

NLP 7

Spracovanie prirodzeného jazyka

Peter Bednár

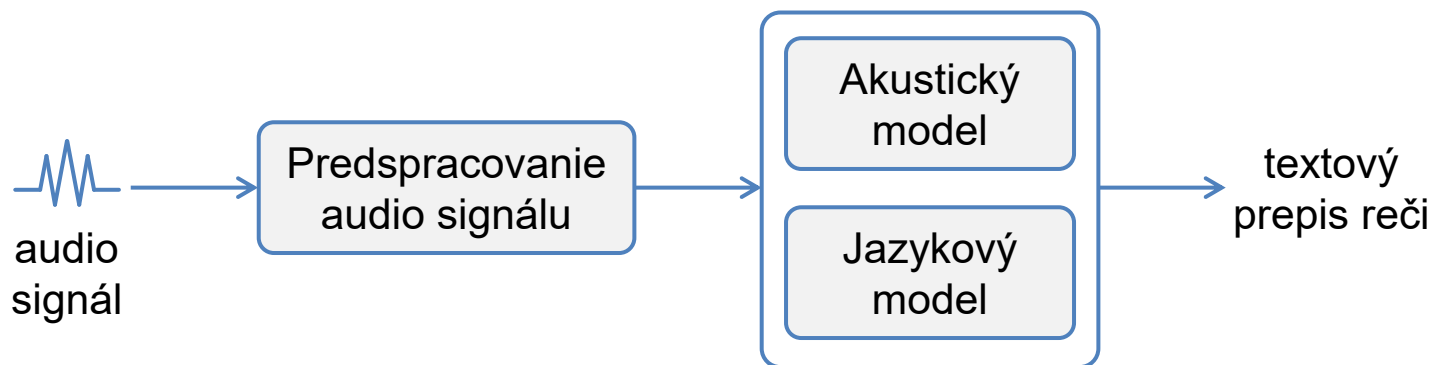
Fonetická reprezentácia a rozpoznávanie reči

Rozpoznávanie hovorenej reči

- Zložitosť úlohy na úrovni audio signálu:
 - rôzne hlasy, rýchlosť hovorenia, artikulácia, prízvuk, rušné prostredie
- Nejednoznačné mapovanie hlások na znaky
 - závislé od kontextu, spodobovanie
- Ľudský mozog využíva pri rozpoznávaní hlások a slov zo zvuku ich jazykové vlastnosti (syntax a význam)

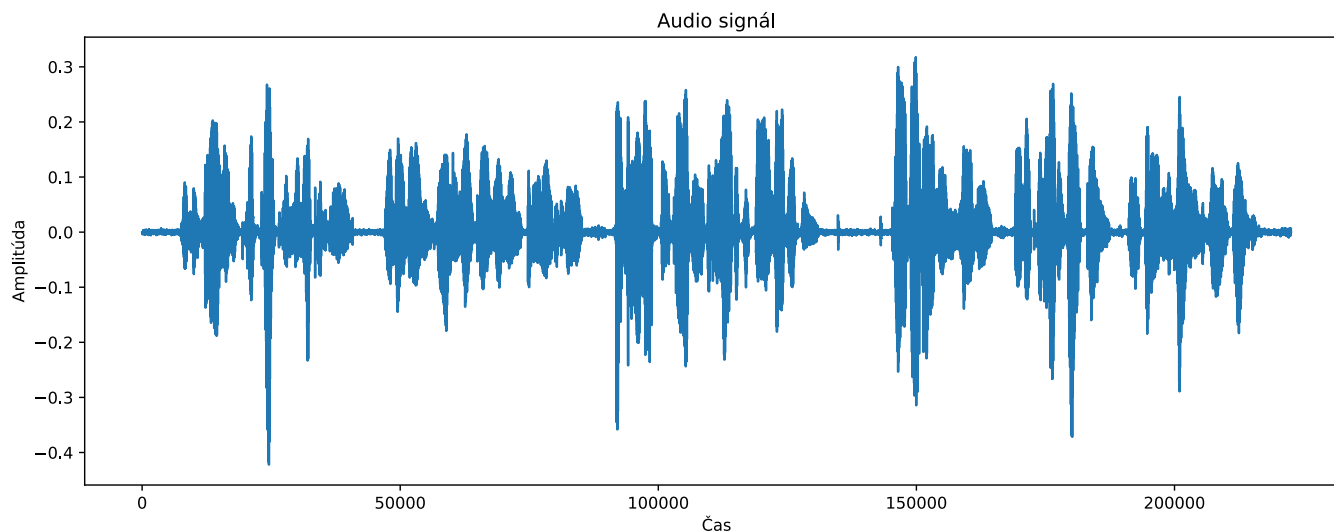
Hlboké učenie pre rozpoznávanie reči

- Audio signál je reprezentovaný ako časová sekvencia
 - mapovanie sekvencií s rozdielnou dĺžkou
 - **kodér/dekodér architektúra**



