

Databáze + Automaticky učené řečové jednotky + Projekty

Jan Černocký ÚPGM FIT VUT Brno, cernocky@fit.vutbr.cz

FIT VUT Brno

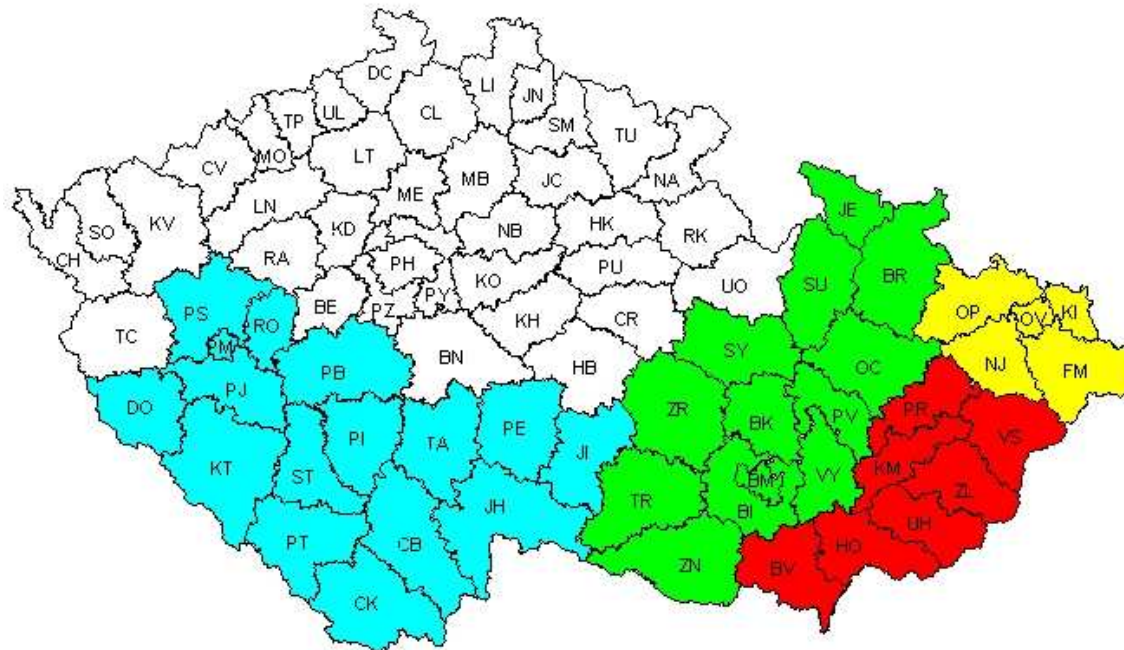
Sběr velkých řečových databází

Proč ?

⇒ rozpoznávače se musí trénovat na popsanych řečových datech.

Požadavky

- Prostředí musí odpovídat reálným podmínkám (auto, obývač, tlf. linka).
- Pokrytí různých kategorií mluvčích (pohlaví, dialekt, věk).



Nahrávané položky

- číslovky, klíčová slova, povely
- řetězce číslic, čísla
- jména (města, lidé)
- hláskování
- foneticky vyrovnaná slova/věty:

Odzbrojením, které je klíčovým bodem dohody, se neobtěžují.

Muž totiž nepřerušil klasickou onkologickou léčbu.

Mám neseriózního jednání dost, poznamenala.

Je ženatý, má tři děti a je vynikajícím hráčem bridže.

Určitě neuhodnete, z čeho to je, prohlásil sebevědomě.

Bridž totiž hrají dvě dvojice proti sobě.

Aprílové počasí provázelo včerejší program mítinku.

Mezi sólisty nové inscenace se objeví řada hostů.

Řekl to ve čtvrtek člen vedení belgické strany zelených.

Po městě jezděte tramvají, ta je ekologická.

DB - projekty - vše společně s ČVUT Praha

- Číslovky (1999) - 1227 mluvčích, telefon, ca 7 min/mluvčí, pouze číslovky, čísla, řetězce čísel + spec. znaky. Finance od Siemens AG R&D Mnichov. Výzkumná a výuková práva.
- SpeechDat-East (1999-2000) - 1052 mluvčích, telefon, ca 15 min/mluvčí, všechny typy položek. Finance od EU - 4. rámcový program, INCO Copernicus, spolupráce s Matra⇒Lernout&Hauspie⇒ScanSoft. Plná práva.
- SpeeCon (2003) - 600 mluvčích (z toho 50 dětí), ca 40 min/mluvčí, prostředí Office, Entertainment, Public, Car. 4-kanálové nahrávání, speeconí “bedna”, notebook, 2×VXPocket2.
- TEMIC2 (teď!) - 600 mluvčích, ca 40 min/mluvčí, různá auta, 2 kanály. nahrávání na DAT, pak přepis.

