

Eduard Bejček

Propojování valenčních slovníků

Seminář formální a aplikované lingvistiky
19. října 2015, ÚFAL, Praha

Ručně zpracované slovníky

- drahé na výrobu
- náročné na údržbu
- stále užitečné
- mnoho existujících
- žádný slovník neobsahuje veškerou informaci

Propojování lexikografických zdrojů

- spojení
 - spojuje výhody obou
 - oba slovníky obohacuje
 - testuje konsistenci, porovnává přístupy
- automatické spojení
 - levné, opakovatelné

Přehled práce

- Valence
- VALLEX a PDT-Vallex
 - ukázka, rozdíly, historie
- Postup porovnávání slovníků
- Společný formát
- Postup porovnávání lexikálních jednotek
- Příklad
- Vyhodnocení a závěr

Valence

- schopnost plnovýznamového slova otevřít určitý počet pozic pro syntakticky závislé členy

Na včerejší demonstraci štvál politik občany proti uprchlíkům .

Na včerejší demonstraci štvál politik policii na odpůrce marně .

Na včerejší demonstraci štvál politik občany [svou xenofobií] .

Na včerejší demonstraci štvál policii na politikovi jeho rasismus .

Valence

- schopnost plnovýznamového slova otevřít určitý počet pozic pro syntakticky závislé členy

[LOC] ACT₁ PAT₄ EFF_{proti+3}
Na včerejší demonstraci štvál politik občany proti uprchlíkům .

Na včerejší demonstraci štvál politik policii na odpůrce marně .

Na včerejší demonstraci štvál politik občany [svou xenofobií] .

Na včerejší demonstraci štvál policii na politikovi jeho rasismus .

Valence

- schopnost plnovýznamového slova otevřít určitý počet pozic pro syntakticky závislé členy



Na včerejší demonstraci štval politik občany [svou xenofobií] .

Na včerejší demonstraci štval policii na politikovi jeho rasismus .

Valence

- schopnost plnovýznamového slova otevřít určitý počet pozic pro syntakticky závislé členy

[LOC]		ACT ₁	PAT ₄	EFF _{proti+3}	
Na včerejší demonstraci	štval	politik	občany	proti uprchlíkům .	~ podněcovat proti někomu
[LOC]		ACT ₁	PAT ₄	[EFF _{na+4}] [MANN]	
Na včerejší demonstraci	štval	politik	policii	na odpůrce marně .	~ pronásledovat, hnát
[LOC]		ACT ₁	PAT ₄	[MEANS ₇]	
Na včerejší demonstraci	štval	politik	občany	[svou xenofobií] .	
Na včerejší demonstraci štval policii na politikovi jeho rasismus .					

Valence

- schopnost plnovýznamového slova otevřít určitý počet pozic pro syntakticky závislé členy

[LOC]		ACT ₁	PAT ₄	EFF _{proti+3}	
Na včerejší demonstraci	štval	politik	občany	proti uprchlíkům .	~ podněcovat proti někomu
[LOC]		ACT ₁	PAT ₄	[EFF _{na+4}] [MANN]	
Na včerejší demonstraci	štval	politik	policii	na odpůrce marně .	~ pronásledovat, hnát
[LOC]		ACT ₁	PAT ₄	[MEANS ₇]	
Na včerejší demonstraci	štval	politik	občany	[svou xenofobií] .	~ zlobit, dopalovat
[LOC]		PAT ₄	[REG _{na+4}]	ACT ₁	
Na včerejší demonstraci	štval	policii	na politikovi	jeho rasismus .	

Valence

- schopnost plnovýznamového slova otevřít určitý počet pozic pro syntakticky závislé členy

[LOC]		ACT ₁	PAT ₄	EFF _{proti+3}	
Na včerejší demonstraci	štval	politik	občany	proti uprchlíkům .	~ podněcovat proti někomu
[LOC]		ACT ₁	PAT ₄	[EFF _{na+4}] [MANN]	
Na včerejší demonstraci	štval	politik	policii	na odpůrce marně .	~ pronásledovat, hnát
[LOC]		ACT ₁	PAT ₄	[MEANS ₇]	
Na včerejší demonstraci	štval	politik	občany	[svou xenofobií] .	~ zlobit, dopalovat
[LOC]		PAT ₄	[REG _{na+4}]	ACT ₁	
Na včerejší demonstraci	štval	policii	na politikovi	jeho rasismus .	~ zlobit, dopalovat

Přehled práce

- Valence
- **VALLEX a PDT-Vallex**
 - ukázka, rozdíly, historie
- Postup porovnávání slovníků
- Společný formát
- Postup porovnávání lexikálních jednotek
- Příklad
- Vyhodnocení a závěr

VALLEX a PDT-Vallex

- dva slovníky valence pro češtinu, vycházejí z FGD, vytvořené na ÚFALu, uložené v XML
- musí to být snadné

VALLEX a PDT-Vallex

- dva slovníky **valence** pro **češtinu**, vycházejí z **FGD**, vytvořené na **ÚFALu**, uložené v **XML**
- musí to být snadné

Sloveso *rozdat*

rozdat-1 **ACT**₁ **ADDR**_{3,mezi+4} **PAT**₄

rozdělit, dát darovat

rozdat mezi děti bonbony;

rozdat dětem sešity

rozdat-1 **ACT**₁ **PAT**₄ **ADDR**₃

dát

rozdat majetek chudým

Sloveso *rozdat*

rozdat-1 ACT₁ ADDR_{3,mezi+4} PAT₄

rozdělit, dát darovat

rozdat mezi děti bonbony;

rozdat dětem sešity

rozdat-1 ACT₁ PAT₄ ADDR₃

dát

rozdat majetek chudým

rozdat-2 ACT₁ PAT₄ ADDR₃

postupně každému dát, co mu náleží

rozdat vysvědčení / karty ke hře;

rozdat žákům testy

Sloveso obracet

ACT(1) PAT(4) +^DIR()

otáčet; působit změnu polohy / orientace; měnit směr
obracet auto / kolonu / seno / loď na bok; o. skříň ke zdi

ACT(1) PAT(4) EFF(k+3, na+4, v+4)

proměňovat

obracel nepřátele v prach; obracel pohany
na křesťanství; obracel lidi k životu

ACT(1) PAT(k+3,na+4) DPHR(pozornost,zájem,zřetel)

zaměřovat

obracet pozornost / zájem / zřetel

ACT(1) PAT(4) MANN()

převracet

obracel vše vzhůru nohama / naruby

ACT(1)

měnit mínění

Pavel najednou rychle obracel

ACT(1) PAT(4) ?EFF(v+4)

měnit

chabá koruna obracela vývoj obchodní bilance;

jeho slova vypjaté diskuse vždy obracela v prudké polemiky

ACT(1) PAT(4) EFF(v+4)

obracejí naše výroky v pravý opak

ACT(1) PAT(4) DPHR(naruby)

obracet vše naruby

ACT(1) CPHR([pozornost,...].4) DIR3(*)

obracet pozornost jinam

ACT(1) PAT(4)

obracet skříň;o. auto;ke zdi.DIR3

ACT(1) PAT(4) ADDR(na+4)

obracela svůj vztek na
dávného přítele

Sloveso obracet

ACT(1) PAT(4) +^DIR()

otáčet; působit změnu polohy / orientace; měnit směr
obracet auto / kolonu / seno / loď na bok; o. skříň ke zdi

ACT(1) PAT(4) EFF(k+3, na+4, v+4)

proměňovat

obracel nepřátele v prach; obracel pohany
na křesťanství; obracel lidi k životu

ACT(1) PAT(k+3,na+4) DPHR(pozornost,zájem,zřetel)

zaměřovat

obracet pozornost / zájem / zřetel

ACT(1) PAT(4) MANN()

převracet

obracel vše vzhůru nohama / naruby

ACT(1)

měnit mínění

Pavel najednou rychle obracel

ACT(1) PAT(4) ?EFF(v+4)

měnit

chabá koruna obracela vývoj obchodní bilance;

jeho slova vypjaté diskuse vždy obracela v prudké polemiky

ACT(1) PAT(4) EFF(v+4)

obracejí naše výroky v pravý opak

ACT(1) PAT(4) DPHR(naruby)

obracet vše naruby

ACT(1) CPHR([pozornost,...].4) DIR3(*)

obracet pozornost jinam

ACT(1) PAT(4)

obracet skříň;o. auto;ke zdi.DIR3

ACT(1) PAT(4) ADDR(na+4)

obracela svůj vztek na
dávného přítele

Sloveso obracet

ACT(1) PAT(4) +^DIR()

otáčet; působit změnu polohy / orientace; měnit směr
obracet auto / kolonu / seno / loď na bok; o. skříň ke zdi