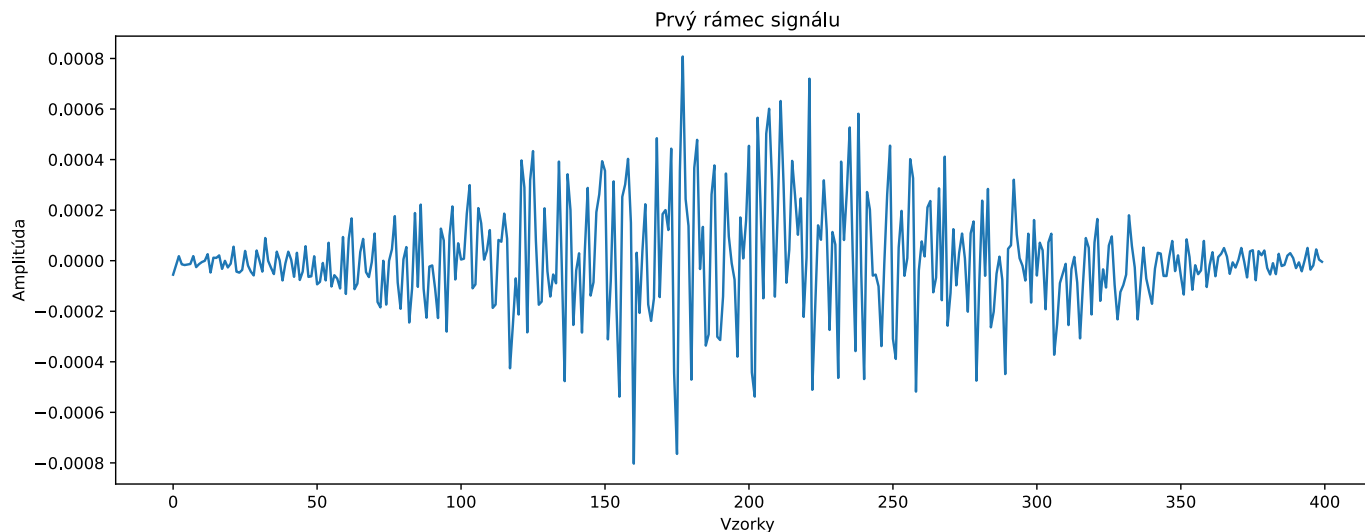
Predspracovanie audio signálu (1)

- Zvuk sa prenáša akustickými vlnami a je zaznamenaný ako časová sekvencia, štandardná vzorkovacia frekvencia pre rozpoznávanie reči je 16 kHz, pre kvalitné audio nahrávanie 44,1 kHz
- Príklad: 222 562 vzoriek, vzorkovacia frekvencia 16 kHz, 13,9 s



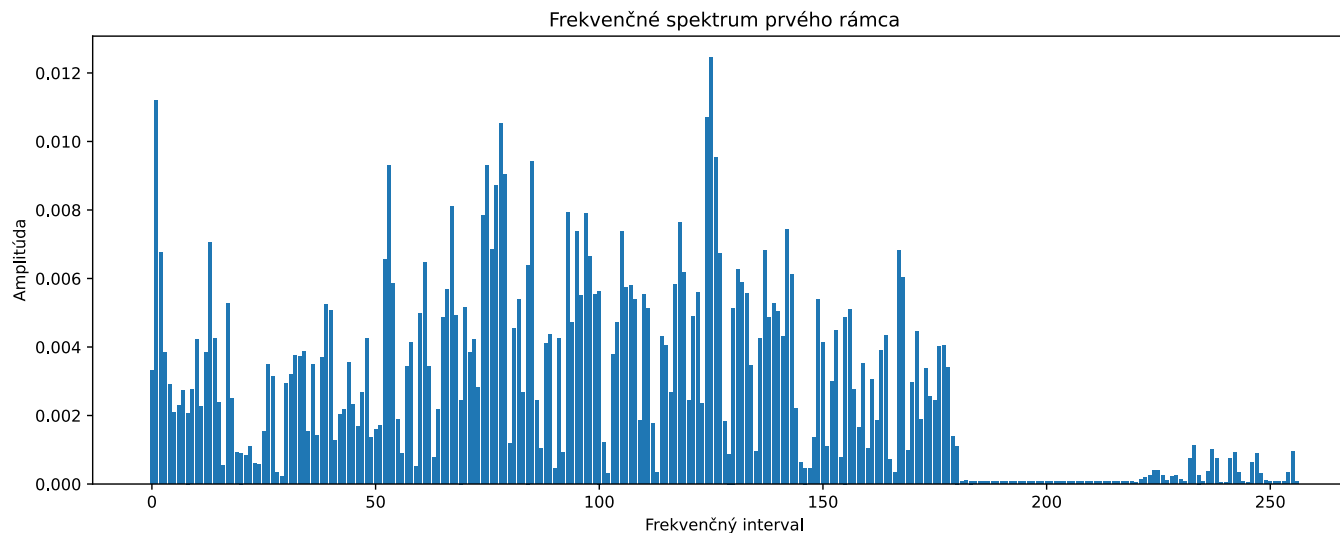
Predspracovanie audio signálu (2)

- Postupnosť sa rozdelí na krátke prekrývajúce sa rámce, najčastejšie 20-40 ms
- Na hodnoty rámca sa aplikuje Hammingov filter, ktorý vyhladí okrajové hodnoty
- Príklad: prvý rámec 25 ms / 400 vzoriek, posun 10 ms



Predspracovanie audio signálu (3)

- Signál sa prevedie do frekvenčnej domény **rýchlou Fourierovou transformáciou** (FFT - *Fast Fourier Transform*)

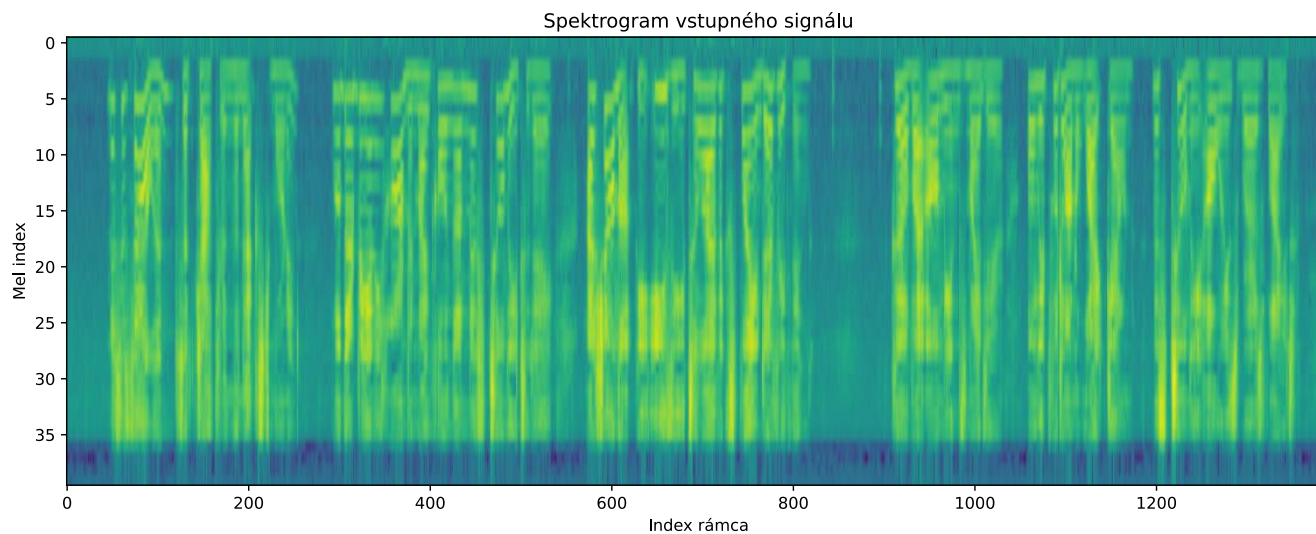


Predspracovanie audio signálu (4)

