

**Explanace  
morfematické struktury  
českého stupňování  
na základě frekvence tvarů  
v Českém národním korpusu**

**Jan Křivan**

jan.krivan@seznam.cz

*Workshop Žďárek,  
Hořovice, 27. listopadu 2009*

# Hlavní body příspěvku

- **explanace** – odpovědi na otázky, proč určité jazykové struktury vypadají právě tak, jak vypadají;
  - předpokladem explanace jsou dobré **deskripce** a **generalizace** (zákonitosti, limity variability, univerzálie)
  - je snaha najít věrohodná vysvětlení, proč a za jakých okolností dochází na jedné straně k podobnostem ve strukturách různých jazyků, na druhé straně čím jsou naopak způsobeny rozdíly. Funkční typologové tyto principy nacházejí **ve způsobu užívání jazyka** (tzv. *usage-based approach*), nikoliv např. uvnitř gramatické struktury samotné.
- **Příklad:**
  - otázka: *Proč je v různých jazycích (včetně češtiny) tendence, že morfologicky značený superlativ je formálně složitější (delší) než komparativ, a ten zase složitější než pozitiv?*
  - odpověď na tuto otázku = explanace, v závěru prezentace.
- **práce s jazykovým korpusem**
  - budou představeny některé výsledky frekvenční analýzy českého stupňování
- **vliv frekvence na strukturu českého stupňování**
  - průběžně bude ukázáno, že formální struktura (její pravidelnosti i nepravidelnosti) českého stupňování může být vhodně vysvětlitelná **frekvencí**

# Explanační principy I – ikonicita

## Ikonicita (ikoničnost)

- ikonicita jako princip vysvětluje formální strukturu jazykových jednotek podle její podobnosti se strukturou významovou; přijímá se předpoklad, že formální struktura odráží strukturu významovou
- tyto podobnosti jsou různého charakteru, rozlišují se:
  - **ik. syntagmatického izomorfismu** (tendence, aby jedné formě odpovídal jeden význam);
  - **ik. susedství** (formy, které k sobě přináležejí sémanticky, se vyskytují vedle sebe)
  - **ik. pořadí** (pořadí forem odpovídá pořadí zkušenosti; vlastně linearita)
  - **ik. opakování** (opakování formy signalizuje opakování významu, např. reduplikace).
  - \* **ik. kvantity** (větší množství významu je vyjádřeno větším množstvím formy, např. *pozitiv – komparativ – superlativ* v latině, ale i češtině)
  - \* **ik. komplexity** (složitější významy jsou vyjádřeny složitějšími formami, pro stupňování platí obdobně)
  - \* **ik. koheze** (významy, které jsou si bližší sémanticky, jsou vyjádřeny více kohesními formami)
- doplňujícím principem k ikonicitě je **příznakovost**
- ikonicita se často klade do protikladu k **arbitrárnosti**