ACT(1) PAT(4) EFF(k+3, na+4, v+4)

proměňovat

obracel nepřátele v prach; obracel pohany
na křesťanství; obracel lidi k životu

ACT(1) PAT(k+3, na+4) DPHR(pozornost, zájem, zřetel)

zaměřovat

obracet pozornost / zájem / zřetel

ACT(1) PAT(4) MANN()

převracet

obracel vše vzhůru nohama / naruby

ACT(1) —○

měnit mínění

Pavel najednou rychle obracel

ACT(1) PAT(4) ?EFF(v+4)

měnit

chabá koruna obracela vývoj obchodní bilance;
jeho slova vypjaté diskuse vždy obracela v prudk

ACT(1) PAT(4) EFF(v+4)

obracejí naše výroky v pravý opak

ACT(1) PAT(4) DPHR(naruby)

obracet vše naruby

**ACT(1) CPHR([pozornost,...].4)
DIR3(*)**

obracet pozornost jinam

ACT(1) PAT(4)

obracet skříň; o. auto; ke zdi. DIR3

mezianotátorská shoda průměrně 92 %

(na slovesech se šesti LU jen 76 %)

VALLEX a PDT-Vallex: historie

- valenční slovník součástí FGP
- práce začaly na začátku tohoto století
 - VALLEX zhruba 2001, podrobně, pomalu
 - nestačil potřebám rozrůstajícího se projektu PDT 2.0
 - PDT-Vallex zhruba 2001: nejprve jen jako seznam použitých rámců
- proto vznikají oba slovníky paralelně:
VALLEX „do hloubky“, PDT-Vallex „do šířky“

VALLEX a PDT-Vallex: rozdíly dnes

VALLEX 2.5	PDT-Vallex 2.0
4.800 slovesných lemmat	7.100 slovesných lemmat
11.200 lexikálních jednotek (LU)	11.900 lexikálních jednotek (LU)
kompletně zpracovaná nejčastější slovesa	valenční rámce, které se objevují v PDT / PCEDT
doplňující syntaktické informace (reciprocita, kontrola, reflexivita, ...)	provázání lexikálních jednotek s daty PDT / PCEDT
popisuje celé lexémy	varianty a vidové protějšky zvlášť i substantiva, adjektiva, adverbia
pouze slovesa	

- shodné/společné:
 - vychází z FGP
 - 3500 slovesných lemmat
 - XML formát

VALLEX a PDT-Vallex: rozdíly dnes

VALLEX 2.5	PDT-Vallex 2.0
4.800 slovesných lemmat	7.100 slovesných lemmat
11.200 lexikálních jednotek (LU)	11.900 lexikálních jednotek (LU)
kompletně zpracovaná nejčastější slovesa	valenční rámce, které se objevují v PDT / PCEDT
doplňující syntaktické informace (reciprocita, kontrola, reflexivita, ...)	provázání lexikálních jednotek s daty PDT / PCEDT
popisuje celé lexémy	varianty a vidové protějšky zvlášť
pouze slovesa	i substantiva, adjektiva, adverbia

- shodné/společné:
 - vychází z FGP
 - 3500 slovesných lemmat
 - XML formát

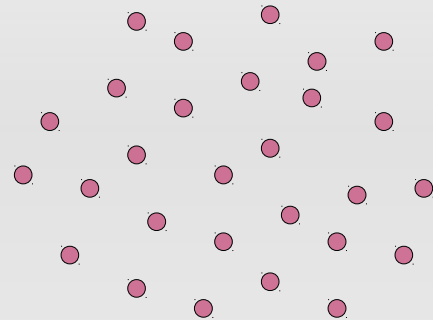
Přehled práce

- Valence
- VALLEX a PDT-Vallex
 - ukázka, rozdíly, historie
- **Postup porovnávání slovníků**
- Společný formát
- Postup porovnávání lexikálních jednotek
- Příklad
- Vyhodnocení a závěr

