Système linguistique et calculabilité des langues slaves de l'Ouest (Nord et Sud) pour leur traitement automatique

Patrice Pognan
Professeur à l'INALCO – associé à l'équipe PLIDAM
Chargé de cours à l'Université Paris Sorbonne – membre de l'équipe STIH

TALN '2015
SLR

Tasla – Caen, 22 juin 2015

Répartition des langues slaves

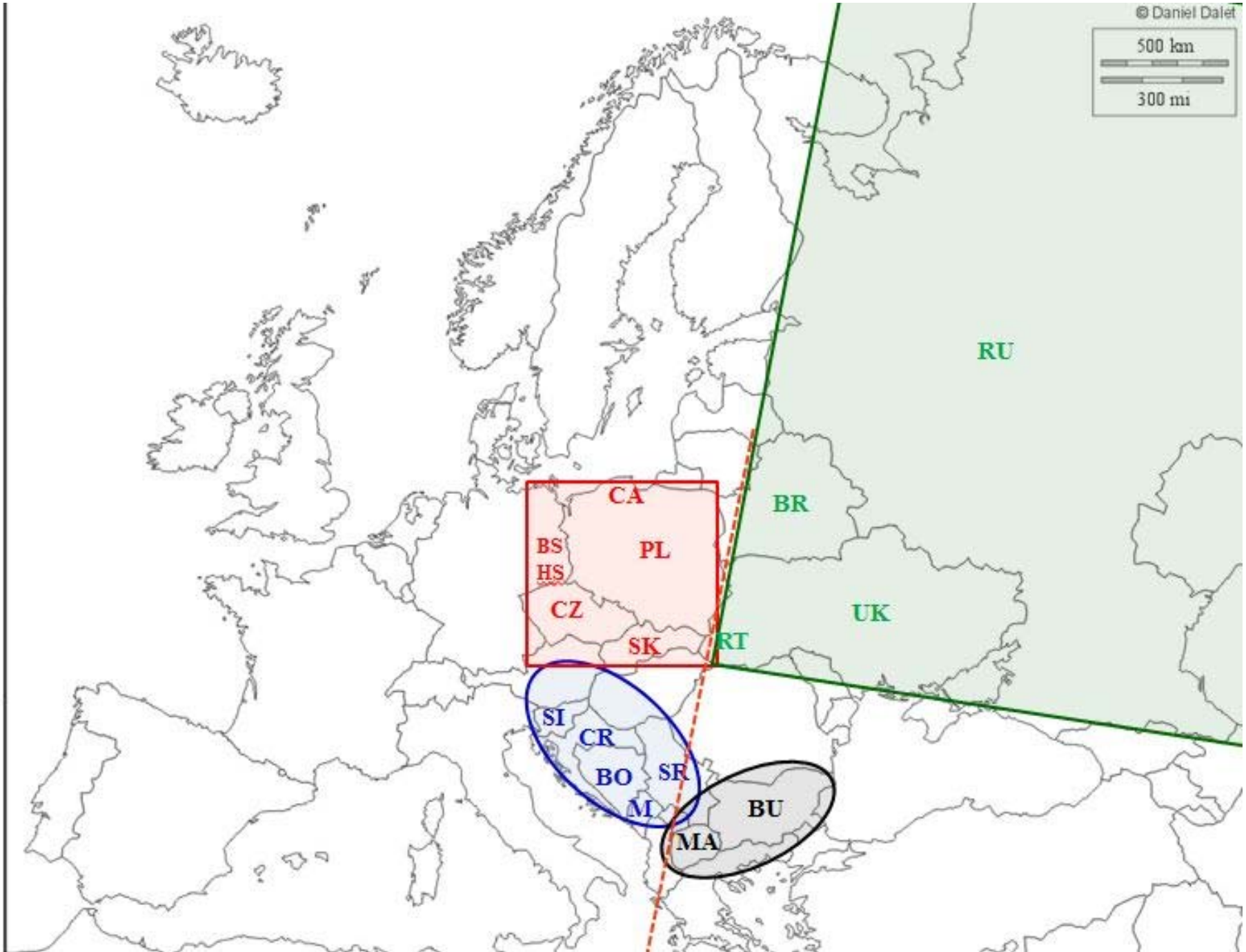
Langues slaves:

de l'Ouest:
CA: cachoube
PL: polonais
BS: bas-sorabe
HS: haut-sorabe
CZ: tchèque
SK: slovaque

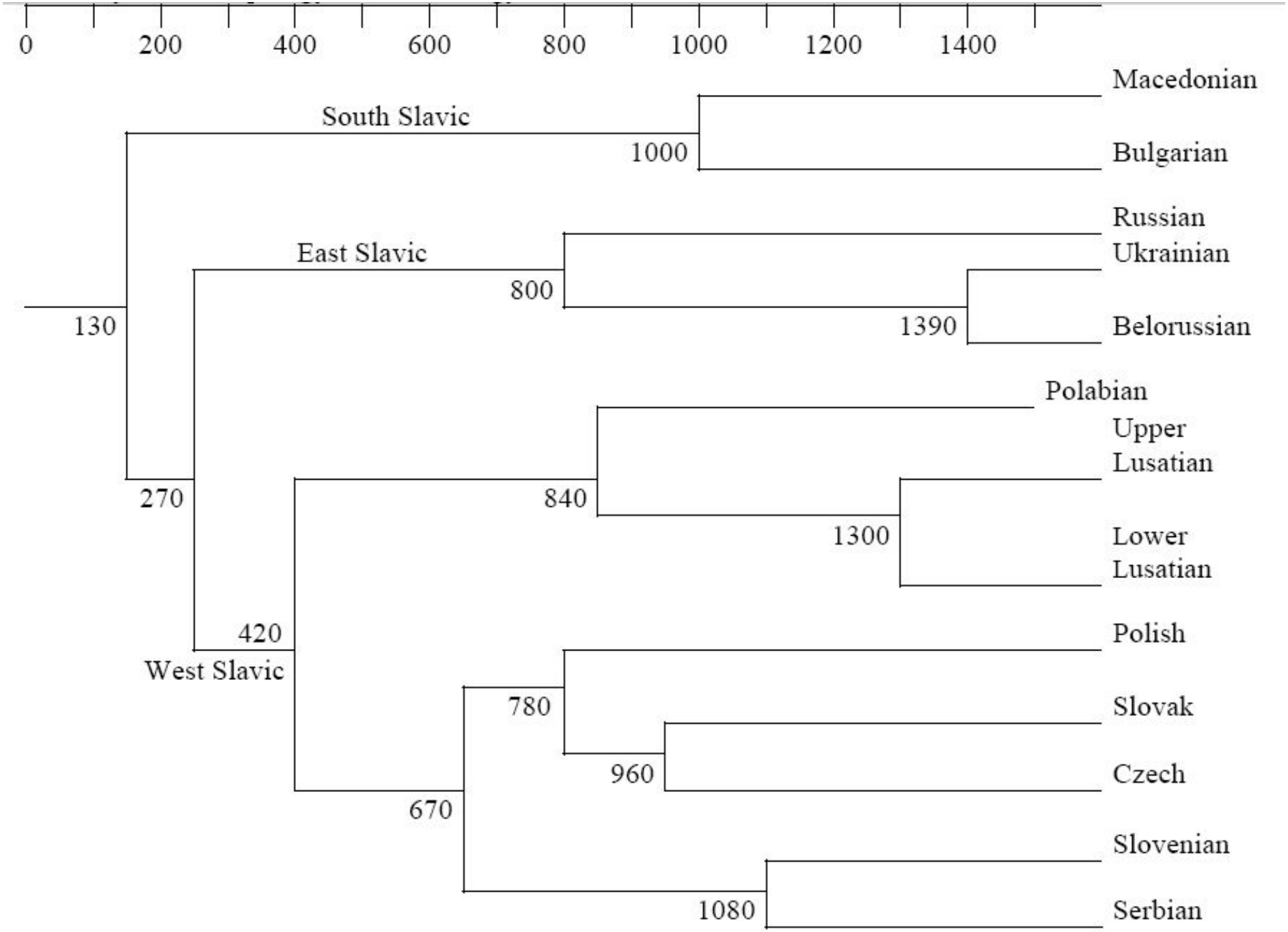
du Sud-Ouest:
SI: slovène
CR: croate
BO: bosniaque
M: monténégrin
SR: serbe

Du Sud-Est:
MA: macédonien
BU: bulgare

De l'Est:
RU: russe
BR: biélorusse
UK: ukrainien
RT: ruthène

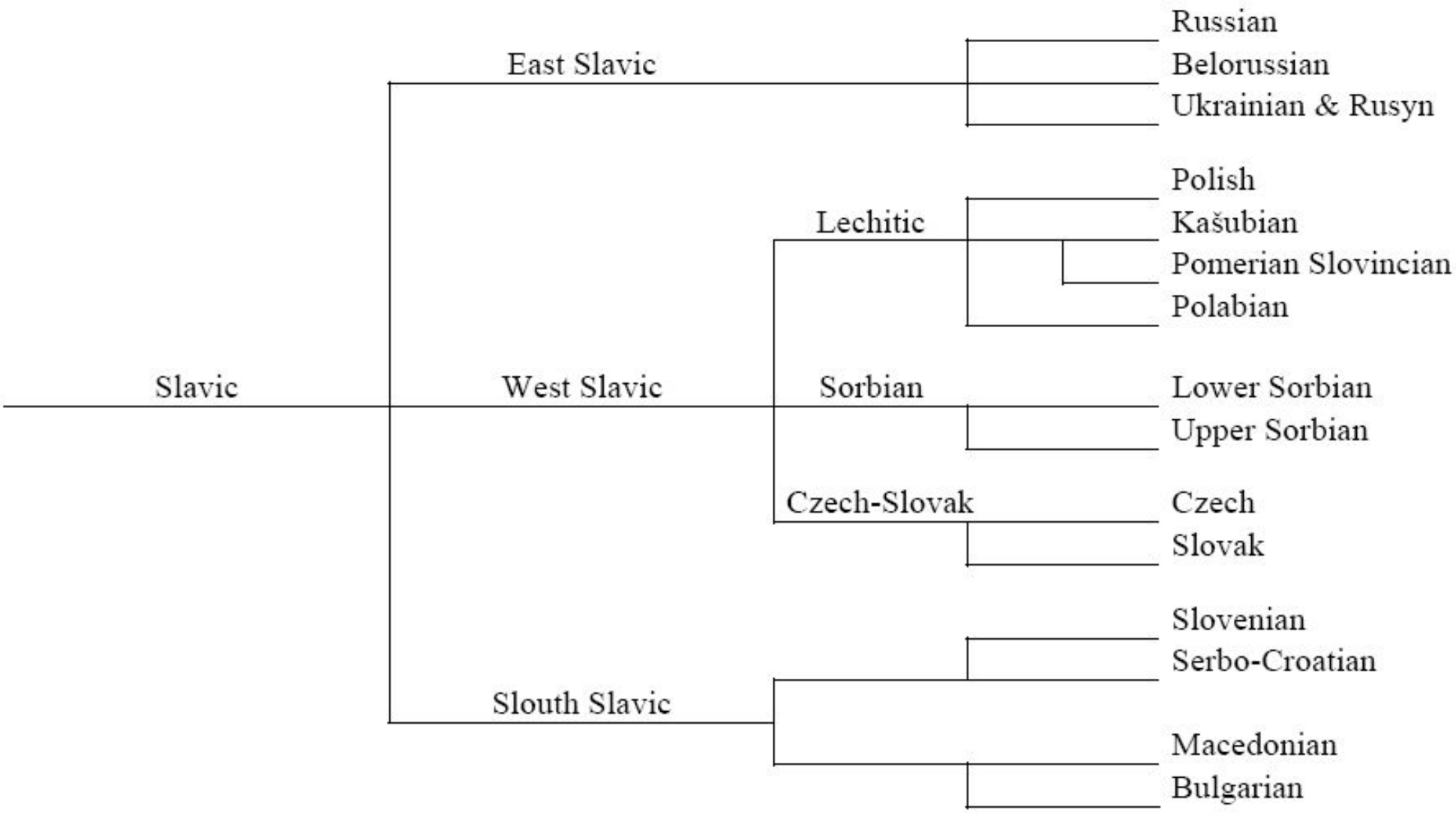


L’hypothèse de Starostin



Par rapport à la répartition traditionnelle

12.1. According to the traditional model the Slavic languages are divided into three parts (cf. e.g. J. & B. Grimes 1996, 58):



Le système linguistique des langues slaves de l’Ouest

Le système linguistique des langues slaves de l'Ouest

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.