- Frekvencie spektra sa premapujú na **mel stupnicu** a hodnoty amplitúdy sa logaritmicky transformujú
 - mel stupnica logaritmicky rozdeľuje intervaly frekvencií a priradzuje im index
 - človek nerozlišuje frekvencie rovnako (rozdiel medzi 100 – 200 Hz vnímame výraznejšie než medzi 1100 – 1200 Hz)
 - hlasitosť (ktorá je daná veľkosťou amplitúdy) tiež vnímame logaritmicky

Predspracovanie audio signálu (5)

- Po prepočítaní každého rámca dostaneme spektrogram vstupného signálu
- Príklad: 40 melových pásiem $\times$ počet rámcov

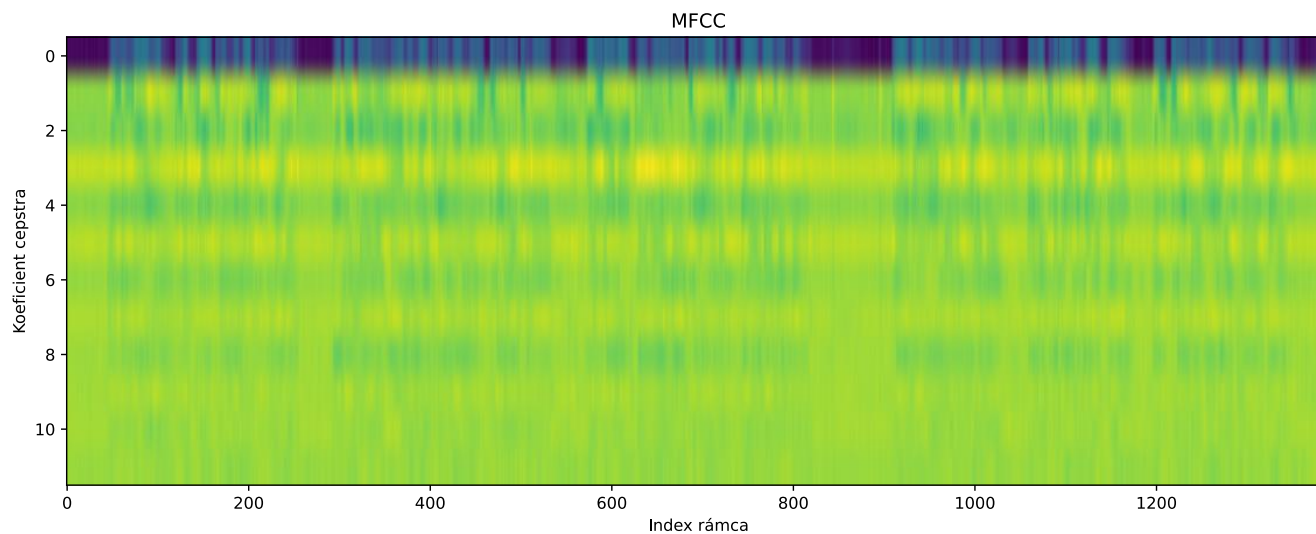


Predspracovanie audio signálu (6)

- Pri rozpoznávaní reči nás zaujímajú najmä prechody medzi frekvenciami
 - pomocou diskkrétnej kosínusovej transformácie (DCT – *Discrete Cosine Transform*) získame **kepstrum** – spektrum spektra

Predspracovanie audio signálu (7)