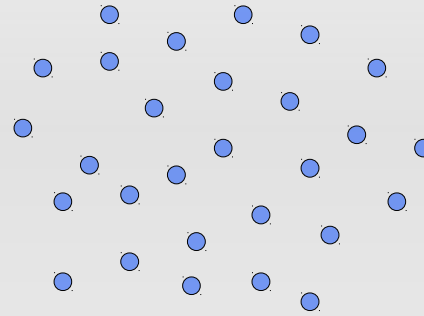
Propojení VALLEXu a PDT-Vallexu

slovesa

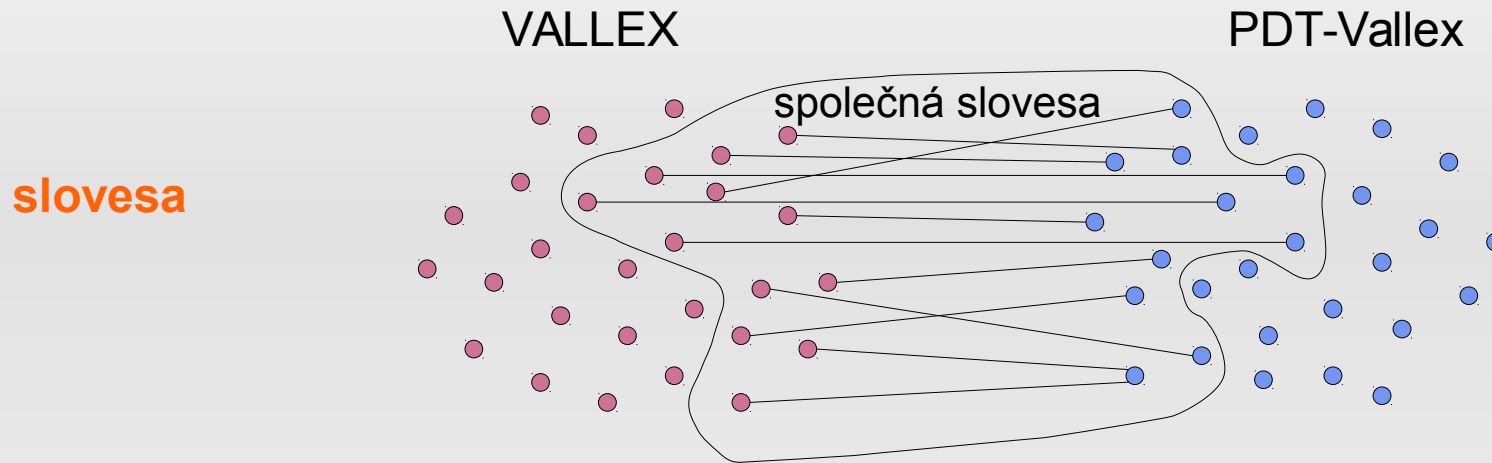
VALLEX



PDT-Vallex

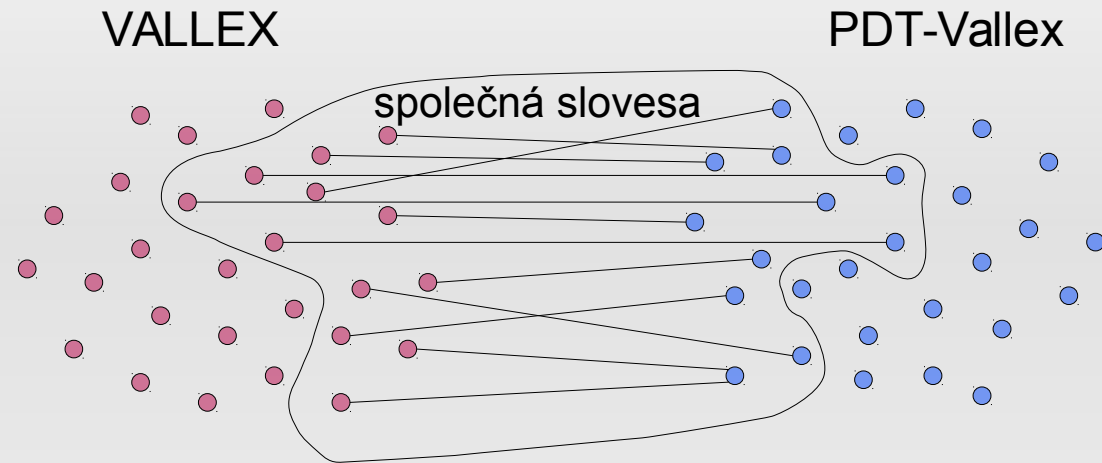


Propojení VALLEXu a PDT-Vallexu



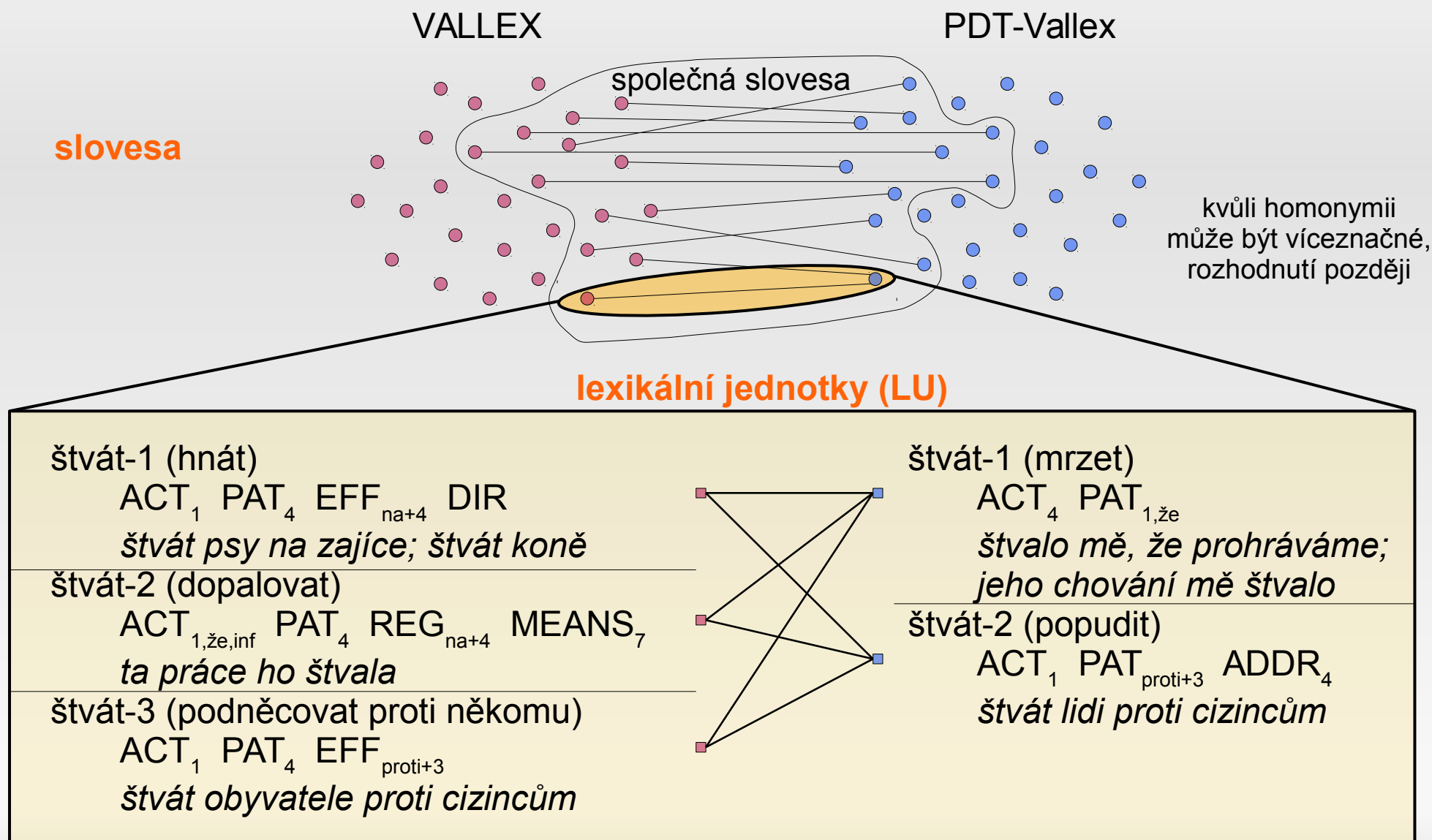
Propojení VALLEXu a PDT-Vallexu

slovesa

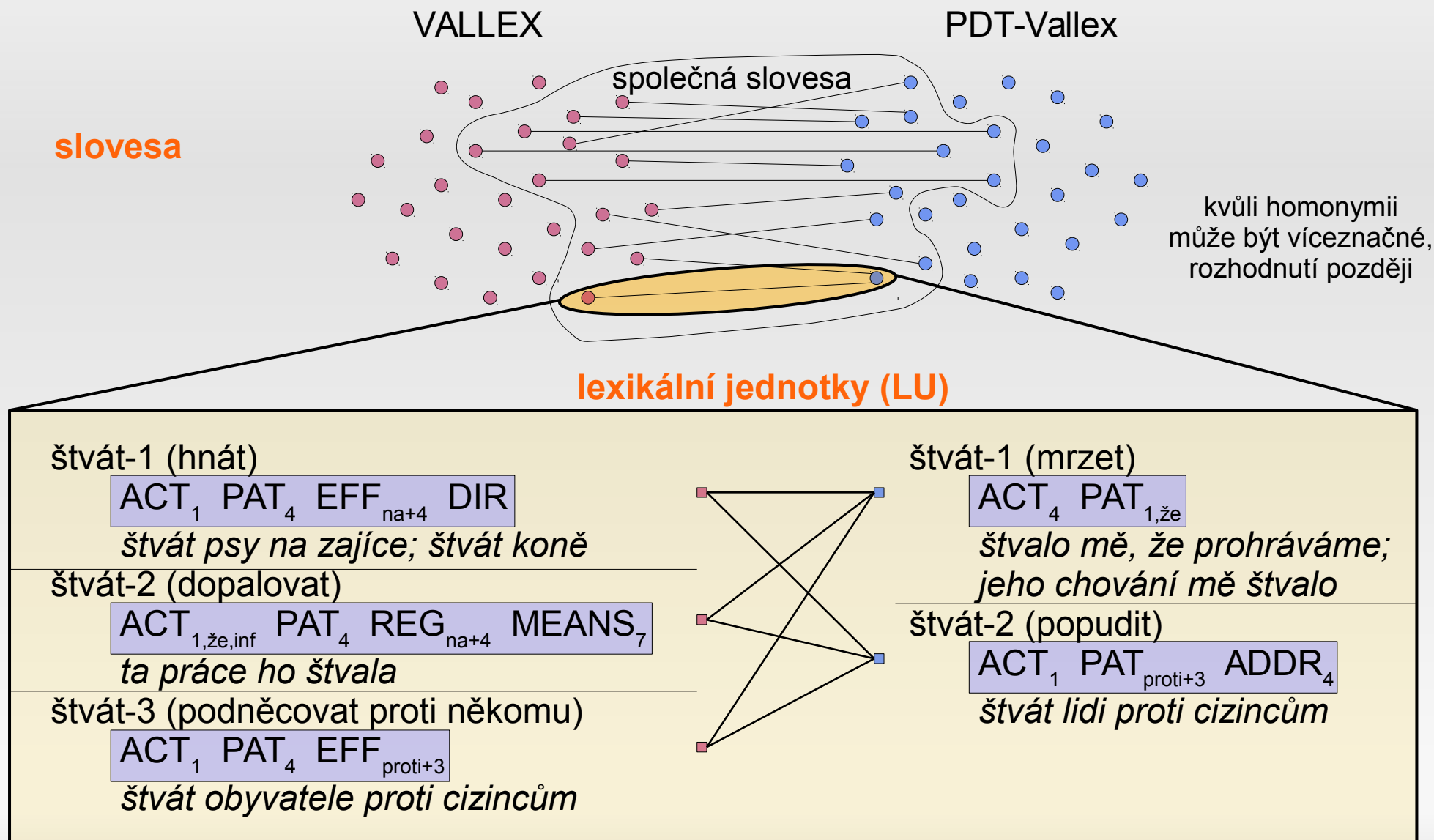


kvůli homonymii
může být víceznačné,
rozhodnutí později

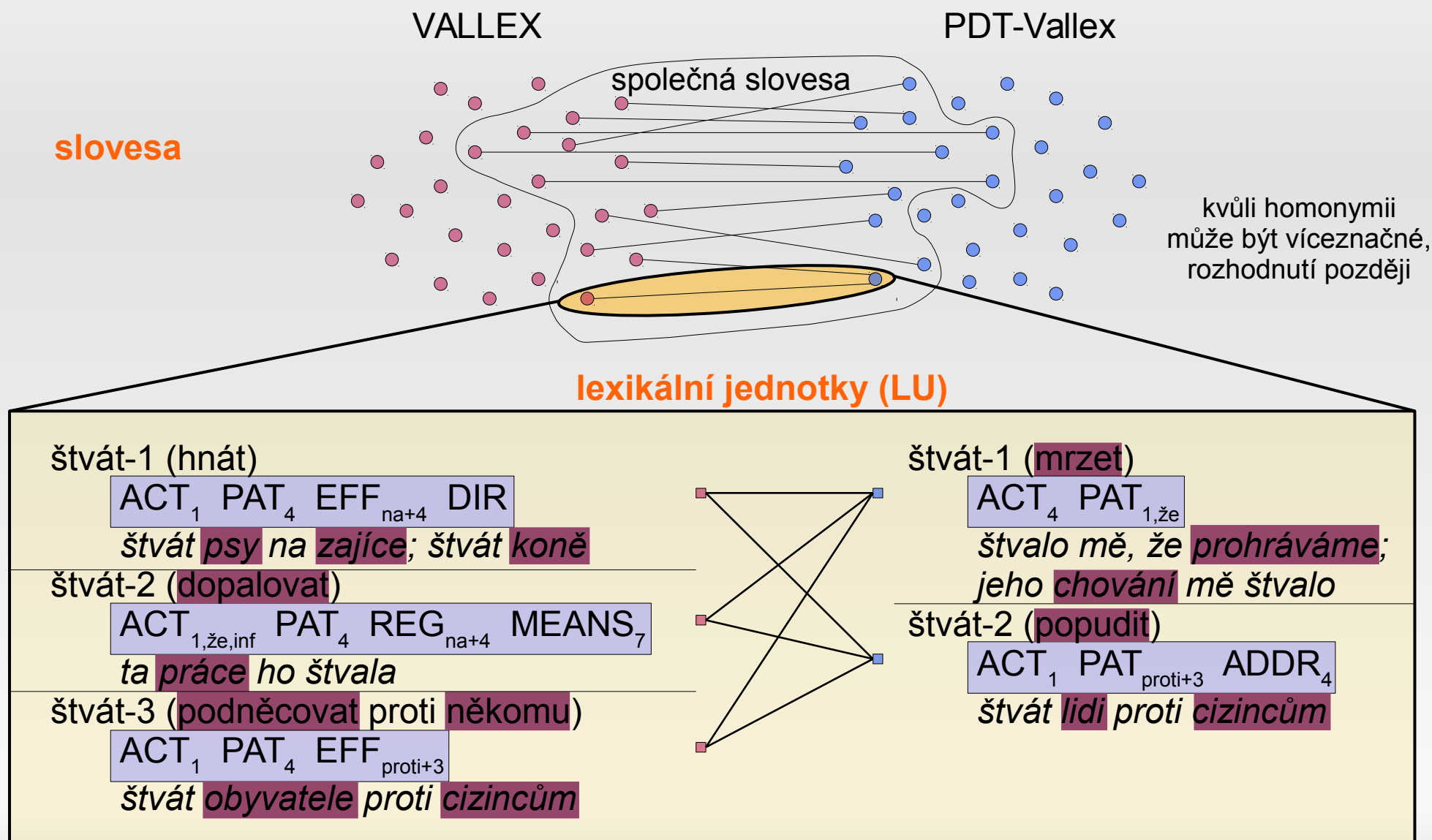
Propojení VALLEXu a PDT-Vallexu



Propojení VALLEXu a PDT-Vallexu



Propojení VALLEXu a PDT-Vallexu



Přehled práce

- Valence
- VALLEX a PDT-Vallex
 - ukázka, rozdíly, historie
- Postup porovnávání slovníků
