

## **GÉNÉRATION AUTOMATIQUE DE LEXÈMES SLAVES À PARTIR DE LEURS RACINES HISTORIQUES : UNE DES BASES DE L'ENSEIGNEMENT MULTILINGUE DES LANGUES SLAVES DE L'OUEST (NORD ET SUD)**

L'article que nous offrons en l'honneur de notre collègue Vladimir Pogačnik concerne le système linguistique du groupe de langues slaves du Nord-Ouest (bas-sorabe, haut-sorabe, tchèque, slovaque, polonais) étendu au groupe slave du Sud-Ouest (slovène, croate), système confronté avec celui du russe, témoin des langues slaves de l'Est.

### **PROLÉGOMÈNES**

Les travaux que nous présentons sont motivés par une série de recherches qui, cumulées les unes aux autres, nous ont menés petit à petit vers une meilleure compréhension du système linguistique slave dans ses dimensions cumulées diachronique et synchronique.

C'est dans un premier temps l'expérience de l'analyse automatique de la langue tchèque (textes médicaux, scientifiques et techniques actuels) par des techniques heuristiques déterministes basées sur une reconnaissance de formes linguistiques (analyse sans dictionnaire) qui fait apparaître la calculabilité du tchèque.

Ensuite, l'enseignement de la grammaire historique du tchèque, essentiellement de son évolution phonologique, a conduit à considérer l'évolution des langues slaves du Nord-Ouest et les traits qui les distinguent au fur et à mesure les unes des autres.

Sur cette base, c'est la comparaison entre elles, au niveau synchronique, qui permet d'élaborer une calculabilité relative de chacune de ces langues pour un certain nombre de phénomènes.

Dès lors, nous avons suffisamment d'éléments pour aborder une définition extensive du système linguistique slave. La diversité des phénomènes au sein des langues slaves du Nord-Ouest, grâce notamment aux sorabes qui apportent un duel complet et les temps simples du passé (imparfait et aoriste), fait que le slovène, pris en considération ultérieurement, s'est « glissé » dans la matrice systémique élaborée (et dans les interfaces homme – machine afférents, par exemple dans nos bases de données) sans le

---

\* *Adresse de l'auteur* : INALCO, 65, rue des Grands Moulins, 75013 Paris, France. Mél : patrice.pognan@gmail.com ;

\*\* *Adresse de l'auteur* : ÚFAL MFF UK, Malostranské náměstí 25, CZ–11800, Prague, République tchèque. Mél : panevova@ufal.mff.cuni.cz

moindre aménagement de cette dernière. Seule la forme d'un futur périphrastique sur la base du participe passé actif (ppa) [bom potoval *je voyagerai*] a été renforcée à côté du polonais où cette forme apparaît en parallèle avec la forme munie d'un infinitif, la forme avec ppa ayant depuis longtemps disparu en tchèque et en slovaque. Ayant introduit le slovène dans nos travaux, la prise en compte du croate est venue naturellement.

Cet ensemble de langues que nous dénommons, à la suite de Starostin, langues slaves de l'Ouest a la caractéristique d'être – par hasard – l'ensemble des langues slaves écrites en caractères latins et appartenant désormais toutes à l'Union européenne.

## **FINALITÉ DE CETTE RECHERCHE**

La génération automatique de lexèmes pour chacune des langues slaves de l'Ouest considérées à partir d'une racine commune vieux-slave ou supposée proto-slave permet de vérifier la calculabilité du système diachronique slave de l'Ouest, c'est-à-dire de vérifier la calculabilité de chacun des phénomènes connus de la grammaire historique et d'en apprécier les conditions de déclenchement. Nous débutons seulement l'appréciation de l'extension des phénomènes, absolue (nombre de racines concernées) et relative (pourcentage d'exception). Par contre, nous avons déjà vérifié par le calcul de l'ensemble des phénomènes phonologiques mis en jeu l'absolue nécessité de respecter scrupuleusement l'ordre chronologique dans lequel ces phénomènes sont donnés par les spécialistes de la grammaire historique. Dans le cas contraire, le programme de génération produit des erreurs ou des résultats aberrants.

Ce programme permet donc de vérifier la validité des lois phonologiques slaves et par conséquent de rendre plus précises les caractéristiques de la calculabilité. Ces résultats en diachronie permettent, par leur prise en compte en synchronie, d'améliorer l'efficacité des analyses automatiques.

Ces résultats permettent aussi une appropriation plus facile d'une langue ou du groupe de langues par la compréhension intime du système linguistique slave. Nous en usons largement dans nos cours de licence et de master de langue et de grammaire tchèques à l'INALCO et à la Sorbonne.

C'est pourquoi nous insisterons dans le présent article sur l'utilisation pédagogique des résultats de cette recherche qui viennent en plus d'autres modes de présentation du multilinguisme slave basés, eux-aussi, sur le système linguistique.

## **LE SYSTÈME LINGUISTIQUE SLAVE**

Nous donnerons ici les caractéristiques principales des langues slaves, tant dans la diachronie que la synchronie. Il s'agit, en quelque sorte, de l'écrasement de l'espace-temps en une présentation maximale de tous les phénomènes, historiques ou contemporains. En d'autres termes, c'est la dimension la plus large du système qu'il convient d'enseigner et d'offrir à l'aide de divers outils dont des outils de présentation.