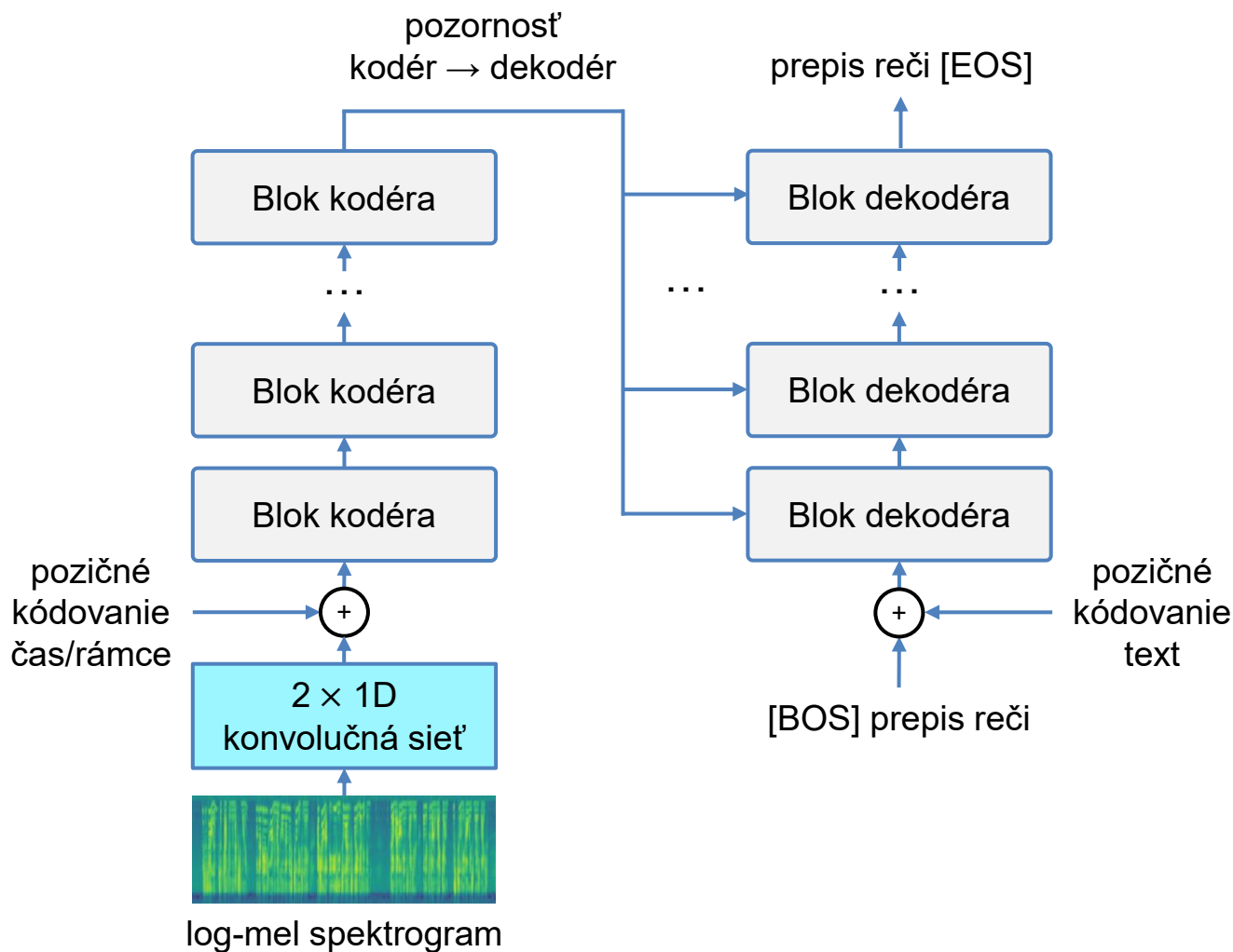
- Výsledná reprezentácia – MFCC – *Mel-Frequency Cepstral Coefficients*
- Príklad: 12 MFC koeficientov $\times$ počet rámcov



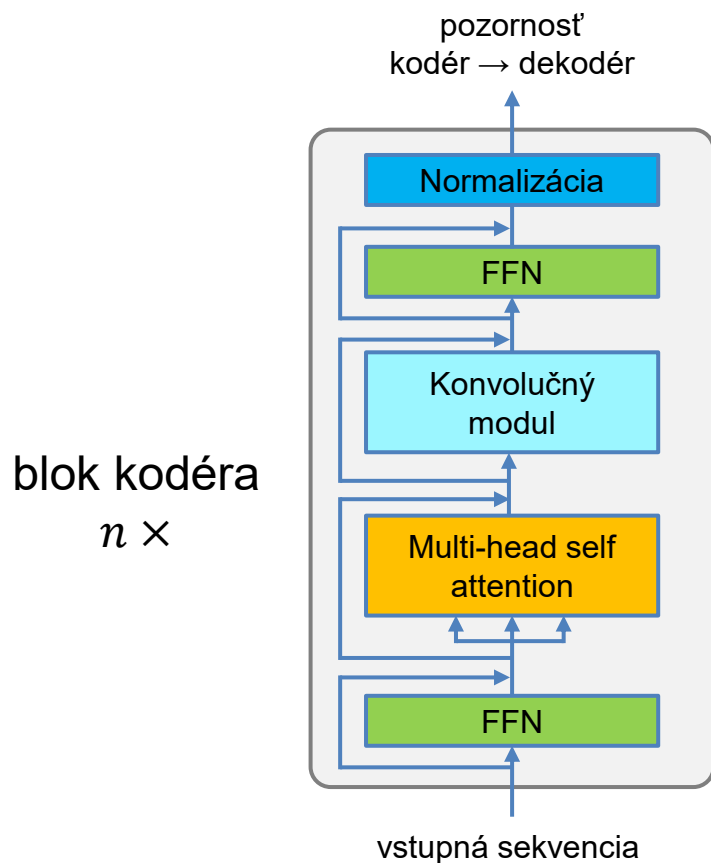
Metódy rozpoznávania reči

- **Akustický model**
 - rozpoznávanie hlások
 - lokálne príznaky extrahované konvolučnou vrstvou
 - kontextové závislosti
 - Conformer – kombinácia konvolučnej a samo-pozornostnej vrstvy
- **Jazykový model**
 - spájanie hlások do slov a slov do viet/vyjadrení
 - mapovanie sekvencia-sekvencia s rôznou dĺžkou
 - kontextové závislosti
 - Transformer – kódér/dekodér architektúra

Transformer pre rozpoznávanie reči



Conformer – konvolučná vrstva + pozornosťná vrstva



- Lokálne závislosti

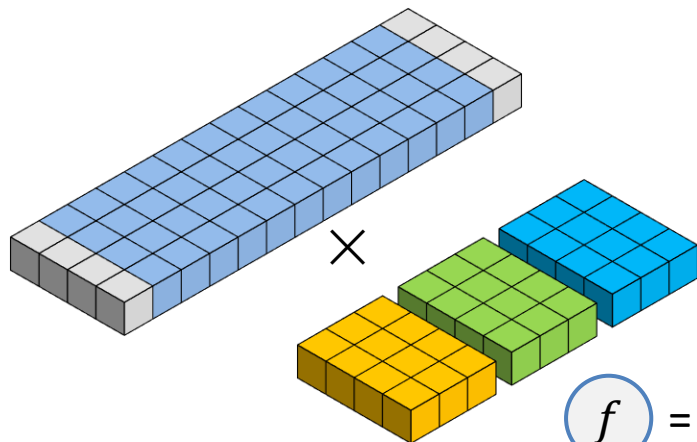
- rozpoznávanie príznakov zo zvuku
- konvolučný modul zložený z *depth-wise* a *point-wise* konvolúcie

- Vzdialené závislosti

- fonetické zloženie slova, syntax, sémantika
- samo-pozornosťná vrstva

Depth-wise a point-wise 1D konvolúcia (1)

