

- **Přednáška #1 07/10/02**

- **Úvodní info**

- Bude nám možná chybět něco z
 - Teorie automatů
 - Statistiky/Pravděpodobnost
- Zkouška
 - Písemná: jedna velká otázka a několik malých (vysvětlení základních pojmů na pár vět).
- ***P íšt v bec není p ednáška (všichni lingvisti jsou na Malt).***
- Co je to *po íta ová* lingvistika?
 - Nemusí být nutně na počítači.
 - Sgall: *komputa ní* lingvistika (angl. computational linguistics)
 - 50. léta: *matematická* lingvistika
 - 60., 70. léta: *statistická* lingvistika (skutečně statistika frekvence hlásek apod., něco jiného než teď, kdy se statistické metody také používají, ale neříká se tomu tak – *natural language processing*)
- Podobory a aplikace
 - Morfologie – tvary
 - Syntaxe – stavba vět
 - Sémantika – vazby ve sdělení
 - Automatický překlad – kombinuje různé jiné obory
 - Rozpoznání mluvené řeči (ve spojení s jinými obory – fyzikou(?))
 - Vyhledávání, zpracování textů, zjišťování klíčových slov
 - Psycholingvistika
 - Korpusy – mají u nás velkou tradici
 - Kontrola překlepů a (do jisté míry) i gramatiky.
- Literatura
 - Úvod do teoretické ... – zatím existuje jen první svazek, druhý v přípravě
- Pozor na *bramborové* a *švestkové* knedlíky!
- Přirozený jazyk má problém víceznačnosti. Syntaxe nám nesděljuje vše.
- Volný pořádek slov v češtině může ovlivňovat význam.
- „garden-path sentences“ (kde posledních pár slov mění význam)
- Páry vycházejí z lesa.
- Tatáž přednáška bude i v angličtině (více méně totéž, stejný kód), nejdřív koncem října.

- **Erasmus Mundus**

- Nejen po Evropě (tzn. i neevropští studenti, ale školy jsou evropské)
- Týká se magisterského studia, dostane tituly z obou škol (na každý studuje rok).
- LCT – Language and Com. Tech.
 - Saarbrücken – Německo (velmi dobrá lingvistika i informatika)
 - Groningen – Nizozemí
 - Bolzano – It.
 - Nancy – Fr.
 - (Malta – ne)
- V lednu před přihláškou do mag. studia. Celý program v **angli tin** .
- Stipendium v Erasmu asi 300-400.
- Jedna diplomová práce a pokud možno jedna obhajoba (tam nebo tady) plus státnice v ČR.

