# 9 Lingvistické problémy ve strojovém překladu

Systémy strojového překladu se málokdy hlásí k určité lingvistické teorii, možné důvody:

- lingvistické teorie se nezabývají všemi, ale jen "lingvisticky pozoruhodnými" problémy
- lingvistické teorie se orientují na několik málo jazyků, zejména angličtinu
- lingvistické teorie věnují jen malou pozornost problémům překladu
- tvůrci systémů strojového překladu mají většinou praktický cíl a volí řešení, která jsou reálná, bez ohledu na teoretický rámec

> *... vývoj strojového překladu dokazuje, alespoň podle mého názoru, že platí dva (stěží slučitelné) principy, které by se daly bez obalu vyjádřit takto: 'V podstatě libovolná teorie, jakkoli nesmyslná, se může stát základem nějakého fungujícího systému.' a 'Úspěšné systémy málokdy fungují na základě teorie, k níž se hlásí.'* Yorick Wilks (1989)

## 9.1 Analýza

### 9.1.1 Morfologie

- slovník všech tvarů nebo slovník kmenů
- rozpoznávání neznámých slov (u flektivních jazyků lze někdy určit slovní druh a morfologické kategorie, jinak název, pokus o "překlad" transliterací)
- ohýbání – flexe
- slovotvorba: odvozování – derivace, skládání – kompozice (*Donaudampfshiffahrtsgesellschaft, Staatsexamen*)

### 9.1.2 Lexikální víceznačnost

– je dána vícenásobnou interpretací jednoho tvaru slova ve slovníku nebo při morfologické analýze

- morfologická víceznačnost: *hrad, kuropění, jarní*
- slovnědruhová víceznačnost: *loučí, ženou, kose, bez, mukl, spíš, round*
- homonymie (homografie): *los, travička, cenit, kynout, proudit* (*být/bít* )
- polysémie: *kniha, kanál, koruna, studnice, probíhat*

Morfologickou a slovnědruhovou víceznačnost lze řešit podle morfologických a syntaktických kategorií v kontextu. Homonymie a polysémie v rámci jednoho slovního druhu vyžaduje obvykle řešení založené na sémantických údajích (např. sémantické rysy v rámci slovesa). V některém kontextu však selže každé řešení, které bere v úvahu jen jednu větu a nemá k dispozici fakta o světě. (*Máme doma pěknou travičku.*)

### 9.1.3 Strukturní víceznačnost

– více možností reprezentace syntaktické struktury. Vzniká v průběhu syntaktické analýzy, v případě možnosti použití více pravidel pro daný řetězec.

*I saw the man in the park with a telescope. Nemoc probíhá bez příznaků. Kouckých loví tlouště na višni. Flying planes can be dangerous. Time flies like an arrow. He noticed her shaking hands. Ústav po dohodě s děkanem fakulty nabídl rozšířené užití programu v loňském letním semestru i pro vyšší ročníky, což bylo vyhlášeno při přednáškách z klinické anatomie. Hlasitá rána v Neapoli (pro)budila cizince.*

Možnosti řešení:

- využití morfologických a syntaktických informací – např. subkategorizační, t.j. syntaktické rámce sloves *Pověsila deštník na věšák.*
- využití sémantických informací – např. selekční omezení, t.j. sémantické rámce sloves, nebo "tématické role" *Paní domácí peče kuře.*
- využití předchozího kontextu – *Viděli ho často s Marií.*
- využití znalostí o světě – *Viděli velbloudy z mikrobusu. Viděli velbloudy z cirkusu. Viděli velbloudy z modelíny.*
- výběr nejpravděpodobnější možnosti
- dotaz (u interaktivních systémů)
- ponechání víceznačnosti nerozhodnuté

### 9.1.4 Odkazování

– zájmena, elipsy; gramatická a textová koreference

*Opice mi vzala čepici, zrovna když jsem ji fotil.*
*Opice mi vzala čepici, zrovna když jsem ji měl na hlavě.*