Klasická 1D konvolúcia

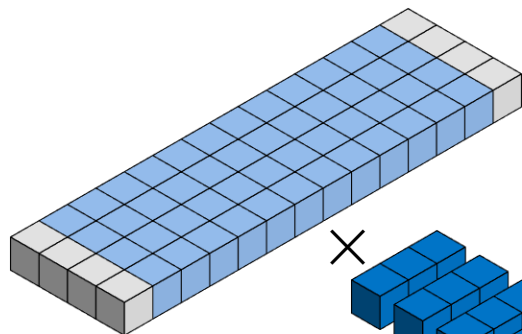


vstup: dĺžka sekvencie + okraj $\times 1 \times$
počet kanálov

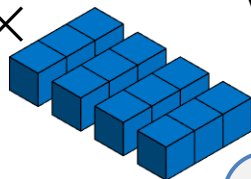
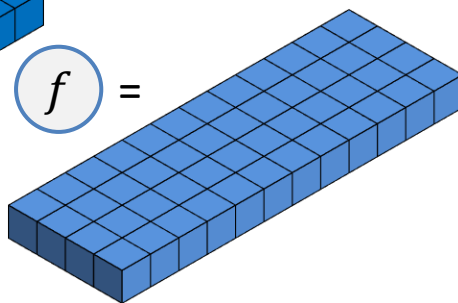
výstup: dĺžka sekvencie $\times 1 \times$
počet filtrov

počet parametrov filtra:
šírka okna $\times 1 \times$ počet
kanálov (voliteľne + 1 pre
bias)

Depth-wise a point-wise 1D konvolúcia (2)



×

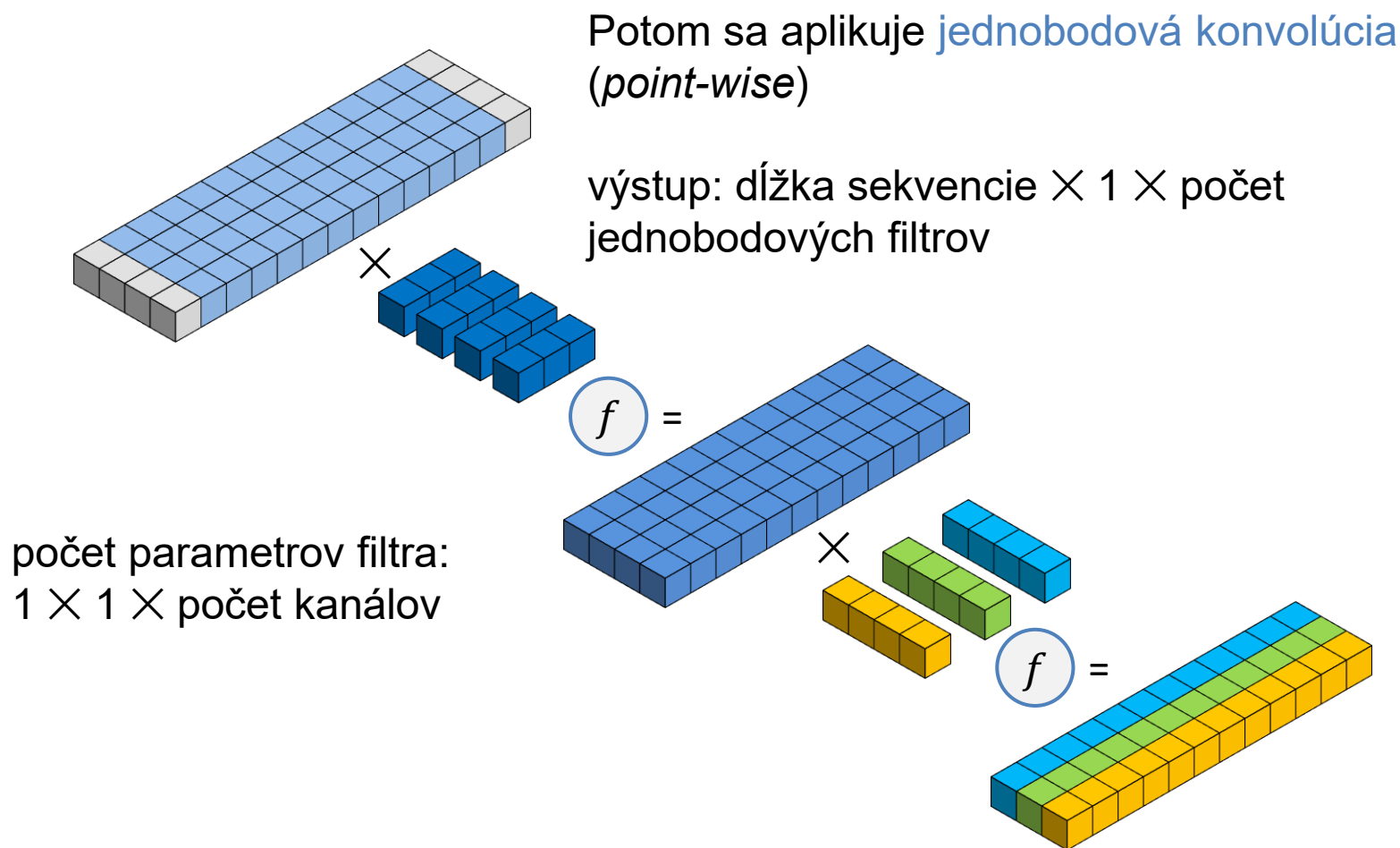
 $f =$ 

Najprv sa aplikuje **konvolúcia do hĺbky** (*depth-wise*)

Filtre sa aplikujú samostatne na každom kanáli
výstup: dĺžka sekvencie $\times$ 1 $\times$ počet kanálov

počet parametrov filtra:
šírka okna $\times$ 1 $\times$ 1

Depth-wise a point-wise 1D konvolúcia (3)



Conformer – konvolučný blok

- Konvolúcia do hĺbky najprv vyextrahuje vzory v jednotlivých frekvenčných intervaloch s nelineárnou aktiváciou
- Jednobodová konvolúcia ich lineárne agreguje a znova nelineárne transformuje

