontrée par les conséquences produites par l'adoption d'un code graphique plutôt que d'un autre, ce qui peut être donné par la comparaison du tchèque et du slovaque, langues appartenant au même sous-système.

En tchèque (*cf.* ci-dessus), le « e mouillé » est représenté par un « e » surmonté d'un háček (« ě »). Un adjectif tchèque est un mot qui possède une désinence toujours longue (désinence constituée par une diphtongue ou possédant une voyelle longue), ce qui est le cas fondamentalement aussi en slovaque. En slovaque, le « ě » n'existe pas, car il est supposé que les suites « de », « te » et « ne » sont suffisamment caractéristiques pour exprimer de manière non ambiguë une prononciation « dě », « tě » et « ně ». D'autre part, le slovaque possède une loi dite de rythme qui empêche la succession de deux syllabes longues (la seconde est abrégée) et ceci quelle que soit la valeur de la longueur, par exemple valeur de désinence d'adjectif. Le croisement de ces phénomènes met en évidence des résultats très différents en tchèque et en slovaque. Soient les deux suites qui se correspondent en tchèque et en slovaque [tab. 10].

Tableau 10 – La mouillure du « e » en tchèque – comparaison avec le slovaque

	adjectif masculin	adjectif féminin	adjectif neutre	adverbe
tchèque	krásný	krásná	krásn**é**	krásn**ě**
slovaque	krásny	krásna	krásne	krásne
	beau	*belle*	*beau*	*de belle manière*

Alors qu'en tchèque la graphie est classificatoire, en slovaque la « calculabilité » n'est plus assurée. On se retrouve dans une situation où la machine est confrontée à une ambiguïté systémique et où les apprenants étrangers sont en difficulté : il est nécessaire de comprendre le texte pour pouvoir l'interpréter (dans un cas, adjectif avec application de la loi de rythme, dans l'autre cas, adverbe où il y a économie du « ě » derrière « n ») et pour pouvoir le lire à haute voix.

6 Outils de comparaison

Parmi les éléments indispensables de comparaison se trouvent les inventaires des racines de chacune des langues. Un choix judicieux de racines, en particulier de nature verbo-nominale, peut avoir un effet bénéfique sur l'appropriation de la langue par l'apprenant, autochtone ou étranger. Il est aussi la base du traitement automatique de la dérivation.

Pour que ces inventaires de racines puissent être comparables, il est nécessaire que la définition du découpage morphématique afférent le soit aussi. Or, c'est un lieu où les visions nationales sont très variées. Elles sont appuyées très généralement sur un héritage indo-européaniste où une racine ne peut pas être uniquement consonantique. Elles mettent souvent en avant la primauté de la désinence face à la racine pour montrer une appartenance à une personne, par exemple la valeur de « m » en tant que référent de la première personne.

Nous présentons deux exemples où notre conception du découpage morphématique (qui jusqu'à preuve du contraire offre une analyse homogène) n'est pas en accord avec des présentations traditionnelles courantes. Il s'agit en premier lieu, en slovaque, de la conjonction d'une vision prioritaire des désinences de personnes en conjugaison liée à la présence supposée d'un morphème zéro, l'ensemble provoquant une conception où le verbe slovaque possède deux radicaux au présent.

Une majorité de grammaires donne le découpage de gauche (il s'agit du verbe être au futur) où, pour montrer les valeurs génériques des indices personnels, on fait apparaître deux thèmes de conjugaison non seulement de manière inutile et non justifiée, mais qui, de plus, multiplie le nombre de racines et provoque des difficultés supplémentaires chez les apprenants qu'ils soient natifs ou étrangers. Nous donnons, en colonne de droite, le découpage que nous utilisons dans l'analyse des langues slaves de l'Ouest.

Tableau 11 – Découpage morphématique – exemple du slovaque

singulier			**pluriel**		
1	**bude – m**	**bud – em**	**1**	**bude – me**	**bud – eme**
2	**bude – š**	**bud – eš**	**2**	**bude – te**	**bud – ete**
3	**bude**	**bud – e**	**3**	**bud – ú**	**bud – ú**

Le second cas concerne la définition de l'infixe d'itération en tchèque qui est lié à une conception non consonantique des racines et à la non primauté des racines dans une conception morphématique.