Mluvčí z Východní Moravy a Slezska needed ! Platíme 200,- za max. hodinovou session :-)

- pořizování multimodálních meeting dat - viz Petr Jenderka.

DB - anotace

zapsat, co mluvčí **skutečně** řekl.

The screenshot shows the Transcriber software window titled "Transcriber /tmp/honza/00176/00176S0.CSO". The interface includes several text input fields and a control panel on the right. The "Sheet text" field contains the sentence: "Navíc si myslím, že stát by měl kauzy medializovat co nejméně." The "Transcript" field shows the phonetic transcription: "[spk] navíc si myslím že stát by měl kauzy (medializovat/medyjalizovat) co nejméně". Below this is a row of buttons for various transcription symbols: _ab..AB, AB.._ab, Digits, Numbers, Money, Date (ledna), Date (první), Year, Time, Spelling (BÉ), Spelling (B), Empty, One space, and Erase symb. The "Phonetic transcript" field contains: "[spk] navíc sy myslým že stát by mñel kAzy medyjalyzovat co nejméñe". The "Comment" field is empty. The "Help" section at the bottom lists keyboard shortcuts for various functions. The status bar at the bottom shows "FTP state: Disconnect", "Call: 11/10", and "File: 521/520".

Transcriber /tmp/honza/00176/00176S0.CSO

Preannotation note:

Sheet text:
Navíc si myslím, že stát by měl kauzy medializovat co nejméně.

Transcript: A ä E ë O ö U ü
[spk] navíc si myslím že stát by měl kauzy
(medializovat/medyjalizovat) co nejméně

_ab..AB AB.._ab Digits Numbers Money Date (ledna) Date (první)
Year Time Spelling (BÉ) Spelling (B) Empty One space Erase symb

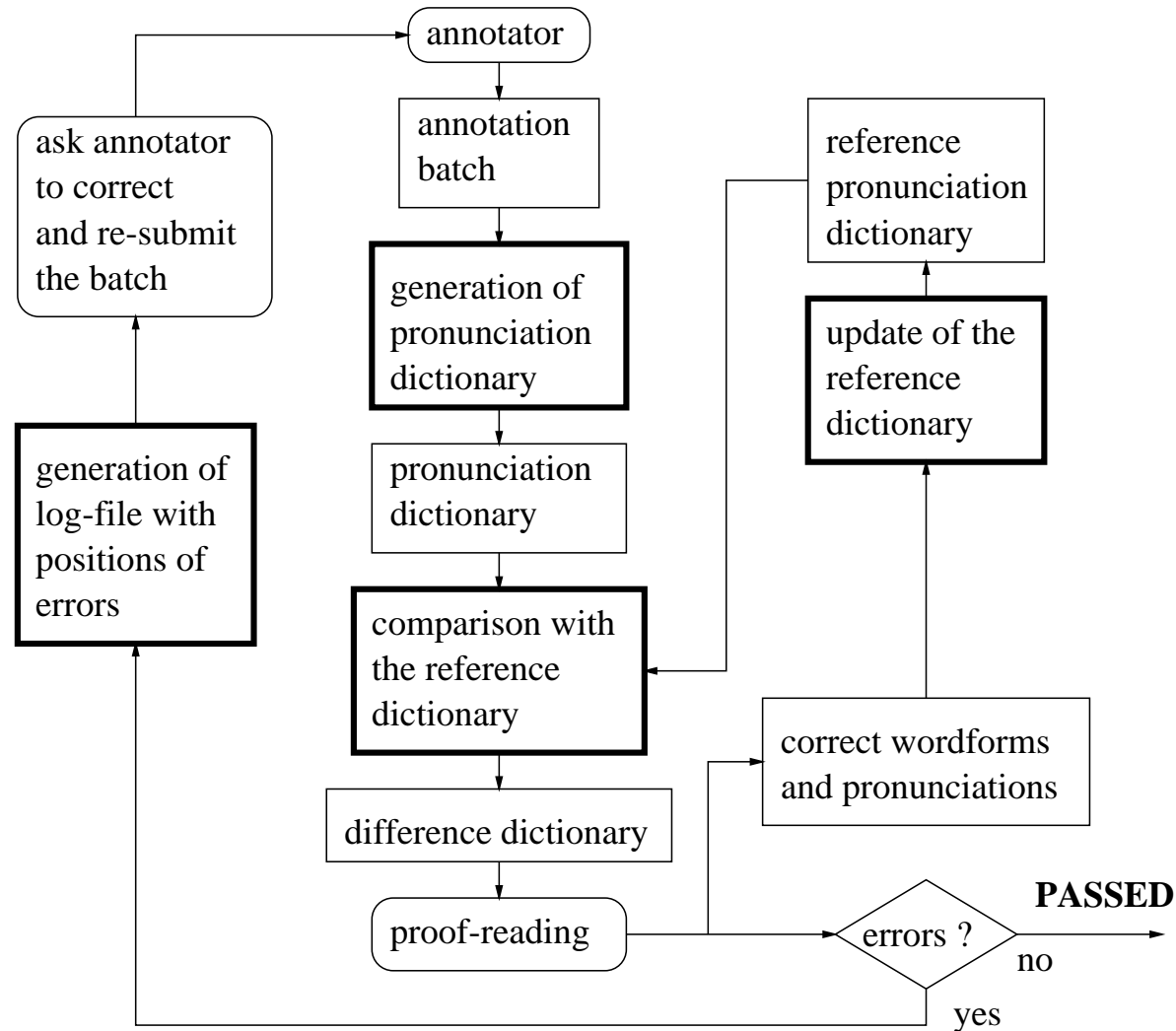
Phonetic transcript:
[spk] navíc sy myslým že stát by mñel kAzy medyjalyzovat co nejméñe