- **Přednáška #3 07/10/16 (druhá nebyla)**

- **Slovníček**

- Pády
 - 1. nominativ
 - 2. genitiv
 - 3. dativ

- 4. akuzativ
- 5. vokativ
- 6. lokál/lokativ
- 7. instrumentál
- Číslo: sg., pl., příp. duál
- Rody: masculinum (inanimatum), animatum (mužský životný), f., n.
- Druhy: substantivum, adjektivum, pronomen, numerale, verbum, adverbium, prepozice (příp. postpozice), konjunkce (spojka), partikule, interjekce (citosl.)
- Větné členy: subjekt, predikát (přísudek), objekt, adverbiale (př. určení, např. adverbiale loci), atribut (přívlastek)
- Deklinace – skloňování, konjugace – časování
- Futurum, presens, participium, perfektum, pluskvamperfektum ...
- Aspekt – vid: imperfektiva/perfektiva
- Aktivum/pasivum
- Indikativ, imperativ, kondicionál
- **Zpracování textu**
 - Pozn.: Nebudeme se zabývat rozpoznáváním mluvené řeči, soustředíme se na text.
 - Nejvíce u nás na univerzitách v Plzni a Liberci (MFF spolupracuje s Plzní)
 - **Definice.** *Morfém* je nejmenší jednotka nesoucí význam (kmen, předpona, přípona, koncovka).
 - Kmen (a předpona, přípona?) = morfém *lexikální*, koncovka = morfém *gramatický*
 - Jazyky morfologii mohou mít a nemusí (v různé míře)
 - aglutinační (finština, maďarština) – nejbohatší
 - čeština – velmi bohatá
 - angličtina – prakticky bez morfologie
 - **Definice.** *Alternace* – změna v morfému při ohýbání, např. dům (domu), švec (ševce), ..., nejvíce alternací: matka (matek, matce, matčin, možná ?matička, ?matečná, ?mateřská)
 - **Definice.** *Alomorf* – varianta morfému
 - Morfologická analýza
 - př.: otec: {lemma: otec, subst., anim., nom., sg., zákl. tvar}, (lemma: otéci, ...)}
 - Analyzátor prof. Hajiče(?): pevná délka značka 15 znaků
 - př: *nejnezajímav jším*: {lemma: zajímavý, popis: "AAFP3----3N----", kde - jsou mezery}
 - Jednotlivé pozice:
 - 1. druh (Adjektivum)
 - 2. poddruh
 - 3. rod (Femininum)
 - 4. číslo (Plurál)
 - 5. pád (dativ (3))
 - ...
 - 3 = stupeň
 - N = „ne“
 - např. slovesná osoba (tady mezera),
 - Značkování – *tagging*. Vybírá takovou (*jednu*) značku, která odpovídá kontextu (*desambiguace* – to je fakt česky)
 - Úspěšnost (včetně „jasných“ slov)
 - Pro angličtinu: 98-99 %
 - Pro češtinu: 95-96 % (u nejasných asi 80 %, čili jeden případ z pěti chyba)
 - Nehodí se pro kontrolu pravopisu (vybere jednu variantu, splete se, ...)
 - Přístupy
 - statistický
 - syntaktický (vyřadí se syntaktické nesmysly, nevede nutně jen na jednu variantu)
 - kombinace? – marginální zlepšení, syntakticky špatné bývá statisticky málo významné, ale dá se využít pro kontrolu *gramatiky*(!).
 - quest.ms.mff.cuni.cz
 - **Přednáška #4 07/10/23**
 - **Analýza resp. lemmatizace vs. generování**

- *Lemmatizace* se dá použít pro vyhledávání (zlemmatizuju texty).
- Stejně tak se dá použít *generování* (vygeneruju si tvary od zadaného lemmatu a ty hledám v nezměněném textu).
- *Lemmatizace/Analýza*
 - Hledám rozdělení daného tvaru na platnou předponu, lemma, příponu, koncovku (odpovídající vzoru).
- *Generování*
 - Vyberu k lemmatu správnou koncovku (podle jeho vzoru).
 - Přístup bez vzorů
 - Tvary se generují bez použití vzorů. Pomocí transformací podle pravidel.
 - Př. Two-level morphology z 80. let, Finové, **Xerox**, (Karttunen, K~)
- **Morfologická typologie jazyků**

V tšinu jazyk nejde jednoduše za adit,

 - Izolační (analytické)
 - 1:1 = morfém : slovo
 - Příklady: čínština, vietnamština, částečně angličtina (množná čísla, tvary sloves, ...)
 - Kontrola spellingu pomocí prostého výčtu povolených slov.
 - Aglutinační
 - Vrstvení koncovek
 - Př.: ugrofinské, částečně japonština
 - Flektivní
 - Jedna koncovka nesoucí více informací.
 - Př.: stará řečtina, latina, slovanské, část. germánské
 - Polysyntetické (někdy se neuvádí)
 - Spojování více kmenů do jednoho tvaru.
 - Př.: eskymácké, pův. indiánské
- **Aplikace morfologie**
 - Kontrola překlepů
 - Hledání chyb
 - Běžný přístup: Lemmata + koncovky, obdobně jako morfologická analýza
 - Alternativní přístupy:
 - Hledání nepovolených kombinací hlásek v daném jazyce – neodhalí všechny chyby.
 - Dokáže odhalit chyby i v neznámých slovech.
 - Návrhy oprav
 - Pomocí metrik (typicky se nabízí opravy se vzdáleností 1):
 - Levenshtein
 - počítá se vložení, odstranění písmene, přehození dvou sousedních, nahrazení
 - Vylepšení: započtení pravděpodobnosti chyb (blízkost písmen na klávesnici apod.)
 - Přímé opravy (bez ptaní):
 - Seznam známých (i syntaktických/gramatických chyb)
 - Např. „ke mě“ → „ke mně“ (téměř vždy správná oprava) – standardně v MS Wordu.
 - Aplikace vyvinuté na Matfyzu
 - ASIMUT

