

# **Potřebují informatici umět komunikovat?**

Předmět Základy odborného stylu  
(VB000)

K. Pala, Centrum ZPJ FI MU  
jaro 2015

# Pozitivní komunikace

- Výchozí axiom: nelze nekomunikovat
- Naše civilizace stojí na komunikaci – je to základ pro vědu, techniku, kulturu, veškeré znalosti lidstva
- Musí splňovat jisté standardy, aby byla efektivní
- Pro komunikaci potřebujeme vhodná pravidla, normy
- Většina lidské komunikace probíhá v přirozeném jazyce
- Má dvě základní podoby: mluvenou a psanou
- Která podoba je prvotní?
- Potřeba norem pro přechod od mluvené podoby k psané

# Mluvený a psaný jazyk

- Mluvený jazyk je primární (cca 5000 jazyků světa)
- Psaný jazyk je až sekundární – je to paměť lidstva
- Komunikace v přirozeném jaz. vyžaduje přiřazení zvuků jazyka (hlásek) psaným znakům (písmenům)
- Úplná reprezentace zvuků v jednotlivých jazycích – International Phonetic Alphabet (IPA)
- Reprezentace zvuků – různé pravopisné systémy
- Jsou základem veškeré civilizace a kultury obecně
- Abecední písma – latinka, cyrilice a mnohá další
- Ideografická písma (logografická) – čínština, cca 50 tis. znaků (základ 9-12 tis. znaků)
- Slabičná písma – japonština, korejština

# Přiřazení zvuků znakům

- Kolik hlásek (fonémů) má čeština?
- Kolik písmen má čeština?
- Kolik hlásek (fonémů) je v angličtině?
- Kolik písmen je v angličtině?
- 40-42: 36 (piš, jak slyšíš)
- 40-44: 26
- Uvedená čísla naznačují povahu daného pravopisného systému – fonetický (čeština) vs. historický (angličtina)
- Spřežkové a diakritické systémy (v češtině)

# Funkce pravopisných systémů

- **Zaznamenávací** – aby se to dobře psalo
- **Vybavovací** – aby se to dobře četlo
- Která funkce je důležitější? Vyváženost?
- Posílíme-li jednu funkci, oslabíme druhou a naopak
- Jaká je situace v praxi?
- Historické systémy (typicky angličtina, 14. stol.)
- Fonetické systémy (čeština, slovenština)
- Který systém je obtížnější k naučení?
- Spřežkové systémy – stará čeština, dnes zčásti i polština
- Lze získat přes milion liber za úspěšnou reformu anglického pravopisu? (nadace Bernarda Shawa, 1910)
- Pravopisné systémy jsou velmi konzervativní

# Pravopisné chyby

- Jazyková správnost a gramotnost – podmínka úspěšné komunikace a profesionální úspěšnosti obecně
- Pravopis není gramatika (popis struktury jazyka)
- V čem se nejvíc chybuje? (styl, interpunkce)
- Typy pravopisných chyb
- Překlepy (*prgram, studijní, ...*)
- Morfologické chyby – koncovky (*hloupejma*)
- Syntaktické chyby – shoda (psaní y/i), vazby (valence)
- Styl a stylistické chyby – *provedení nařízení*
- Typografické chyby – mezery, pomlčky, spojovníky, uvozovky, fonty, jednopísmenové předložky na konci řádků
- Spisovnost : nespisovnost, formálnost : neformálnost

# Opravování chyb v textech

- Snahou je, aby v našich textech bylo málo chyb
- Lidé – korektoři nejsou dokonalí
- Softwarové nástroje odhalující chyby (překlepy) v textech: různé typy korektorů
- Kde je lze najít? – v produktech, jako jsou Microsoft Office (Word), Open Office aj.
- Co dovedou? Mohou být chytřejší než uživatel?
- Opravují jednotlivá slova v textu obsahující chyby jako (*chitří, právda, mluvým, ...*)
- Korektory překlepů opravují jen jednotlivá slova, slovní spojení (v kontextu) *šikovná student nikoli*

# Pravopisné korektory

- Jak jsou tyto nástroje konstruovány?
- Hrubá síla nebo inteligence?
- Kolik slov má čeština? Kolik slovních tvarů?
- PSJČ – cca 250 000 základních tvarů slov
- Slovních tvarů v češtině je kolem 60 milionů
- Morfologická analýza a morfologický analyzátor – může sloužit jako korektor překlepů
- Pro češtinu – morf. analyzátor majka, cca 400 000 českých kmenů (ukázka v nástroji DebDict, vytvořen v Centru ZPJ, používá jej firma Seznam)



# Gramatické korektory

- Dovedou opravovat slovní spojení v kontextu, např. *studentk šel do školi když pršeló.*
- Na rozdíl od korektoru překlepů gram. korektor podtrhává zeleně a opravy jen doporučuje
- Chyby v gramatické shodě a vazbách
- Upozorňují také na chyby v interpunkci
- Povaha české interpunkce je syntaktická
- Gramatický korektor k dispozici jen v české verzi Wordu (Oliva, Květoň, Petkevič)
- Aplikace Grammaricon od Lingey – kvalita?