Partant du fait que toutes les descriptions grammaticales ne sont que le reflet de la conception que nous nous faisons des langues que nous parlons ou que nous connaissons, nous allons décrire la logique et les raisonnements qui sous-tendent notre conception. De manière générale également, tout l'enseignement du tchèque donne le segment « va » comme infixe d'itération. Il n'en est rien pour nous et d'ailleurs, les segments que nous donnons (au nombre de 4) répondent à la structure majoritaire de la suffixation qui est voyelle – consonne(s). Nous prendrons une entrée du dictionnaire Vallex (dictionnaire de valences verbales) comme premier pas de démonstration : *klouzat*^{impf}, *klouznout*^{pf}, *klouzávat*^{iter}. Nous avons ici en entrée la forme imperfective « *klouzat* » glisser, la forme perfective « *klouznout* » et la forme itérative (donc imperfective) « *klouzávat* ». Si nous mettons ces trois formes en commun, nous obtiendrons une structure cohérente en ordonnant les morphèmes radical et suffixaux de la manière suivante :

Tableau 12 – Exemple tchèque de « klouzat »

pf	klouz		nout
ipf			at
iter		áv	at

Cet exemple a pour but de faire apparaître la dimension de la racine et de l'infixe d'itération. Nous pouvons confirmer la forme de cet infixe d'itération en partant d'autres situations. Prenons le mot étranger « *telefon* ». Nous pouvons former le verbe imperfectif « telefonovat » pour lequel nous donnerons pour le moment le découpage suivant : « *telefon -ovat* ». La prise en compte de la forme itérative « *telefonovávat* » apporte par rapport à « *telefonovat* » une remarque intéressante : « ovat » n'est pas un, mais deux suffixes entre lesquels s'insère l'infixe « áv » qui est ainsi confirmé. De même, si nous prenons le verbe polonais « *wychować* » (élever) (*wy-chow-ać*), nous aurons la forme itérative en « *wychowywać* », le seul segment différenciant les deux verbes est le segment -yw-, d'où « *wy-chow-yw-ać* ». L'alternance de voyelles

en polonais (de « a » à « y ») devrait être suffisante pour empêcher de produire des racines en « chowa » et « chowy ». L'infixe d'itération polonais « yw » confirme bien la position prise en tchèque. L'itération en tchèque est exprimée par la série d'infixes à voyelle longue « -áv- », « -év- », « -ív- » et « -ýv- » et non pas par l'unique infixe « -va- ».

Pour un couple de verbes tels que « *dát* » / « *dávat* » donner, le découpage généralement proposé est :

« *dá-t* » et « *dá-va-t* »,

où l'on voit que le deuxième découpage est en contradiction avec les découpages que nous préconisons. Pour être en conformité, nous devons donner « *d-áv-at* » où nous obtenons bien l'infixe d'itération « -áv- » et le suffixe d'infinitif en « at ». Dès lors, pour rester cohérent, il convient de découper « *dát* » en « *d-át* » (un infinitif monosyllabique prend toujours une longueur). Nous obtenons la racine de donner sous la forme « D », ce qui n'est généralement pas reconnu par les milieux de la slavistique.

Nous allons maintenant confirmer le découpage de la racine « KLouZ » dans les verbes « *klouznout* », « *klouzat* » et « *klouzávat* » glisser en les comparant à d'autres mots de la même famille « *kluzák* » planeur, « *kluzký* » glissant, gluant et « *kluziště* » patinoire. Il convient ici de rappeler que « ou » en tchèque provient de la diphtongaison d'un « u long » (« ú ») commun au tchèque et au slovaque provenant lui-même de la perte de nasalisation de la nasale « ǫ ». Sous la forme de l'opposition synchronique « ou/u », nous n'avons en réalité qu'une opposition de longueur « ú/u ».

Nous allons découper les trois nouveaux exemples respectivement en « *kluz-ák* », « *kluz-k-ý* » et « *kluz-išt-ě* » (nous pourrions également écrire « *kluz-išť-e* », puisqu'il y a une règle de réécriture de « ť » + « e » en « tě »). Nous obtenons donc deux variantes de la racine « klouz » et « kluz », ce qui confirme bien que le découpage des verbes était juste au niveau de la racine.

Nous donnons ci-dessous le tableau de la racine « KL{u/ou}Z » avec les formes nominales, puis des exemples de découpage en slovaque et en polonais [tab. 13].