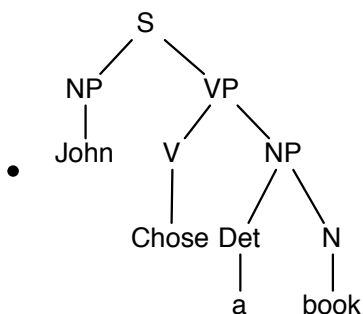
Automatická Selekce Informací Metodou Úplného Textu

 - 80. léta, publikace (brožurky) v 1990.
 - Neexistovala morfologická analýza, ale bylo třeba vyhledávat v databázích textů.
 - Dvě části
 - Vyhledávací část (nezajímavá)
 - Dotaz uživatele (např.): *vzdálenost! , odstup! -3- rodinný! -1- domek!*
 - -3- znamená v téže větě (jiná č.: vedle sebe, blízko sebe, v odstavci, ...)
 - čárka nebo
 - ! skloňování
 - Žádná kontrola gramatiky (př. rodinný domku)
 - Předpokládá se strukturovanost textu (kapitoly, odstavce, věty).

- Jazykový modul
 - Základ: Retrogradní slovník dr. Slavíčkové)
 - 40 tis. hesel
 - Společné vlastnosti stejně končících slov – spolehlivé u podst. a příd. jmen
 - Příklady slov:
 - -ý: tvrdé přídavné jméno
 - Ale úterý, prý, čehý
 - vydáno 1975
 - Příklad:
 - cokoli na
 - *hromosvod*
 - Podle brožurky:
 - w = hromosvod
 - konec(w) = 2: NE
 - konec(w) = a, t, k, e, ...: NE
 - Kontrola asi 30 výjimek jako dikobraz – nic
 - Připiš vzor N1 – hrad, kámen, sklep
 - Kódování diakritiky: čárka (2), háček (3), kroužek (7), ostatní čísla pro ruštinu apod.
- **Přednáška #5 07/10/30**
 - **Aplikace morfologie / Aplikace vyvinuté na Matfyzu (pokračování):**
 - Problémy ASIMUTu
 - Vzor nejde vždy poznat podle koncovky
 - Příklad: -ík: životné i neživotné (zhruba stejně početné, a to obě dost) (právník/trávník)
 - Proto je zbytečné dělit vše zbytečně najemno → raději se koncovky přegenerovávají
 - Nikdy nebyl příliš využíván
 - Vznikal na sálových počítačích, později byly k dispozici jiné morfologické slovníky (90. léta).
 - Pokus o časování sloves: nepřilíš úspěšný – příliš velká víceznačnost → větší přegenerování
 - Navíc ne tak důležité jako jména.
 - MOSAIC / MOZAIKA (v angl. akronym Morphological ... Condensation)
 - podobný asimutu
 - vybírá klíčová slova a třídí podle důležitosti ze strukturovaného (kapitoly, odstavce, ...) textu
 - ideálně vybírá i úseky textu („condensation“)
 - zakončení slova (nejen kmen) může nést sémantickou informaci
 - en: -ity, -ness; -er, -or
 - cs: -graf -skop -metr (u cizích slov – často přístroje), -ič, -ač, -čka, -ér (české často přístroje); -ací, -ecí (účely) apod.
 - důležité pojmy používají v textu vícekrát, ale v různých formách
 - „Operační zesilovač Tesla KC-asdasdasdafskgjlskdj; tento operační zesilovač; zesilovač, ...“
 - Je třeba nejen lemmatizaci, ale i nějakou syntaktickou analýzu pro rozeznání delších termínů a jejich vzájemného přiřazení.
 - Přiřazení důležitosti podle pozice ve struktuře
 - Obyčejný odstavec uprostřed: skóre 1, odstavce na konci/začátku kapitoly/textu → vyšší skóre (při každém „kroku“ se zdvojnásobuje.
 - Normalizace (nezávislost na délce textu): absolutní počty bodů se upraví, aby ten nejlepší pojem měl 100.
 - Autor: dr. Kirschner (dnes v důchodu), stejná doba jako ASIMUT. Implementace v 80. letech, pokusy na právnických a později hornických textech (hu?!). Zase je na to brožurka (1983).
 - Pro celý jazyk (cs) by bylo třeba databáze asi tisíce zakončení slov.
 - (Dnes jediná) výhoda: Rozpoznání slov, která nejsou ve slovníku (nových).
 - Dnes by se použil nějaký statistický přístup.
 - **Syntaxe**
 - Používané datové typy
 - Závislostní strom
 - Starší přístup
 - Panini