Comment:

Help:
F1 - Help F5 - speaker noise [spk] F9 - *abc Ctrl-X - abc -> (abc/abc)
F2 - OK and next F6 - intermittent noise [int] F10 - ** Ctrl-F7 - sedm -> (sedm/sedum)
F3 - Back F7 - filled noise [fil] F11 - ~abc Ctrl-F8 - osm -> (osm/osum)
F4 - Play F8 - stationary noise [sta] F12 - abc~ Shift-Ctrl-F7 - sedm -> sedum

FTP state: Disconnect Call: 11/10 File: 521/520

DB - anotace - kontrola

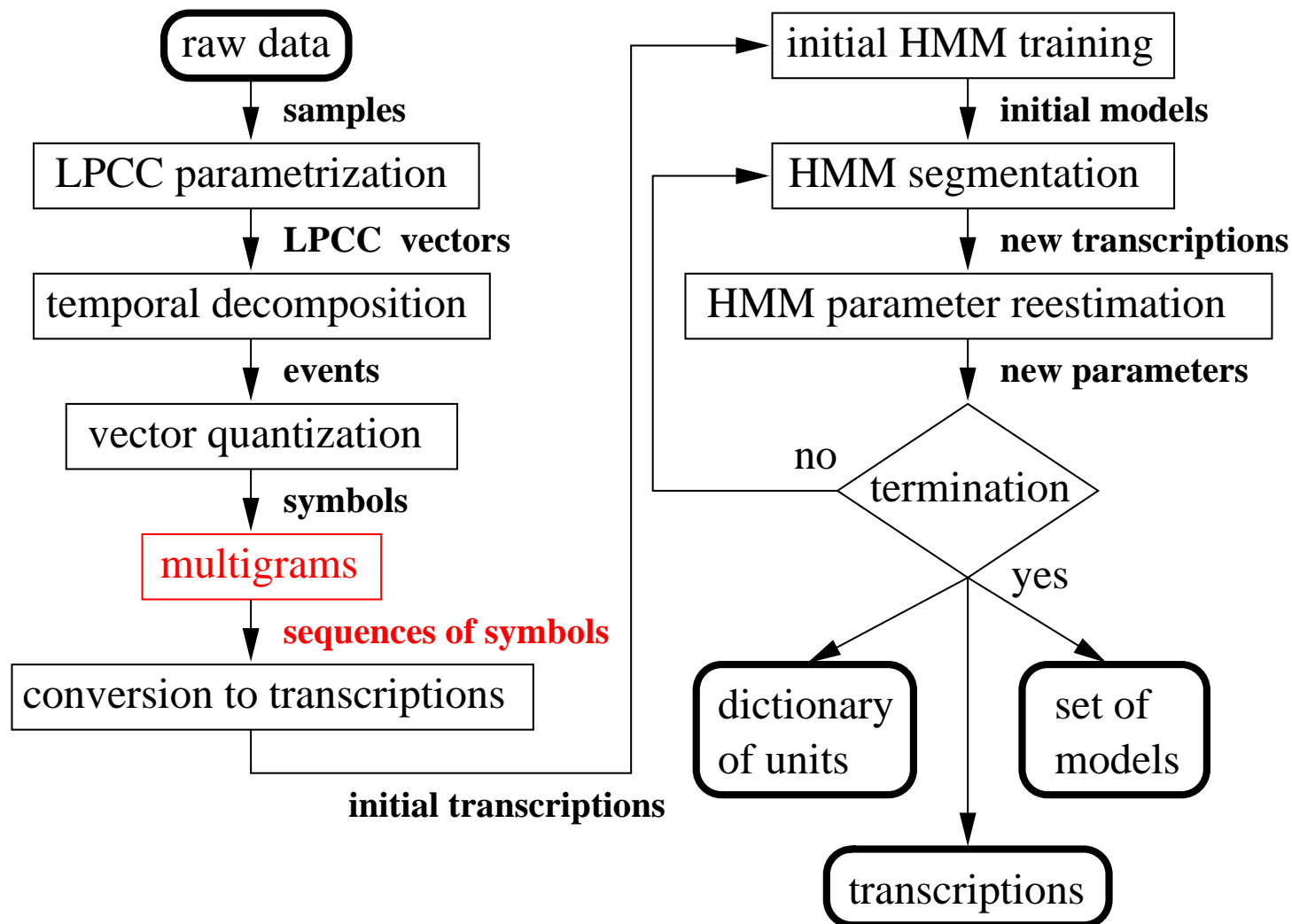


Automaticky naučené řečové jednotky

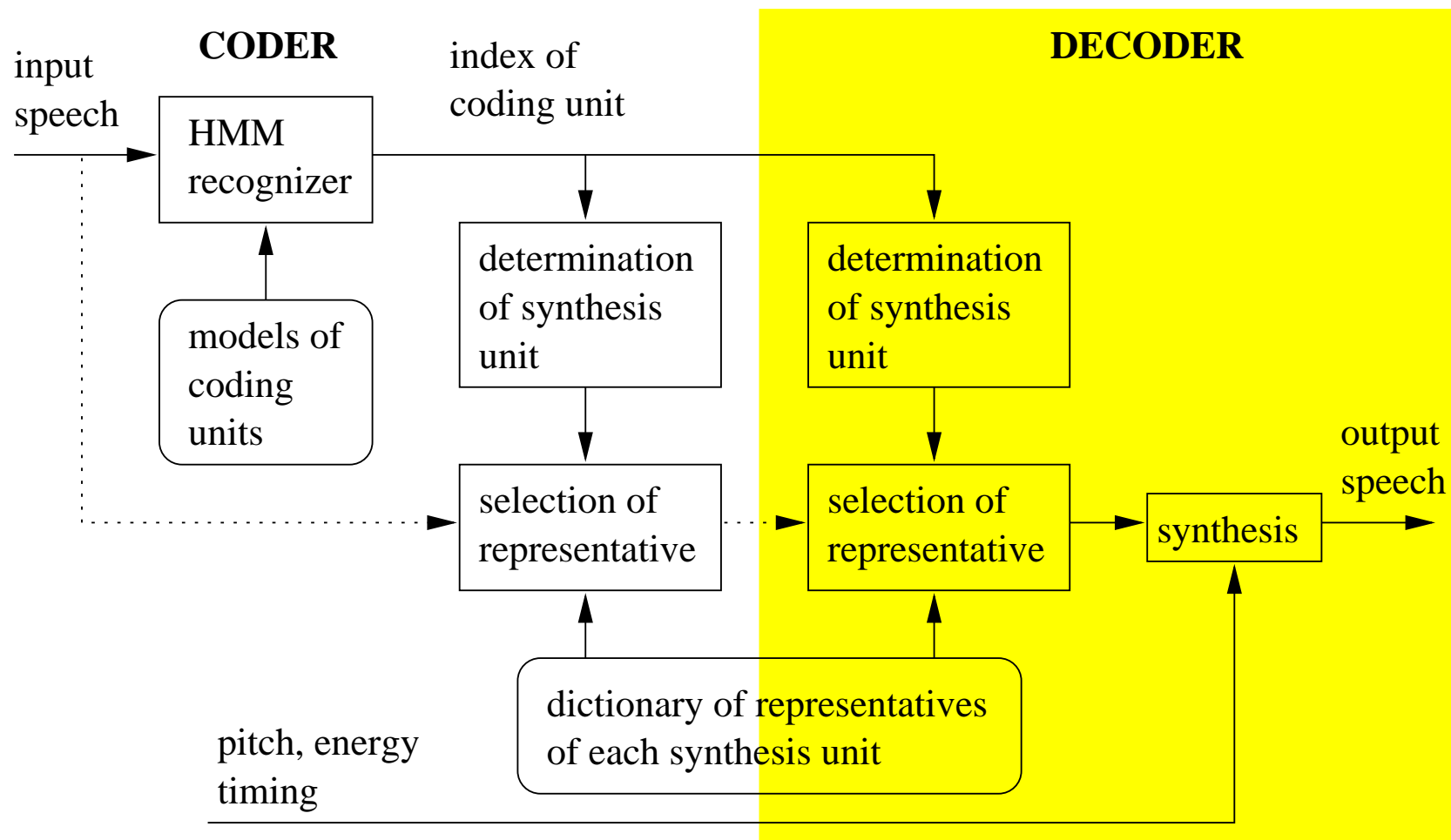
Proč?