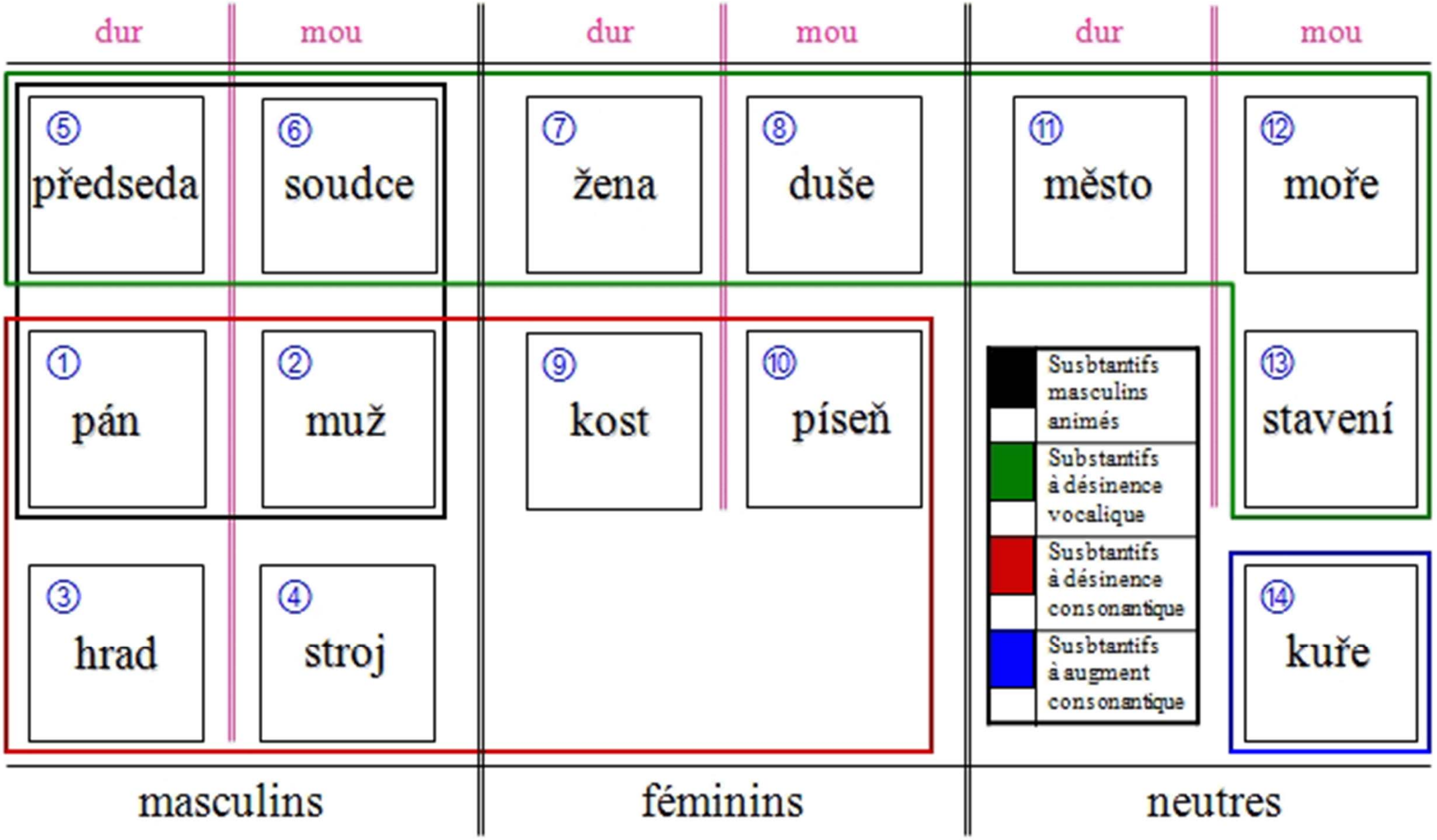
Le système linguistique (et tout particulièrement sa composante synchronique) est présenté suivant une dichotomie système nominal / système verbal en raison de particularités très différentes. Cette distinction, qui nous semble valide pour la description et la compréhension des langues slaves, n'est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

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.

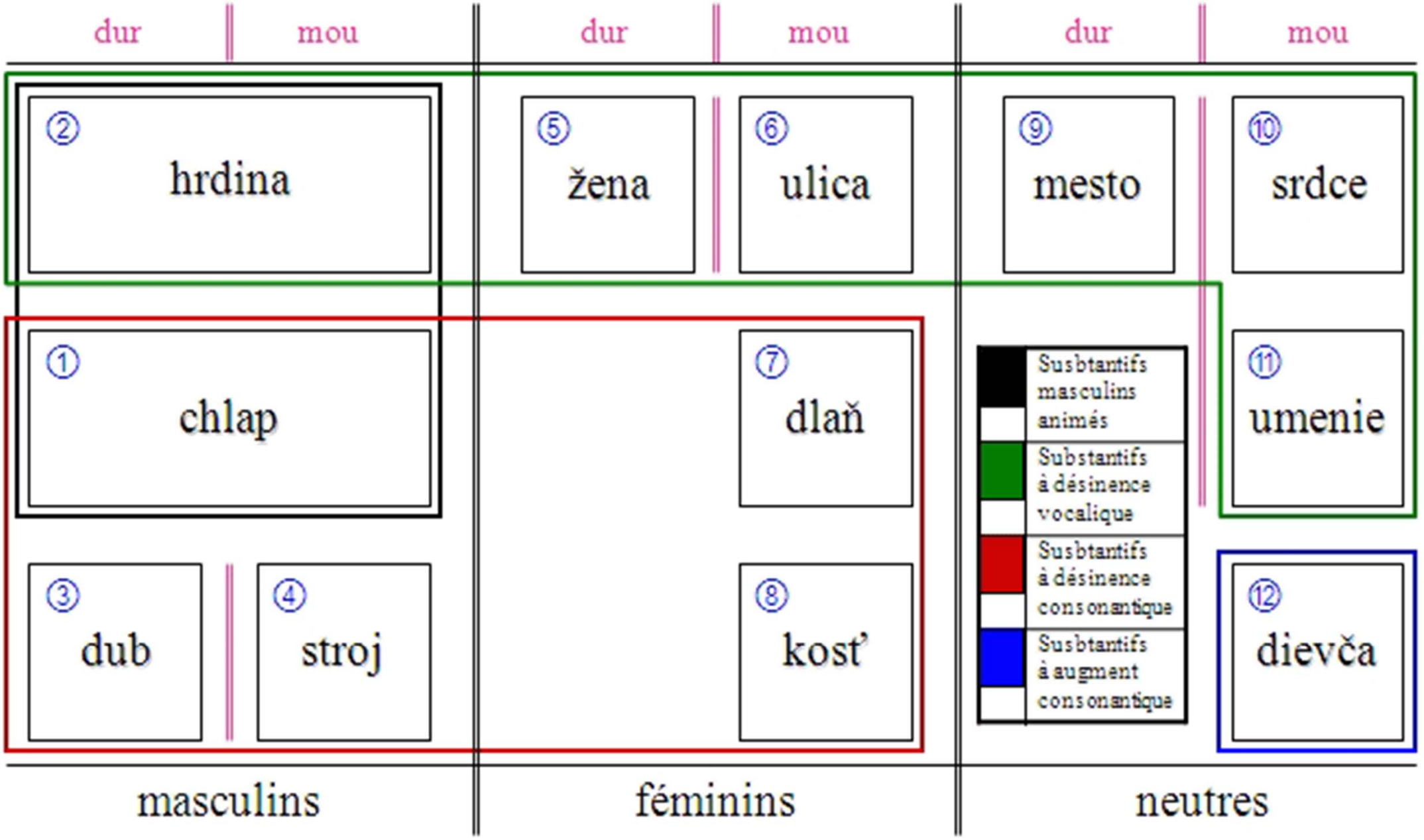
Système linguistique slave nominal

- flexion externe: marques casuelles en fin de mot
(sauf bulgare et macédonien)
- trois genres: masculin, féminin, neutre
- trois sous-genres masculins:
masculins humains
masculins animaux
masculins inanimés
- trois nombres: singulier, duel, pluriel
- sept cas: nominatif, génitif, datif, accusatif, vocatif, locatif,
instrumental
- terminaison vocalique / consonantique // augment consonantique
- opposition « dur / mou »

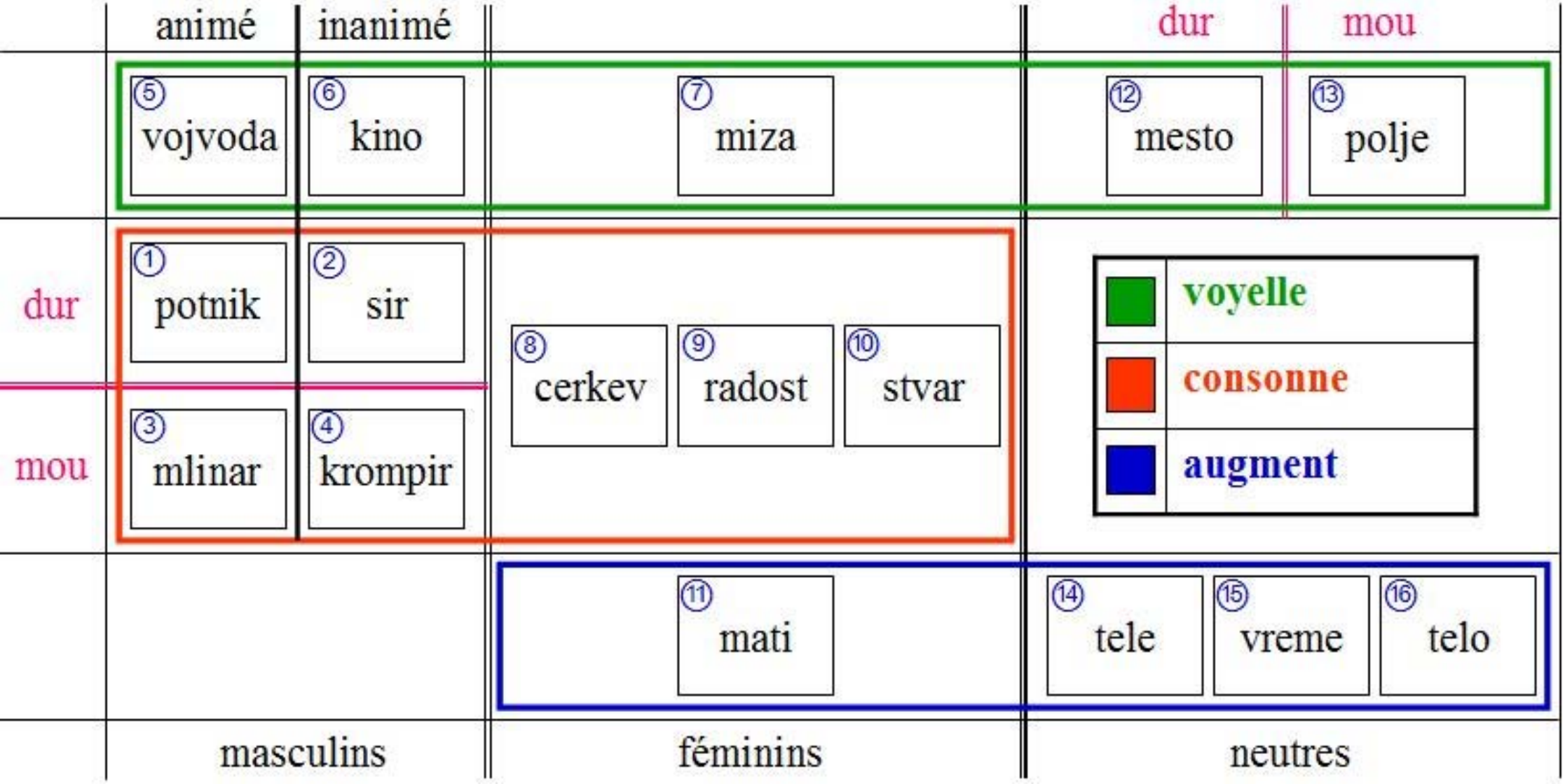
Paradigmes substantivaux



Paradigmes substantivaux



Paradigmes substantivaux



Système linguistique slave nominal (adjectif)