Les langues slaves sont des langues indo-européennes marquées par une flexion externe, c'est-à-dire avec des désinences en fin de mot qui donnent la fonction (sauf en

ce qui concerne le bulgare et le macédonien, devenues des langues isolantes). Elles possèdent trois genres, masculin, féminin et neutre, ainsi que trois sous-genres masculins qui se sont développés dans le temps: masculin humain, masculin animal et masculin inanimé. On peut trouver également des compromis avec deux sous-genres où le masculin humain et le masculin animal sont regroupés en un seul masculin animé (comme en tchèque) face à un masculin inanimé.

Depuis des temps immémoriaux, les langues slaves ont trois nombres (singulier, duel et pluriel) et sept cas (nominatif, génitif, datif, accusatif, vocatif, locatif et instrumental). En dehors de l'aspect, quasiment définitoire des langues slaves, celles-ci possèdent des temps simples du passé (imparfait et aoriste) et des temps composés (passé composé, plus que parfait, conditionnel, futur à base de passé,...).

Le duel reste vivant en slovène et dans les langues sorabes. Dans des langues comme le tchèque, la présence du duel permet d'appréhender globalement les quelques termes qui l'utilisent encore et qui possèdent en plus un pluriel appliqué à d'autres significations que celles afférentes à des organes ou parties du corps. Le haut et le bas-sorabe conservent encore les temps simples du passé, aoriste et imparfait.

*a. expression du système linguistique par des interfaces homme-machine*

Nos bases de données pour les langues slaves reflètent l'intégralité du système linguistique et chaque langue vient s'inscrire dans ce cadre général.

Nous donnerons ci-dessous l'exemple du lexème verbal pour la formation du passé composé, du conditionnel,... en ce qui concerne le tchèque, le slovaque et le slovène. On remarque pour le tchèque l'absence de duel et l'opposition masculin animé / masculin inanimé. Les formes de pluriel possèdent des désinences (formes courtes) qui distinguent les genres (« i » pour le masculin animé, « a » pour le neutre et « y » pour le féminin et le masculin inanimé).

présent	imparfait	aoriste	participe passé actif	passé composé	plus que parfait	conditionnel	impératif	formes impersonnelles		
<i>participe passé actif</i>										
			singulier				duel			
			m. animé	<u>zaručoval</u>				m. humain	<input type="text"/>	
			m. animal	<input type="text"/>				m. animal	<input type="text"/>	
			m. inanimé	<u>zaručoval</u>				m. inanimé	<input type="text"/>	
			féminin	<u>zaručovala</u>				féminin	<input type="text"/>	
			neutre	<u>zaručovalo</u>				neutre	<input type="text"/>	
									pluriel	
									m. animé	<u>zaručovali</u>
									m. animal	<input type="text"/>
									m. inanimé	<u>zaručovaly</u>
									féminin	<u>zaručovaly</u>
									neutre	<u>zaručovala</u>

Par contre, le slovène possède le duel, mais ne présente qu'un masculin unique. Les désinences sont plus ambiguës qu'en tchèque du fait de la présence du duel. En tchèque, il existe une ambiguïté naturelle entre féminin singulier et neutre pluriel (désinence en « a »). Cette ambiguïté demeure en slovène, mais il faut y rajouter le masculin du duel. Le féminin et le neutre du duel apportent en plus une ambiguïté avec le masculin pluriel (désinence en « i »).

présent	imparfait	aoriste	participe passé actif	passé composé	plus que parfait	conditionnel	futur ppa	impératif	formes
---------	-----------	---------	-----------------------	---------------	------------------	--------------	-----------	-----------	--------

**participe passé actif**

singulier		duel		pluriel	
m. animé	potoval	m. humain	potovala	m. animé	potovali
m. animal		m. animal		m. animal	
m. inanimé		m. inanimé		m. inanimé	
féminin	potovala	féminin	potovali	féminin	potovale
neutre	potovalo	neutre	potovali	neutre	potovala

Le slovaque s'inscrit dans le même cadre que le tchèque, mais il ne différencie plus les formes de pluriel où « i » est la désinence des masculins animé et inanimé, du féminin et du neutre.

présent	imparfait	aoriste	participe passé actif	passé composé	plus que parfait	conditionnel	impératif	formes imperson
---------	-----------	---------	-----------------------	---------------	------------------	--------------	-----------	-----------------

**participe passé actif**

singulier		duel		pluriel	
m. animé	zaručoval	m. humain		m. animé	zaručovali
m. animal		m. animal		m. animal	
m. inanimé	zaručoval	m. inanimé		m. inanimé	zaručovali
féminin	zaručovala	féminin		féminin	zaručovali
neutre	zaručovalo	neutre		neutre	zaručovali

Les extraits présentés ci-dessus montrent l'usage pédagogique que l'on peut faire de l'interface des bases de données.

### b. présentations (powerpoint) du système linguistique

De la même manière, nous présenterons ci-dessous l'exemple de la comparaison de substantifs à déclinaison consonantique avec un augment. Nous prenons le modèle actif en tchèque, le modèle des « petits d'animaux » (*kuře, poulet*) qui développe un augment en « t », autrefois en « nt ».

Nous avons choisi le mot tchèque « tele » (*veau*), « celo » en haut-sorabe et « cielę » en polonais. L'ordonnancement des fiches en Powerpoint est totalement conservé dans l'interface des bases de données. La comparaison des différentes langues apportera des connaissances historiques supplémentaires. Ainsi, l'examen des trois langues permet-il de supposer que l'augment du pluriel a vraisemblablement toujours été dur « t ». Par contre, on apprend, grâce au haut-sorabe, que l'augment du duel est le même que celui du singulier. Par la comparaison avec le haut-sorabe et le polonais qui possèdent un segment mouillé au singulier et au duel, respectivement « ec̣ » et « eci », on peut supposer – le tchèque étant une langue qui a beaucoup dépalatalisé – que le segment

tchèque « et » est un segment originellement mouillé en « et' ». Enfin, la forme du mot en haut-sorabe et en polonais doit nous faire envisager une forme primitive tchèque où le « t » de « tele » était également mou.