# Explanační principy II – frekvence

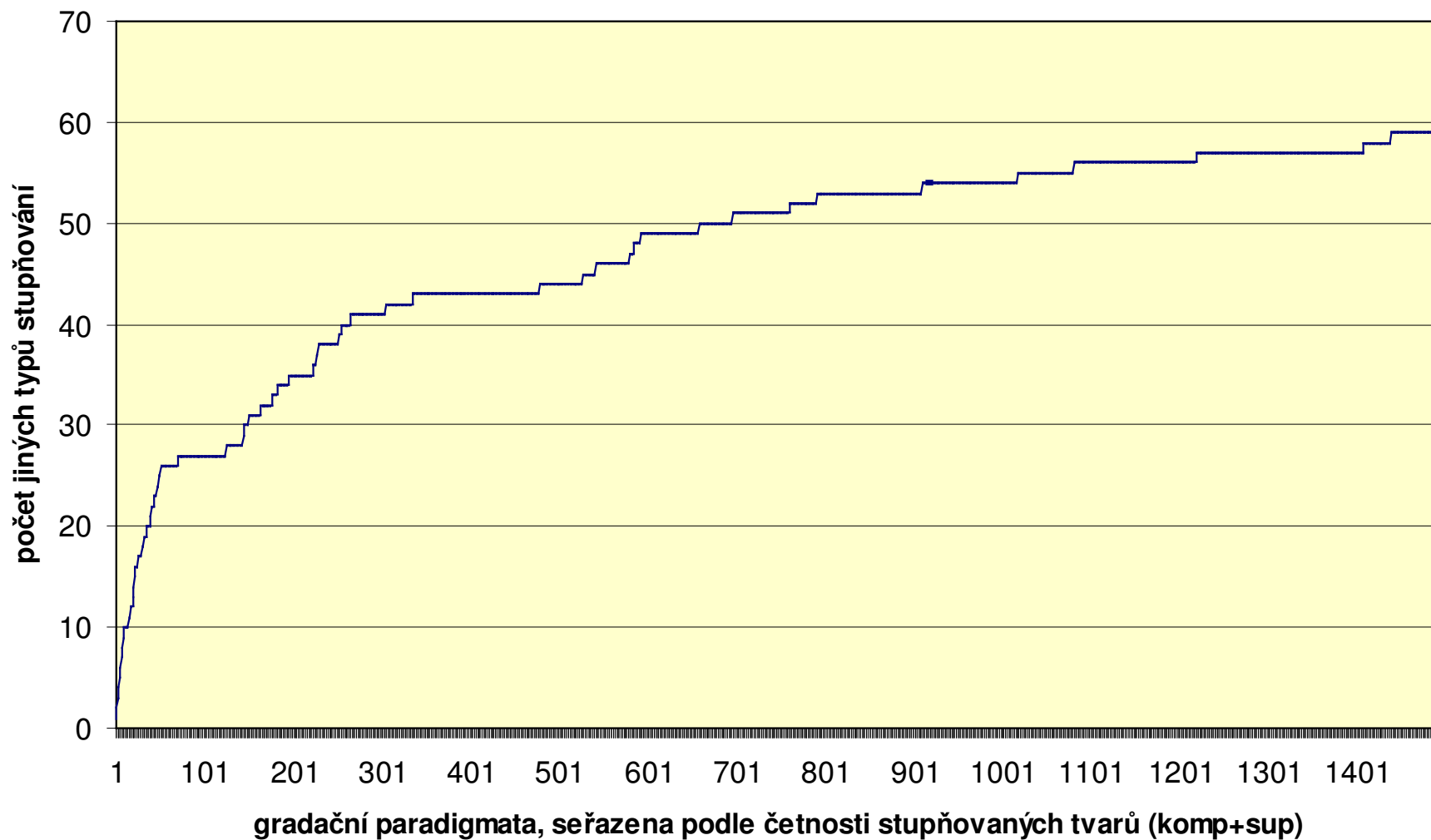
## Frekvence

- frekvence umožňuje vysvětlovat průběh diachronních změn (a synchronní dynamický stav);
- od 70. let 20. století: přínos **J. Bybee** (2006, 2007).

## Rozlišení token × type:

- frekvence **tokenová**
  - počet výskytů jednotky v textu, např. největší frekvenci ze všech komparativů má v češtině komparativ od *velký*, tedy *větší*
- frekvence **typová**
  - týká se pouze kategorií (jazykových vzorců); počet jednotek, které jsou reprezentovány daným vzorcem; např. komparativ utvořený příponou *-ější* má v češtině nejvyšší typovou frekvenci (v SYN > 99,95 % všech gradací); typová frekvence je jedním z faktorů **produktivity**

## Počet jiných typů stupňování než -ější v závislosti na pořadí gradačních paradigmat podle absol. četnosti stupňovaných tvarů



Pouhých **60** stupňovacích paradigmat jiného typu představuje v korpusu SYN **>1,6 mil. výskytů**, zatímco zbylých **>5300** stup. paradigmat typu -ější představuje jen **900 tis. výskytů**