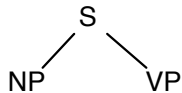
- Jako se kreslí v češtině na základce/gymplu, jen podmět se dává „pod“ přísudek.
- Existují různé varianty.
- Pár slov: Řídící (nahore) a závislé (dole) slovo.
- Uzly: Slova, jako atribut (uzlu nebo hrany) jeho role.
- Předložky tvoří obvykle s podstatným jménem jednu jednotku stejně slovesné tvary.
- Naopak lze i dělat uzly pro vše – včetně interpunkce.
- Problém: Koordinace, př. „Petr a Pavel“ – závislost? – třeba zvolit reprezentaci ve stromu.
 - Možnost 1: Zvolit koordinační spojku jako řídící (dá se na ní pověsit třeba Petr, Pavel i Jiří ze spojení „Petr, Pavel i Jiří“), to ale zcela nemusí neodpovídat tomu, co je skutečně v jednotce podstatné.
 - Možnost 2: Řídící je to, co je poslední – i tak je ale třeba dodat informaci o tom, že jednotku „vlastně“ spolupřidává i ostatní závislá slova...
- Problém: Předložkové vazby (předložky závislé na jménech nebo na opak?)
- Složkový strom
 - N. Chomsky
 - Derivační strom(?) – Automaty a překladače (?)
 - Příklad
 - „Petr šel včera do kina.“
 - S → NP VP (noun phrase, verb phrase)
 - NP → N (noun, v tomto případě nerozvitě)
 - VP → VP PP (prepos. phrase)
 - VP → V Adv
 - PP → Prep N
 - V → „šel“ (terminální)
 - N → „Petr“, „kino“ (terminální)
 - Prep → „kdo“ (terminální)
 - Adv → „včera“ (terminální)
 - Nezachycuje se shoda členů.
 - Poměrně složitá reprezentace stromem (mnoho neterminálních uzlů)
- Do jisté míry vzájemně převoditelné (ve složkovém chybí závislost).
 - Např. závislost se dá ve složkové rep. kreslit šikmo (jinak vertikálně, třeba V → VP)
- Výhoda (?) závislostního stromu
 - Vhodnost pro jazyky, kde tolik nezáleží na pořadí.
 - „Soubor se nepodařilo otevřít.“
 - Nepodařilo se → otevřít → soubor (cik cak „neprojektivní konstrukce“)
 - Někdy se mohou závislosti i protnout, pokud dodržíme pořádek slov.
 - Dá se buď rozkřížit bez ohledu na pořadí nebo nechat.
 - Nedá se ale vůbec analyzovat bezkontextovou gramatikou (složkový strom), která spojuje, co je za sebou. – Musí se používat tech. finty.
- Ani jeden z datových typů není ideální, ale dají se nějak použít oba.
- **Přednáška #6 07/11/06**
 - **Syntaxe (pokračování)**
 - Ještě neprojektivní konstrukce (které se blbě zapisují do složkového stromu)
 - „Soubor se nepodařilo otevřít.“
 - „Vánoční nadešel čas.“
 - V Pražském závisl. korpusu: asi 4 % hran neprojektivních, takže asi 15 % neprojektivních vět. Problém.
 - Neprojektivní ve větě může být i více:
 - Tuto knihu jsem se mu rozhodl dát.
 - Neprojektivní vazby:
 - dát: knihu
 - dát: mu
 - „Vzor“: řídící sloveso a infinitiv s vazbami → neprojektivní konstrukce, jde jich takhle vytvořit v češtině dost.
 - Mnoho závislých členů může mít například sloveso *p ed lat* (co, na co, komu).

- Pozn.:
 - *Tuto knihu jsem se mu rozhodl dát v pátek.*
 - „v pátek“ nemusí být nutně neprojektivita (může se vztahovat k *dát*)
- V angličtině se zdá (prý), že mohou být neprojektivity nejvýše dvě (a ještě nezávislé).
- Ještě větší problém, pokud chceme brát v úvahu i chybné věty (korektura pravopisu apod.):
 - *Které d v ata cht la dostat ovoce?* (Která děvčata, nebo které ovoce?)
 - Výsledek: Dobře neprojektivně vs. špatně projektivně.
 -
- Neprojektivity brání ve zpracování bezkontextovou gramatikou.
-
- Historie
 - Deskriptivismus
 - Bloomfield 1933
 - Především skutečně popis, především mizejících (indiánských) jazyků
 - Analytická syntax
 - Jespersen 1937
 - Logický přístup
 - Ajdukiewicz 1935
 - Transformační gramatika
 - Noam Chomsky 1957: Syntactic Structures (není rodilý Američan → Chomsky/Čomsky)
 - Idea: Příbuzné jevy, například odvozenost tázací věty od oznamovací.
 - Komponenty
 - ?Fonologická/fonetická (nezajímá nás)
 - Generativní
 - Generuje z věty strom
 - Věta se rozdělí na fráze → (složkovému) stromu se říká phrase marker, frázový ukazatel
 - Transformační
 - John chose a book.