- opposition forme longue / forme courte

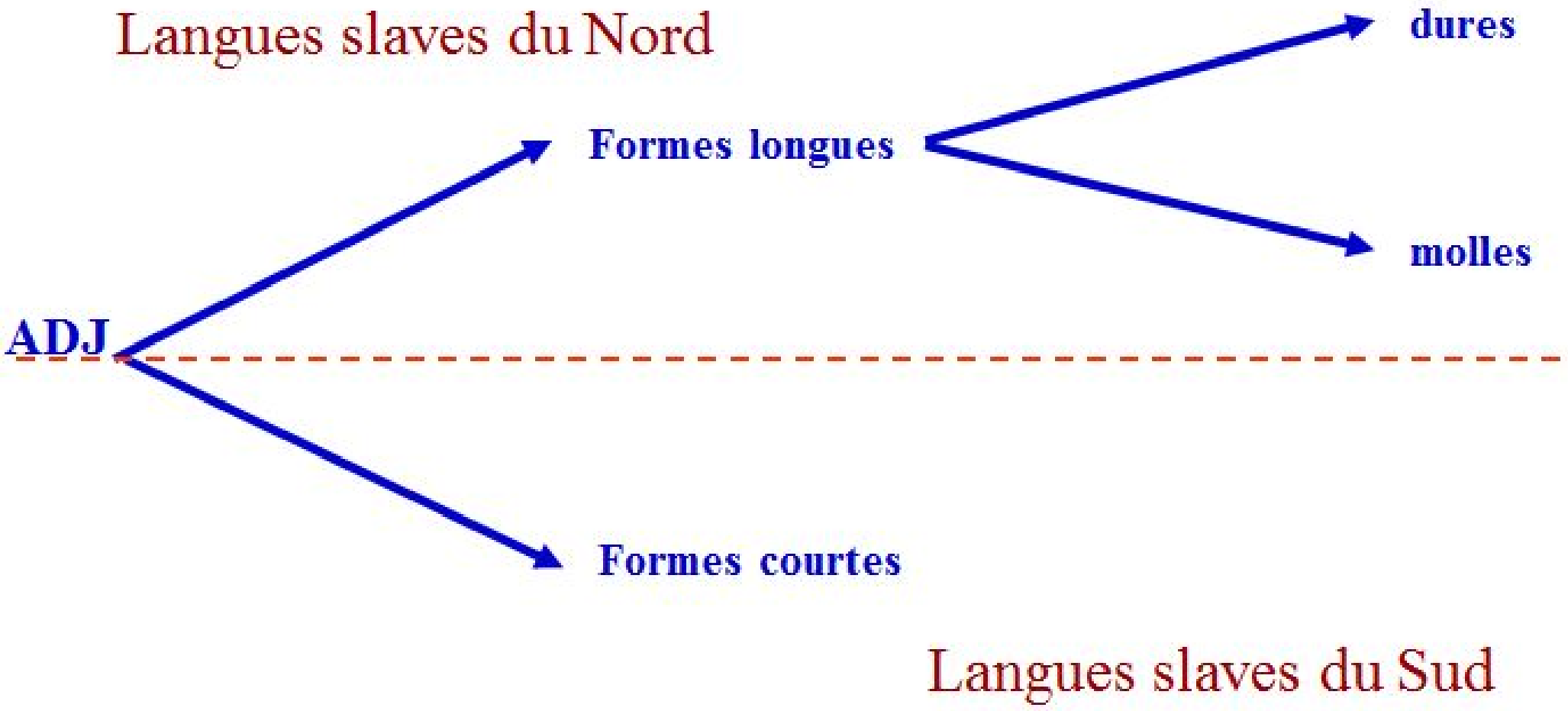
- opposition de la forme longue: dure / molle

Les langues slaves du Nord possèdent principalement des formes longues, les formes courtes y apparaissent de manière annexe. Au sein de ces langues du Nord, la proportion entre adjectifs durs et adjectifs mous varie considérablement: le slovaque a un modèle mou « cudzí » très minoritaire.

Les langues slaves du Sud sont marquées par les formes courtes.

Les différences de comportement au niveau des adjectifs auront des conséquences considérables au niveau des formes impersonnelles des verbes.

Système linguistique slave nominal (adjectif)



Système linguistique slave verbal

- trois genres: masculin, féminin, neutre
- trois sous-genres masculins:
masculins humains
masculins animaux
masculins inanimés
- trois nombres: singulier, duel, pluriel
- trois personnes (1, 2, 3)

- quatre valeurs aspectuelles:
perfectif, imperfectif, bi-aspectuel, itératif

Système linguistique slave verbal

formes personnelles

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			

Système linguistique slave verbal

- infinitif
- supin

- présent
(présent sémantique si imperfectif,
futur sémantique si perfectif)

- impératif

Système linguistique slave verbal

- deux passés simples:

- aoriste

- imparfait

- trois composés à partir du participe passé actif (participe parfait de Vaillant) par ajout d'auxiliaires:

- passé (composé)

- plus que parfait

- conditionnel (composé libre ou agglutiné)

Système linguistique slave verbal

Temps simples en bas-sorabe:

Infinitif:	kupowaś	***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	kupowašćo
3.	kupowašo	kupowaštej	kupowachu

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

Système linguistique slave verbal

- un futur simple: le présent morphologique du perfectif

- deux futurs composés:

- avec le lexème du passé (participe passé actif)

- avec le lexème de l'infinitif

Système linguistique slave verbal

Polonais – futurs composés

***Futur avec le participe passé actif en –l*:**

Singulier

masculin hum.	masculin animal	masculin inan.	féminin	neutre
1 będę miłował	1 będę miłował	1 będę miłował	1 będę miłowała	1 będę miłowało
2 będziesz miłował	2 będziesz miłował	2 będziesz miłował	2 będziesz miłowała	2 będziesz miłowało
3 będzie miłował	3 będzie miłował	3 będzie miłował	3 będzie miłowała	3 będzie miłowało

Duel

masculin hum.	masculin animal	masculin inan.	féminin	neutre
1	1	1	1	1
2	2	2	2	2
3	3	3	3	3

Pluriel

masculin hum.	masculin animal	masculin inan.	féminin	neutre
1 będziemy miłowali	1 będziemy miłowały	1 będziemy miłowały	1 będziemy miłowały	1 będziemy miłowały
2 będziecie miłowali	2 będziecie miłowały	2 będziecie miłowały	2 będziecie miłowały	2 będziecie miłowały
3 będą miłowali	3 będą miłowały	3 będą miłowały	3 będą miłowały	3 będą miłowały

Système linguistique slave verbal

Polonais – futurs composés

Futur avec l'infinitif:

Singulier

	masculin hum.	masculin animal	masculin inan.	féminin	neutre
1	będę miłować	będę miłować	będę miłować	będę miłować	będę miłować
2	będziesz miłować	będziesz miłować	będziesz miłować	będziesz miłować	będziesz miłować
3	będzie miłować	będzie miłować	będzie miłować	będzie miłować	będzie miłować

Duel

	masculin hum.	masculin animal	masculin inan.	féminin	neutre
1					
2					
3					

Pluriel

	masculin hum.	masculin animal	masculin inan.	féminin	neutre
1	będziemy miłować	będziemy miłować	będziemy miłować	będziemy miłować	będziemy miłować
2	będziecie miłować	będziecie miłować	będziecie miłować	będziecie miłować	będziecie miłować
3	będą miłować	będą miłować	będą miłować	będą miłować	będą miłować

Système linguistique slave verbal

formes impersonnelles

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			

Système linguistique slave verbal

***- gérondif présent (sur* imperfectif)**

- adjectif issu du gérondif présent

- gérondif passé (sur perfectif)

- adjectif issu du gérondif passé

(pour rappel): participe passé actif / « parfait » (en -l)

- adjectif issu du participe passé actif

- participe présent passif (en-m, obsolète)

- adjectif issu du participe présent passif

- participe passé passif (en -n ou en-t)

- adjectif issu du participe passé passif

- substantif verbal issu du participe passé passif

Système linguistique slave verbal

tchèque - gérondif passé:

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		

Système linguistique slave verbal

tchèque – adjectif issu du gérondif passé:

Adjectif issu du gérondif passé [pf]:

Singulier

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

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			nakupovavší		
2			nakupovavších		
3			nakupovavším		
4			nakupovavší		
5			nakupovavší		
6			nakupovavších		
7			nakupovavšími		

Système linguistique slave verbal

	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

	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	---	---

Qu’est-ce que 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.

Deux exemples de calculabilité synchronique

1. Les emprunts
2. la morphologie tchèque

Les emprunts

Reconnaissance automatique des emprunts (origine gréco-latine)

permet le « calcul » des adjectifs dérivés des substantifs en –ičnost.

Liste de prépositions et de conjonctions.

Substantifs en –ost.

permet le calcul automatique des radicaux d'adjectifs correspondants.

Substantifs verbaux.

Adjectifs durs, y compris:

ceux issus d'un participe passé passif
et les adjectifs prédicatifs en -telný

Adjectifs mous, y compris:

les gérondifs
et les adjectifs verbaux de but

Suffixation substantivale

p. ex. suffixes exprimant le lieu (-ště).

Les emprunts

Les emprunts, essentiellement ceux d'origine gréco-latine, sont reconnus:

1. par une structure graphémique différente.

2. et par le non respect des « lois » tchèques.

Les emprunts - structure graphémique différente

- f, x, q, w

- tchèque, slovaque et haut-sorabe: **„g“ ⇨ („γ“) ⇨ „h“**
évolution phonologique (entre fin du 10ème siècle et fin du 14ème siècle)
⇨ **t**out mot contenant un « g » est d'origine étrangère:
inte**g**rovat embar**g**o pro**g**ram

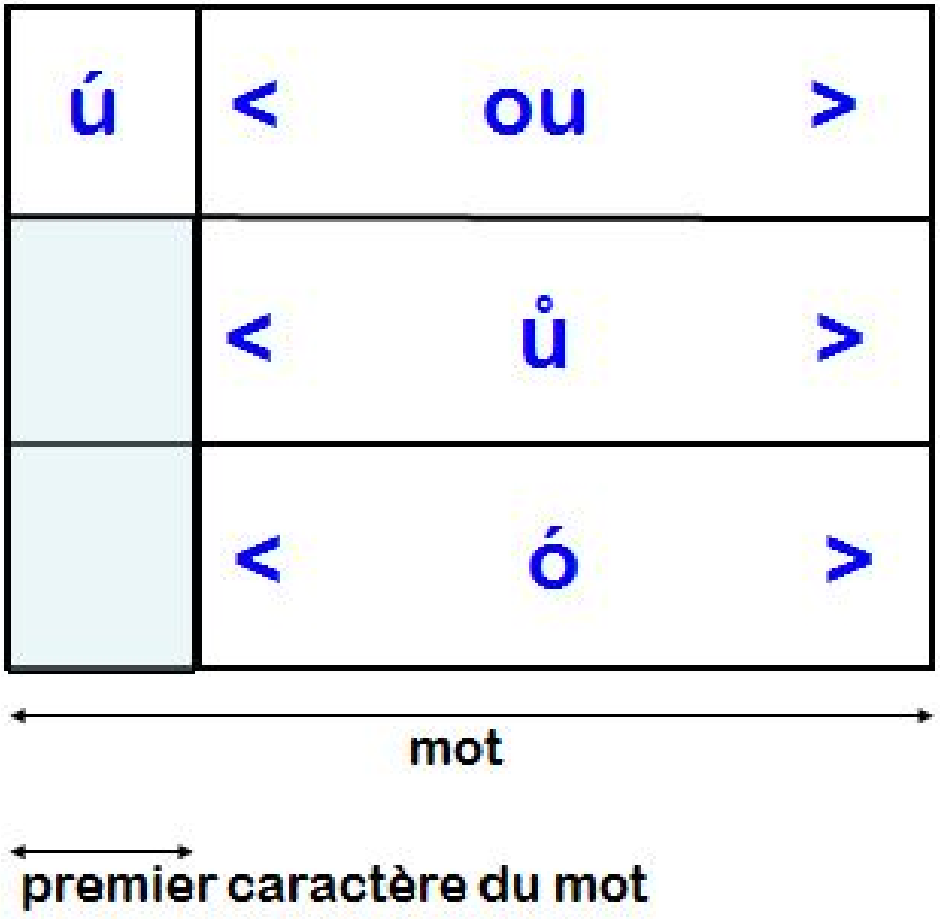
Les emprunts - structure graphémique différente

Évolution du o long tchèque: **„ó“ – „uo“/„uó“ – „ů“.**
Il en résulte que le « o » long dans un texte tchèque
marque toujours une origine étrangère.