Déclinaison d'un substantif tchèque neutre à flexion consonantique

tele ( <i>veau</i> )	singulier	duel	pluriel
<b>nominatif</b>	tel e		tel at a
<b>génitif</b>	tel et e		tel at
<b>datif</b>	tel et i		tel at ům
<b>accusatif</b>	tel e		tel at a
<b>vocatif</b>	tel e		tel at a
<b>locatif</b>	tel et i		tel at ech
<b>instrumental</b>	tel et em		tel at y

Déclinaison du même substantif en haut-sorabe

ćelo ( <i>veau</i> )	singulier	duel	pluriel
<b>nominatif</b>	ćel o	ćel ec i	ćel at a
<b>génitif</b>	ćel ec a	ćel ec ow	ćel at ow
<b>datif</b>	ćel ec u	ćel ec omaj	ćel at am
<b>accusatif</b>	ćel o	ćel ec i	ćel at a
<b>vocatif</b>			
<b>locatif</b>	ćel ec u	ćel ec omaj	ćel at ach
<b>instrumental</b>	ćel ec om	ćel ec omaj	ćel at ami

Nous apprenons aussi par le polonais que le support vocalique de l'augment était à l'origine une nasale, demeurée en polonais.

Déclinaison du même substantif en polonais

cielę ( <i>veau</i> )	singulier	duel	pluriel
<b>nominatif</b>	ciel ę		ciel ęt a
<b>génitif</b>	ciel ęci a		ciel ąt
<b>datif</b>	ciel ęci u		ciel ęt om
<b>accusatif</b>	ciel ę		ciel ęt a
<b>vocatif</b>	ciel ę		ciel ęt a
<b>locatif</b>	ciel ęci u		ciel ęt ach
<b>instrumental</b>	ciel ęci em		ciel ęt ami

# GÉNÉRATION AUTOMATIQUE DE LEXÈMES SLAVES À PARTIR DE LEURS RACINES HISTORIQUES

## a. *l'évolution phonologique*

Nous avons abordé l'évolution phonologique des langues slaves de l'Ouest par l'intermédiaire de l'évolution du tchèque que nous connaissons le mieux et qui est divisée en trois périodes. La première d'entre elles concerne des phénomènes très importants communs à l'ensemble des langues slaves de l'Ouest. Plus on remonte les deuxième et troisième périodes, plus on se situe dans la différenciation de langues proches. Ainsi, la distance entre le tchèque et le slovaque est-elle donnée par les 7 derniers phénomènes.

- Du protoslave à la fin du 10<sup>ème</sup> siècle :
  - Métathèse des liquides (évolution de: **tort, tolt, tert, telt**)
  - Contraction
  - Evolution des nasales
  - Disparition et vocalisation des jers
- De la fin du 10<sup>ème</sup> siècle à la fin du 14<sup>ème</sup> siècle :
  - Passage g → h
  - Evolution du r mouillé en ř
  - Transformations 'a → ě + 'u → í
  - Dépalatalisation
- De la fin du 14<sup>ème</sup> siècle à la fin du 16<sup>ème</sup> siècle :
  - Transformation du « u long » en diphtongue: ú → ou
  - Contraction ie → í
  - Contraction uo → ů
  - Changement aj → ej

## b. *organisation du programme*

Le programme lit des racines, actuellement proposées sur liste, à terme par une interaction avec l'utilisateur. Ces racines, extraites principalement du dictionnaire de Derksen, sont écrites en majuscules pour le cadre consonantique, en minuscules pour les voyelles.

Avant de commencer ses traitements, le programme exécute quelques préparatifs :

- les signes diacritiques non nécessaires, notamment sur la désinence, sont supprimés.
- le programme compte le nombre de jers, vérifie leur rang pair ou impair et les marque en tant que jer dur ou jer mou.
- il fournit la racine aménagée en caractères minuscules et réalise une première impression de contrôle.

De même, après le traitement de la phonologie, le programme réalise encore quelques aménagements, dont la correspondance « w » (bas-sorabe, haut-sorabe, polonais) et

« v » (tchèque, slovaque, slovène, croate) et la réécriture en cyrillique du russe et du serbe (cyrilliques différents).

Le traitement lui-même suit scrupuleusement, comme nous l'avons souligné plus haut, l'enchaînement des phénomènes, mais doit les traiter de manière plus fine et plus structurée.

En première période (jusqu'à la fin du 10<sup>ème</sup> siècle), ce sont les phénomènes les plus marquants et les plus décisifs qui sont traités :

Après les métathèses en « e » et les métathèses en « o », c'est la contraction qui est étudiée. Le traitement des nasales est beaucoup plus délicat. Il montre, pour la nasale « ǫ », la nécessité de prendre en compte dans les calculs les tons/accents du vieux slave : ton ascendant long, ton descendant long, voyelle non accentuée longue, ton ascendant bref, ton descendant bref, voyelle non accentuée brève, pour pouvoir obtenir une génération structurée et cohérente de la vocalisation de la nasale (cf. c2). Pour la nasale « ę » (cf. c3), c'est le contexte qui est déterminant. Le traitement des jers doit être catégorisé en suppression des jers impairs durs, vocalisation des jers pairs durs et des jers pairs mous distinguée dans 3 langues slaves, le russe, le bulgare et le macédonien. Les jers impairs mous ne disparaissent pas toujours sans laisser de trace comme les durs correspondants, mais produisent souvent une forme modifiée (plus ou moins palatalisée) de la consonne précédente, ce qui conduit à prendre en compte chaque type de consonne précédent un jer mou impair (nous y reviendrons plus bas).

