

## **MŮŽE BÝT FONOLOGIE UŽITEČNÁ ETYMOLOGII?**

- Ano, ale obvykle pouze diachronní fonologie, resp. procesuální fonologie (jak se hlásky mění)
- Ale co synchronní fonologie?

### **Co ze synchronní fonologie lze využít v etymologii:**

- A) Rozbor fonologické stavby určitého stavu jazyka  
např. Mathesiovo pravidlo
- B) Informace o fonologické stavbě současných jazyků  
např. vyhýbání se stejnému místu artikulace CVC
- C) Metody synchronní fonologie  
např. výpočet fonotaktické pravděpodobnosti, poměr mezi doloženou a očekávanou frekvencí

## Základní otázka etymologie: domácí či přejaté slovo?

- Intuitivní a triviální fakt: přejatá slova (PS) vypadají jinak než slova domácí (DS)
- Rozdíl se netýká jen pravopisu, ale i výslovnosti, což zahrnuje nejen výskyt fonémů, ale také stavbu slov
- PS sice podstupují adaptaci, ale stále mohou nést stopy své cizosti, byť jsou jinak zcela adaptovaná
- Budou mít totiž fonologické vlastnosti odlišné od domácích slov
- Příklad: *bažant* (2 slabiky, /nt/ na konci)

## **Slova se mohou lišit v tom**

- A) Jaké fonologické prvky obsahují
- B) Jakou mají společné fonologické prvky frekvenci

## **Dva typy fonologických prvků**

- A) Fonémy
- B) Kombinace fonémů

## **Dva druhy frekvence**

- A) Celková frekvence prvku bez ohledu na kontext/pozici
- B) Frekvence prvku v daném kontextu/pozici

## Fonologický prvek 1: Frekvence fonémů v kontextu

- Celková frekvence fonémů je příliš hrubá, a proto volíme frekvenci fonémů v nějakém kontextu
- Před a za každým fonémem se vyskytuje konsonant (C), vokál (V), nebo hranice slova (#)
- Tedy tyto možnosti:

#_#	#_C	#_V
C_#	C_C	C_V
V_#	V_C	V_V
- Kontexty nezohledňují délku slova a konkrétní kombinace fonémů

## Fonologický prvek 2: Fonémy v kombinaci (bifony)

- Kombinace fonémů mohou být různé, proto volíme bifony
- Bifon = posloupnost dvou fonémů vedle sebe
- Bifony se překrývají:
  - /koza/ : bifony /ko/, /oz/, /za/
- Bifony se vyskytují v určité pozici/pořadí
  - /ko/ = 1. bifon
  - /oz/ = 2. bifon
  - /za/ = 3. bifon
- Využívají se např. při výpočtu fonotaktické pravděpodobnosti slov a při automatické syntéze řeči

## Vyjádření rozdílu mezi PS a DS

- Inspirace obvyklou metodou užívanou ve fonologii: rozdíl mezi doloženou a očekávanou frekvencí (D/O)

Doložená frekvence fonémů pro kontext #\_C

	<i>Frekvence v PS</i>	<i>Frekvence v DS</i>	<i>Součet</i>
/k/	385	410	795
...	...	...	...
<i>Součet</i>	6 699	8 837	15 536
%	43 %	57 %	100 %

- Pokud se PS a DS od sebe neliší, bude platit stejný poměr pro jednotlivé fonémy
- Očekávaná frekvence /k/ pro PS je tedy 342,7977 (= 0,43×795)

## Poměr mezi doloženou a očekávanou frekvencí

- Poměr  $D/O$  vyjádřený procentuálně:  $(D - O)/O \times 100$
- Čím je tato hodnota větší, tím se prvek  $X$  objevuje častěji v určitém kontextu/pozici, než bychom očekávali
- Takový prvek je tedy více preferovaný v PS
- Pro příklad výše: 12,31 %

<i>Hodnota poměru (<math>x</math>)</i>	<i>Výskyt prvku v PS</i>
-1	nevyskytuje se
$-1 < x < 0$	méně často, než očekáváme
0	stejně často, jak očekáváme, <i>nebo</i> nedoložen ani v PS, ani DS
$0 < x < \infty$	častěji, než očekáváme

## Poměr mezi doloženou a očekávanou frekvencí (pokr.)

- Poměr lze vypočítat pro oba typy prvků, fonémy a bifony (= hodnota rozdílu)
- Zároveň lze vypočíst průměrnou hodnotu rozdílu pro určité slovo podle počtu prvků (= průměrná hodnota rozdílu, PHR)
- Čím je PHR větší, tím více slovo obsahuje prvků preferovaných v PS
- Dva druhy PHR: PHR-kontext a PHR-bifon
- Zajímají nás pouze případy, kdy se PHR-kontext a PHR-bifon shodují (= **míra cizosti**)