„u“ long: ú

„o“ long tchèque: ó

„o“ long étranger: ó



„acetón”, „ambiciózní”, „chór”, „cirhóza”, „medailón”, „mykóza”, „ozón”, „penzión”, „prognóza”, „sezóna”, „skleróza”, „viróza” ...

Les emprunts - structure graphémique différente

A cause:

— du phénomène de contraction (entre le 9ème siècle et la fin du 10ème siècle)

— de la disparition de la diphtongue "ie" transformée en "í":

"trpie"	➪	"trpí" (*il(s)/elle(s) souffre(nt)*)...
"miera"	➪	"míra" (*mesure*)
"viera"	➪	"víra" (*foi*)
"bielý"	➪	"bílý" (*blanc*)

En dehors de la diphtongue « ou » issue du « u » long (« ú »), toute succession de deux voyelles marque un mot étranger

eufon**ie**

histor**ie**

Les emprunts - structure graphémique différente

- à une douzaine d'exceptions près, tous les termes débutant par "a",
- à une exception près, tous les termes débutant par "e",
- à une / deux exception près, tous les termes débutant par "i".

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' ".

aplikace
embargo
integrovat

Les emprunts - structure graphémique différente

1. $\mathscr{V}$ **N** $\mathscr{C}$

sauf à la jointure de morphèmes préfixe - racine, racine - suffixe, ...
où "$\mathscr{C}$" représente toute consonne sauf: {MPB}

k**onstant**a

2. $\mathscr{V}$ **M** $\mathscr{C}$

où "$\mathscr{C}$" = {PBF}

embargo
l**ymf**a, n**ymf**a

Il n'y a pas de "**M**", car le tchèque ne maintient pas la gémination lors des emprunts.

Les emprunts - non respect des « lois » tchèques

Oppositions « dur – mou »

1. $C_D + V_D \Rightarrow C_D + V_D$

2. $C_M + V_M \Rightarrow C_M + V_M$

3. $C_D + V_M \Rightarrow C_M + V_M$

4. $C_M + V_D \Rightarrow C_M + V_M$

Les emprunts - non respect des « lois » tchèques

règle 3: une consonne dure se transforme en consonne molle devant une voyelle molle

Exemples de palatalisations non respectées:

hi, ki, ri: **hi**sto**ri**e

he, ke, re: ko**ke**tovat, **re**servace

règle 4: une voyelle dure se transforme en voyelle molle après une consonne molle

Le mot: **cy**kl

qui ne respecte pas cette règle est étranger.

La morphologie tchèque

Reconnaissance automatique des emprunts (origine gréco-latine)

permet le « calcul » des adjectifs dérivés des substantifs en –ičnost.

Liste de prépositions et de conjonctions.

Adjectifs durs, y compris:

ceux issus d'un participe passé passif
et les adjectifs prédicatifs en -telný

Adjectifs mous, y compris:

les gérondifs
et les adjectifs verbaux de but

Substantifs en –ost.

permet le calcul automatique des radicaux d'adjectifs correspondants.

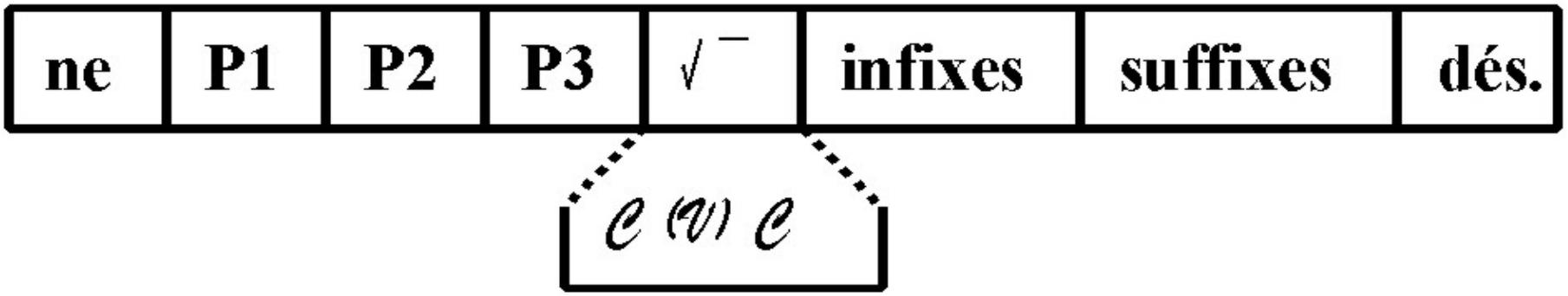
Substantifs verbaux.

Suffixation substantivale

p. ex. suffixes exprimant le lieu (-ště).

La morphologie tchèque

Structure du mot tchèque:



Grammatème:

E	TSS	DERIV	CATLEX	GRAMM (genre, nombre, cas)
P1	V		SUBST	M*NA,FS*,FPNA,...
Q	¬V		ADJ	
L			PREP...	
No				

La morphologie tchèque

Dans une langue à flexion externe
telle que le tchèque

les valeurs de catégorisation lexicale
des mots autochtones
sont situées à la **fin**

mais aussi au **début**

et même au **centre du mot**!

La morphologie tchèque

substantifs	
-ů	masculin génitif pluriel
-atům	neutre datif pluriel
-ům	masculin / neutre datif pluriel
-ami **-ěmi** **-emi**	féminins instrumental pluriel
adjectifs durs	
-ý	3 exceptions (prý, úterý, čehý)
-ého **-ých** **-ými**	sans exception (renvoient à environ 16000 adjectifs)

La morphologie tchèque

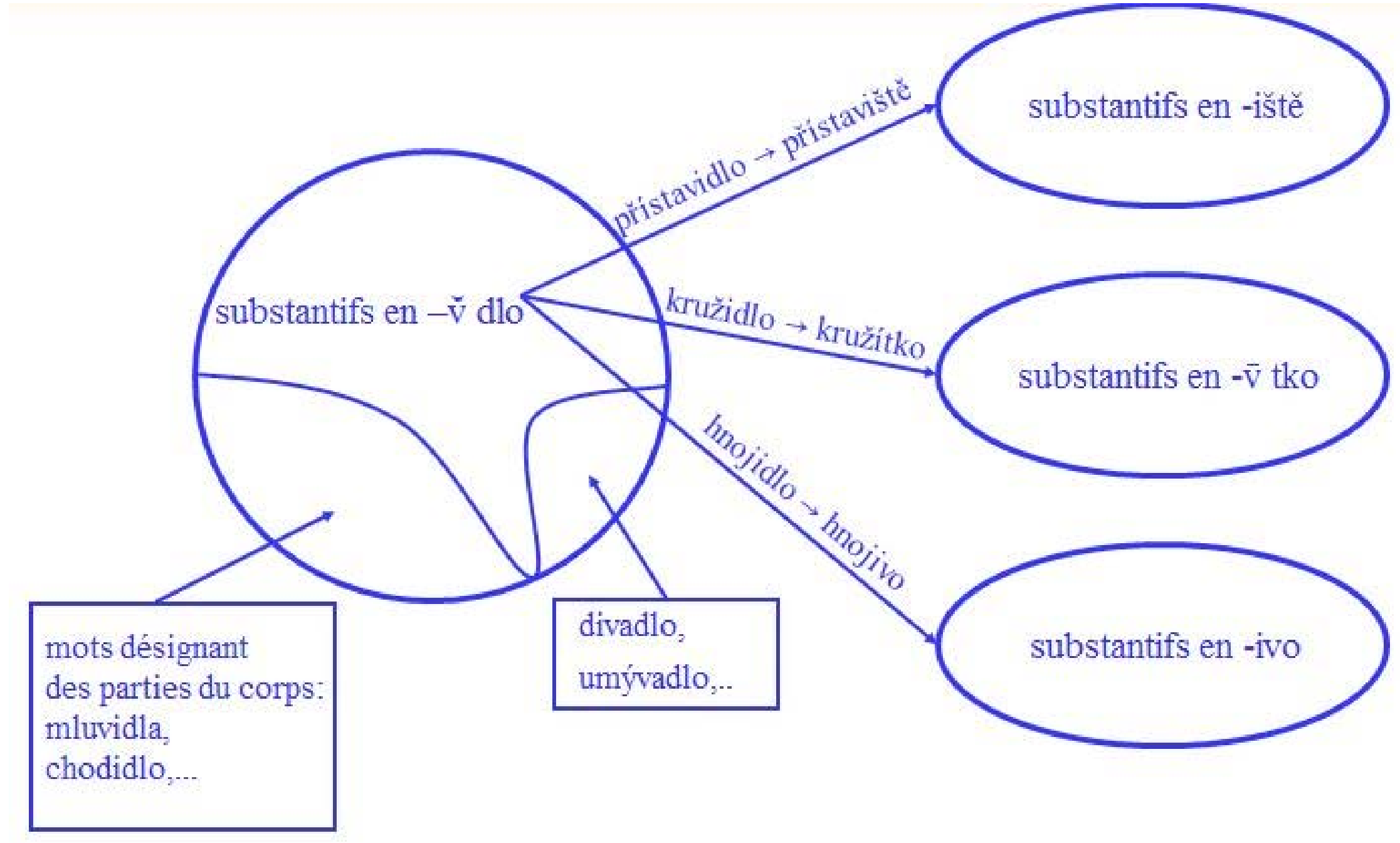
Bares note le passage des termes suffixés en « -*v*dlo »:
- exprimant **un lieu** vers une suffixation en "-ště":
"přistavidlo" ➠ "přistaviště" (*embarcadère*),
- nommant **un petit instrument** vers une suffixation en « -*v*tko »:
"kružidlo" ➠ "kružítko" (*compas*),
- désignant **des substances** (généralement en liquide ou en granulé) vers une suffixation en "-ivo":
"hnojidlo" ➠ "hnojivo" (*engrais*).

Il en résulte qu'en dehors de mots isolés tels que "divadlo" (*théâtre*),... et une petite série de termes désignant **des organes**:

"mluvidla" (*organes phonatoires*),
"kusadla" (*mandibules*),
"tykadla" (*antennes*),
"chodidlo" (*plante du pied*),...,
"rodidla" (*organes génitaux*),
"makadla" (*palpes*),
"chapadlo" (*tentacule*),

les mots terminés par "-ǔ dlo" renvoient à un ensemble productif représentant **des machines**:
"letadlo" [let – a-dlo]: *une machine permettant de* voler ➠ avion,
"vozidlo" [voz – i-dlo]: *une machine permettant de* transporter ➠ véhicule
"čerpadlo" [čerp – a-dlo]: *une machine permettant de* puiser ➠ pompe,...

La morphologie tchèque



La morphologie tchèque

Le suffixe « iště » forme des mots dont le genre est neutre.

Il marque, à 3 exceptions près, **un lieu** qui dans l'immense majorité des cas est un lieu ouvert.

Une spécification ultérieure est apportée par le radical du mot:

"rašeliniště" est un lieu où l'on trouve "rašelina" (*tourbe*), donc une *tourbière*,

"prameniště" est *un lieu où se trouvent des sources*, par exemple celles de la Vltava.

Souvent le radical est de nature verbale, exprimant l'action qui se déroule dans ce lieu:

"hřiště" est un lieu où l'on peut "hr" (de "hrát") *jouer*: (*un terrain de jeu*).

La morphologie tchèque

Les suffixes « árna » et « írna » renvoient à **des lieux où l'on a l'habitude de faire un certain type d'action:**

"tančírna" où l'on peut "tanč-it" (*danser*): *la salle de danse*

où l'on fabrique, vend ou consomme certains produits:

"lékárna" (de "lék" *médicament*): *la pharmacie*,

"mlékárna" (de "mléko" *lait*) est une *laiterie* ou une *crémerie*,

"kavárna", lieu où l'on consomme du café (dérivé de "káva", *le café*),

"tavírna" (de "tavit" *fondre*) est une *fonderie*

"ocelárna" (de "ocel" *acier*) est une *aciérie*, ...

La morphologie tchèque

Le suffixe "-ovna" représente très généralement **un lieu fermé**:

"pracovna", lieu où l'on peut "prac-ovat" (*travailler*), est *une salle de travail*,

"knihovna", (de "knih-a", *livre*) est *une bibliothèque*,...