Vyhodnotenie metód rozpoznávania reči

- Porovnanie vygenerovaného prepisu s referenčným prepisom nahrávky

Základné metriky

WER

word error rate – podiel slovných chýb (počet zámen slova + vynechaní slov + vložení slov), minimálny počet chýb sa určí zarovnaním prepisov pomocou editačnej vzdialenosti

CER

character error rate – rovnaký princíp na úrovni znakov

exact match

podiel výpovedí, ktorých prepis sa úplne zhoduje s referenciou

Vyhodnotenie metód rozpoznávania reči

Ďalšie hodnotené vlastnosti

- Rýchlosť spracovania – RTF
 - *real-time factor*
 - čas spracovania delený dĺžkou nahrávky, ak $RTF < 1$ spracovanie rýchlejšie než v reálnom čase
- Latencia
 - oneskorenie priebežného alebo konečného prepisu, najmä pri streamovanom rozpoznávaní
- Robustnosť
 - presnosť pri šume, rôznych prízvukoch, dialektoch a hovoriacich
- Výpočtové nároky
 - využitie pamäte, CPU/GPU

Morfosyntaktická analýza

Morfosyntaktická analýza

- Viacúlohové učenie
 - vytvorí sa jeden model pre riešenie viacerých úloh naraz
- Lematizácia
 - priradí slovám základný tvar - *lemu*
- Morfologické značkovanie
 - priradí slovám slovný druh (podstatné meno, prídavné meno, sloveso, ...) a morfologické kategórie (rod, pád, číslo, slovesný vid, životnosť, ...)
- Syntaktické parsovanie
 - rozdelí vety na frázy a priradí syntaktické roly (podmet, prísudok, predmet, ...)

Morfosyntaktická analýza

- Vstup:
 - text segmentovaný na vety a rozdelený na postupnosť tokenov pre každú vetu
- Výstup:
 - pre každé slovo základný tvar, slovný druh a morfológické kategórie
 - závislostný strom, ktorý priradí slovám syntaktické roly

Morfosyntaktická analýza (1)

Lematizácia

1. Vytvorenie zoznamu editačných pravidiel, ktoré transformujú daný tvar slova na základný tvar
2. Klasifikačná úloha
 - model každému slovu na vstupe priradí editačné pravidlo pre jeho transformáciu do základného tvaru

Morfologické značkovanie

- Klasifikačná úloha
 - model každému slovu na vstupe priradí slovný druh a hodnoty jednotlivých morfológických kategórií

Morfosyntaktická analýza (2)

Syntaktické parsovanie

- Klasifikačná úloha
 1. pre každú dvojicu slov sa klasifikuje, či medzi nimi existuje závislostná hrana
 2. ak áno, klasifikuje sa, aké je jej označenie

Vytvorenie množiny editačných pravidiel

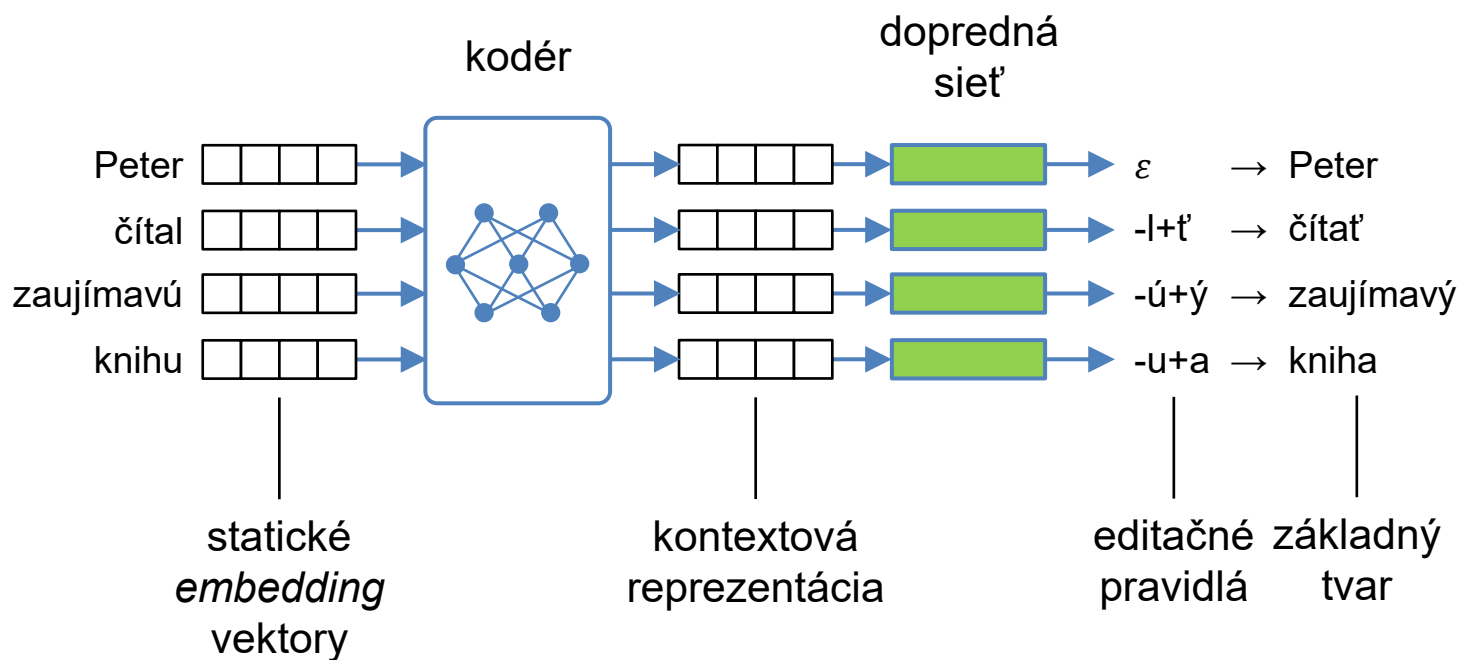
Pre každé slovo v trénovacej množine sa vykoná:

1. Zarovnanie na najdlhší spoločný reťazec so základným tvarom
2. Určenie minimálnej postupnosti znakových operácií, ktoré transformujú pôvodný tvar slova na základný tvar
 - vloženie / odstránenie / zámena znaku
 - výpočet pomocou editačnej vzdialenosti

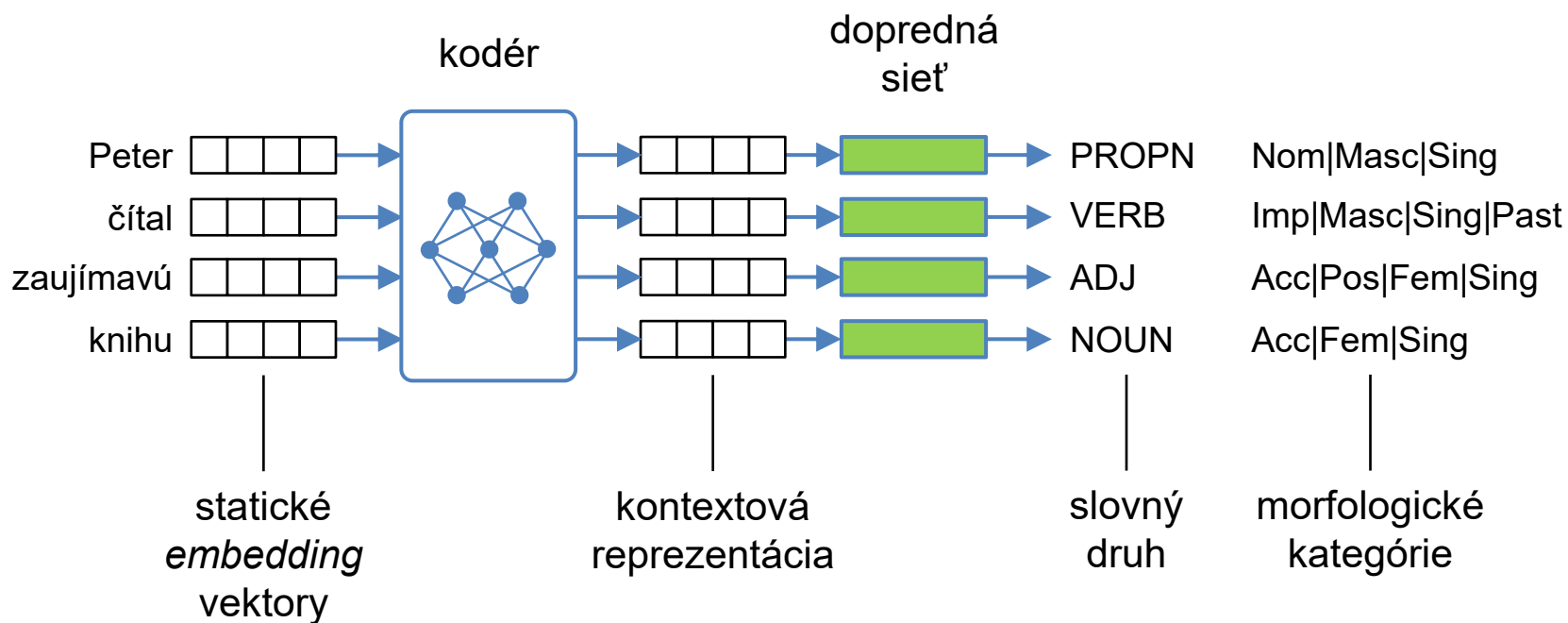
Vytvorenie množiny editačných pravidiel

Tvar slova	Lema	Spoločná časť	Editačné pravidlo
domy	dom	dom	-y
domov	dom	dom	-ov
stromy	strom	strom	-y
ženy	žena	žen	-y+a
pracoval	pracovať	pracova	-l+t'
mestách	mesto	mest	-ách+o
mesto	mesto	mesto	ε (prázne pravidlo)
otcov	otec	ot	-cov+ec
najkrajší	krásny	kr	-naj -ajší+ásny
		...	

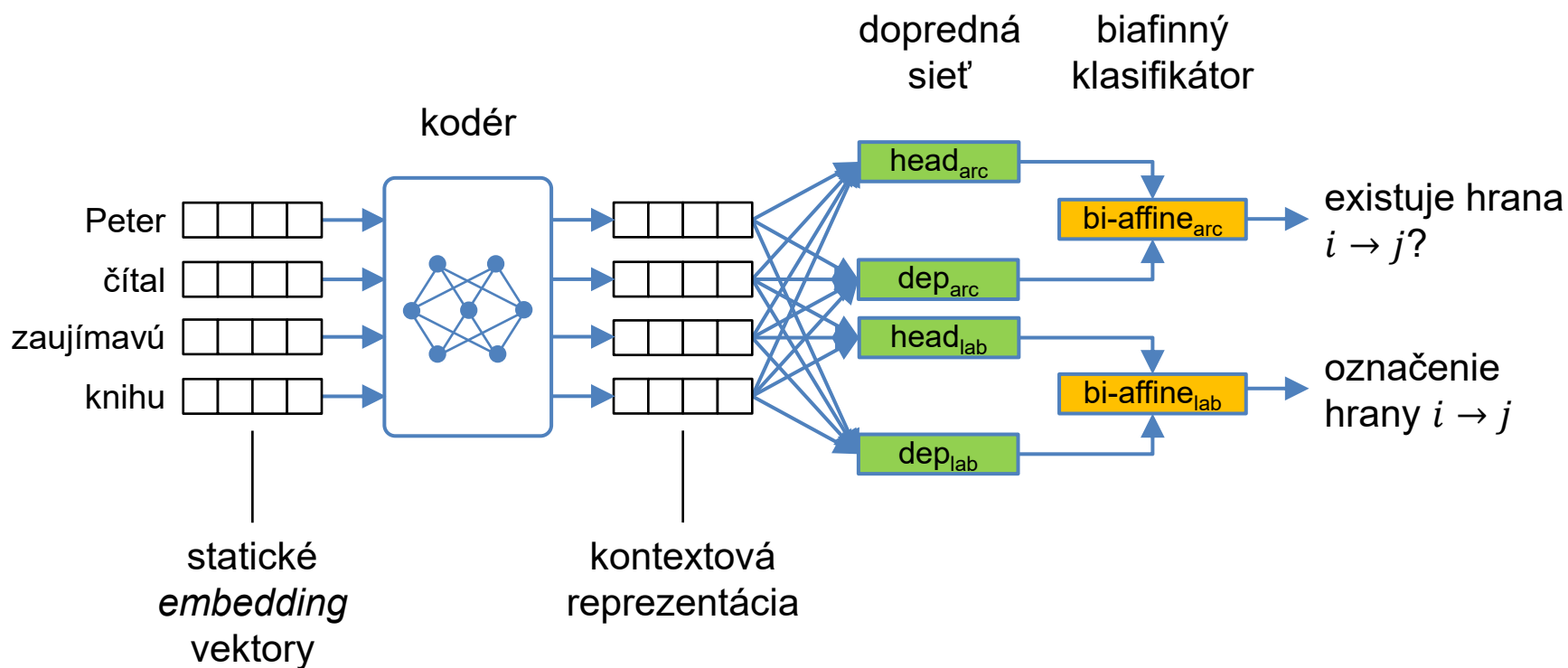
Lematizácia a morfológické značkovanie (1)



