
Obsah

Předmluva	3
1 Presumpce viny	5
1.1 Chyby dělají všichni	5
1.2 Co všechno se dá zkazit	5
1.3 Nenápadně, leč závažně mimo	12
1.4 Rukopis strojového překladače	12
1.5 Proč bychom to mohli rovnou chtít vzdát	15
2 Anatomie strojového překladače	17
2.1 Předzpracování a závěrečné úpravy	17
2.1.1 Vliv tokenizace a segmentace na další komponenty systému	18
2.1.2 Vliv předzpracování na lidské hodnocení	20
2.2 Jakou hloubku porozumění simulovat?	20
2.3 Nejmenší překladová jednotka	21
2.4 Pravidlové a statistické systémy	24
2.4.1 Rysy pravidlového systému	24
2.4.2 Co dělá překladač statistickým	25
2.4.3 Srovnání pravidlových a statistických systémů	26
3 Opisování vítáno	29
3.1 Slovník versus korpus	29
3.2 Postup budování korpusu	30
3.3 Velikost a vyváženost korpusu pro účely překladu	32
3.4 Zarovnávání dokumentů	33
3.5 Zarovnání po větách	34
3.6 Zarovnání po slovech	34

3.7	Zarovnání po větných členech	37
4	Věty možné a nemožné	41
4.1	Úloha jazykového modelování	41
4.2	N-gramový jazykový model	42
4.3	Vyhlažování n-gramových modelů	43
4.4	N-gramy slov či značek?	45
4.5	Syntaktické jazykové modely	47
4.6	Pokuta za slovo	48
5	Fráze, kam se podíváš	49
5.1	Komponenty frázového překladu	49
5.2	Tabulka frází	50
5.2.1	Postupy vybudování frázové tabulky	53
5.2.2	Pravděpodobnosti překladového modelu	54
5.3	Příprava možností překladu	56
5.3.1	Faktorové modely	58
5.4	Prohledávání stavového prostoru	62
5.5	Skórování hypotéz	64
5.5.1	Lokální a nelokální rysy	65
5.5.2	Prohledávání jako hypergraf	65
5.6	Klady a zápory frázového překladu	67
6	Obhajoba větného rozboru	69
6.1	Stromy závislostní a stromy složkové	70
6.1.1	Neprojektivita	73
6.1.2	Míry neprojektivity	75
6.2	Gramatiky	76
6.3	Stromové gramatiky	78
6.3.1	Stromové substituční gramatiky	79
6.3.2	Stromové adjunkční gramatiky	81
6.3.3	Adjunkce ve stromech složkových a závislostních	82
6.4	Unifikační gramatiky	82
6.5	Stromy povrchové a hloubkové	83

7	Stromy v překladu	87
7.1	Vliv směru překladu na zachycení větné struktury	87
7.2	Hierarchický překlad	88
7.2.1	Extrakce hierarchických frází	89
7.2.2	Hierarchický překlad jako SCFG	91
7.2.3	Jazykový model v hierarchickém překladu	93
7.3	Synchronní stromové substituční gramatiky	93
7.4	Problémy syntaktického překladu	95
7.5	Hlubkový překlad	98
7.5.1	Překlad přes t-rovinu	98
7.5.2	Předpoklad izomorfismu stromů	100
7.5.3	Formémy místo hlubkových rolí	100
7.5.4	Skrytý stromový markovovský model	100
7.5.5	Slabiny a přínosy TectoMT	103
8	V papoušcích zaručeně nejdelší	105
8.1	Proč měřit strojový překlad	105
8.2	Ruční hodnocení kvality překladu	105
8.3	Nevýhody ručního hodnocení	109
8.4	Kolik je „správných překladů“	109
8.5	Dosažitelnost referenčních překladů	111
8.6	Automatické metody měření kvality překladu	112
8.6.1	BLEU a jeho slabiny	113
8.7	Intervaly spolehlivosti pro automatické metriky	116
8.8	Závažnost chyb	118
9	Víc hlav víc ví	121
9.1	Tichá pošta vs. konzilium	121
9.1.1	Konzilium	123
9.2	Ladění vah	126
9.3	Kombinace systémů	129
9.3.1	Řazení systémů za sebe	129
9.3.2	Hlasování o slovech	130

10 Síla spolupráce	133
10.1 Překlad o závod	133
10.2 Volně šiřitelný kód	134
10.3 Volně šiřitelná data	135
10.4 Reprodukovatelnost až na kost	136
10.5 Veřejná rozhraní	136
Závěr: čeká to na vás	137
Literatura	139
Seznam obrázků	149
Seznam tabulek	151
Rejstřík	153
Slovníček anglických termínů	158