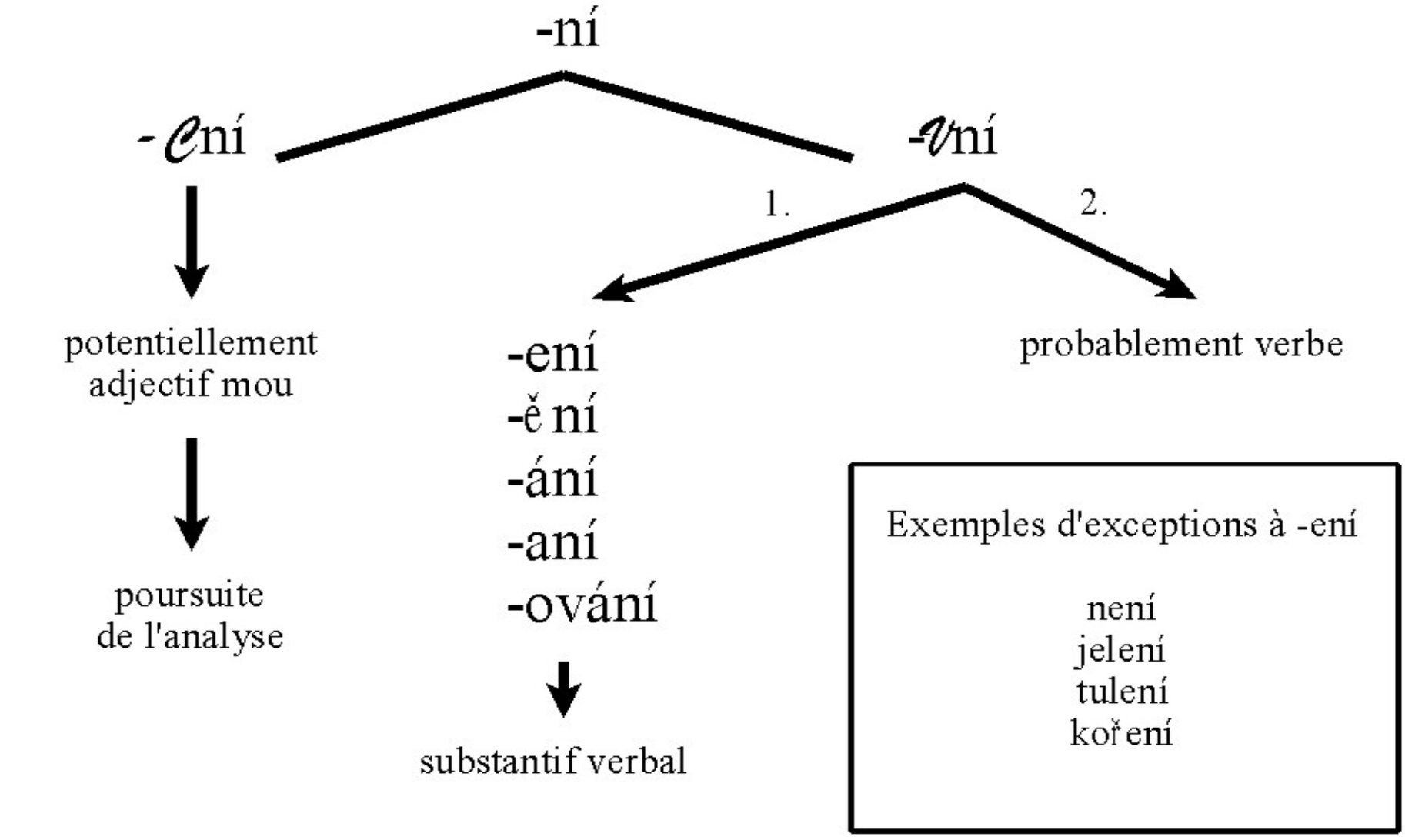
Le suffixe "-kyně" représente **des agents féminins** à une exception près ("jeskyně": *la grotte*) et une partie importante d'entre eux renvoie à **des professions**:

"soudkyně" (*la juge*).

Le suffixe "-ost" indique des substantifs de **qualité** dérivés d'un adjectif...

La morphologie tchèque

Patrons pour la reconnaissance de catégories lexicales

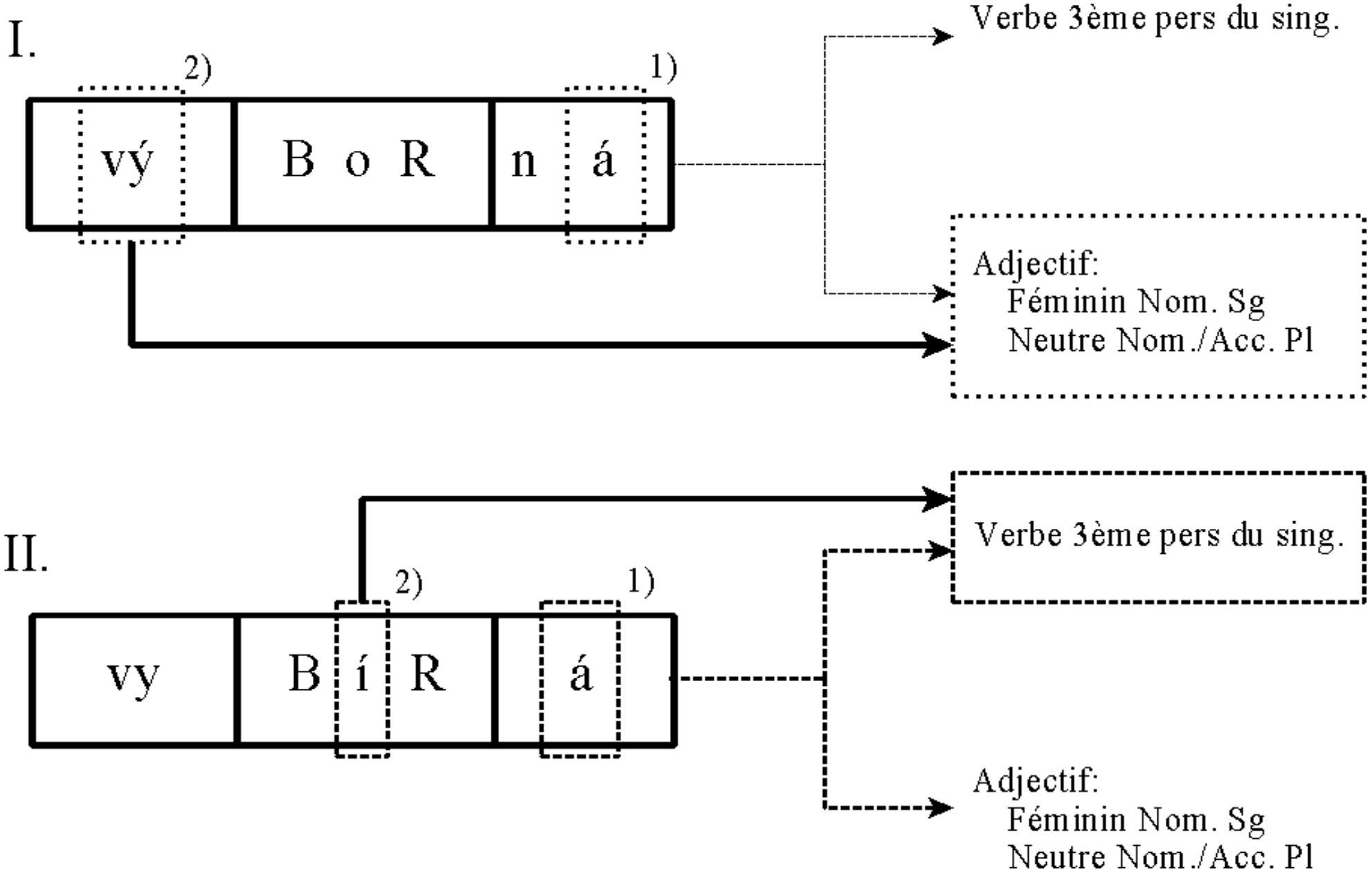


La morphologie tchèque

Préfixes à voyelle longue:
dů-, ná-, před-, prů-, pů-, ú-, vý-, zá-

Préfixe 1 long + √ brève	Préfixe 1 bref + √ longue
DERIVATION NON VERBALE	**DERIVATION VERBALE**
sauf une dizaine d'exceptions	valeur indéterminée
výpočtu	**vypočtu**
le calcul *(au génitif ou datif singulier)*	*je calculerai*
règle non réversible	

La morphologie tchèque



La morphologie tchèque

```
2 Protoz*e
2 zleps*ova/ni/m          Pl V   SUBST        NSI.NPD.
2 a                             COORD
2 zdokonalova/ni/m        Pl V   SUBST        NSI.NPD.
2 klasicke/                  ^V  ADJ          MPNA.FSGDL.FPNA.NSNA.
2 souc*a/stkove/             ^V  ADJ          MPNA.FSGDL.FPNA.NSNA.
2 za/kladny
2 z                             PREP         **G.                    G
2 diskre/tni/ch           E  ^V  ADJ          *PGL.
2 funkc*ni/ch             E  ^V  ADJ          *PGL.
2 elektronicky/ch         E  ^V  ADJ          *PGL.
2 prvku@                     ^V  SUBST        MPG.                        G
2 bylo
2 moz*no
2 dosa/hnout
2 jen
2 di/lc*i/ch
2 zleps*eni/              Pl V   SUBST        NSNGDAL.NPNGA.
2 a                             COORD
2 zdokonaleni/            Pl V   SUBST        NSNGDAL.NPNGA.
2 s                             PREP         **IG.                   IG
2 omezeny/m                  V   ADJpp        MSI.NSI.*PD.
2 technicko-ekonomicky/m     ^V  ADJ          MSI.NSI.*PD.
2 pr*i/nosem
2 ,
```

La morphologie tchèque

```
2 projevily
2 se
2 koncem
2 s*edesa/ty/ch          No ^V ADJ          *PGL.
2 let
2 na                        PREP          **LA.                LA
2 pracovis*ti/ch         L  ^V SUBST        NPL.                     G
2 v                         PREP          **LA.                LA
2 za/kladni/m               ^V ADJ         MSLI.NSLI.*PD.
2 a                         COORD
2 aplikovane/m           E     V  ADJpp     MSL.NSL.
2 vy/zkumu
2 u                         PREP          **G.                 G
2 sve*tovy/ch               ^V ADJ         *PGL.
2 vy/robcu@                 ^V SUBST       MPG.                     G
2 elektronicky/ch        E  ^V ADJ         *PGL.
2 souc*a/stek
2 prvni/                    ^V ADJ         M*NA.FS*.FPNA.N*NA.
2 snahy
2 o                         PREP          **LA.                LA
2 vytvor*eni/            Pl V  SUBST       NSNGDAL.NPNGA.
2 pr*edpokladu@             ^V SUBST       MPG.                     G
2 pro                       PREP          **A.                 A
2 pr*echod
2 na                        PREP          **LA.                LA
2 novou                     ^V ADJ         FSAI.
2 ,
2 kvalitativne*
2 vys*s*i/
2 u/roven*
2 elektroniky            E
2 --
2 integrovanou           E     V  ADJpp     FSAI.
2 elektroniku            E
2 .
```

calculabilité comparée des langues slaves de l’Ouest

Vers la calculabilité comparée des langues slaves de l'Ouest

La calculabilité d'un texte écrit commence au niveau du système d'écriture et de ses codes.
Ainsi, là où le tchèque possède les graphèmes « i », « í », « y » et « ý », le polonais ne possède que « i » et « y » (absence de longueur) et le slovène « i » (non maintien de l'opposition des deux « i »).
L'ambiguïté augmente lorsque le nombre de formes diminue.

Dans les deux suites qui se correspondent en tchèque et en slovaque, en tchèque la graphie est classificatoire, en slovaque la « calculabilité » n'est plus assurée :

	adjectif masculin	adjectif féminin	adjectif neutre	adverbe
tchèque	krásný	krásná	krásné	krásně
slovaque	krásny	krásna	krásne	krásne
	beau	*belle*	*beau*	*de belle manière*

Vers la calculabilité comparée des langues slaves de l'Ouest

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.