Lematizácia a morfologické značkovanie (2)



Syntaktické parsovanie



Bi-afinný klasifikátor

- Modeluje vzťah medzi nadradeným a závislým slovom na základe ich kontextových reprezentácií
- Pre vstupný vektor závislého slova $\mathbf{d}_i$ a nadradeného slova $\mathbf{h}_j$ sa vypočíta predikcia podľa:

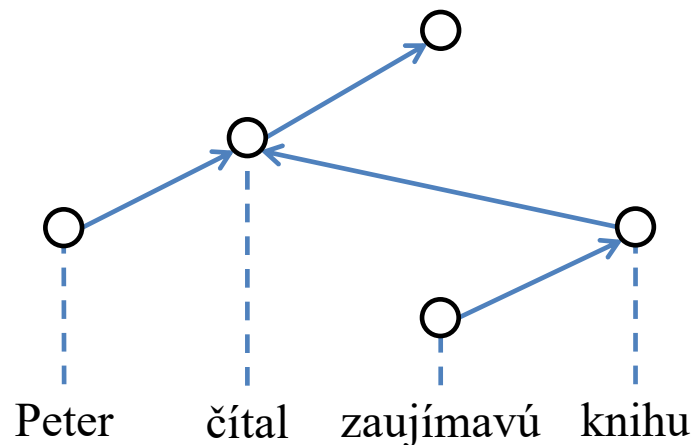
$$y_{i,j} = \mathbf{d}_i^T \mathbf{U} \mathbf{h}_j + \mathbf{W} \begin{bmatrix} \mathbf{d}_i \\ \mathbf{h}_j \end{bmatrix} + \mathbf{b}$$

kde $\mathbf{U}_{l \times d \times d}$, $\mathbf{W}_{l \times 2d}$ a $\mathbf{b}_l$ sú parametre nastavené pri učení, l je počet tried a d je rozmer vstupných vektorov $\mathbf{d}_i$ a $\mathbf{h}_j$ (pre predikciu hrany $l = 1$, pre predikciu označenia hrany $l = \text{počet označení}$)

Vytvorenie závislostného stromu

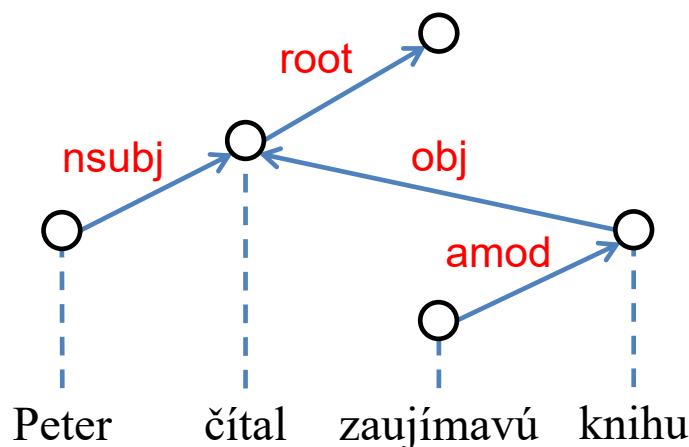
- Pri parsovaní sa vypočíta predikcia bi-affine_{arc}($\mathbf{d}_i, \mathbf{h}_j$) pre každú kombináciu hrán $i \rightarrow j$ a $j \rightarrow i$, získame maticu ktorá reprezentuje úplne prepojený vážený graf
- *Edmondsov algoritmus*
 - nájde maximálny strom, ktorý pokrýva všetky uzly (*maximum spanning tree*)

	ROOT	Peter	čítal	zaujímalú	knihu
Peter	0.2	-	8.7	0.3	1.1
čítal	9.2	1.2	-	0.2	1.4
zaujímalú	0.1	0.3	2.0	-	9.1
knihu	0.4	0.7	8.9	1.8	-



Výsledná klasifikácia označení hrán

- Následne sa určí pre každú hranu $i \rightarrow j$ jej označenie (tzn. syntaktická rola závislého slova) podľa predikcie $\text{bi-affine}_{\text{lab}}(\mathbf{d}_i, \mathbf{h}_j)$



Vyhodnotenie morfosyntaktickej analýzy

Lematizácia

správnosť *accuracy* - priradenie základných tvarov

Morfologické značkovanie

správnosť priradenie slovných druhov (POS – *part of speech*)

presnosť, návratnosť, priradenie jednotlivých morfologických kategórií

F_1 miera

Syntaktické parsovanie

UAS *unlabeled attachment score* - podiel slov, ktoré majú správne priradený nadradený uzol nezávisle od označenia hrany

LAS *labeled attachment score* - podiel slov, ktoré majú správne priradený nadradený uzol aj označenie hrany