Au-delà des « comparés », nous avons également besoin de « comparants » sous des formes diverses, essentiellement de tableaux ou d'interfaces dans nos bases de données. Pour pouvoir comparer des choses aussi élémentaires que des déclinaisons, nous sommes obligés de trancher parmi les diverses traditions, n'ayant pas toutes le même ordre des cas. Nous avons choisi l'ordre tchèque, parce qu'il est exprimé sous forme numérale qui est une pré-formalisation pour les traitements automatiques.

Nous allons montrer sous forme de tableau la variation de calculabilité pour un phénomène qui relève vraisemblablement plus de la langue que du système graphique : il s'agit, en slave, de la couverture par une consonne d'une éventuelle voyelle de tête. La situation de chacune des langues donne une calculabilité différente dans la reconnaissance automatique des emprunts.

20. Découpage que propose Jozef Mistrík, dans *Moderná slovenčina* (*Le slovaque moderne*) (p. 134).

Tableau 13 – Tableau de la racine « KLu/ouZ »

tch.	préfixe	racine	infixe	suffixe	suffixe désinentiel
pf		KLouZ		n	out
ipf					at
iter			áv		at
		KLuZ		k	ý
				ák	
				išt	ě
sk	préfixe	racine	infixe	suffixe	suffixe désinentiel
iter		CHyT	áv [20]		ať
pol	préfixe	racine	infixe	suffixe	suffixe désinentiel
	wy	CHoW			ać
iter	wy		yw		ać

Une tendance naturelle des langues slaves est de ne pas mettre de voyelle en tête de mot. Déjà dans *Dějiny české řeči a literatury* (1792), Josef Dobrovský notait : « *Slovan nechává samohlásku 'a' na začátku slova zřídka kdy, 'e' pak nikdy bez joty. Říman říká 'est', Slovan 'jest'* ». En général, cette voyelle de tête est « couverte » par « j » pour « a », « e » et « i » et par « w » pour « o » et « u ». C'est ce qui fait la différence des troisièmes personnes du singulier du présent du verbe être en français et dans les langues slaves de l'Ouest : « est » face à « *jest* » (forme aujourd'hui ancienne en tchèque et en slovaque).

Si nous analysons le tableau [fig27] :

voyelle en premier caractère d'un mot					
haut sorabe	1	ja	a		agronom
		je	e		energija
		ji	i		integrować
		wo	o	wokno	organ
		wu	u	wučbnica	unija
tchèque	0,6	ja	a		apatie
		je	e		embargo
		ji	i		ideologie
				okno	
				učebnice	
slovaque	0,5		i	istota	

Figure 27 – Voyelle en premier caractère

nous voyons que le haut-sorabe suit pleinement ce principe : contrairement aux autres langues il ne laisse pas un mot slave débuter par « o » ou « u ». Il possède « *wokno* » et non pas « *okno* » (fenêtre), « *wučbnica* » et non pas « *učbnica* » (manuel). Il en résulte une règle très simple : tout mot débutant par une voyelle est d'origine étrangère, sauf exception due à la non-métathèse, par exemple « *ert* » (bouche). Le tchèque ne prend aucune mesure pour « o » et « u » et il y a des interférences pour « i » en slovaque, p. ex. « *istota* » au lieu de « *jistota* » en tchèque (sûreté, sécurité).

On remarquera aisément dans un dictionnaire de quelque langue slave que ce soit que les mots autochtones débutant par « a » sont des mots outils en nombre réduit (moins de deux dizaines).

Un de nos outils, couramment employé dans nos bases de données, est un onglet spécialisé dans la comparaison de langues deux à deux prenant en compte la diachronie, mais aussi des phénomènes observés en synchronie. Nous donnons un exemple tchèque – slovène [fig. 28].