Soient les suites suivantes:

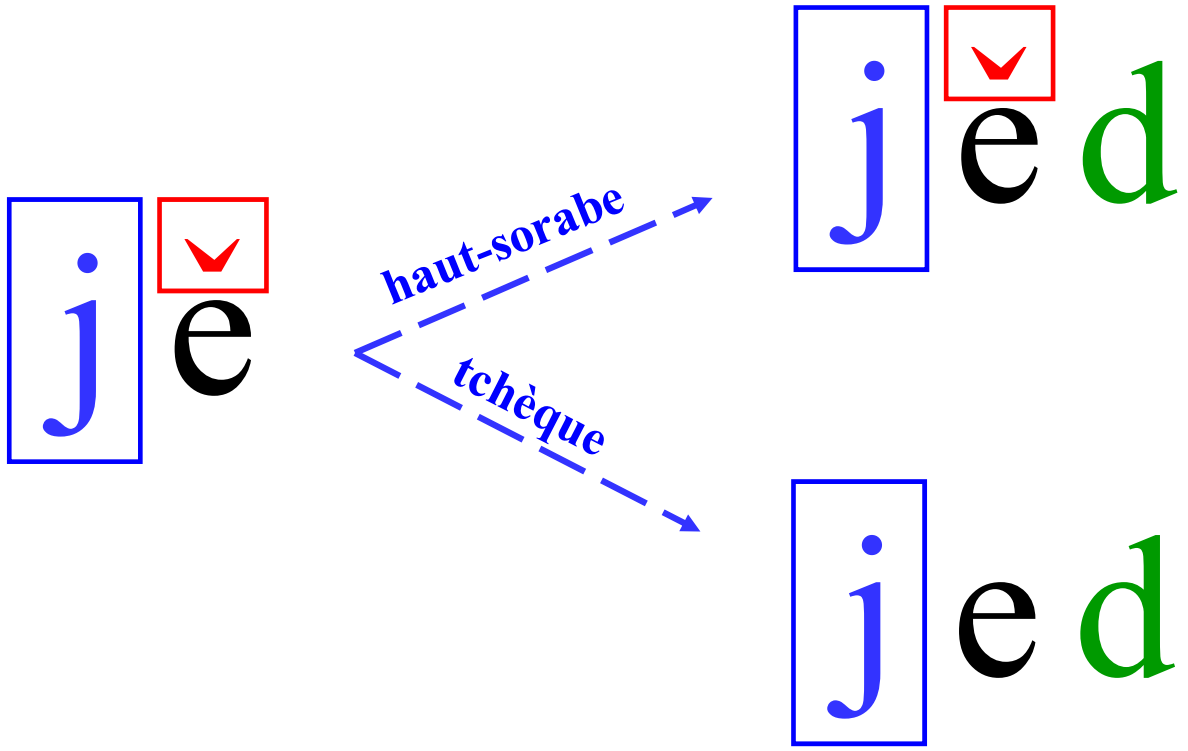
tchèque:	« blízký, blízká, blízké »	« blízko »	« blíže »
croate:	« blizak, bliska, blisko »		« bliže »
polonais:	« bliski, bliska, bliskie »	« blisko »	« bliżej »

En croate, la forme « bliže » (*plus proche*) peut se comprendre par rapport à la forme masculine de l'adjectif.

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.

Vers la calculabilité comparée des langues slaves de l'Ouest

Double mouillure en haut-sorabe impossible en tchèque:



Vers la calculabilité comparée des langues slaves de l'Ouest

Une tendance naturelle des langues slaves est de ne pas mettre de voyelle en tête de mot. 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).

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	

Vers la calculabilité comparée des langues slaves de l'Ouest

Métathèse des liquides:

haut sorabe	0,5	kruwa	ert	žałza
tchèque	1	kráva	ret	žláza
slovaque		krava		
polonais		krowa		
bas sorabe	?	krowa		

Vers la calculabilité comparée des langues slaves de l'Ouest

Opposition g – h:

haut sorabe	1	ratarske graty
tchèque		integrovaný Hotel Gomel
slovaque		*výjimky*: mozog
polonais	0	
bas sorabe		

Vers la calculabilité comparée des langues slaves de l'Ouest

Succession de deux voyelles:

tchèque	0,9	kromě « ou » histor**ie**
slovaque	0,5	kromě « ou », « ia », « ie », « iou », « iu » demokra**cia**

Du passé au présent: quelques lois de calcul

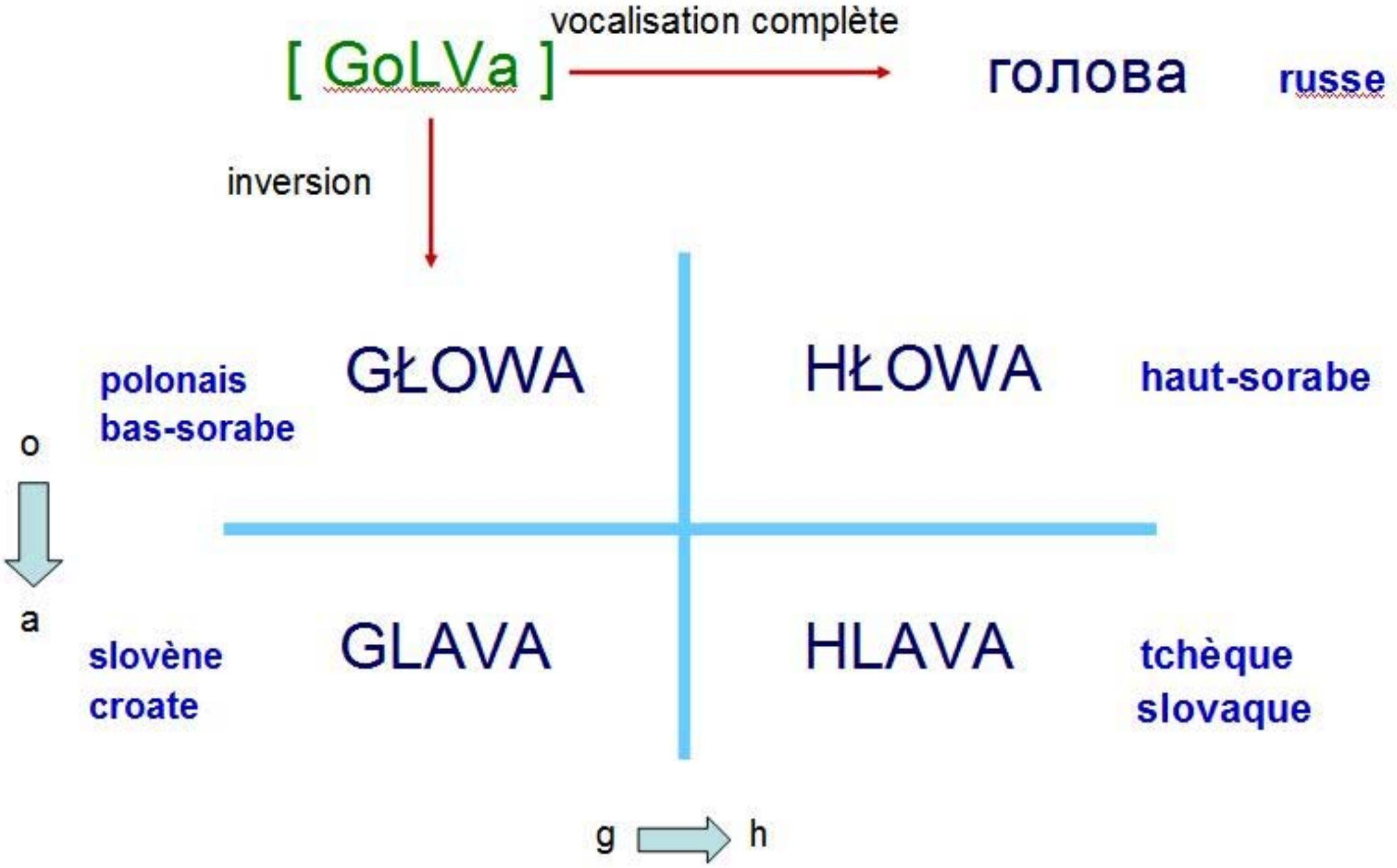
vérifiées sur la base du PHC Barrande 2010-11 « génération automatique de lexèmes slaves à partir de racines, liée à la connaissance de la morpho-phonologie diachronique »,

publiées dans: Pognan, P., Panevová, J.: 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). Ljubljana : Linguistica n° 52, pp 59-75, 2013

Diachronie: l'évolution diachronique du tchèque

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

Diachronie: la métathèse en « o »

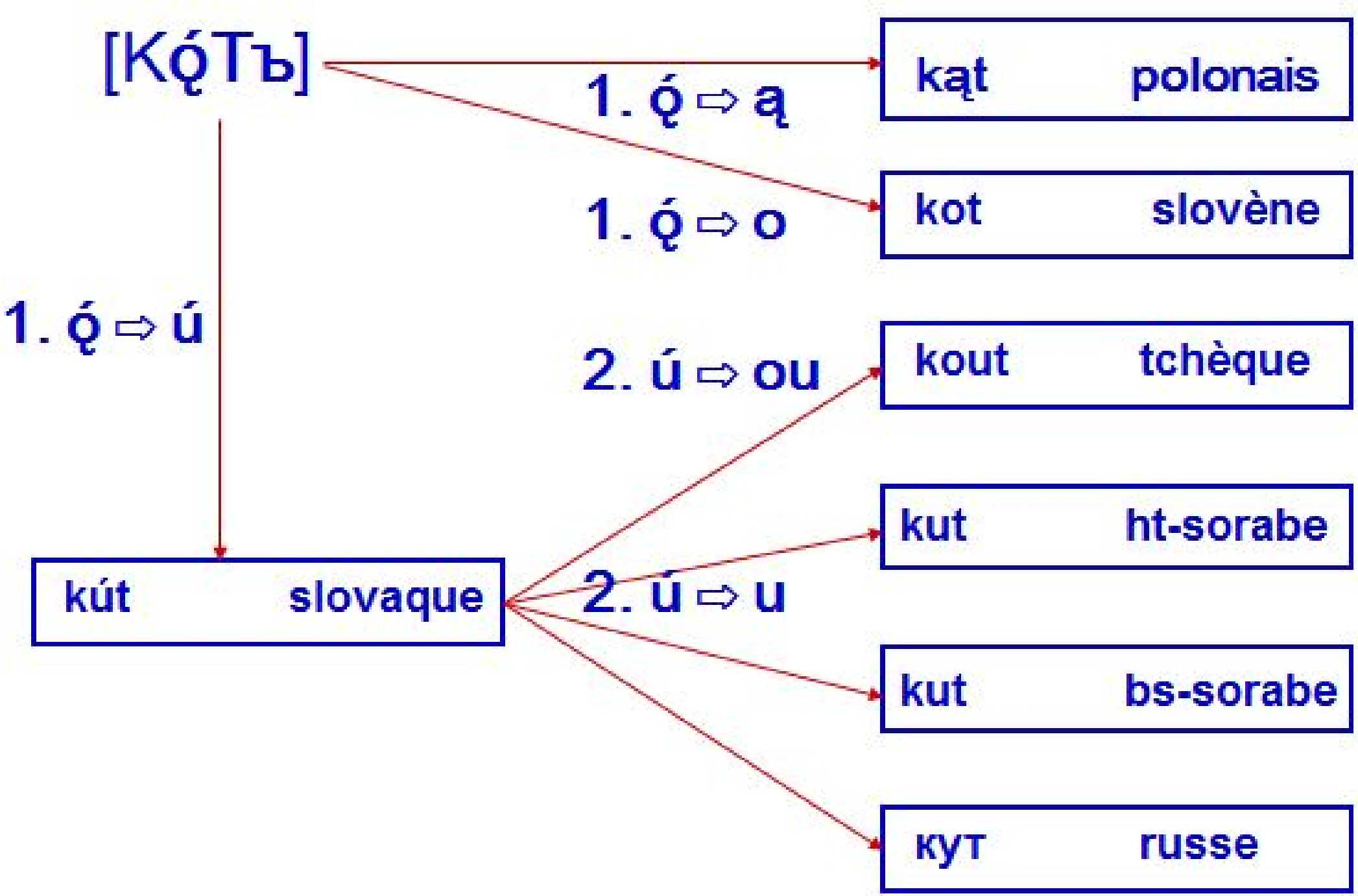


Diachronie: la métathèse en « o »

GoLVă

fon. jev kořen	ruština golva	polština golva	čeština golva	slovenšt. golva	dolní_luž golva	horní_luž golva	slovinšt. golva	chorvat, golva	srbština golva
fon. jev méth -o-	ruština golova	polština głova	čeština glava	slovenšt. glava	dolní_luž głova	horní_luž głova	slovinšt. glava	chorvat, glava	srbština glava
fon. jev g / h	ruština golova	polština głova	čeština hlava	slovenšt. hlava	dolní_luž głova	horní_luž hłova	slovinšt. glava	chorvat, glava	srbština glava
fon. jev v / w	ruština golova	polština głowa	čeština hlava	slovenšt. hlava	dolní_luž głowa	horní_luž hłowa	slovinšt. glava	chorvat, glava	srbština glava
fon. jev cyrill + x	ruština голова	polština głowa	čeština hlava	slovenšt. hlava	dolní_luž głowa	horní_luž hłowa	slovinšt. glava	chorvat, glava	srbština глава

