**CJBB75**

 **29. 4. Je možné kombinovat empirický výzkum a introspekci**

Máme kousek porušeného textu (dvě slova, z nichž známe buď poslední písmeno, nebo poslední dvě písmena). Které z následujících dvojic mohou/nemohou být českými větami (jeden z členů dvojice je verbum finitum). Dokažte to.

a)

|  |  |
| --- | --- |
|  ý |  ůl |

b)

|  |  |
| --- | --- |
|  š |  ě |

c)

|  |  |
| --- | --- |
|  í |  m |

d)

|  |  |
| --- | --- |
|  u |  č |

e)

|  |  |
| --- | --- |
|  t |  r |

f)

|  |  |
| --- | --- |
|  é |  ou |

g)

|  |  |
| --- | --- |
|  p |  s |

Důkaz proveďte 1) empiricky na základě analýzy dat z korpusu a 2) analyticky na základě znalostí české gramatiky.

Řešení:

Máme kousek porušeného textu (dvě slova, z nichž známe buď poslední písmeno, nebo poslední dvě písmena). Které z následujících dvojic mohou/nemohou být českými větami (jeden z členů dvojice je verbum finitum). Dokažte to.

Např.

|  |  |
| --- | --- |
|  s |  á |

|  |  |
| --- | --- |
|  ***pe***s |  ***štěk***á |

a)

|  |  |
| --- | --- |
|  ý |  ůl |

b)

|  |  |
| --- | --- |
|  š |  ě |

c)

|  |  |
| --- | --- |
|  í |  m |

d)

|  |  |
| --- | --- |
|  u |  č |

e)

|  |  |
| --- | --- |
|  t |  r |

f)

|  |  |
| --- | --- |
|  é |  ou |

g)

|  |  |
| --- | --- |
|  p |  s |

Důkaz proveďte 1) empiricky na základě analýzy dat z korpusu a 2) analyticky na základě znalostí české gramatiky.