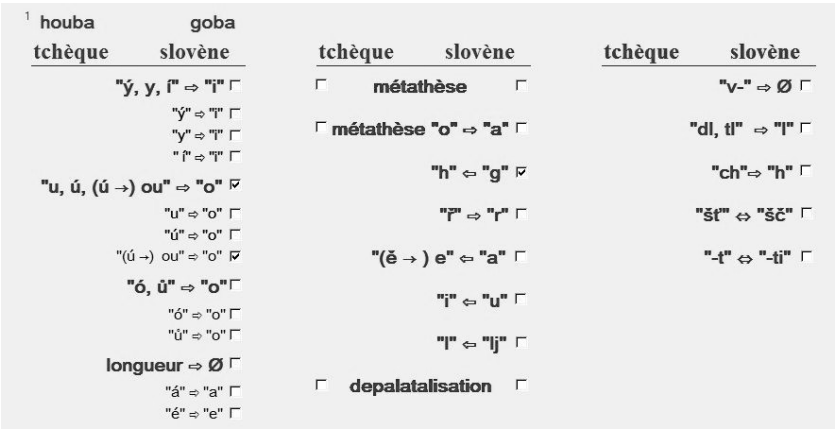


Figure 28 – Tableau comparatif tchèque – slovène

En première colonne, l'absence de longueur en slovène et les transformations vocaliques entre tchèque et slovène sont présentées en détail. La seconde colonne reprend des tables historiques les phénomènes qui marquent la forme des deux langues et la troisième colonne présente des différences plus spécifiques au couple tchèque – slovène.

Les informations consignées dans cette table permettent de générer automatiquement la forme slovène sur la base de la forme tchèque.

Le « h » tchèque découle du « g » slave qu'il convient de rétablir en slovène. Ainsi, « *houba* » (champignon) devient-il « gouba ». Le slovène a transformé les « u » sous toutes ses formes (y compris « ou », diphtongaison du « ú » slovaco-tchèque) en « o » : « gouba » devient alors « *goba* ».

Remarquons que la génération ne peut se faire pour ce mot que dans le sens tchèque vers slovène. Dans l'ordre slovène – tchèque, tout « o » slovène n'est pas issu d'un « u » alors que toute forme « u » tchèque devient « o » en slovène.

Nous avons présenté quelques éléments qui participent à la définition du système linguistique des langues slaves de l'Ouest, d'abord en synchronie, puis à l'aide d'évolutions phonologiques fondamentales et avons enfin abordé la question des codes écrits et les différences de calculabilité qui peuvent en découler.

L'ensemble de ces connaissances devrait pouvoir, à notre avis, aider à l'apprentissage d'une langue et à une meilleure compréhension des autres langues du groupe.

« *Dobra kniha druhdy ryzy złoto přesaha* » (« Un bon livre vaut parfois mieux que de l'or pur ») est la publicité faite par « Smolerjec kniharnja », la librairie sorabe de Bautzen / Budyšin. La compréhension est ici assurée surtout par des unités lexicales qui possèdent une généralité assez vaste au sein des langues slaves.

Cependant l'intercompréhension, même à l'écrit, peut être entravée ou même empêchée par plusieurs facteurs :

- on se trouve confronté à des unités lexicales inconnues de la langue que l'on connaît : par exemple, « *kobieta* » (femme) en polonais, « *blido* » (table) en haut-sorabe… ;
- le mot a un sens complètement différent de celui que l'on connaît dans sa langue de référence, par exemple « *otrok* », esclave en tchèque, est un enfant en slovène.



Figure 29 – Dobra kniha

Si on se réfère à une matrice faite de préfixations – racine – infixations – suffixations, cette matrice n'est pas toujours stable dans le temps même au sein de la même langue. Ainsi, peut-on noter des variations telles que « *náhled* » pour exprimer « l'opinion », mot formé sur l'une des racines de la vision « HLeD », jusqu'au milieu du XIX^{e} siècle, alors que depuis les Tchèques utilisent « *názor* » formé du même préfixe, mais avec une autre racine de la vision « ZR » sous sa forme nominale « ZoR ».

De manière beaucoup plus fréquente, on trouve des « discordances » dans la matrice entre langues slaves. Pour monter dans un moyen de transport, les Tchèques ont

« *nástup* » et « *výstup* » pour en descendre. La racine « STu/ouP » est la même racine que celle de « step » en anglais. Elle exprime le fait de faire un pas. Les Croates ont « *ulaz* » et « *izlaz* », ce qui, en situation, par exemple écrit sur un autobus, sera immédiatement compris, mais prêtera à rire ou au moins à sourire, la racine « LaZ » rappelant « LeZ » en tchèque (ramper).

La compréhension d'une autre langue peut être compliquée par des glissements de sens au sein de cette langue que celui qui n'en est pas porteur peut ignorer. Par exemple, en tchèque, sur la racine « ZR », nous pouvons avoir le verbe « *zřít* » (voir, apercevoir) qui ne laissera rien soupçonner d'un sens déviant par exemple à l'imperfectif avec la variante adéquate de la racine « ZíR » et le verbe « *zírat* » (regarder bouche bée).

Ceci veut dire que les phénomènes généraux que nous avons présentés précédemment doivent être aussi complétés par des connaissances plus spécifiques situées en dehors des régularités.