

Dvě zprávy Dany Konečné o strojovém překladu

Ondřej Mrázek

Filozofická fakulta
Masarykova univerzita, Brno

27. dubna 2015

První pokus se strojovým překladem v Československu (1960)

- 12. ledna 1960 proveden první pokus strojového překladu na SAPO
- 2 roviny projektu:
 - jazyková příprava překladu (P. Sgall, P. Novák, B. Palek, D. Konečná)
 - naprogramování překladu (K. Korvasová)
- místo pokusu: [Výzkumný ústav matematických strojů](#)

Průběh

- zpracování tří anglických vět:
 - sestavení slovníku těchto vět (40 slov)
 - pro každé české slovo vybrán jeden český ekvivalent (neřešena víceznačnost)
 - anglický slovník „přepsán“ na dřevné štítky:
 - = pro 1 an. slovo: štítek s an. slovem + štítek s metainformacemi (sl. druh an. slova a ml. kategorie cz. ekvivalentu)
- uložení slovníku do paměti stroje tak, aby bylo možné vyhledávání (pomocí aritm. operací)

Řešení

- předpoklad: stroj nerozumí obsahu věty
- vytvoření pravidel pro sestavení cz. věty za pomoci:
 - údajů získaných z an. věty
 - údajů o české mluvnici
- převedení ling. pravidel do strojově zpracovatelných instrukcí

Problémy

- kapacita paměti (jaké metainformace z české mluv. stavby ukládat)
- určení anglických složených slovesných tvarů
- pravidla pro převod an. slova, jež může být subs. i adj., na subs., nebo adj.
- zjištění informací, podle kterých stroj pozná mluvnické kategorie a syntaktickou fci slova

Výstup

- absence tiskařského zařízení = výstup v děrných štítcích
- mimo překladu základních vět i překlad jednoduchých vět vzniklých kombinací slov ze základních vět
- nulová chybovost překladu (ke spis. normě)
- získání zkušeností z přípravy strojového překladu
- ověření si správnosti dosavadního metodického postupu

Ke zkoumání češtiny z hlediska strojového překladu (1960)

- nárůst zájmu o problematiku strojového překladu (dále s. p.)
- v ČSR jsou pro rozvoj s. p. dobré podmínky (SAPO, kvalitní technický výzkum)
= ideální vlastnosti pro rozvoj nových ling. metod nutných pro přípravu s. p.
- potřeba sjednotit bádání v ČSR se SSSR a USA (zatím je ČSR pozadu)

Minulost a současnost s. p.

- dříve pouze úsek textu určeného pro překlad do jazyka experimentátorů
= sestavena pravidla pouze pro daný úsek, převod z jednoho jazyka do druhého
- dnes snaha o vytvoření převodního jazyka, tzn. možnost překládat do a z desítek jazyků
- kvalitní výzkum díky sovětským pracovištím

Převodní jazyk

- podrobné rozборы jednotlivých jazyků, a to podle nové metodiky
- vytvoření jednoznačných alg. pro daný jazyk (z rozboru), tzn. taková pravidla, jež může počítač sám použít a která postihnou všechny (téměř) případy, které mohou nastat
- pravidla jsou získána pouze z formy jazyka
- pravidla pro vstupní jazyk = pravidla **analýzy**
- pravidla pro výstupní jazyk = pravidla **syntézy**
- analýza má z textu získat veškeré informace potřebné pro překlad do kteréhokoliv jiného jazyka
- syntéza z informací vytváří text v mezích normy daného jazyka

Podrobněji

- pro obě roviny (analýza, syntéza) je potřeba pracovat s mluvnicí i lexikem
- zpracování jazyka pro s. p. se liší podle jeho typologie
- čeština patří mezi flektivní jazyky
= z tvaru slova lze často vyčíst slovnědruhovou příslušnot, někdy i syntaktickou fci
- pro č. j. jsou proto důležité koncovky slov a rozbor jejich distribuce
- mezi problémy patří:
 - různé koncovky jednoho pádu
 - stejné koncovky pro více pádů
 - hláskové alternace v kmeni slova
- díky flektivnosti češtiny jsou pomocí koncovek vyjadřovány i synt. fce (shoda, rekce)

Podrobněji

- údaje ze slovosledu jsou v případě č. j. (flekt. j.) pouze druhotným zdrojem
- řešení problémů prostupuje všemi jazykovými plány, tzn. při přípravě pro s. p. se neřeší zvlášť např. flexe, slovosled aj.
- pravidla pro syntézu reflektují současné mluvnicе češtiny (jak se co tvoří?); je potřeba pouze doplnit všechny možné případy a řešení zapsat úsporně

Podrobněji

- např. pro člověka dostatečný deklinační vzor „růže“ musí být pro stroj (kvůli alternacím) rozveden jako „sestra“, „ruka“, „noha“, „žena“, „rosa“ aj.
- jinde je zase dovoleno zjednodušení zápisu (např. u některých dublet nemusí být oba tvary,...)
- potřeba zpřesnit dále pravidla pro shodu a rekcii i pravidla AVČ
- = pro syntézu většinou „pouze“ upravujeme již existující pravidla

Analýza

- ve zprac. analýzy nemá jazykověda tradici, tzn. téměř pro každý jev je potřeba pracně získat podklady a navrhnout řešení pro s. p.
- důležité teoretické práce M. Dokulila, Fr. Daneše a K. Hausenblase
- využívání poznatků SSSR o s. p. ruštiny (protože je typologicky podobná češtině, lze využít některých nalezených výsledků)

Problémy analýzy

- mnohovýznamovost lexikální (ale i na rovině syntaxe, např. spojka může být členská i větná, aj.)
- jednoznačnosti se dosahuje pomocí rozboru kontextu věty dané i vět okolních
- např. pád substantiva je ovlivně slovesem apod.
- důležitost statistických výskytů, ty však dosud chybí (v použitelné míře)
- platí zásada: při s. p. se nevychází z toho, co je v jazyce možné, ale z toho, co se v textech opravdu vyskytuje
- potřeba sestavit frekvenci jevů a přihlížet k ní v rámci algoritmů (samotná vyšší frekvence ale nemá nutně přednost)

Shrnutí

- příprava s. p. v SSSR i v ČSR se zaměřuje na texty odborné (praktické využití s. p.; ustálená terminologie; jednodušší problematika syntaktická a frazeologická)
- z a do češtiny se překládají texty elektrotechnické a matematické
- rozbor mluvnice pokročil, chystá se zaměření na lexikum (a stručně i stylistiku)
- dobré vyhlídky do budoucna
- ze začátku budou překlady chybové, ale nebude složité upravit algoritmy a eliminovat chyby
- přesto dnes nelze předvídat konečnou hranici uplatnění s. p.

Zdroje:

KONEČNÁ, Dana. Ke zkoumání češtiny z hlediska strojového překladu. *Naše řeč*. [online]. 1960, roč. 43, č. 5–6 [cit. 25. dubna 2015]. Dostupné z: <http://nase-rec.ujc.cas.cz/archiv.php?art=4757>

KONEČNÁ, Dana. První pokus se strojovým překladem v Československu. *Naše řeč*. [online]. 1960, roč. 43, č. 3–4 [cit. 25. dubna 2015]. Dostupné z: <http://nase-rec.ujc.cas.cz/archiv.php?art=4743>