1. V korpusu vyhledáme všechny sekvence tvaru (Implicitní atribut - word) *.\*ý .\*ůl .* Získáme tak dvojice jako ***zlatý důl, kulatý stůl, okřídlený vůl***. P-filtrem budeme hledat dvojice takové, že jeden z členů bude (Implicitní atribut – pos) sloveso. Buď *. V*, nebo *V .* . V obou případech nenalezneme ani jeden doklad. Empiricky jsme dokázali, že v korpusu není ani jeden doklad posloupnosti vyhledaných slovních tvarů taková, že by jeden z členů dvojice byl slovesný tvar (tudíž nemůže být ani verbum finitum).
2. Český slovesný tvar určitý (ani neurčitý) nemůže končit na příslušné řetězce. Určité tvary mohou končit pouze na vokál *–a, -á, -e, -ě, -i, -í, -y, -o, -u (-ou)* a na téměř všechny konsonanty (*?x, ?q, ?w*). Pokud končí na *–l*, pak se může jednat buď o l-ové příčestí (ve 3. os. je tvarem verb. fin.), pak ale mohou před –l předcházet z vokálů pouze *[aáeěiyu]*, nebo o imperativ, pak ovšem před *-l* může předcházet pouze krátký vokál (*val,vel, šil,vol, bul, nemyl se*).
3. V korpusu vyhledáme všechny sekvence tvaru (Implicitní atribut - word) *.\*š .\*ě .* Získáme tak dvojice jako *příliš pozdě, příliš vážně,* ale i *chceš, mě, máš ještě*. P-filtrem budeme hledat dvojice takové, že jeden z členů bude (Implicitní atribut – pos) sloveso. Buď *. V*, nebo *V .* . V korpusu nenalezneme doklady (*Miloš nechtě, ...*) (*chceš, mě, máš ještě*). Empiricky jsme dokázali, že v korpusu se vyskytují posloupnosti vyhledaných slovních tvarů takové, že první z členů dvojice je slovesný tvar (navíc i verbum finitum). Případy opačné v korpusu doloženy nejsou (chyby morf. značkování). Přechodník není verbum finitum.
4. Český slovesný tvar určitý může ovšem končit na příslušné řetězce. Určité tvary mohou končit jak na *–š* (2. os. sg. ind. préz. akt., 2. os. imperativu), tak na *–*ě (pouze 3. os. sg. ind. préz. akt. slovesa *stonat*).
5. V korpusu vyhledáme všechny sekvence tvaru (Implicitní atribut - word) *.\*í .\*m .* Získáme tak dvojice jako (***mají zájem, není problém***)*,* ale i (*domácí tým, informační centrum*). P-filtrem budeme hledat dvojice takové, že jeden z členů bude (Implicitní atribut – pos) sloveso. Buď *. V*, nebo *V .* . V korpusu nenalezneme doklady (***mají zájem, není problém***) (***další prosím***). Empiricky jsme dokázali, že v korpusu se vyskytují posloupnosti vyhledaných slovních tvarů takové, že první z členů dvojice je slovesný tvar (navíc i verbum finitum) i případy opačné.
6. Český slovesný tvar určitý může ovšem končit na příslušné řetězce. Určité tvary mohou končit jak na *–í* (3. os. sg./pl. ind. préz. akt), tak na *–m* (pouze 1. os. sg. ind. préz. akt. nebo 2. os. sg. imperativu).
7. V korpusu vyhledáme všechny sekvence tvaru (Implicitní atribut - word) *.\*u .\*č .* Získáme tak dvojice jako (***jsou pryč, tužku strč***)*,* ale i (***dřevěnou tyč, jízdou napříč***). P-filtrem budeme hledat dvojice takové, že jeden z členů bude (Implicitní atribut – pos) sloveso. Buď *. V*, nebo *V .* . V korpusu nenalezneme doklady (***jedou pryč***) (***chvilku plač***). Empiricky jsme dokázali, že v korpusu se vyskytují posloupnosti vyhledaných slovních tvarů takové, že první z členů dvojice je slovesný tvar (navíc i verbum finitum) i případy opačné.
8. Český slovesný tvar určitý může ovšem končit na příslušné řetězce. Určité tvary mohou končit jak na *–u* (3. os. sg./pl. ind. préz. akt), tak na *–č* (2. os. sg. imperativu).
9. V korpusu vyhledáme všechny sekvence tvaru (Implicitní atribut - word) *.\*t .\*r .* Získáme tak dvojice jako (***ksicht sežer, jest režisér***)*,* ale i (***dávat pozor, dvacet liber***). P-filtrem budeme hledat dvojice takové, že jeden z členů bude (Implicitní atribut – pos) sloveso. Buď *. V*, nebo *V .* . V korpusu nenalezneme doklady (***ksicht sežer***) (***jest režisér***). Empiricky jsme dokázali, že v korpusu se vyskytují posloupnosti vyhledaných slovních tvarů takové, že první z členů dvojice je slovesný tvar (navíc i verbum finitum) i případy opačné.
10. Český slovesný tvar určitý může ovšem končit pouze na *–r* (2. os. sg. imperativu). Na *–t* může končit jeden jediný český slovesný tvar určitý, a to arch. tvar 3. os. sg. ind. préz. akt. slovesa *být - jest*.
11. V korpusu vyhledáme všechny sekvence tvaru (Implicitní atribut - word) *.\*é .\*ou .* Získáme tak dvojice jako (***lidé nepřijdou***)*,* ale i (***také jednou***). P-filtrem budeme hledat dvojice takové, že jeden z členů bude (Implicitní atribut – pos) sloveso. Buď *. V*, nebo *V .* . V korpusu nenalezneme doklady (***ženské nepojedou***). Empiricky jsme dokázali, že v korpusu se vyskytují posloupnosti vyhledaných slovních tvarů takové, že druhý z členů dvojice je slovesný tvar (navíc i verbum finitum), nikoli případy opačné.
12. Český slovesný tvar určitý nemůže ovšem končit na řetězec *-é*. Určité tvary mohou končit na *–ou* (3. os. pl. ind. préz. akt).
13. V korpusu vyhledáme všechny sekvence tvaru (Implicitní atribut - word) *.\*p .\*s .* Získáme tak dvojice jako (***chlap zhas, nákup vodnes, kup voves*** )*,* ale i (***postup přes***). P-filtrem budeme hledat dvojice takové, že jeden z členů bude (Implicitní atribut – pos) sloveso. Buď *. V*, nebo *V .* . V korpusu nenalezneme doklady (***kup******voves***). Empiricky jsme dokázali, že v korpusu se vyskytují posloupnosti vyhledaných slovních tvarů takové, že druhý z členů dvojice je slovesný tvar (navíc i verbum finitum) i případy opačné (***nákup vodnes***) (některé jsou chyby ve značkování *p okus = pokus*).
14. Český slovesný tvar určitý může ovšem končit na příslušné řetězce. Určité tvary mohou končit jak na *–p,* tak na *–s* (2. os. sg. imperativu), na *–s* mohou navíc končit tvary s volným morfémem *–s* (***stop napsals***). Na oba konsonantu mohou končit nespisovné tvary l-ovýchpříčestí mask. sg. (*zhas, nakop*)

**Úkol na 6. 5. 2015**

**Najděte (empiricky) 10 případů takových, že dvě slova za sebou končí na trojici shodných písmen a zároveň jde o základní skladebnou dvojici ve tvaru substantivum+verbum finitum.**

**Příklad: <Slůně/NNNS1.\* stůně/V......3.\*>.**

**Popište, jak jste postupovali. Pokud nedokážete nalézt doklady empiricky, použijte korpus a popište, jak jste postupovali.**

**Použijte korpusu a vyhledejte k empiricky zjištěným 10 případům, jaké další bigramy značek mohou mít takto zakončené dvojice tvarů.**

**Příklad: Kolkolem <vůně/NNF.\* vůně/NNF.\*> a ženy alkohol čpí Děti si sliní prst zkouší odkud vítr vane Kdo ví ?**