Les deuxième et troisième périodes sont traitées au sein du programme de manière bien moins complexe.

La deuxième période (jusqu'à la fin du 14<sup>ème</sup> siècle) voit la transformation systématique de « g » en « h » en haut-sorabe, tchèque et slovaque, la transformation « a → ě » et « u → i » uniquement en tchèque, ce qui donne à cette langue un caractère particulier face à toutes les autres langues slaves. Le tchèque a beaucoup dépalatalisé les labiales et les dentales, se mettant ainsi dans une position intermédiaire entre les langues slaves du Nord-Ouest, toutes mouillées comme les langues slaves de l'Est et le slovène qui est très peu mouillé.

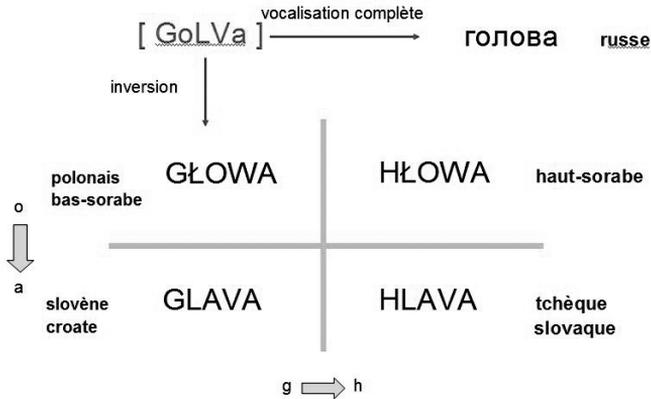
La troisième période donne lieu à des phénomènes plus limités concernant essentiellement le tchèque : la diphtongaison « ú → ou », la contraction de « ie → í » et surtout la transformation du o long en diphtongue « uo », puis en u long « ů ». Ces deux dernières transformations sont très utiles dans la reconnaissance automatique des emprunts en tchèque moderne.

### *c. quelques exemples*

Nous souhaitons montrer par quelques exemples choisis l'usage qui peut être fait des résultats de ce programme pour la compréhension et l'assimilation du système slave, pour l'accès à un multilinguisme slave de l'Ouest.

c1. métathèse

Nous donnerons en premier un exemple de métathèse en « o » (GoLVa tête) avec un schéma montrant l'effet du phénomène sur les différentes langues :



A partir de la racine « GoLVa », le russe construit par une vocalisation complète une forme particulière « GoLoVa » par l'application de la règle :

$$(C) o \{L | R\} C = (C) o \{L | R\} o C$$

Par contre, pour les autres langues, il y a une vraie métathèse avec inversion de la voyelle et de la liquide « l » ou « r » :

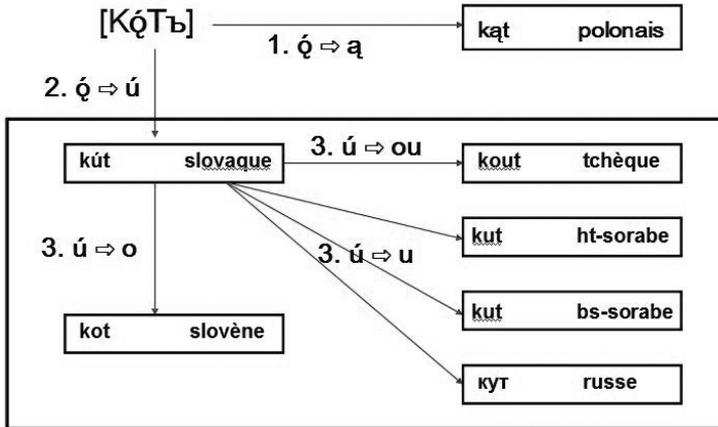
$(C) o \{L | R\} C = (C) \{L | R\} o C$ , soit une forme GLoVa à partir de laquelle on obtiendra les formes actuelles correctes par une ou plusieurs transformations: réécriture du « l dur » et du « w » pour le polonais. Dans le dessin ci-dessus, en allant de gauche à droite, on exerce une transformation de « g » à « h » et allant de haut en bas, une transformation de « o » à « a » ( $((C) o \{L | R\} C = (C) \{L | R\} a C)$ ), ce qui nous donne l'ensemble des formes voulues.

On peut voir à chaque ligne de résultat le passage du programme dans un phénomène phonologique donné et la transformation effectuée sur les lexèmes des différentes langues:

GoLVá										
GoLVa										
phénom	russe	polonais	tchèque	slovaque	b_sorabe	h_sorabe	slovène	croate	serbe	
racine	golva	golva	golva	golva	golva	golva	golva	golva	golva	golva
phénom	russe	polonais	tchèque	slovaque	b_sorabe	h_sorabe	slovène	croate	serbe	
méth -o-	golova	glova	glava	glava	glova	glöva	glava	glava	glava	glava
phénom	russe	polonais	tchèque	slovaque	b_sorabe	h_sorabe	slovène	croate	serbe	
g / h	golova	glova	hlava	hlava	glöva	hlöva	glava	glava	glava	glava
phénom	russe	polonais	tchèque	slovaque	b_sorabe	h_sorabe	slovène	croate	serbe	
v / w	golova	glöva	hlava	hlava	glöva	hlöva	glava	glava	glava	glava
phénom	russe	polonais	tchèque	slovaque	b_sorabe	h_sorabe	slovène	croate	serbe	
cyrill + x	голова	glöva	hlava	hlava	glöva	hlöva	glava	glava	glava	rnava

### c2. vocalisation des nasales

De manière identique, nous montrons ci-dessous le calcul effectué sur une nasale en « q » avec un accent ascendant long à partir de la racine « KqTъ » *coin*.