- **Společný formát**
- Postup porovnávání lexikálních jednotek
- Příklad
- Vyhodnocení a závěr

Formáty obou slovníků

- oba XML
- PDT-Vallex: lemmata
- VALLEX: lexémy (sdružují varianty i vidy)
 - tj. pohromadě např.
vystříhovat^{impf1}, *vystříhávat* / *vystříhávat*_i^{impf2},
vystříhnout^{pf1}, *vystříhat* / *vystříhat*_i^{pf2},
vystříhovávat^{iter}
 - *vystříhat*_{||} se je zvlášť stejně jako v PDT-Vallexu
 - většina LU společná, některé specifické

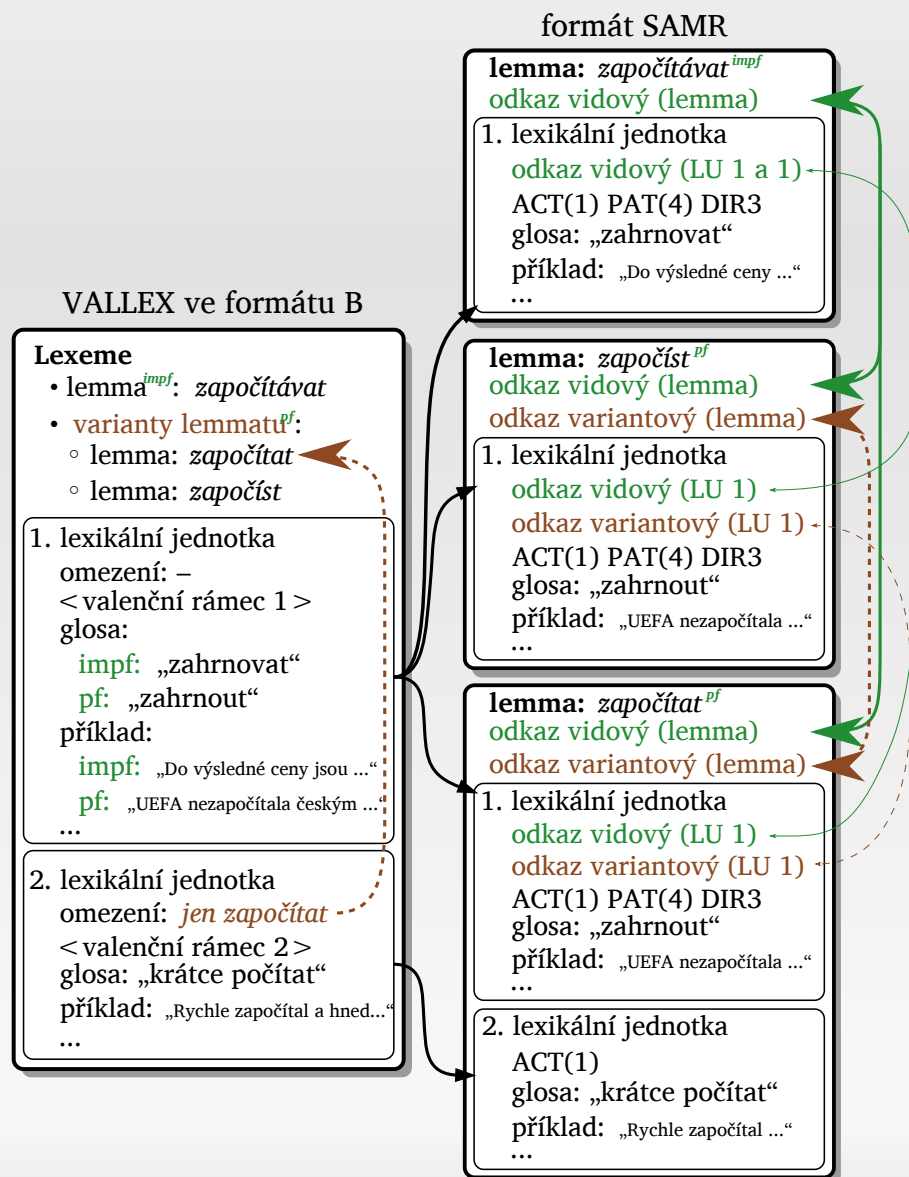
Formáty obou slovníků

- morfemické varianty doplnění
 - VALLEX: čitelný výpis
 - proti+3, po zádech
 - PDT-Vallex: struktura a-roviny
 - .3[proti], mráz.S1 po-1[záda:P6]