- Transformace na *A book was chosen by John.*
 - strukturní index
 - strukturní změna
 - NP1 - V - NP2 ⇒ NP2 - was - V+en - by + NP1
- Transformační teorie je ale příliš silná, transformace má sílu Turingova stroje. ⇒ Změny.
- 1965, Chomsky: Aspects of the theory of Syntax
 - Standard theory
 - 1968 Extended standard theory
 - 80. léta: Government-Binding (GB, nadřazení a vázání)
 - Myšlenka: univerzální gramatika, kterou máme všichni vrozenou, jednotlivé jazyky této univerzální gramaticy jen dodávají parametry.
 - zač. 90. let: Teorie minimalismu (zatím platná)
 -
- Dělení (z dřívějšíka, ale Chomsky začlenil do své teorie)
 - povrchová rovina (*analytická*)
 - V povrchové rovině může být ve větě vynechaný podmět apod.

- hloubková rovina (*tektogramatická*)
 - Věta musí obsahovat vše, co má.
 - Jednomu zápisu na povrchové rovině může odpovídat několik zápisů na hloubkové a naopak.
- Jiná alternativní teorie: Lexical Function Grammar (LFG)
 - Kaplan, Bresnanová
 - rozdělení
 - c-structure (constitutional) – složkový strom
 - f-structure (functional??) – sturture (feature structure – atributy, př. Lex. hodnota, číslo apod.)
 - Pozn
- Jiná alternativní teorie: Tree Adjoining Grammars
 - prof. Joshi
 - Množina elementárních stromů, např.:



- Množina pravidel na spojování stromů.
- Jak ale dělat vlastní analýzu?
 - Čistá bezkontextový gramatika
 - Nezachycuje shodu. Je třeba upravit/rozšířit apod. Samotná prepisovací pravidla nestačí.
 - zač. 70. let, Woods: Rozšířené přechodové sítě (Augmented Transition Networks)
 - Příklad
 - The girl saw a boy
 - Gramatika (pro ilustraci)
 - $S \rightarrow NP VP$
 - $NP \rightarrow Det NO$
 - $VP \rightarrow V [NP]$
 - Sítě stavů
 - $s_0 \rightarrow seek NP \rightarrow s_1 \rightarrow seek VP \rightarrow s_1$
 - $np_0 \rightarrow cat Det \rightarrow np_1 \rightarrow cat N \rightarrow np_2$
 - $vp_0 \rightarrow cat V \rightarrow vp_1 \rightarrow (seek NP \mid jump) \rightarrow vp_2$
 - V podstatě přepis bezkontextové gramatiky(?).
 - Princip grafového analyzátoru (chart parser).
 - Alain Colmerauer: Q-systémy
 - Formalismus pro jeden z prvních sys. automatického překladu (METEO, pro meteorologické zprávy).
 - Jediná nevýhoda oproti (trochu pozdějším) rozšířeným přechodovým sítím: Špatně prezentován. Publikace zprvu jen ve francouzštině – nikdo to nečetl! (→ Převálcované sítěmi).
 - Vlastní Q-systémy příště.
- **Přednáška #7 07/11/13 – chyběl jsem**
- **Přednáška #8 07/11/20**
 - **Funkční generativní popis**
 - Sgall, Hajičová, Panevová, (?)Piřka(?)
 - Pražská škola, Roman Jakobson
 - Snaha o objektivní lingvistiku bez „názorových“/„pocitových“ dohadů
 - Stratifikační teorie, (5, případně 4) rovin
 - tektogramatická rovina (povrchová, syntaktická)
 - rovina větné vazby (někdy se vypouští)
 - morfematická r.
 - fonologická r.
 - fonetická r.
 - Důraz na valence/vazby (~předmět u slovesa)
 - V tektogramatické rovině