- DB projekty mají rozpočty v milionech, jednodušší by bylo zapíchnout počítač do rádia a jen nahrávat.
- jenže pro trénování klasických řečových jednotek — fonémy, trifóny — potřebujeme **anotace** (nejvíce peněz a největší opruz. . .).
- pro některé aplikace (kódování, identifikace jazyka) nepotřebujeme vůbec souvislost s textem.
- ⇒ daty řízené metody pro učení jednotek.

Návrh, jak na to (Černockého PhD)

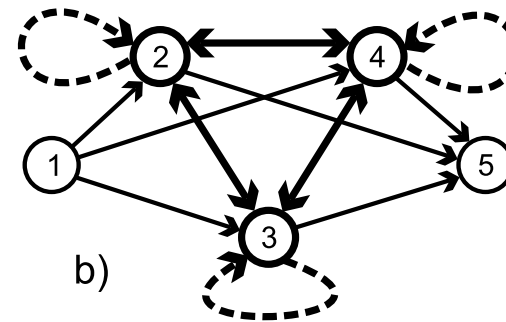
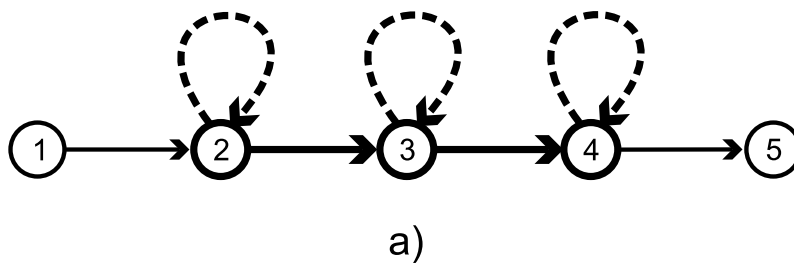


Aplikace I. – Kódování VLBR



Pokračování

- Petr Motlíček - minimalizace přechodů mezi jednotkami - diphone-like jednotky.
- Igor Szoke - HNM syntéza, EHMM pro koherentnější postup trénování jednotek.



<http://www.fit.vutbr.cz/~szoke/speech/index.html>

Problém: na toto nám nikdo nedává peníze :-)

EU projekty – M4

Multimodal Meeting Manager - EU IST 5th PCRD (10 partnerů z Evropy a USA)

- Development of a "smart" meeting room, collection and annotation of a multimodal meetings database.
- Analysis and processing of the audio and video streams.
- Integration and structuring using the output of the various recognizers.
- Demonstrator.

Speech@Brno tasks:

- down-scaled meeting room with hyperbolic mirror - data collection and annotation.
- LVCSR, phoneme recognition, feature extraction.

<http://www.m4project.org>

EU projekty – AMI

Augmented MultiParty Interaction - EU IST 6th PCRD (16 partnerů z Evropy a USA včetně průmyslových (např. Philips Smart Display) a W3C).

- Multimodal input interface
- Integration of modalities and coordination among modalities
- Meeting dynamics and human-human interaction modelling
- Content abstraction (multimodal information indexing, summarising, and retrieval)
- Technology transfer
- Training activities, including an international exchange programme.

Speech@Brno task:

- Keyword detection / Acoustic event spotting (with supporting technologies)

<http://www.amiproject.org>