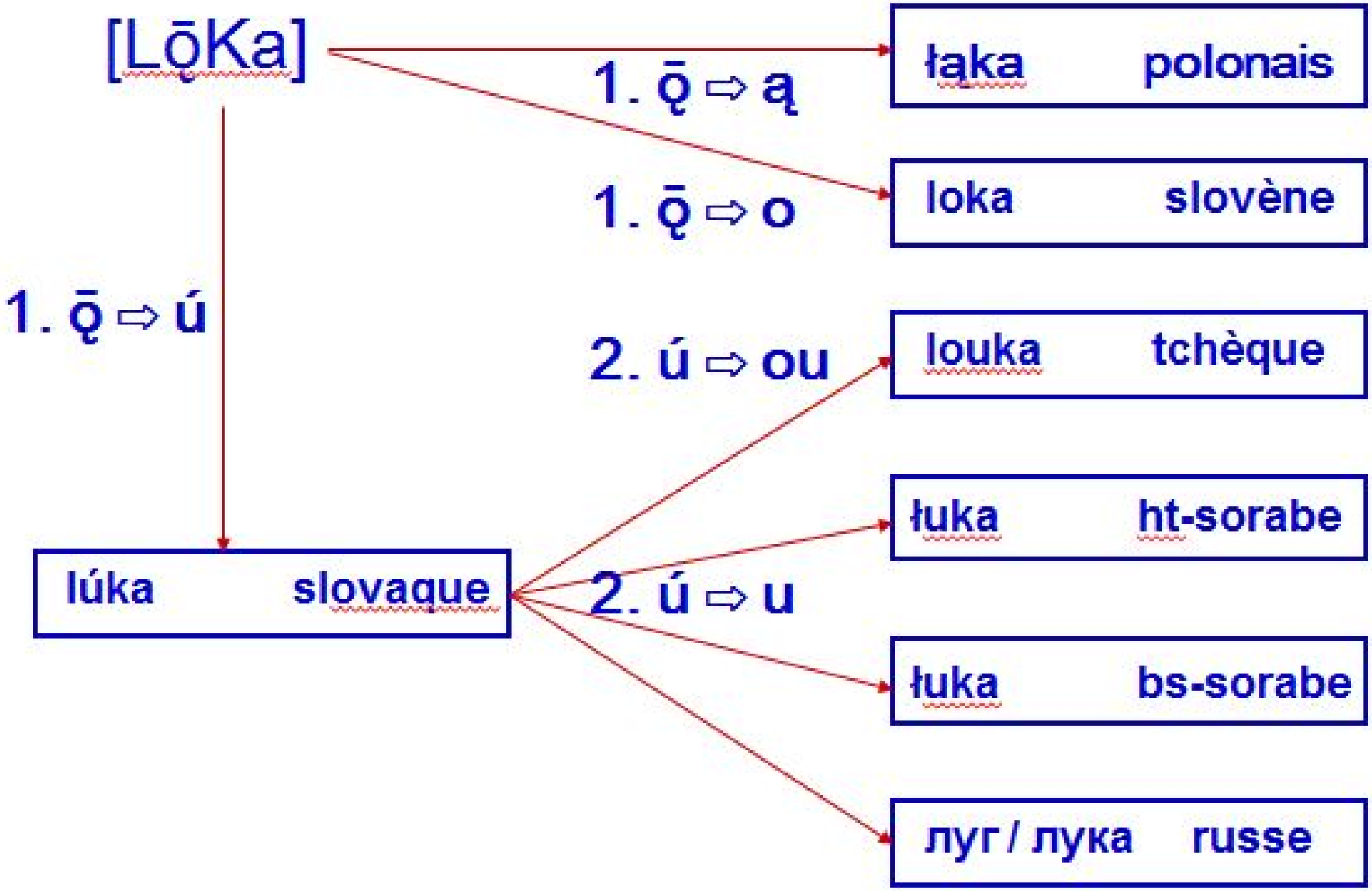
Diachronie: transformation de la nasale « ǫ » ascendante longue



Diachronie: transformation de la nasale « ǫ » ascendante longue

KǫTъ									
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat,	srbština
kořen	kǫtъ	kǫtъ	kǫtъ	kǫtъ	kǫtъ	kǫtъ	kǫtъ	kǫtъ	kǫtъ
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat,	srbština
ǫ / ą	kutъ	kątъ	kútъ	kútъ	kutъ	kutъ	kotъ	kutъ	kutъ
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat,	srbština
ъ / 0	kut	kąt	kút	kút	kut	kut	kot	kut	kut
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat,	srbština
ú ⇨ ou	kut	kąt	kout	kút	kut	kut	kot	kut	kut
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat,	srbština
cyrill + x	кут	kąt	kout	kút	kut	kut	kot	kut	кут

Diachronie: transformation de la nasale « ǫ » inaccentuée longue



Diachronie: transformation de la nasale « ǫ » inaccentuée longue

LǭKa

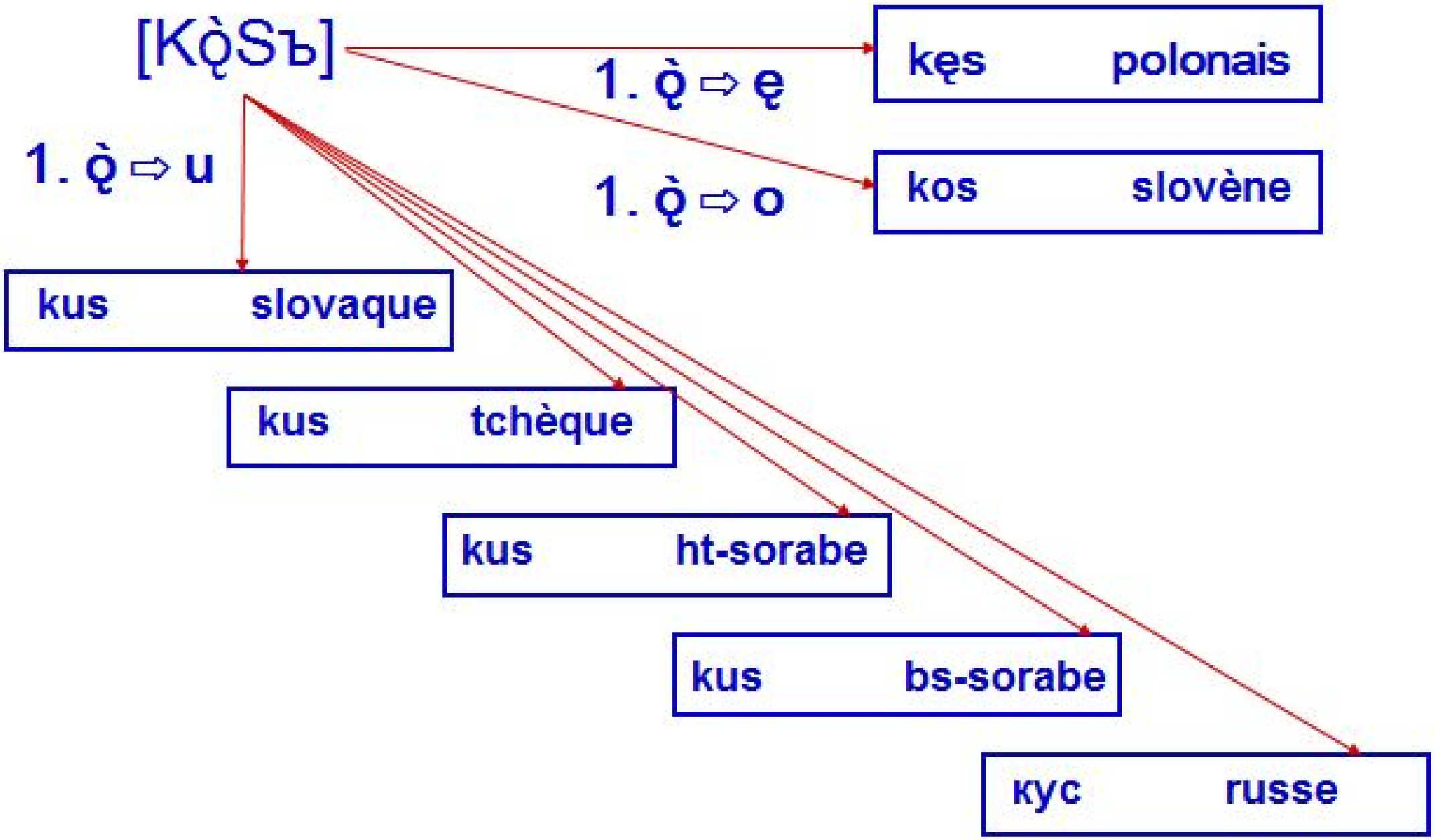
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat,	srbština
kořen	lǭka	lǭka	lǭka	lǭka	lǭka	lǭka	lǭka	lǭka	lǭka

fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat,	srbština
ǭ / ą	luka	łąka	lúka	lúka	luka	luka	loka	luka	luka

fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat,	srbština
ú ⇨ ou	luka	łąka	louka	lúka	luka	luka	loka	luka	luka

fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat,	srbština
cyrill + x	лука	łąka	louka	lúka	luka	luka	loka	luka	лука

Diachronie transformation de la nasale « ǫ » descendante longue



Diachronie transformation de la nasale « ǫ » descendante longue

KǫSъ
KǫS

phénom	russe	polonais	tchèque	slovaque	b_sorabe	h_sorabe	slovène	croate
rac minus	kǫs	kǫs	kǫs	kǫs	kǫs	kǫs	kǫs	

phénom	russe	polonais	tchèque	slovaque	b_sorabe	h_sorabe	slovène	croate
(u'ǫ', u'ę')	kus	kęs	kus	kus	kus	kus	kus	

phénom	russe	polonais	tchèque	slovaque	b_sorabe	h_sorabe	slovène	croate
g vers h	kus	kęs	kus	kus	kus	kus	kus	

phénom	russe	polonais	tchèque	slovaque	b_sorabe	h_sorabe	slovène	croate
(u'u', u'o')	kus	kęs	kus	kus	kus	kus	kos	

phénom	russe	polonais	tchèque	slovaque	b_sorabe	h_sorabe	slovène	croate
(u'ь', u'')	kus	kęs	kus	kus	kus	kus	kos	

phénom	russe	polonais	tchèque	slovaque	b_sorabe	h_sorabe	slovène	croate
cyrill + x	кус	kęs	kus	kus	kus	kus	kos	

Diachronie: transformation de la nasale « ę »

PęTь

fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat,	srbština
kořen	pętь	pętь	pętь	pętь	pętь	pętь	pętь	pętь	pętь
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat,	srbština
ę-2	pjatь	piętь	pětь	pätь	pětь	pjetь	petь	petь	petь
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat,	srbština
tь	pjatь	pięć	pět	päť	pěś	pjeć	petь	pet	petь
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat,	srbština
ь	pjatь	pięć	pět	päť	pěś	pjeć	pet	pet	pet
fon. jev	ruština	polština	čeština	slovenšt.	dolní_luž	horní_luž	slovinšt.	chorvat,	srbština
cyrill + x	пять	pięć	pět	päť	pěś	pjeć	pet	pet	пет