## Testovací materiál: čeština

- *Fonologický korpus češtiny*, databáze domácích slov, DS (33 566 položek) a přejatých slov, PS (36 234 položek)
- PS = slova vyskytující se ve slovnících cizích slov
- Databáze rozděleny na tyto soubory:
  - S1 = 20 000 náhodně vybraných DS
  - S2 = 20 000 náhodně vybraných PS
  - S3 = zbytek DS + zbytek PS (28 800 slov)
- S1 a S2 sloužil k výpočtu D/O pro fonémy a bifony
- Pro slova v S3 se vypočítala jejich PHR-kontext a PHR-bifon („míra cizosti“)
- Ze souboru S3 vyřazeno 8 jednohláskových slov (neobsahují bifony)

## Výsledek

- Materiál: S3, tj. 28 792 slov rozdělených na domácí a přejatá
- Predikce: u přejatých slov bude míra cizosti větší než 0; u domácích menší než 0

<i>Míra cizosti (PHR-kontext a PHR-bifon)</i>	<i>Počet slov</i>	<i>Z toho přejatých slov</i>
> 0	13 989	97 %
< 0	12 867	9 %
není shoda	2 936	59 %

## Závěr

- Zvolenou metodou lze poměrně přesně určit přejatá a domácí slova
- 97 % slov správně rozpoznáno jako přejatá
- 91 % slov správně rozpoznáno jako domácí

## Aplikace na staroslověštinu

- *Etymologický slovník jazyka staroslověšského* (ESJS) obsahuje 2 483 hesel (lemmat)
- U nich uvedeny tři typy původu:
  - A) Domáci neutrální slova zděděná z praslovanštiny
  - B) Neutrální slova přejatá z cizích jazyků (bez ohledu na dataci)
  - C) Onomatopoická, interjekcionální, lalická, expresivní slova („elementárně příbuzná“)
- Pro zjednodušení budeme A + C pokládat za domácí
- Některá slova mají nejasnou etymologii: uvažuje se jak o domácím, tak o přejatém původu
- Zohledněny pouze hlavní etymologické výklady (vyjádřené vzorci a/nebo psané větším písmem, tedy ne petitem)
- Všechny etymologie zohledněny pouze u hesel, u kterých jsou všechny výklady petitem

## Rozdělení stsl. slov podle původu

<i>Typ</i>	<i>Počet slov</i>	<i>%</i>
Domácí (S1)	1 770 <sup>[*]</sup>	71 %
Přejatá (S2)	590	24 %
Nejasná (S3)	123	5 %

[\*] Z toho 158 onomatopoeických atd. slov

- S1 a S2 použity pro výpočet D/O pro fonémy a bifony
- S3 použit pro výpočet míry cizosti

## Výsledky pro soubory S1 (domácí) a S2 (přejaté)

- Pro kontrolu byla nejprve vypočítána míra cizosti pro S1 a S2
- Predikce: u přejatých slov bude míra cizosti větší než 0; u domácích menší než 0

<i>Míra cizosti (PHR-kontext a PHR-bifon)</i>	<i>Počet slov</i>	<i>Z toho přejatých slov</i>
> 0	695	74 %
< 0	1 186	2 %
není shoda	473	12 %

## Závěry

- Predikce se potvrdila, byť ne 100%
- Ne zcela triviální výsledek (dvě metody)
- Přesto existuje velký počet PS, která jsou rozpoznána jako domácí
- Nicméně čím větší míra cizosti, tím větší pravděpodobnost, že je slovo přejaté

<i>Míra cizosti (PHR-kontext a PHR-bifon)</i>	<i>Počet slov</i>	<i>Z toho přejatých slov</i>
> 0	695	74 %
> 0,1	542	83 %
> 0,2	394	88 %
> 0,3	265	90 %

## Výsledky pro soubor S3 (slova s nejasnou etymologií)

- 123 slov
- Míra cizosti  $> 0$ : 32 slov (pravděpodobnost cizího původu 74 %?)
- Míra cizosti  $> 0,3$ : 6 slov (pravděpodobnost cizího původu 90 %?)
- Míra cizosti  $< 0$ : 50 slov (pravděpodobnost cizího původu 2 %?)

## Závěry

- Představená metoda může sloužit pro automatickou detekci přejatých slov
- Metoda nemůže a ani nemá nahradit jiné, etymologické způsoby určení původu
- Je to však ukázka toho, že fonologická analýza synchronní stavby jazyka může být pro etymologii prospěšná
- A může se stát podpůrným argumentem tehdy, kdy tradiční etymologické metody nejsou s to jednoznačně určit původ slova

## Další možnosti využití synchronní fonologie

- Vyhýbání se stejnému místu artikulace  
neutrální × příznaková slova  
morfémy × hranice morfémů
- Zvukový symbolismus:  
/r/ se vyskytuje ve slovech spojených s významem „drsnost“  
(332 jazyků z 84 jaz. rodin; statisticky signifikantní; diachronně stabilní pro IE jazyky)