# Gramatické korektory

- Co umí nástroj Grammaticon (od firmy Lingea)  
– falešné hlášky
- Co gramatické korektory nedovedou?
- Nakolik se na ně lze spolehnout?
- Evaluační parametry: přesnost a pokrytí
- Přesnost (úspěšnost) se dnes pohybuje do 70 %, jde o velmi těžkou úlohu
- Principy fungování gramatických korektorů?
- Automatická syntaktická analýza – parsery (Set),
- Plus další vhodná heuristická pravidla

# Nebojme se pravidel českého pravopisu

- Knižní podoba Pravidel českého pravopisu
- Internetová jazyková příručka – ÚJČ a FI MU
- Má dvě části – slovníkovou a normativní
- Slovníková část pokrývá cca 60 000 čes. slov
- Běží na serveru Centra ZPJ, denně více než 30 tis. přístupů (ukázka)
- Příručka obsahuje automatickou morfologii a je nově doplněna o dva normativní slovníky
- Používá se ve standardně nejen ve školách

# Pravidla českého pravopisu

- Je český pravopis obtížný?
- Kombinace fonetického a historického principu, fonetický převažuje
- V čem se nejvíc chybuje?
- Stylistické chyby – cca 23 %, jde tu o formulaci myšlenek, ta je nesnadná
- Interpunkční chyby – asi 20 %
- Ostatní – překlepy, y/i, velká písmena, tvary mne/mě, typografické chyby aj.
- Reformy českého pravopisu

# Současný stav I

- Poslední reforma čes. pravop. proběhla v r. 1993
- Pokus o tzv. demokratizaci pravopisu?
- Úprava psaní slov cizího původu (kurs/kurz)
- Ref. vedla k malé pravopisné válce (filos/zofie)
- Výsledek: špinavý kompromis (fonetizace)
- (ne)úspěšnost reformy si lze ověřit na datech – dnes na velkých souborech textů – korpusech

Jaký je tedy současný stav? – je vidět, že norma je rozkolísaná – nejhorší možný výsledek

# Současný stav II

- Korpusy (CzTenTen12) to potvrzují  
kurs: 65,360 vs. kurz: 581,913, feminismus:  
11,895 vs. feminizmus: 332
- Rozkolísanost standardů je obecně nežádoucí,  
komplikuje plynulost komunikace, reforma??
- Bezbolestná úprava by byla ú/ů
- Za problém se pokládají velká písmena,  
volnost je značná
- Pokud jde o y/i, situace pro úpravu není zralá
- Případná inspirace slovenštinou

# Vyhlídky a perspektivy

- Institucí, která se stará o jazykovou kulturu v ČR, je Ústav pro jazyk český AV ČR
- Komunikuje s veřejností prostřednictvím Jazykové poradny (a nyní též IJP)
- Jejich aktuální přístup k problematice českého pravopisu je spíše liberální
- Lze očekávat nějaké reformy?
- Pro absolventy FI je potřebné, aby se v čes. prav. standardně orientovali a psali bez chyb

# Další nástroje pro práci s jazykem

- Elektronické slovníky české – nástroj DebDict
- Vícejazyčné slovníky – např. produkty firmy Lingea
- Google Translator
- České překladače: Eurotran, PC Translator – málo kvalitní
- Problematika strojového překladu
- Morfologické a syntaktické analyzátoři
- Dialogové systémy
- Porozumění přirozenému jazyku (extrakce inf.)



# Počítače a PJ

- Komunikace mezi člověkem a počítačem je dnes primárně jednocestná
- Její kvalita ve skutečnosti závisí na tom, jak dobře uživatel zná programové vybavení svého počítače (jeho OS)
- Počítače s námi zatím nedovedou komunikovat
- Potřeba dvoucestnosti – zde se dostáváme k umělé inteligenci – její součástí je mj. zpracování PJ (modelování funkcí lidského mozku)

# Počítače a PJ 2

- Reprezentace znalostí o světě, inference (logika), zvládnutí PJ (gramatika) – tři součásti
- Dialogové systémy je musí obsahovat
- Turingův test, Eliza, chatboty, každoroční soutěž o Loebnerovu cenu, roboty
- Zpracování mluvené řeči – diktovací systémy
- Dovedou přepisovat zvuky na znaky – pro češtinu: Newton Technologies, Dictate 4
- U těchto systémů zatím nejde o porozumění PJ

# Mgr. Jiří Materna, Seznam

- **Vysokou školu jste neprošel pouze u počítače, ale zahrál jste si například i ve fakultním divadle. Jaké to bylo, když se informatici pustili do divadla?**

Celá řada lidí si myslí, že divadlo na fakultě informatiky nemá co dělat, ale já jsem přesvědčený o opak. Problémem českého školství je obrovská pasivita studentů a jejich neschopnost prezentovat své myšlenky na veřejnosti. Je to dáno především tím, že k tomu nejsou vedeni již od útlého věku. Často