

# Slovotvorná analýza

PLIN059

Mgr. Dana Hlaváčková, Ph.D.

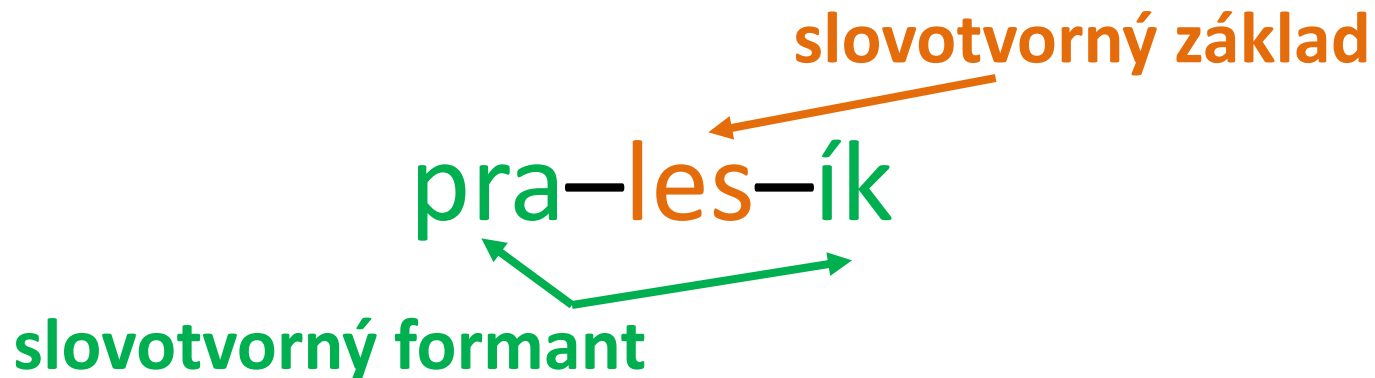
Mgr. Jakub Machura, Ph.D.

# Slovotvorná analýza

- segmentace slova na
  1. (slovotvorný) základ
  2. (slovotvorný) formant
- chceme získat/zjistit
  - základové slovo
  - postup a způsob, jak vzniklo slovo nové

# Slovotvorná analýza

- segmentace slova na
  1. (slovotvorný) základ
  2. (slovotvorný) formant
- chceme získat/zjistit
  - základové slovo
  - postup a způsob, jak vzniklo slovo nové



# Tvoření slov v češtině

- **derivace** – odvozování, tvoření slov pomocí slovotvorných formantů (kombinace morfémů)
- **kompozice** – skládání na úrovni slovních základů
  - konekt – *život-o-pis, knih-tisk*
- **abreviace** – zkracování

# Segmentace slova

- **slovotvorný základ** (kmen, kořen)
  - celé základové slovo: do-dělat > děla
  - slovní tvar: zmizel-ý > zmizel
  - kmen: sladk-ost > sladký
  - kořen: slad-it > sladký

# Segmentace slova

- slovotvorný základ (kmen, kořen)
- **afixi:**
  - prefix: před základem slova (*při-jít*)
  - sufix: za základem slova (*učí-tel*)
  - infix: (*mal-il-inký*)
  - interfix (*hlad-o-mor*)
  - koncovka – součástí sufixu, sama sufixem, mění se při flexi (*zdrav-í*)
  - postfix – za koncovkou, při flexi se nemění (*jaký-si, jakého-si*)

# Slovotvorné postupy

- **derivace**
  - prefixace/deprefixace – připojení prefixu (*dělat – při-dělat*)
  - sufixace/desufixace – připojení sufixu (*tábor-ník*)
  - resufixace – náhrada jednoho sufixu druhým (*vlas-atý – vlas-áč*)
  - postup **prefixačně-sufixační** (*při-hloup-lý*)
  - diskrétní formant *se, si*
- **hláskové alternace v kmenovém základu** – změna kvality konsonantů, změna kvantity vokálů
  - *plech – plíšek, brát – branec*

# Typologie změn – změna významu nebo slovního druhu

- **mutace** – významová změna a změna slovního druhu  
(*bílit* – *bělidlo*; *vrtat* – *vrtáčka*)
- **transpozice** – pouze změna slovního druhu bez významové směny  
(*bílit* – *bílení*; *starý* – *stáří*)
- **modifikace** – změna významového příznaku  
(*bílit* – *obílit*; *mrak* – *mráček*)



# Počítačové zpracování

- slovo a jeho segmenty = řetězce znaků
- **derivace** – připojení/odpojení, záměna části řetězce (prefix, sufix)
- při automatickém zpracování nezáleží na směru převodu (slovo základové – odvozené)
- možnost pracovat s hláskovými alternacemi
- **slovotvorná Ajka** (CZPJ)
  - zachyceny jednoduché případy derivace (*učit – učitel, boj – bojiště*)
  - bez alternací, významově chybné dvojice (*míč – mít*)

# Deriv

- webové rozhraní – **Deriv**
- autor Pavel Šmerk (1. verze Radek Sedláček)
- generování seznamů slov (prefix, sufix)
- generování dvojic slov (až n-tic) na základě přidání, odebrání, záměny prefixu/sufixu
- omezení pomocí morfologické značky (atributivní systém)
- hláskové alternace – regulární výrazy
- propojení s DEBDictem, KonTextem a SketchEnginem (SYN2000, czTenTen12)
- **Derivanče** (derivační analyzátor češtiny)
- <https://nlp.fi.muni.cz/projekty/derivance/>

# DeriNet

- <http://ufal.mff.cuni.cz/derinet/search>
- vizualizace slovotvorných vztahů a slovotvorného hnízda
- Zdeněk Žabokrtský, Magda Ševčíková, Jonáš Vidra
- vyhledává v 1 mil. slovních jednotek z MorfFlexu
- není zahrnuto skládání slov
- podporováno korpusovými daty
- zapojen morfologický analyzátor, derivační a gramatická pravidla

# Morfio

- ÚČNK, <https://morfio.korpus.cz/>
- autoři – Václav Cvrček, Pavel Vondříčka
- vyhledávání přímo v korpusu  
[SYN2005/SYN2010/SYN2015/Araneum](#)
- společná báze, specifické formanty
- omezení pomocí morfologické značky (poziční systém)
- možnost hledání alternací (uživatelské rozhraní)

# Úkoly

Z čeho je odvozeno substantivum *knihovna*?

Vzniklo sloveso *černat* z adjektiva *černý*, nebo naopak?