# Explanační principy III – frekvence

## Rozlišení absolutní × relativní

- frekvence **absolutní**
  - např. nejvíce superlativů je utvořeno od adjektiva *velký*: *největší*
  - efekty absolutní tokenové frekvence: redukující, konzervující, autonomní
- frekvence **relativní**
  - poměr absolutní frekvence jedné kategorie vůči absolutní frekvenci jiné kategorie;
  - např. relativní frekvence superlativu vůči komparativu; pro adj. *různý*: *nejrůznější* je v SYN celkem 32155 výskytů, *různější* pouze 9. > Relativní frekvence superlativu vůči komparativu je u *různý* výrazně větší než u jiných adjektiv.
- podle **M. Haspelmatha** (2008a, b) lze frekvencí vhodně vysvětlit i struktury, které se jindy vysvětlují ikonicitou (viz výše ozn. *hvězdičkou*);
  - Podle něj vždy platí, že **kratší, jednodušší a více kohezní formy se vyskytují signifikantně častěji než** odpovídající **delší, složitější a méně kohezní formy**.
  - **Relativní frekvencí** lze vysvětlit některé **asymetrie**. Např. v súdanském jazyce *Krongo* je klasická distinkce nezačtený *sg* × začtený *pl*. Ale pro některá jména (např. ‚*list*‘) je *pl* nezačtený a *sg* je začten zvláštním singulativním afixem. Proč? U těchto jmen je relativní frekvence *pl* : *sg* vyšší než u jiných jmen.

# Efekty tokenové frekvence (Bybee, 2006, 2007)

## Redukující efekt

- vysokofrekvenční slova podstupují fonetickou redukci.
- např.: na místo tvaru *stařejší* vznikla vývojem redukovaná podoba > *starší*; tyto podoby se však nevytvořily u méně frekventovaných adjektiv

## Konzervující efekt

- vysokofrekvenční nepravidelné struktury mají tendenci nepodléhat vlivu produktivních vzorců; jsou automatizované, uživatelé jazyka je mají tzv. pevně zakořeněné, *entrenched*.
- např. v češtině se zachovává nepravidelné stupňování *velký-větší-největší*; *nízký-nižší-nejnižší*, ale i výše uvedené *starý-starší-nejstarší* apod.

## Autonomie

- extrémní případ konzervujícího efektu: některá slova zcela ztratí svoji původní strukturu a oddělí se od původního paradigmatu, vzniká supletivismus.
- u č. stupňování *lepší* původně (v psl.) komparativ k *lepší*.
  - brzký – dřívější – nejdřívější (? brzcí/brzčejší)

# Ukázka konzervujícího efektu I

- **neproduktivní nepravidelné** (supletivní nebo původně redukované) stupňované formy:

SYN, poř.			pozit.	komp.	sup.	komp. + sup.
1	velký	větší	559170	180896	238780	<b>419 676</b>
2	dobry	lepší	254145	115872	166102	<b>281 974</b>
4	malý	menší	274988	78920	23967	<b>102 887</b>
7	špatný	horší	85711	36748	29925	<b>66 673</b>
3	vysoký	vyšší	197525	122063	108259	<b>230 322</b>
32	hluboký	hlubší	35459	9062	3200	<b>12 262</b>
6	nizký	nižší	52387	61429	17787	<b>79 216</b>
8	blizký	bližší	46554	13633	46416	<b>60 049</b>
44	uzký	užší	23574	6065	1794	<b>7 859</b>
10	dlouhý	delší	131874	41129	6334	<b>47 463</b>
15	široký	širší	46471	21022	4271	<b>25 293</b>
20	těžký	težší	99989	11235	10176	<b>21 411</b>