Společný formát SAMR

- bližší PDT-VALLEX_u
- VALLEX rozgenerovaný podle jednotlivých lemmat
 - pomocí odkazů (mezi lemmaty i mezi LU) zachována informace o vidových protějšcích i o variantách
- zápis forem není (snadno) převeditelný
 - ponechán odlišně pro každý slovník
 - informace se zanedbá teprve při zpracování

VALLEX transformovaný do SAMR



V PDT-Vallexu byly
potřeba jen drobné změny

Universální formáty slovníků

- Linguistic Annotation Framework
- Morpho-syntactic Annotation Framework
- Lexical Markup Framework
- meta-formáty, dají se uzpůsobit různým účelům
- zdá se, že se příliš nepoužívají
- pro naše účely též neefektivní (malý užitek, hodně práce s přizpůsobením – ta nebyla třeba)

Přehled práce

- Valence
- VALLEX a PDT-Vallex
 - ukázka, rozdíly, historie
- Postup porovnávání slovníků
- Společný formát
- **Postup porovnávání lexikálních jednotek**
- Příklad
- Vyhodnocení a závěr

Propojení VALLEXu a PDT-Vallexu

VALLEX

lexikální jednotky (LU)

PDT-Vallex

štvát-1 (hnát)

ACT₁ PAT₄ EFF_{na+4} DIR

štvát psy na zajíce; štvát koně

štvát-2 (dopalovat)

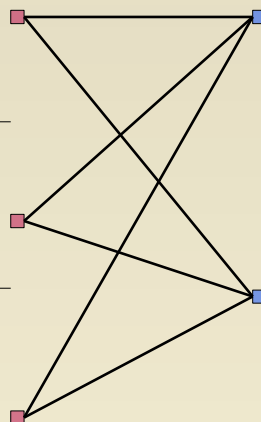
ACT_{1,že,inf} PAT₄ REG_{na+4} MEANS₇

ta práce ho štvála

štvát-3 (podněcovat proti někomu)

ACT₁ PAT₄ EFF_{proti+3}

štvát obyvatele proti cizincům



štvát-1 (mrzet)

ACT₄ PAT_{1,že}

*štválo mě, že prohráváme;
jeho chování mě štválo*

štvát-2 (popudit)

ACT₁ PAT_{proti+3} ADDR₄

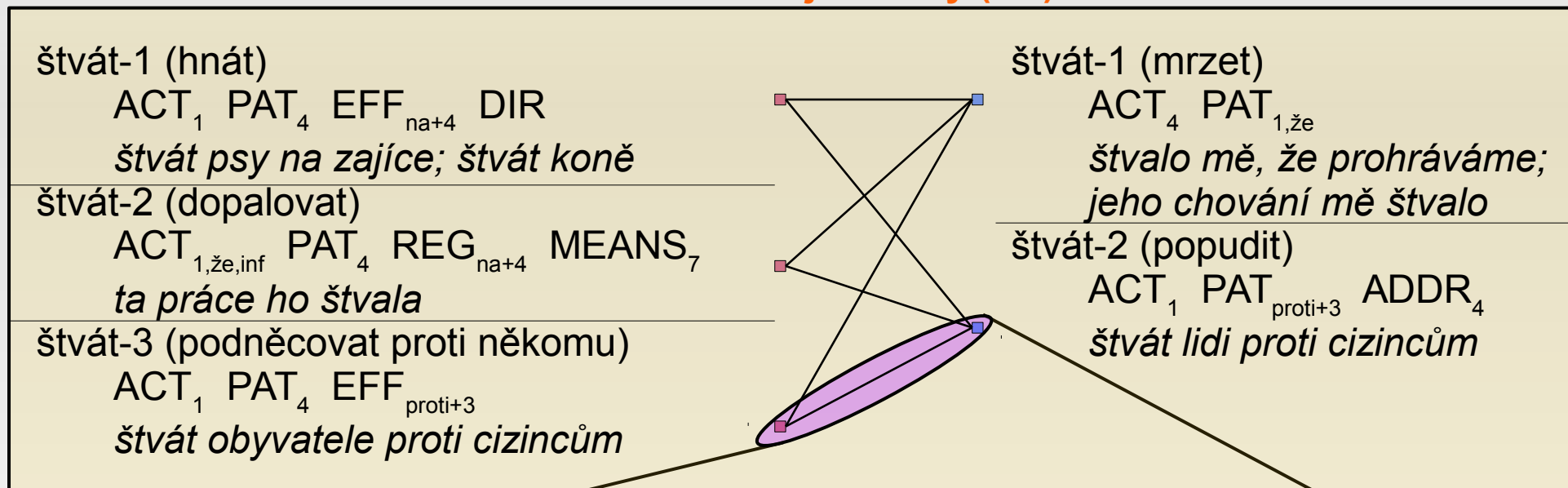
štvát lidi proti cizincům

Propojení VALLEXu a PDT-Vallexu

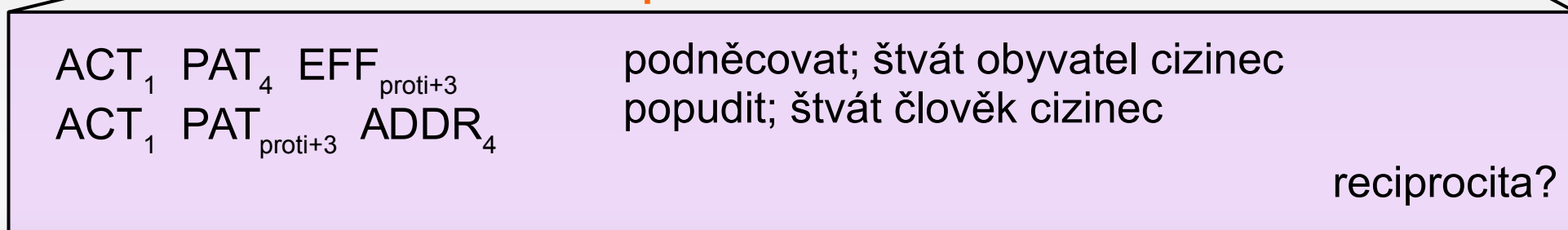
VALLEX

PDT-Vallex

lexikální jednotky (LU)