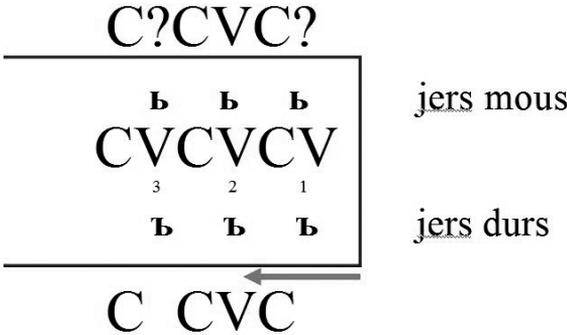
Ce schéma fait apparaître deux opérations premières : celle qui forme le polonais et celle qui forme le slovaque. Toutes les autres langues du groupe sont générées sur la base du slovaque par une transformation appropriée. C’est notamment au niveau de la vocalisation de cette nasale que le slovène possède « o » face au « u » des autres langues slaves. Les résultats du programme font clairement apparaître le respect chronologique des opérations, ainsi la forme « ou » du tchèque n’apparaît qu’après une étape commune avec le slovaque en « ú » :

	Kq̇Tъ									
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat.	srbština	
kořen	kq̇tъ	kq̇tъ	kq̇tъ	kq̇tъ	kq̇tъ	kq̇tъ	kq̇tъ	kq̇tъ	kq̇tъ	
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat.	srbština	
q̇ / q̄	kutъ	kq̄tъ	kútъ	kutъ	kutъ	kutъ	kotъ	kutъ	kutъ	
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat.	srbština	
ъ / 0	kut	kq̄t	kút	kut	kut	kut	kot	kut	kut	
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat.	srbština	
ú ⇨ ou	kut	kq̄t	kout	kút	kut	kut	kot	kut	kut	
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat.	srbština	
cyrill + x	кут	kq̄t	kout	kút	kut	kut	kot	kut	кут	

### c3. vocalisation des jers

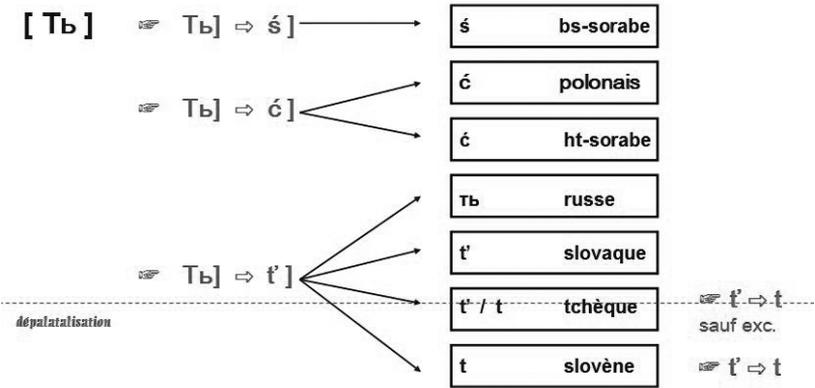
Le calcul des jers est relativement compliqué et il occupe une partie importante du programme. Nous donnons d’abord le schéma qui reflète la loi de Havlík qui permet d’expliquer la chute et la vocalisation des jers. Le programme mis en place a montré la

nécessité de préciser la loi de Havlík en ce qui concerne les jers mous impairs qui, en fait, ne disparaissent pas toujours.



La loi de Havlík stipule un comptage des jers (éléments de position vocalique dans une structure supposée consonne-voyelle-consonne-voyelle... du vieux slave) à partir de la fin du mot. Les séries impaires sont dites faibles et sont censées disparaître (cela est vrai pour le jer dur: on le voit sur l'exemple précédent où le jer dur de fin de mot, donc de rang 1 disparaît sans laisser de trace) tandis que la série paire est vocalisée. Chaque langue slave vocalise suivant ses lois propres. Ainsi, seuls le russe, le macédonien et le bulgare spécialisent la vocalisation des deux jers: « e » pour le jer mou, « o » en russe et en macédonien pour le jer dur, le bulgare conservant le signe dur. Les autres langues slaves ne savent pas distinguer la nature du jer sous-jacent : le tchèque et le polonais ont « e » pour les deux jers, le croate « a ». La situation du slovène est compliquée du fait que cette langue vocalise le jer mou en « a » ou en « e » et le jer dur également en « a » ou en « e ». Le slovaque central (on distingue un slovaque occidental, central et oriental) est particulièrement complexe avec une vocalisation en « a », « e » ou « o ».

Les jers mous en position impaire nécessitent un traitement approprié pour chaque combinaison consonne - jer mou. Nous donnons ci-dessous l'exemple de la combinaison « T-jer mou ».



Le slovène durcit le « t mou », le tchèque a durci la majorité des labiales et dentales molles à des degrés divers: le « ě » est certainement moins souvent dépalatalisé que « d' » et « t' » (on le voit, par exemple, avec le paradigme de flexion nominale du féminin « píseň »). Pour t-fer mou, nous devons rajouter un filtre pour le tchèque constitué des exceptions qui conservent la mouillure, par exemple « pouť » *pèlerinage* (dérivé de *chemin, route*).