- Agens/Aktor (obvykle odpovídá podmětu ve větněvazebné rovině)
- Patiens (obvykle předmět ve VV rovině)
- Adresát
- Origo
- Efekt
- Aktant
- Volná doplnění
- Vazby: povinné/nepovinné (ne zcela zřejmé, třeba „telefonovat [komu?]“)
- Valenční slovník (počítačově použitelný) až v 90. letech:
 - „Valex“ od Lopatkové a Žabokrtského, bude vydán knižně
 - V Brně jiný, ne tak dobrý.
- Tektogramatická vrstva: Jádro výpovědi, téma, réma, podrobněji v sémantice.
- **Kontrola gramatiky**
 - (Opakovalo se to, o čem se mluvilo dříve. Že je třeba nechat projít i podivné, ale potenciálně správné, věty.)
 - Zvláště v případě češtiny je nutné rezignovat na mnoho potenciálních chyb.
 - Přístupy
 - Chybové vzorky
 - Zvláště pro jazyky s pevným slovosledem.
 - Příklad:
 - Po podmětu musí být přísudek, který je ve shodě.
 - V češtině: „ke mě“, „aby jsi“
 - Matfyz
 - RFOGD (Robust Free-Order Dependency Grammar)
 - Formalismus pravidel, kde je přímo označené, kde by se mohly objevovat chyby.
 - Vyvinuto asi před asi 14, 15 lety.
 - A – to co je vlevo
 - B – to co je vpravo
 - X – se z toho vytvoří
 - Příklad
 - ; 6. pravidlo – vytvoření predložkové fráze
 - PROJEKT TRUE
 - A.syntel = prep
 - B.syntel = Noun
 - A.case ? B.case Case_Dis_Prep_Noun
 - IF B.wel = prn THEN
 - B.pprep ? yes NonprepFormOfPronoun
 - ELSE ENDIF
 - X := B
 - X.syntel := prepfr
 - X.prep := A.lexf
 - END_P
 - Pravidla s otazníkem jsou měkká pravidla
 - Nejdříve konzervativně otazníky jako rovnítko a bez neprojektivit – rychlé.
 - Pak nový průběh – volněji, hledá se, kde je chyba, bere se také v úvahu neprojektivita.
 - Může se stát, že chyba není ani v této fázi odhalena.
 - Nebo příliš mnoho chyb, neprojektivit.
 - Poslední fáze – zároveň chybné, neprojektivní atd.
 - LanGR
 - Květoň 2003, Petkevič
 - Vlastnosti
 - primárně pro desambiguaci č. morfologie

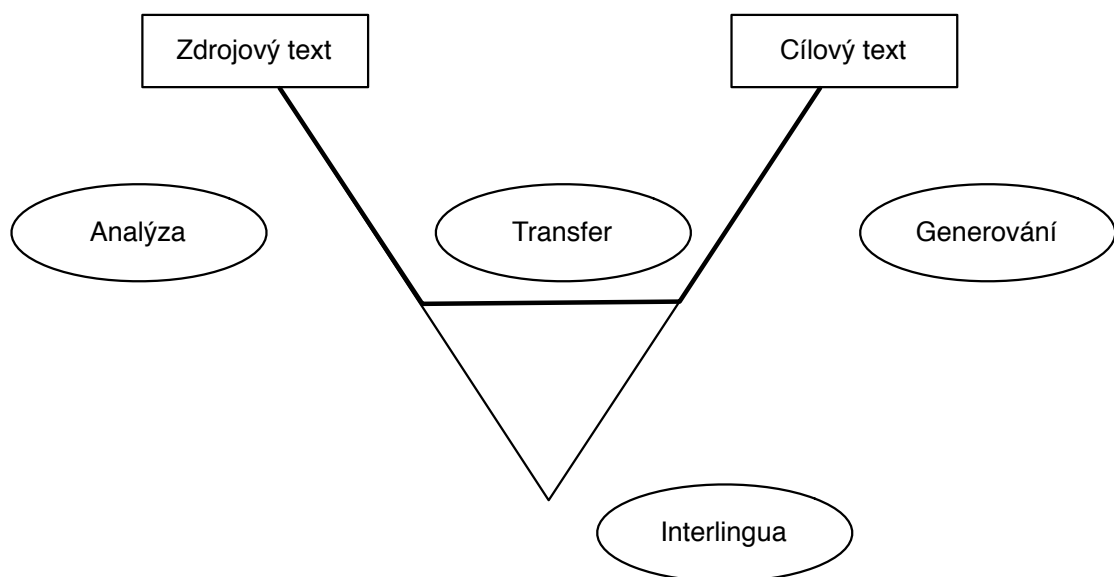
- pozitivní a neg. desam. pravidla
 - tedy i pravidla pro nutně špatné konstrukce
- pravidla mohou mít neomezený kontext
- redukční metoda – snaha udržet 100 % přesnost
 - Jako důkaz sporem. Když zmizí všechny možnosti – jde o chybu (je třeba ještě zjistit kde).
- pravidla psána ručně, avšak na základě dat z korpusu
- pravidla vzájemně nezávislá
- 4 části: kontext, desambiguační část, report a akce
- V další fázi se pomocí vzorců hledá, kde chyba vlastně je, když už o ní víme.
- Použit v korektoru v MS Office 2007 (ke stažení pro 2003).