společné informace o LU

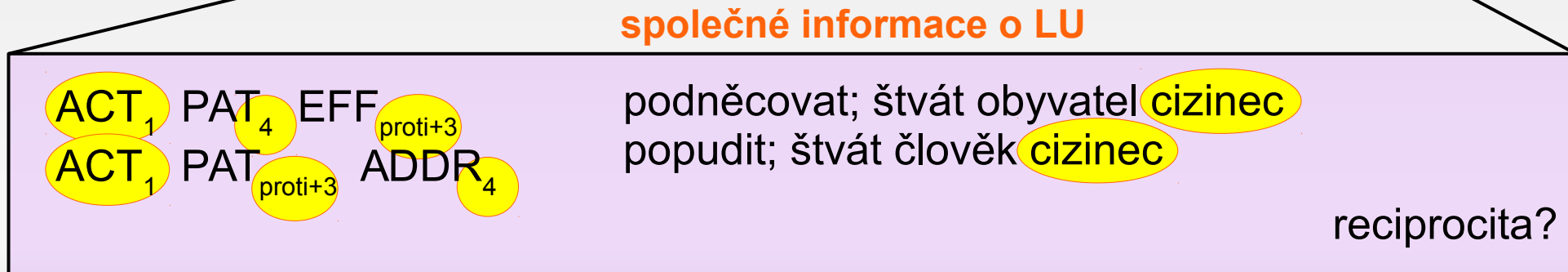
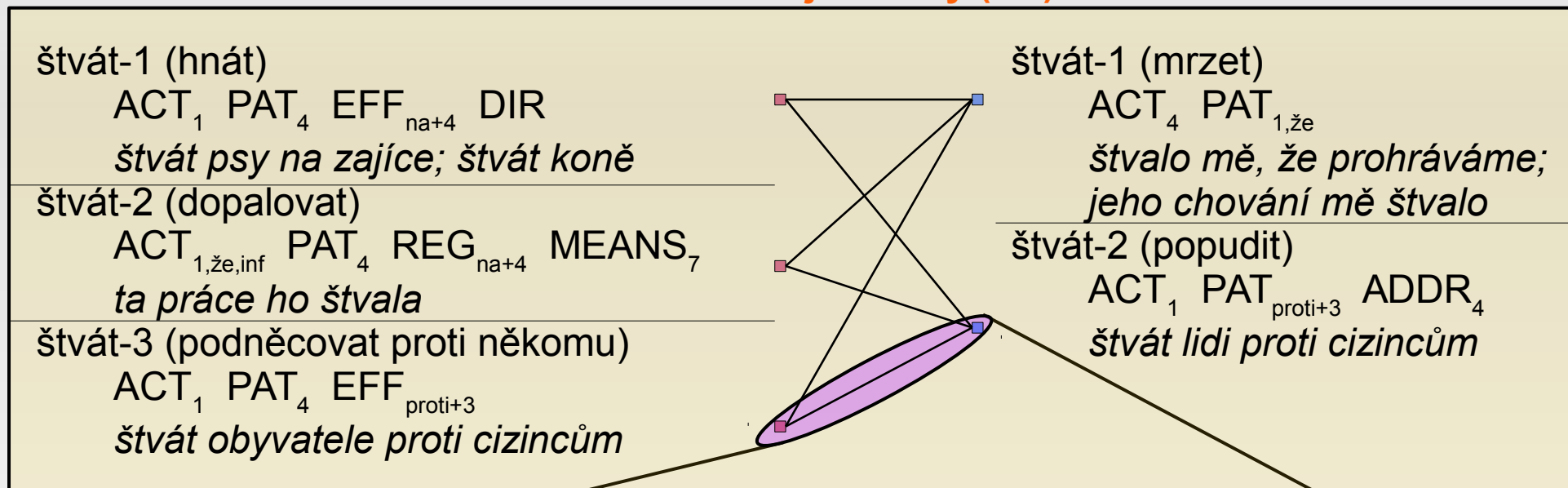


Propojení VALLEXu a PDT-Vallexu

VALLEX

PDT-Vallex

lexikální jednotky (LU)

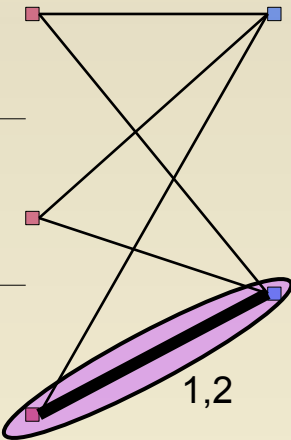


Propojení VALLEXu a PDT-Vallexu

VALLEX

lexikální jednotky (LU)

PDT-Vallex

štvát-1 (hnát) ACT₁ PAT₄ EFF_{na+4} DIR <i>štvát psy na zajíce; štvát koně</i> štvát-2 (dopalovat) ACT_{1,že,inf} PAT₄ REG_{na+4} MEANS₇ <i>ta práce ho štvála</i> štvát-3 (podněcovat proti někomu) ACT₁ PAT₄ EFF_{proti+3} <i>štvát obyvatele proti cizincům</i>		štvát-1 (mrzet) ACT₄ PAT_{1,že} <i>štválo mě, že prohráváme; jeho chování mě štválo</i> štvát-2 (popudit) ACT₁ PAT_{proti+3} ADDR₄ <i>štvát lidi proti cizincům</i>
---	---	--

- ohodnocení páru podle podobnosti – přidělení *Score*
- porovnání každých dvou lexikálních jednotek

Propojení VALLEXu a PDT-Vallexu

VALLEX

lexikální jednotky (LU)

PDT-Vallex

štvát-1 (hnát) ACT₁ PAT₄ EFF_{na+4} DIR <i>štvát psy na zajíce; štvát koně</i>		štvát-1 (mrzet) ACT₄ PAT_{1,že} <i>štválo mě, že prohráváme; jeho chování mě štválo</i>
štvát-2 (dopalovat) ACT_{1,že,inf} PAT₄ REG_{na+4} MEANS₇ <i>ta práce ho štvála</i>		štvát-2 (popudit) ACT₁ PAT_{proti+3} ADDR₄ <i>štvát lidi proti cizincům</i>
štvát-3 (podněcovat proti někomu) ACT₁ PAT₄ EFF_{proti+3} <i>štvát obyvatele proti cizincům</i>		

- ohodnocení páru podle podobnosti – přidělení *Score*
- porovnání každých dvou lexikálních jednotek
- odstranění párů, které nepřesáhnou práh
- ze zbytku páry s nejvyšším *Score* hladovým algoritmem...

Propojení VALLEXu a PDT-Vallexu

VALLEX

lexikální jednotky (LU)

PDT-Vallex

štvát-1 (hnát) ACT ₁ PAT ₄ EFF _{na+4} DIR <i>štvát psy na zajíce; štvát koně</i>		štvát-1 (mrzet) ACT ₄ PAT _{1,že} <i>štválo mě, že prohráváme; jeho chování mě štválo</i>
štvát-2 (dopalovat) ACT _{1,že,inf} PAT ₄ REG _{na+4} MEANS ₇ <i>ta práce ho štvála</i>		štvát-2 (popudit) ACT ₁ PAT _{proti+3} ADDR ₄ <i>štvát lidi proti cizincům</i>
štvát-3 (podněcovat proti někomu) ACT ₁ PAT ₄ EFF _{proti+3} <i>štvát obyvatele proti cizincům</i>		

- ohodnocení páru podle podobnosti – přidělení *Score*
- porovnání každých dvou lexikálních jednotek
- odstranění párů, které nepřesáhnou práh
- ze zbytku páry s nejvyšším *Score* hladovým algoritmem...
...aby pokud možno každá LU patřila do právě jednoho páru

Propojení VALLEXu a PDT-Vallexu

VALLEX

lexikální jednotky (LU)

PDT-Vallex

štvát-1 (hnát) ACT ₁ PAT ₄ EFF _{na+4} DIR <i>štvát psy na zajíce; štvát koně</i>		štvát-1 (mrzet) ACT ₄ PAT _{1,že} <i>štválo mě, že prohráváme; jeho chování mě štválo</i>
štvát-2 (dopalovat) ACT _{1,že,inf} PAT ₄ REG _{na+4} MEANS ₇ <i>ta práce ho štvála</i>		štvát-2 (popudit) ACT ₁ PAT _{proti+3} ADDR ₄ <i>štvát lidi proti cizincům</i>
štvát-3 (podněcovat proti někomu) ACT ₁ PAT ₄ EFF _{proti+3} <i>štvát obyvatele proti cizincům</i>		

rozhodnutí
o homonymech

- ohodnocení páru podle podobnosti – přidělení *Score*
- porovnání každých dvou lexikálních jednotek
- odstranění párů, které nepřesáhnou práh
- ze zbytku páry s nejvyšším *Score* hladovým algoritmem...
...aby pokud možno každá LU patřila do právě jednoho páru