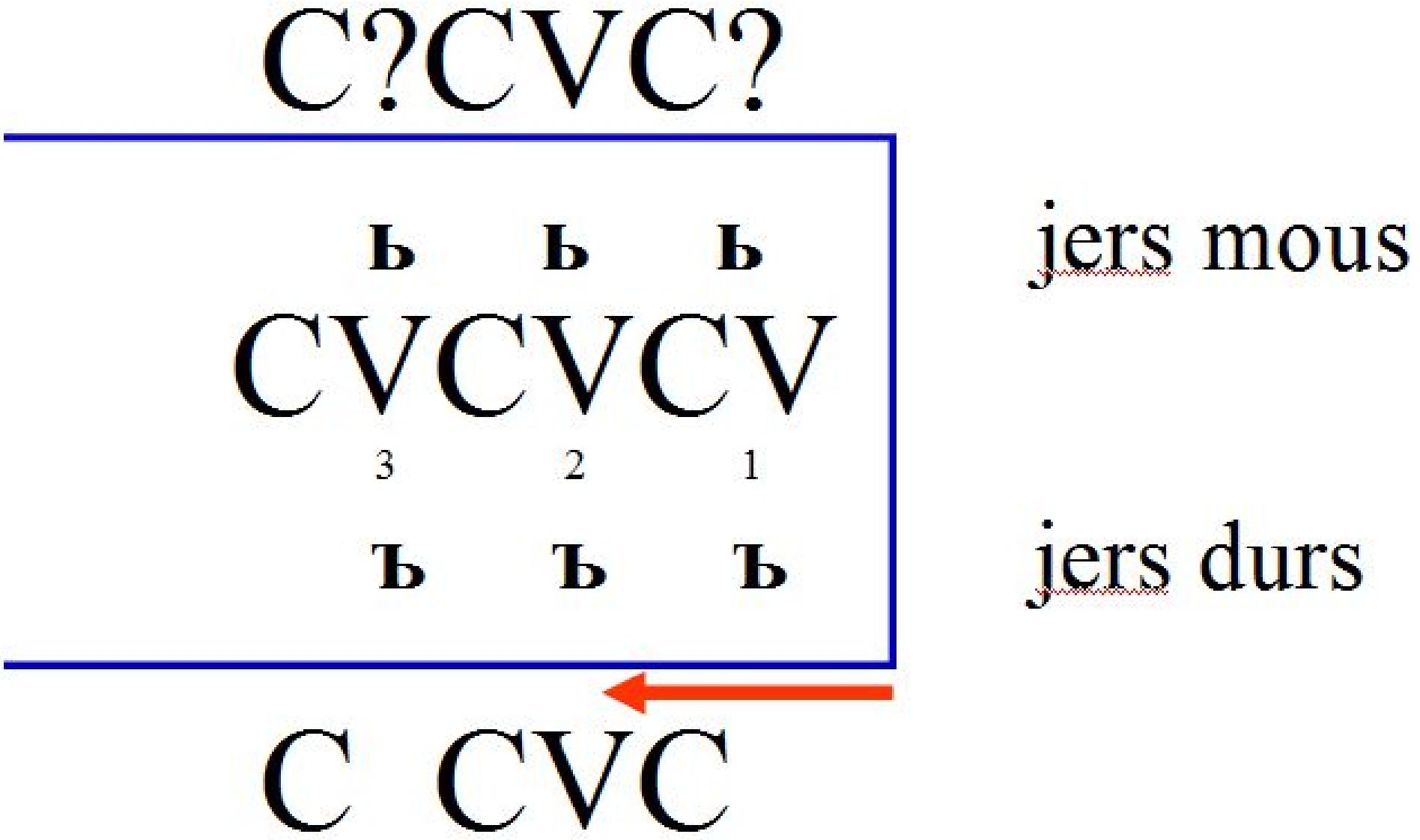
Diachronie: la contraction et ses conséquences

(repris de Marvan, J. (2000): Jazykové milénium. Slovanská kontrakce a její český zdroj. Prague : Academia)

		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	—	—	—	+	+	+	—	—	—	—	3	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		

Diachronie: disparition et vocalisation des « jers »

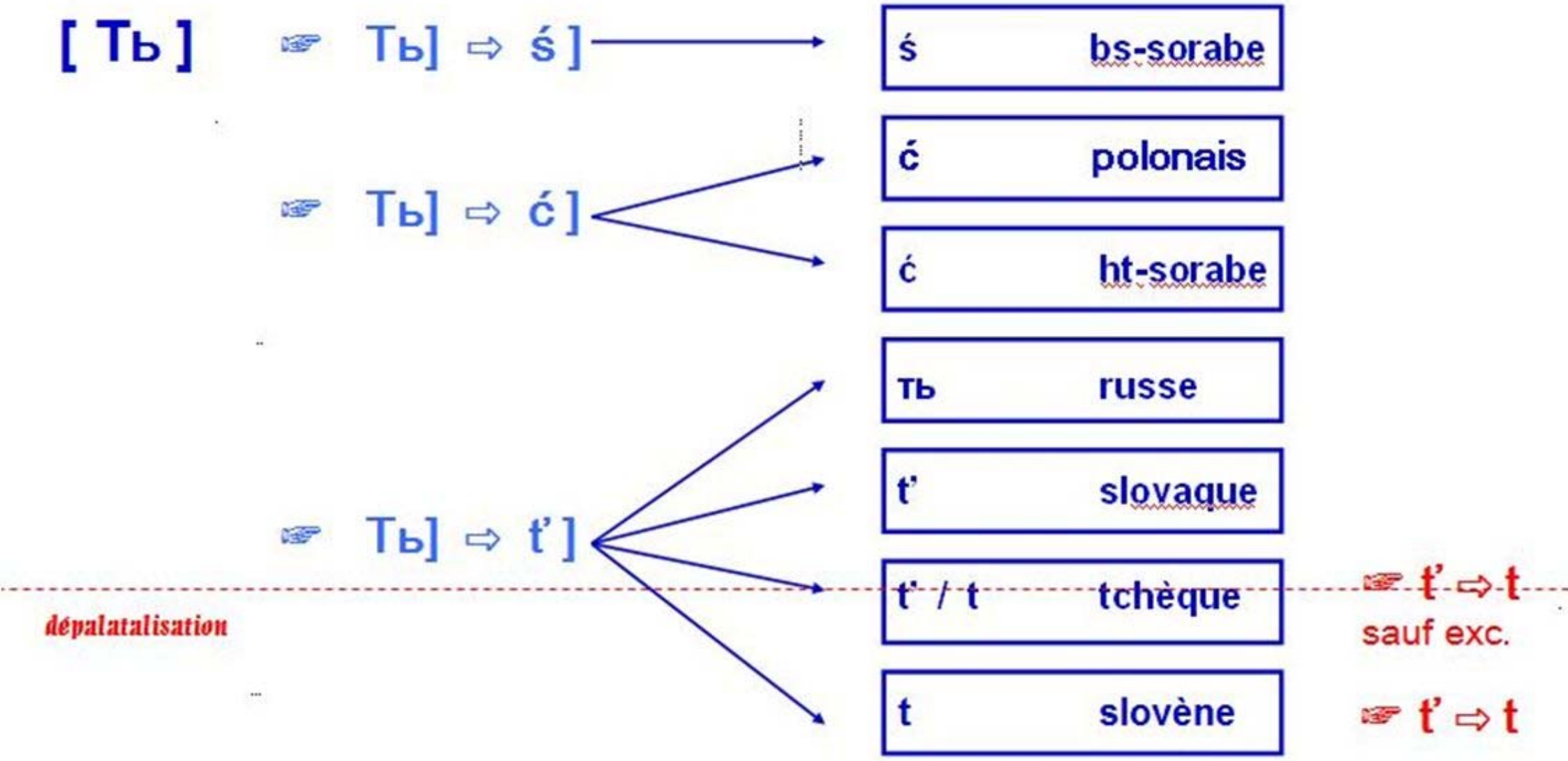
(loi de Havlík)



Diachronie: disparition et vocalisation des « jers »

	ь	ъ
tchèque	e	e
polonais	e	e
bas 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

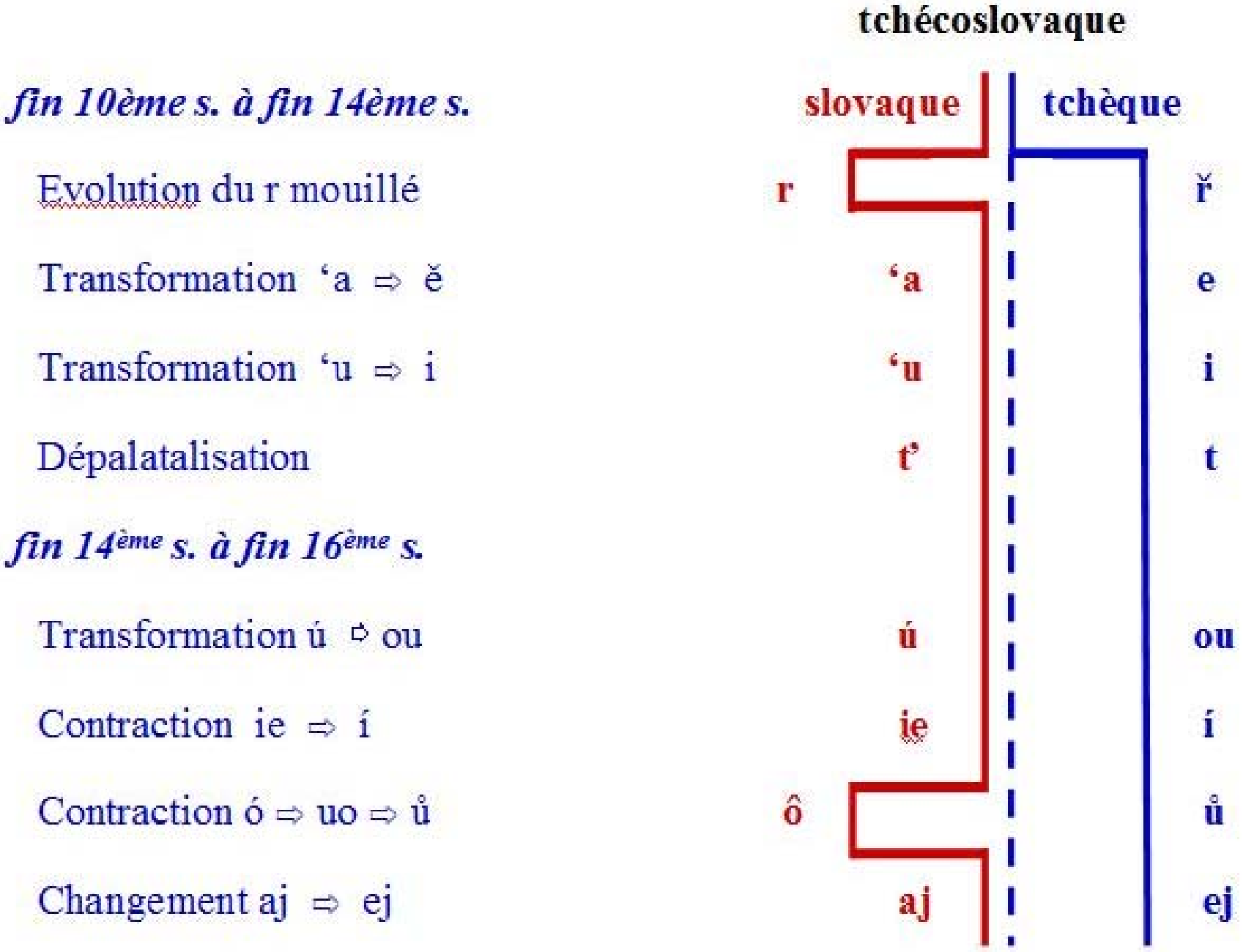
Diachronie: transformations avec la disparition du jer mou impair



Diachronie: transformations consonne molle – voyelle dure

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/i	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

Comparaisons induites grâce à la diachronie:



Comparaisons induites grâce à la diachronie:

Comparaison tchèque – slovène (interface base de données)

[1] houba goba

tchèque	slovène
"ý, y, í" ⇨ "i"	☐
"ý" ⇨ "i"	☐
"y" ⇨ "i"	☐
" í" ⇨ "i"	☐
"u, ú, (ú →) ou" ⇨ "o"	☑
"u" ⇨ "o"	☐
"ú" ⇨ "o"	☐
"(ú →) ou" ⇨ "o"	☑
"ó, ů" ⇨ "o"	☐
"ó" ⇨ "o"	☐
"ů" ⇨ "o"	☐
longueur ⇨ Ø	☐
"á" ⇨ "a"	☐
"é" ⇨ "e"	☐

tchèque	slovène
☐ **métathèse**	☐
☐ **métathèse "o" ⇨ "a"**	☐
"h" ⇦ "g"	☑
"ř" ⇨ "r"	☐
"(ě →) e" ⇦ "a"	☐
"i" ⇦ "u"	☐
"l" ⇦ "lj"	☐
☐ **depalatalisation**	☐

tchèque	slovène
"v-" ⇨ Ø	☐
"dl, tl" ⇨ "l"	☐
"ch"⇨ "h"	☐
"št" ⇔ "šč"	☐
"-t" ⇔ "-ti"	☐

La calculabilité diachronique corrobore l'existence et l'unicité du système linguistique