• Přednáška #N 07/12/04

Minulá nebyla, na jedné jsem chyb I, doplním.

• Automatický překlad přirozených jazyků

- Jeden z prvních nápadů po válce, co dělat s počítači.
- Velmi složitý problém, relativně malý pokrok.
- Překlad není převod řetězců, vyžaduje porozumění.
- Je třeba
 - slovník
 - morfologie
 - „pravidla pravopisu“ (např. o slovosledu, tvoření číslovek apod.)
 - ustálená spojená, idiomy
 - ! reálie a kulturní kontext – to je zvlášť velký problém
- Příklady dalších problémů
 - Různá úroveň víceznačnosti v různých jazycích (např. v angličtině velká, v němčině malá)
 - Nepomůže ani statistika.
 - Složené výrazy
 - left wing tip tank filler gap gasket
 - Mädchenschule není škola obchodu s děvčaty, ale dívčí obchodní škola
 - Závislost na situaci
 - Např. umístění nápisů.
- Prehistorie



- Transfer může být více nebo méně „hluboko“. Interlingua je umělý jazyk.
- Poválečná historie (s počítači, tedy možností aplikace)
 - Od slovníkového k překladu přes morfémový (předpony, přípony, koncovky) k metodám pre-editing, post-editing (činnosti dělané lidmi, typicky hlavně post-editing, dělá se pořád). Pre-editing je vlastně omezení, zjednodušení zdrojového jazyka
 - 50. léta – skutečné začátky oboru a velké nadšení
 - pivotní jazyk (obdoba interliguy, ale jedná se o přirozený jazyk)

- georgetownský experiment (z ruštiny do angličtiny) – překlad jednoduchých vět za účasti reportérů
- Bouřlivý rozvoj
 - ... 1957 Chomsky – Standard theory, 1960 vystřízlivění („The box was in the pen“ – „v ohradě“)
- Vystřízlivění a „doba temna“
 - Zpráva americké vládní komise ALPAC: Nutnost dalších (velkých investic) → propad, konec výzkumu v USA na téměř 20 let. V Evropě a USA výzkum pokračoval (aspoň nějak)
- Vzniklé systémy
 - Velmi často ad hoc poslepované, ale po dlouhém úsilí relativně funkční (evropský SYSTRAN [překlad dokumentů evropské komise) apod.)
 - První úspěšný komerční systém:
 - TAUM-METEO (1976)
 - překlad meteorolog. zpráv en->fr
 - omezený, dobře definovaný jazyk
 - vhodná implementace (Q-systémy)
 - rozeznání textu, který dělá potíže, předá se překladateli
 - Méně úspěšné a neúspěšné
 - EUROTRA (neuskutečněný projekt EU mezi všemi jaz. páry EU)
 - VERBMOBIL (německý překlad mluvené řeči)
- Současné trendy
 - Statistické metody
 - využití paralelních označkových korpusů
 - BLEU score (podobnost referenčním překladům, technické číslo, statistické hodnocení, neměří shodu významu)
 - Nástroje podporující překlad
 - často nabízí dříve použité překlady (překladová paměť), typicky se využijí pro aktualizaci stávajících překladů
 - std. systém TRADOS
 - nezávislé na jazykových párech, jen mechanismus, funguje to až od druhého a dalšího překladu
- **Přednáška #(N+1) 07/12/11**
 - **České systémy na automatický překlad**
 - (slajdy)
 -