Porovnávání valenčních rámců

- oba valenční rámce převedeny do stejné, lineární formy
- sada pravidel určujících, které dvojice valenčních rámců jsou dostatečně podobné
 - ve formě regulárních výrazů (pro dvojici rámců)
 - použití závorkování, předdefinovaných proměnných
 - různá důvěryhodnost podle komplexnosti rámců
- přes 30 pravidel

Přehled práce

- Valence
- VALLEX a PDT-Vallex
 - ukázka, rozdíly, historie
- Postup porovnávání slovníků
- Společný formát
- Postup porovnávání lexikálních jednotek
- **Příklad**
- **Vyhodnocení a závěr**

Ukázka pravidla – úpravy rámce

ACT(1) PAT(4, že, aby, .f) ADDR(3)
?EFF(.4[za-1], .4[{jako, jakožto}:/AuxY])

ACT(1) ADDR(3) PAT(4, inf, aby, at', zda, že) ?EFF(jako+4, za+4)

Ukázka pravidla – úpravy rámce

ACT(1) PAT(4, že, aby, .f) ADDR(3)
?EFF(.4[za-1], .4[jako :/AuxY])

ACT(1) ADDR(3) PAT(4, inf, aby, at', zda, že) ?EFF(jako+4, za+4)

původní stav:

ACT(1) PAT(4, že, aby, .f) ADDR(3)
?EFF(.4[za-1], .4[{jako, jakožto}:/AuxY])

ACT(1) ADDR(3) PAT(4, inf, aby, at', zda, že) ?EFF(jako+4, za+4)

Ukázka pravidla – úpravy rámce

ACT(1) PAT(4, že, aby, .f) ADDR(3)
?EFF(.4[za], .4[jako :/AuxY])

ACT(1) ADDR(3) PAT(4, inf, aby, at', zda, že) ?EFF(jako+4, za+4)

původní stav:

ACT(1) PAT(4, že, aby, .f) ADDR(3)
?EFF(.4[za-1], .4[{jako, jakožto}:/AuxY])

ACT(1) ADDR(3) PAT(4, inf, aby, at', zda, že) ?EFF(jako+4, za+4)

Ukázka pravidla – úpravy rámce

ACT(1) PAT(4, že, aby, .f) ADDR(3)
?EFF(za+4 , jako+4)

ACT(1) ADDR(3) PAT(4, inf, aby, at', zda, že) ?EFF(jako+4, za+4)

původní stav:

ACT(1) PAT(4, že, aby, .f) ADDR(3)
?EFF(.4[za-1], .4[{jako, jakožto}:/AuxY])

ACT(1) ADDR(3) PAT(4, inf, aby, at', zda, že) ?EFF(jako+4, za+4)

Ukázka pravidla – úpravy rámce

ACT(1) PAT(4, že, aby, inf) ADDR(3)
 ?EFF(za+4 , jako+4)
ACT(1) ADDR(3) PAT(4, inf, aby, at', zda, že) ?EFF(jako+4, za+4)

původní stav:

ACT(1) PAT(4, že, aby, .f) ADDR(3)
 ?EFF(.4[za-1], .4[{jako, jakožto}:/AuxY])
ACT(1) ADDR(3) PAT(4, inf, aby, at', zda, že) ?EFF(jako+4, za+4)

Ukázka pravidla – úpravy rámce

ACT(1) PAT(4, že, aby, inf) ADDR(3) ?EFF(za+4, jako+4)

ACT(1) ADDR(3) PAT(4, inf, aby, at', zda, že) ?EFF(jako+4, za+4)

původní stav:

ACT(1) PAT(4, že, aby, .f) ADDR(3)
?EFF(.4[za-1], .4[{jako, jakožto}:/AuxY])

ACT(1) ADDR(3) PAT(4, inf, aby, at', zda, že) ?EFF(jako+4, za+4)

Ukázka pravidla – úpravy rámce

ACT(1) PAT(4, že, aby, inf) ADDR(3) ?EFF(za+4, jako+4)

ACT(1) PAT(4, inf, aby, at', zda, že) ADDR(3) ?EFF(jako+4, za+4)

původní stav:

ACT(1) PAT(4, že, aby, .f) ADDR(3)
?EFF(.4[za-1], .4[{jako, jakožto}:/AuxY])

ACT(1) ADDR(3) PAT(4, inf, aby, at', zda, že) ?EFF(jako+4, za+4)

Ukázka pravidla – úpravy rámce

ACT(1) PAT(4,aby, inf, že) ADDR(3) ?EFF(jako+4,za+4)

ACT(1) PAT(4,aby,at',inf,zda,že) ADDR(3) ?EFF(jako+4,za+4)

původní stav:

ACT(1) PAT(4,že,aby,.f) ADDR(3)
?EFF(.4[za-1],.4[{jako,jakožto}:/AuxY])

ACT(1) ADDR(3) PAT(4,inf,aby,at',zda,že) ?EFF(jako+4,za+4)

Ukázka pravidla

ACT(1) PAT(4,aby, inf, že) ADDR(3) ?EFF(jako+4,za+4)

ACT(1) PAT(4,aby,at',inf,zda,že) ADDR(3) ?EFF(jako+4,za+4)

```
/^ACT\ (. \)
  (\??PAT) \ (
    ($forma) ($formy) ($formy) ($formy) ($formy) ($formy)
  \)
  (.*)
$/
```

```
$vypln = [^()]+
$forma = [^(), ]+
$formy = (,$forma)*
```

Ukázka pravidla

ACT(1) PAT(4,aby, inf, že) ADDR(3) ?EFF(jako+4,za+4)

ACT(1) PAT(4,aby,at',inf,zda,že) ADDR(3) ?EFF(jako+4,za+4)

```
/^ACT\ (. \)
  (\??PAT) \ (
    ($forma) ($formy) ($formy) ($formy) ($formy) ($formy)
  )
  (.*)
$/
```

```
$vypln = [^()]+
$forma = [^(),]+
$formy = (,$forma)*
```

```
/^ACT\ (. \)
  $1 \ (
    ($vypln,)?
    $2 (,$vypln)? $3 (,$vypln)? $4 (,$vypln)?
    $5 (,$vypln)? $6 (,$vypln)? $7 (,$vypln)?
  )
  $8
$/
```

Ukázka pravidla – aplikace

ACT(1) PAT(4,aby, inf, že) ADDR(3) ?EFF(jako+4,za+4) ###

ACT(1) PAT(4,aby,at',inf,zda,že) ADDR(3) ?EFF(jako+4,za+4)

```
/^ACT\ (. \)
  (\??PAT) \ (
    ($forma) ($formy) ($formy) ($formy) ($formy) ($formy)
  \)
  (.*)
###

ACT\ (. \)
$1 \ (
  ($vypln,)?
  $2 (,$vypln)? $3 (,$vypln)? $4 (,$vypln)?
  $5 (,$vypln)? $6 (,$vypln)? $7 (,$vypln)?
\ )
$8
$/
```

$\$vypln = [^()] +$
 $\$forma = [^() ,] +$
 $\$formy = (, \$forma) *$

Ukázka pravidla – aplikace

ACT(1) PAT(4,aby, inf, že) ADDR(3) ?EFF(jako+4,za+4) ###

ACT(1) PAT(4,aby,at',inf,zda,že) ADDR(3) ?EFF(jako+4,za+4)

```
/^ACT\ (. \)
  (\??PAT) \ (
    ($forma) ($formy) ($formy) ($formy) ($formy) ($formy)
  \)
  (.*)
###
  ACT\ (. \)
  $1 \ (
    ($vypln,)?
    $2 (,$vypln)? $3 (,$vypln)? $4 (,$vypln)?
    $5 (,$vypln)? $6 (,$vypln)? $7 (,$vypln)?
  \)
  $8
$/
```