Pěť										
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat.	srbština	
kořen	pěť	pěť	pěť	pěť	pěť	pěť	pěť	pěť	pěť	
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat.	srbština	
ě-2	pjaty	pięć	pět	pät	pět	pjet	pet	pet	pet	
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat.	srbština	
ť	pjaty	pięć	pět	pät'	pés	pjec	pet	pet	pet	
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat.	srbština	
ь	pjaty	pięć	pět	pät'	pés	pjec	pet	pet	pet	
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat.	srbština	
cyrill + x	пятъ	pięć	pět	pät'	pés	pjec	pet	pet	пет	

Dans le résultat ci-dessus (calcul de la vocalisation de la nasale par rapport au contexte), la racine choisie se termine par « t-fer mou », ce qui permet de voir la disparition du fer ou la transformation subie (elle est forte dans l'ensemble polonais – sorabes). On remarquera la conservation de la nasale en polonais.

#### c4. transformations de voyelles dures (a, u) en molles (e, i)

Nous sommes ici dans un phénomène qui, bien qu'appartenant à la seconde période d'évolution phonologique ne concerne plus que le tchèque, donnant à cette langue une forme unique par rapport à toutes les autres langues slaves. Nous donnons un exemple de génération avec « duša » *âme* :

Duša										
Duša										
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat.	srbština	
kořen	duša	duša	duša	duša	duša	duša	duša	duša	duša	
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat.	srbština	
a/u⇒ě/í	duša	duša	duše	duša	duša	duša	duša	duša	duša	
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat.	srbština	
cyrill + x	душа	duša	duše	duša	duša	duša	duša	duša	душа	

Le tchèque transforme le contexte consonne molle – voyelle dure en consonne molle – voyelle molle. Historiquement, l'alternance de voyelles est « a – ě – e » et, de

manière plus restreinte « u – i ». Actuellement, la seule alternance possible est « y – i ». Le programme respecte, ici aussi, l'ordre réel de l'évolution en ayant une étape en « ě » qui disparaît, lorsque le tchèque refusera la présence d'une double mouillure (*jed poison*) contrairement au haut-sorabe qui la conserve (*jěd*). La disparition de mouillure est visible aussi pour « u » : « pl'úca » *poumon* en slovaque, « pljuča » en slovène, mais « plíce » en tchèque.

Nous avons présenté ici quelques exemples de génération qui nous semblaient bien indiqués pour montrer les implications pédagogiques dans l'enseignement des langues slaves que ce soit dans un milieu linguistique slave ou dans un milieu linguistique extérieur au monde slave.

Pour assurer une vision complète des phénomènes examinés, nous devons « revisiter » ce travail en partant aussi du point de vue de chacune des langues slaves concernées. Par exemple, vérifier les conditions de passage du « l dur » au « l barré » en polonais et en sorabe sera intéressant.

## COMPARAISON DE LANGUES SLAVES DEUX À DEUX

Du travail actuel, nous tirons également des processus de comparaison des langues deux à deux dont nous donnerons des exemples dans la comparaison tchèque – slovène et tchèque – slovaque.

### a. comparaison tchèque - slovène

Pour des raisons pratiques (enseignement, traitement automatique), cette comparaison se fait aux deux niveaux diachronique et synchronique.

<i>phénomène concerné</i>	<i>slovène</i>	<i>tchèque</i>
Evolution du « r » mouillé en « ř / r »	r	ř
Non transformation de « a » en « ě »	stopnica	stupnice
Transformation de « u, ú, ou » en « o »		
Transformation de « u, ü » allemands en « o »	šolar	alld: Schüler
Transformation de « ý, y, í, i » en « i »	sin	syn
Métathèse de « o » à « a »	glava	hlava
Maintien du « g » en slovène		
Dépalatalisation	radost	radost
Chute du « v » prothétique	osa	vosa
Correspondance « šč » - « št' »	igrišče	hřiště

Certains phénomènes, notamment la transformation de « u » en « o » en slovène, doivent être revus de manière globale à la lumière des évolutions historiques.

<sup>1</sup> houba	goba	tchèque	slovène	tchèque	slovène	tchèque	slovène
	"ý, y, ě" ⇒ "i" ▢		▢	<b>métathèse</b>	▢		"v-" ⇒ Ø ▢
	"ŷ" ⇒ "r" ▢		▢	<b>métathèse "o" ⇒ "a"</b> ▢			"dl, tl" ⇒ "l" ▢
	"y" ⇒ "r" ▢			"h" ⇔ "g" ▢			"ch" ⇒ "h" ▢
	"ř" ⇒ "r" ▢			"r" ⇒ "r" ▢			"št" ⇔ "šč" ▢
"u, ú, (ú →) ou" ⇒ "o" ▢				"(ě →) e" ⇔ "a" ▢			"-t" ⇔ "-ti" ▢
	"u" ⇒ "o" ▢			"i" ⇔ "u" ▢			
	"ú" ⇒ "o" ▢			"l" ⇔ "lj" ▢			
"(ú →) ou" ⇒ "o" ▢							
"ó, ů" ⇒ "o" ▢				<b>depalatalisation</b>	▢		
	"ó" ⇒ "o" ▢						
	"ů" ⇒ "o" ▢						
<b>longueur</b> ⇒ Ø ▢							
	"á" ⇒ "a" ▢						
	"é" ⇒ "e" ▢						

La présentation ci-dessus est extraite de notre base de données slovène. Elle met en évidence les éléments qui doivent être pris en compte pour un transfert automatique entre les deux langues ou pour expliquer à des apprenants la proximité des deux langues dès lors que l'on connaît les mécanismes du système.