# Ukázka konzervujícího efektu II

– zakončení *b-ý* (původní adjektiva;  
odvozená – např. *dlouhodobý* – se stupňují pravidelně)

poř.			pozit.	komp.	sup.	komp. + sup.	„komp“ v „googlu“
22	slabý	slabší	29502	13968	4006	<b>17974</b>	517 000 / 858
		<i>slabější</i>		2	1	<b>3</b>	2 500 / 264
144	hrubý	hrubší	42526	731	582	<b>1313</b>	62 800 / 897
		<i>hrubější</i>		28	13	<b>41</b>	11 200 / 570
	chabý	chabší	2500	0	0	<b>0</b>	618 / 112
1523		<i>chabější</i>		17	2	<b>19</b>	1 390 / 220

**s klesající absolutní frekvencí stupňovaných tvarů roste relativní podíl produktivního typu -ější**

tuto tendenci lze sledovat i u dalších zakončení

# Některá zjištění nad daty

- nejčastější gradační typ je vzorec s příponou *-ější*
- podle fonologického okolí lze vymezit určité typy – uzavřené třídy, které se skloňují zcela pravidelně nezávisle na frekvenci:
  - 1) zcela bezvýjimečné je stupňování adjektiv zakončených na *-chý*
    - *jednoduchý, tichý, suchý, plochý, plachý* (komp. *š-ší*)
  - 2) podobně bezvýjimečné je stupňování adjektiv zakončených na *-d-ký*
    - *sladký, hladký, řídký, prudký* (komp. *d-ší*)
  - 3) produktivní je stupňování s kmenovou změnou *k > č* a příponou *-í*. Varianty s příponou *-ejší* jsou zcela okrajové (v korpusu od každého jen 0–4 doklady).
    - *lehký, hezký, měkký, tenký, křehký, vlhký, mělký, trpký, křepký, hebký*
    - Mluvnice u některých těchto jmen sice povolují i komparativ *-ější*, ale leckdy se při vyhledávání přes google najdou tyto tvary jen v rámci testů a učebnic...
  - 4) adjektiva zakončená na *-st-ý* se z > 90 % stupňují s příponou *-ší*. Ale zachovávají si vždy i variantu *-ější*; relativní frekvence *-ější* stoupá až u těch nejméně častých (pustý)
    - *čistý, hustý, prostý, tlustý, sprostý, pustý* (ne odvozené *častý*)

# Otázka

- Proč je superlativ formálně složitější a delší než komparativ?

## Explanace frekvencí:

- Protože výskyt superlativů je v češtině méně častý než výskyt komparativů.
- Spočítáme-li výskyt všech komparativů a všech superlativů, je poměr
  - komparativ : superlativ 53,2 % : 46,8 % (což je signifikantní)
- Spočítáme-li průměr relativních frekvencí pro jednotlivé gradační jednotky, vychází:
  - komparativ : superlativ 64 % : 36 % (což je signifikantní)

Tvoření superlativu je zcela pravidelné, proto je namístě kategoriální pohled.

[Ovšem všiml jsem si navrch této tendence: existují-li dvě varianty stupňování, např. *nejsprostější/nejsprostší*, v superlativu se použije ta kratší častěji než v komparativu. Proč?]

# Závěry

- frekvenční analýza poskytuje vhled do utváření jazykových kategorií: frekvencí lze vysvětlovat diachronní vývoj
- ukazuje se, co je v jazyce (v dané kategorii) pevně zakotvené, a co nikoliv
- ale frekvence nemůže být jediným principem, který vysvětlí každou gramatickou strukturu

## Literatura:

Bybee, Joan. (2006). From usage to grammar: the mind's response to repetition.

*Language* 82(4). s. 711–733.

Bybee, Joan. (2007). Introduction. In: Bybee. *Frequency of use and the organization of language*. Oxford: Oxford University Press, s. 5–22.

Haspelmath, M. (2008a). Frequency vs. iconicity in explaining grammatical asymmetries.

*Cognitive Linguistics* 19, vol. 1, s. 1–33.

Haspelmath, M. (2008b). *Creating economical morphosyntactic patterns in language change*. In: Good, Jeff (ed.) *Language universals and language change*. Oxford: Oxford University Press, s. 185–214.

**Díky za pozornost.**

Jan Křivan  
*jan.krivan@seznam.cz*