*Kotě si hrálo s klubkem, neboť se dobře vyspalo.*
*Kotě si hrálo s klubkem, neboť bylo červené.*
*Kotě si hrálo s klubkem, neboť bylo brzy po ránu.*

*Každý student si myslí, že dostane jedničku.*
*Každý student si myslí, že on dostane jedničku.*
*Každý student si myslí, že právě on dostane jedničku.*

*The soldiers shot at the women and some of them fell.*
*The soldiers shot at the women and some of them missed.*

### 9.1.5 Dosah operátorů

*Všechny hospodyně nesnášejí myši.*
$\forall h \forall m(hospodyně(h) \wedge myš(m) \rightarrow \neg snáší(h,m))$ – 'Neexistuje ani jedna hospodyně, která myši snáší.'
$\neg\forall h \forall m(hospodyně(h) \wedge myš(m) \rightarrow snáší(h,m))$ – 'Jenom některé hospodyně snášejí myši.'

*Všechny kabely jsou připojeny ke konektoru.*
$\forall x \exists y(kabel(x) \wedge konektor(y) \wedge je\_připojen(x,y))$
$\exists y \forall x(konektor(y) \wedge kabel(x) \rightarrow je\_připojen(x,y))$

*No smoking seats are available on domestic flights.*
[[ *no smoking* ] *seats* ]
[ *no* [ *smoking seats* ]]
*Staff only behind the counter.*
[ *staff* ] [[ *only* ] [ *behind the counter* ]]
[ *staff only* ] [ *behind the counter* ]

## 9.2 Transfer a interlingva

### 9.2.1 Lexikální nejednoznačnost

- stylistická nejednoznačnost *domicile – byt / bydliště, manželka – kanai / okusan*
- syntaktická nejednoznačnost *know – vědět / znát*
- pojmová nejednoznačnost *řeka – rivière / fleuve, noha – foot / leg, modrý – sinij / goluboj*
- chybějící ekvivalent *protalina, zdejší, tehdejší*

### 9.2.2 Strukturní rozdíly

- relativně jednoznačné případy: postavení adjektiva v češtině a francouzštině, postavení slovesa v angličtině a japonštině: *I like swimming. – Rád(a) plavu.* Horší případy:
- pasivum v angličtině: *The building was constructed in 1923. The books are easily read. Mary was given a book. This bed has been slept in.*
- neosobní konstrukce: v češtině *tančilo se*, v japonštině *Tošio-ga cuma-ni nigerareta.*
- *Il traversa la rue à pied / la rivière (en bateau). Ils partirent par avion de Ruzyně.*
- *Jmenuji se ...*
- slovesné časy, modalita, interakce mezi determinací a aktuálním členěním, číslo
- idiomy (nelze překládat kompozicionálně)

### 9.2.3 Transfer na jaké úrovni?

- morfologie
- syntaktická struktura
- syntaktické funkce
- rovina významu

Lexikální transfer: 1:1, n:1, 1:n
Strukturní transfer: pravidla transferu, která se aktivují řídícím členem konstrukce (*like – líbit se* )
Strukturní interlingva – kompromis mezi (lexikálním) transferem a interlingvou

## 9.3 Syntéza

V přímých systémech: jen morfologická syntéza, lokální slovosledné úpravy
V systémech s transferem se řada operací, které by koncepčně patřily do syntézy, provádějí v transferu, zejména ty operace, které nejsou jednoznačné (volba mezi synonymními konstrukcemi a lexémy).

- syntaktická syntéza
- morfologická syntéza
- jednoznačný postup / výběr z více možností

problém generování velkého počtu výstupů, možnosti řešení:

- zachování syntaktické struktury, kde je to možné (člověk překládá přesně opačně)
- zachování aktuálního členění, transformace mezi slovoslednými a ostatními prostředky pro jeho vyjádření
- výběr nejfrekventovanějších variant, stochastické modelování

stylistické úpravy
reversibilní gramatiky