#### b. comparaison tchèque - slovaque

Nous donnons ci-dessous deux tableaux différents des mêmes correspondances historiques entre tchèque et slovaque :

<i>phénomène phonologique concerné</i>	<i>slovaque</i>	<i>tchèque</i>	<i>traduction</i>
Evolution du « r » mouillé en « ř »	reč	řeč	« discours »
Transformations de « a » en « ě » et de « u » en « i » derrière une consonne molle	ovca pľúca	ovce plíce	« mouton » « poumons »
Dépalatalisation	hovorit'	hovořit	« parler »
Transformation de « ú » en « ou »	súd	soud	« tribunal »
Contraction de « ie » en « i »	miera	míra	« mesure »
Contraction de « uo » issu de « ó » vers « ů » en tchèque et « ó » en slovaque	pôvod	původ	« origine »
Changement de « aj » vers « ej » au sein d'une syllabe	naj-	nej-	préfixe du superlatif

tchécoslovaque	
slovaque	tchèque
<p><i>fin 10<sup>ème</sup> s. à fin 14<sup>ème</sup> s.</i></p> <p>évolution du r mouillé</p> <p>transformation ‘a ⇒ ě</p> <p>transformation ‘u ⇒ i</p> <p>dépalatalisation</p> <p><i>fin 14<sup>ème</sup> s. à fin 16<sup>ème</sup> s.</i></p> <p>transformation ú ⇔ ou</p> <p>contraction ie ⇒ í</p> <p>contraction ó ⇒ uo ⇒ û</p> <p>changement aj ⇒ ej</p>	

Ce dernier tableau va nous donner quelques éléments de conclusion.

En effet, ce tableau réunit les traits de différenciation entre tchèque et slovaque, mais aussi il les organise en fonction de la chronologie de la phonologie historique et en termes de collatéralité. Les deux traits verticaux centraux, celui de gauche pour le slovaque, celui de droite pour le tchèque représentent la forme commune au tchèque et au slovaque jusqu’au moment où est intervenu une transformation.

Si un tel tableau est utile pour l’apprentissage des deux langues concernées, il a aussi une autre vertu intéressante: il infirme le sentiment non fondé que le slovaque serait la « petite sœur » du tchèque ainsi que l’impression que le slovaque s’écartere volontairement du tchèque. Si cela est vrai dans les 30 ou 40 dernières années à un niveau lexical par un choix de mots ou de formes de mots qui « semblent plus slovaques », il n’en est rien au niveau de l’évolution des deux langues où sur 6 siècles et par 8 phénomènes phonologiques, c’est le tchèque qui s’est écarté de la plateforme commune du système linguistique tchécoslovaque...

## Bibliographie

- BARTOŠ, Jozef/Joseph GAGNAIRE (1972) *Grammaire de la langue slovaque*. Bratislava/Paris : Matica slovenská/Institut d’études slaves.
- BIELEC, Dana (1998) *Polish: An Essential Grammar*. Londres : Routledge.
- BLAŽEK, Václav (2005) *On the internal classification of Indo-European languages: survey*. *Linguistica ONLINE*. <http://www.phil.muni.cz/linguistica/art/blazek/bla-003.pdf>.

- DECAUX, Étienne (1978) *Leçons de grammaire polonaise*. Paris : Institut d'études slaves.
- DERKSEN, Rick (2008) *Etymological Dictionary of the Slavic Inherited Lexicon*. Brill : Leiden.
- GRAPPIN, Henri (1963) *Grammaire de la langue polonaise*. Paris : Institut d'études slaves.
- HAVRÁNEK, Bohuslav/Alois JEDLIČKA (1960) *Česká mluvnice*. Prague : SPN.
- HERRITY, Peter (2000) *Slovene: a Comprehensive Grammar*. Londres : Routledge.
- HOLUB, Josef/František KOPEČNÝ (1952) *Etymologický slovník jazyka českého*. Praha : SPN.
- JAMBOROVA-LEMAY, Diana (2003) *Analyse automatique du slovaque. Étude approfondie du système linguistique slovaque et sa reconnaissance d'après la forme dans les textes scientifiques et techniques. Application au machinisme agricole*. Thèse de doctorat. Paris : CERTAL-INALCO.
- JANAŠ, Pětr (2011) *Dolnosrbská mluvnice*. Prague : Karolinum.
- KOMÁREK, Miroslav (1958) *Historická mluvnice česká I. Hlásoklovi*. Prague : SPN.
- KOPEČNÝ, František (1981) *Základní všeslovanská slovní zásoba*. Prague : Academia.
- LAMPRECHT, Arnošt/Dušan ŠLOSAR/Jaroslav BAUER (1986) *Historická mluvnice češtiny*. Prague : Státní Pedagogické Nakladatelství.
- LEČIČ, Rada (2009) *Osnove slovenskega jezika*. Cerkno : Gaya.
- MACHEK, Václav (1997) *Etymologický slovník jazyka českého*. Prague : Nakladatelství LN.
- MAREŠ, František Václav (1999) *Diachronische Phonologie des Ur- und Frühslavischen*. Frankfurt am Main : Peter Lang.
- MARVAN, Jiří (2000) *Jazykové milénium. Slovanská kontrakce a její český zdroj*. Prague : Academia.
- MAZON, André (1952) *Grammaire de la langue tchèque*. Paris : Institut d'Etudes Slaves.
- MEILLET, Antoine/André VAILLANT (1980) *Grammaire de la langue serbo-croate*. Paris : Champion.
- MEŠKANK, Timo (2001) *Retrogradny slownik hornjoserbskeje řeče*. Berlin : Mensch und Buch Verlag.
- MEŠKANK, Timo (2009) *Aussagenstruktur im Sorbischen. Untersuchungen zur Syntax und Satzsemantik*. Hamburg : Verlag Dr. Kovač.
- MISTRÍK, Jozef (1983) *Moderná slovenčina*. Bratislava : Slovenské Pedagogické Nakladateľstvo.
- NAUGHTON, James (2006) *Czech: An Essential Grammar*. Londres : Routledge.
- PIEŃKOS, Elżbieta/Jerzy PIEŃKOS/Leon ZARĘBA/Jerzy DOBRZYŃSKI (2001) *Wielki słownik polsko-francuski – Grand dictionnaire polonais – français*. 5 vol. Varsovie : Wiedza powszechna.
- PLESKALOVÁ, Jana (2001) *Stará čeština pro nefilology*. Brno : Filosofická fakulta Masarykovy University.