$\$vypln = [^()] +$
 $\$forma = [^() ,] +$
 $\$formy = (, \$forma) *$

Ukázka pravidla – aplikace

ACT (1) PAT (4, aby, inf, že) ADDR (3) ?EFF (jako+4, za+4) ###

ACT (1) PAT (4, aby, at', inf, zda, že) ADDR (3) ?EFF (jako+4, za+4)

```
/^ACT\ (. \)
  (\ ?? PAT) \ (
    ($forma) ($formy) ($formy) ($formy) ($formy) ($formy)
  \)
  (. *)
  ###
  ACT\ (. \)
  $1 \ (
    ($vypln, ) ?
    $2 (, $vypln) ? $3 (, $vypln) ? $4 (, $vypln) ?
    $5 (, $vypln) ? $6 (, $vypln) ? $7 (, $vypln) ?
  \)
  $8
$/
```

```
$vypln = [^()] +
$forma = [^(),] +
$formy = (, $forma) *
```

Ukázka pravidla – aplikace

ACT (1) PAT (4, aby, inf, že) ADDR (3) ?EFF (jako+4, za+4) ###

ACT (1) PAT (4, aby, at', inf, zda, že) ADDR (3) ?EFF (jako+4, za+4)

```
/^ACT\ (. \)
  (\ ?? PAT) \ (
    ($forma) ($formy) ($formy) ($formy) ($formy) ($formy)
  \)
  (.*)
###

ACT\ (. \)
$1 \ (
  ($vypln, ) ?
  $2 (, $vypln) ? $3 (, $vypln) ? $4 (, $vypln) ?
  $5 (, $vypln) ? $6 (, $vypln) ? $7 (, $vypln) ?
\ )
$8
$/
```

$\$vypln = [^()] +$
 $\$forma = [^() ,] +$
 $\$formy = (, \$forma) *$

Ukázka pravidla – aplikace

ACT (1) PAT (4, aby, inf, že) ADDR (3) ?EFF (jako+4, za+4) ###

ACT (1) PAT (4, aby, at', inf, zda, že) ADDR (3) ?EFF (jako+4, za+4)

```
/^ACT\ (. \)
  (\ ??PAT) \ (
    ($forma) ($formy) ($formy) ($formy) ($formy) ($formy)
  \)
  (.*)
###

ACT\ (. \)
$1 \ (
  ($vypln, ) ?
  $2 (, $vypln) ? $3 (, $vypln) ? $4 (, $vypln) ?
  $5 (, $vypln) ? $6 (, $vypln) ? $7 (, $vypln) ?
\ )
$8
$/
```

$\$vypln = [^()] +$
 $\$forma = [^() ,] +$
 $\$formy = (, \$forma) *$

Ukázka pravidla – aplikace

ACT (1) PAT (4, aby, inf, že) ADDR (3) ?EFF (jako+4, za+4) ###

ACT (1) PAT (4, aby, at', inf, zda, že) ADDR (3) ?EFF (jako+4, za+4)

```
/^ACT\ (. \)
  (\ ??PAT) \ (
    ($forma) ($formy) ($formy) ($formy) ($formy) ($formy)
  \)
  (.*)
###
```

```
$vypln = [^()] +
$forma = [^(),] +
$formy = (,$forma) *
```

```
ACT\ (. \)
$1 \ (
  ($vypln, ) ?
  $2 (,$vypln) ? $3 (,$vypln) ? $4 (,$vypln) ?
  $5 (,$vypln) ? $6 (,$vypln) ? $7 (,$vypln) ?
\ )
$8
$/
```

Ukázka pravidla – aplikace

ACT (1) PAT (4, aby, inf, že) ADDR (3) ?EFF (jako+4, za+4) ###

ACT (1) PAT (4, aby, at', inf, zda, že) ADDR (3) ?EFF (jako+4, za+4)

```
/^ACT\ (. \)
  (\ ??PAT) \ (
    ($forma) ($formy) ($formy) ($formy) ($formy) ($formy)
  \)
  (.*)
###
```

```
ACT\ (. \)
$1 \ (
  ($vypln, ) ?
  $2 (, $vypln) ? $3 (, $vypln) ? $4 (, $vypln) ?
  $5 (, $vypln) ? $6 (, $vypln) ? $7 (, $vypln) ?
  \)
  $8
$/
```

```
$vypln = [^() ]+
$forma = [^() , ]+
$formy = (, $forma) *
```

Ukázka pravidla – aplikace

ACT (1) PAT (4, aby, inf, že) ADDR (3) ?EFF (jako+4, za+4) ###

ACT (1) PAT (4, aby, at', inf, zda, že) ADDR (3) ?EFF (jako+4, za+4)

```
/^ACT\ (. \)
  (\ ??PAT) \ (
    ($forma) ($formy) ($formy) ($formy) ($formy) ($formy)
  \)
  (.*)
###
```

```
ACT\ (. \)
$1 \ (
  ($vypln, ) ?
  $2 (, $vypln) ? $3 (, $vypln) ? $4 (, $vypln) ?
  $5 (, $vypln) ? $6 (, $vypln) ? $7 (, $vypln) ?
\ )
$8
$ /
```

$\$vypln = [^()] +$
 $\$forma = [^() ,] +$
 $\$formy = (, \$forma) *$

Ukázka pravidla – aplikace

ACT (1) PAT (4, aby, inf, že) ADDR (3) ?EFF (jako+4, za+4) ###

ACT (1) PAT (4, aby, at', inf, zda, že) ADDR (3) ?EFF (jako+4, za+4)

```
/^ACT\ (. \)
  (\ ??PAT) \ (
    ($forma) ($formy) ($formy) ($formy) ($formy) ($formy)
  \)
  (.*)
###
```

```
ACT\ (. \)
$1 \ (
  ($vypln, ) ?
  $2 (, $vypln) ? $3 (, $vypln) ? $4 (, $vypln) ?
  $5 (, $vypln) ? $6 (, $vypln) ? $7 (, $vypln) ?
\ )
$8
$/
```

$\$vypln = [^()] +$
 $\$forma = [^() ,] +$
 $\$formy = (, \$forma) *$

Porovnávání lemmat

- slova v glosse
- slova v příkladech
- lemmatizace, autosémantika, bez slovesa
- poměr počtu v průniku ku sjednocení
- příklad:
 - *obracel* vše *vzhůru nohama* / *naruby* => vzhůru noha **naruby** (jedna ze tří ... 1/3)
 - *obracet* vše *naruby* => **naruby** (1)
 - průměr $(1/3 + 1) / 2 = 2/3$

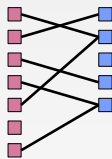
Reciprocita a kontrola

- ve VALLEXu zachycené explicitně
- do PDT-Vallexu je možno je částečně dogenerovat z PDT
- málo časté, nepomohlo

Přehled práce

- Valence
- VALLEX a PDT-Vallex
 - ukázka, rozdíly, historie
- Postup porovnávání slovníků
- Společný formát
- Postup porovnávání lexikálních jednotek
- Příklad
- **Vyhodnocení a závěr**

Vyhodnocení

- ruční anotace 200 sloves
 - s různým počtem LU: od největší skupiny L_1 po L_{26}
 - 2700 možných párů, 500 párů anotátoři zvolili
- úspěšnost zvlášť podle skupin $L_1 - L_{26}$,
vážený průměr podle zastoupení ve slovnících
- baseline:  F_1 score 63
- výsledky:  F_1 score 77 (Prec 78, Rec 76)
 - téměř 2000 LU s Prec 91 z VALLEXu propojených s PDT-Vallexem, zpřístupněno na webu [\(link\)](#)

Splynutí valenčních slovníků?

- následná údržba
- oba slovníky nadále vyvíjeny
 - PDT-Vallex stále do šířky (přibývá anotace PDTSC)
 - VALLEX stále do hloubky (anotace lexikálních alternací, diatezí, analytických predikátů)
- odlišný přístup autorských týmů trvá
- **technicky** by splynutí možné bylo, po řadě rozhodnutí o jednotném zápisu
- nechystá se

- 200 sloves anotovaných v 8000 větách (2005)
- anotace dle lexikálních jednotek VALLEXu 1.0
- změny lexikálních jednotek mezi verzemi nebyly zaznamenávány
- stejným postupem propojen VALLEX 1.0 a 2.5
- původní anotované věty zpřístupněny s vydáním VALLEXu 2.5

Závěr

- automatická metoda pro provazování slovníků
- zpřístupnění 8000 vět z VALEVALu a 75.000 příkladů z PDT a PCEDT pro VALLEX
- opakovatelnost