- POGNAN, Patrice (1998) « Histoire de l'écriture et de l'orthographe tchèques. » *Histoire, Epistémologie, Langage*, 27–62.
- POGNAN, Patrice (2001) « Introduction aux systèmes d'écriture des langues slaves de l'Ouest (polonais, bas-sorabe, haut-sorabe, tchèque, slovaque). » *Slavica occitania*, 283–310.
- POGNAN, Patrice (2007) « Forme et fonction en analyse automatique du tchèque. Calculabilité des langues slaves de l'Ouest. » *BULAG Les langues slaves et le français : approches formelles dans les études contrastives* 32, 13–33.
- POGNAN, Patrice (2011) « Bases de données à pivot français pour l'étude lexicale et grammaticale des langues slaves. » *Lingue e letteratura romanze*, 749–766.
- POGNAN, Patrice (sous presse) Langues slaves de l'Ouest, de la diachronie à la calculabilité. In : *Actes du colloque international "Des langues collatérales en domaine slave"*. Paris : L'Harmattan.
- ROSPOND, Stanisław (2000) *Gramatyka historyczna języka polskiego*. Varsovie : Wydawnictwo Naukowe PWN.
- SCHUSTER-ŠEWC, Heinz (1984) *Gramatika hornjoserbskeje rěče*. Bautzen : Nakładnistwo Domowina.
- SCHUSTER-ŠEWC, Heinz (à partir de 1983) *Historisch-etymologisches Wörterbuch der ober- und niedersorbischen Sprache*. Bautzen : Nakładnistwo Domowina.
- SCHUSTER-ŠEWC, Heinz (2000) *Das Sorbische im slawischen Kontext*. Bautzen : Nakładnistwo Domowina.
- SNOJ, Marko (2009) *Slovenski etimološki slovar*. Ljubljana : Inštitut za slovenski jezik Frana Ramovša.
- STAROSTA, Manfred (1999) *Dolnoserbsko-nimski słownik*. Bautzen : Nakładnistwo Domowina.
- TEŽAK, Stjepko (2000) *Gramatika hrvatskoga jezika*. Zagreb : Školska knjiga.
- WRÓBEL, Henryk (2001) *Gramatyka języka polskiego*. Cracovie : Od Nowa.
- ZAREMBA, Charles (2009) *Éléments de grammaire historique du polonais*. Paris : Institut d'Études Slaves.

### Résumé

#### GÉNÉRATION AUTOMATIQUE DE LEXÈMES SLAVES À PARTIR DE LEURS RACINES HISTORIQUES : UNE DES BASES DE L'ENSEIGNEMENT MULTILINGUE DES LANGUES SLAVES DE L'OUEST

Les travaux que nous présentons sont motivés par une série de recherches qui, cumulées les unes aux autres, nous ont menés petit à petit vers une meilleure compréhension du système linguistique slave dans ses dimensions cumulées diachronique et synchronique. C'est dans un premier temps l'expérience de l'analyse automatique de la langue tchèque (textes médicaux, scientifiques et techniques actuels) par des techniques heuristiques déterministes basées sur une reconnaissance de formes

linguistiques (analyse sans dictionnaire) qui fait apparaître la calculabilité du tchèque. Ensuite, l'enseignement de la grammaire historique du tchèque, essentiellement de son évolution phonologique, a conduit à considérer l'évolution des langues slaves du Nord-Ouest et les traits qui les distinguent au fur et à mesure les unes des autres.

#### Povzetek

### SAMODEJNO TVORJENJE SLOVANSKIH LEKSEMOV IZ NJIHOVIH HISTORIČNIH KORENOV: EDEN OD TEMELJEV ZA VEČJEZIČNO POUČEVANJE ZAHODNOSLOVANSKIH JEZIKOV

Pričujoče besedilo je plod celega zapovrstja raziskav, ki postopoma vodijo k vse boljšemu razumevanju slovanskega jezikovnega sistema, in sicer v njegovih diahronih in sinhronih razsežnostih. V prvi fazi gre za eksperimentalno raziskavo avtomatske analize češčine (sodobna medicinska, znanstvena in tehnična besedila) s pomočjo hevrističnih metod, utemeljenih na prepoznavanju jezikovnih oblik (analiza brez slovarja), ki kaže možnosti za avtomatsko tvorjenje češčine. V nadaljevanju je spoznavanje češke historične slovnice, zlasti v njeni fonološki razsežnosti, pripeljalo do določenih ključnih ugotovitev v zvezi z razvojem severozahodnih slovanskih jezikov in sistemskimi lastnostmi, po katerih se ti med seboj razlikujejo.