
Obsah

I Úvodem	1
II Lingvistický kontext výzkumu anafory a koreference	7
II.1 Kognitivní a diskurzní přístupy k analýze anafory	11
II.2 Anaforické vztahy v kontextu teorie reference	15
II.2.1 Ruská tradice teorie reference	15
II.2.2 Česká teorie reference	22
II.2.3 Teorie reference v pracích německojazyčných slavistů	25
II.2.4 Predikace vs. identifikace jmenné fráze v pozici přísudku	29
II.3 Zpracování koreference v počítačové lingvistice	39
II.3.1 Anotační schéma MUC	39
II.3.2 Anotační schéma ACE	42
II.3.3 Anotační schéma MATE a jeho aplikace (korpusy GNOME a VENEX) . . .	43
II.3.4 Princip anotace koreferenčních vztahů u Müller - Stube (2001)	49
II.3.5 Anotace koreference a asociační anafory v projektu PoCoS	51
II.3.6 Projekt anotace koreference AnCora-CO pro španělštinu	53
II.3.7 Jiná anotační schémata	54
II.4 Celkové zhodnocení teorií a korpusů	57
III Schéma anotace rozšířené koreference v PDT	59
III.1 Teoretické zásady anotace	63
III.1.1 Princip důslednosti	63

III.1.2	Princip dodržování (maximálního) koreferenčního řetězce	64
III.1.3	Princip maximální velikosti koreferující jednotky	66
III.1.4	Princip kooperace se syntaktickou strukturou tektogramatické roviny . .	66
III.1.5	Preference koreference před asociační anaforou	68
III.1.6	Princip rozhodujícího koreferenčního vztahu	69
III.1.7	Princip zvláštní váhy podílu na koherenci textu	71
III.1.8	Omezení počtu vztahů z jednoho uzlu / na jeden uzel	71
III.1.9	Preference anaforického vztahu před kataforickým	73
III.1.10	Princip jednofázové anotace	73
III.2	Formální charakteristika koreferovaných uzlů	75
III.2.1	Komplexní uzel v pozici anaforu	75
III.2.1.1	Sémantické substantivum v pozici anaforu	76
III.2.1.2	Sémantické adjektivum v pozici anaforu	79
III.2.1.3	Sémantické adverbium v pozici anaforu	82
III.2.1.4	Sémantické sloveso jako člen koreferenčního vztahu	84
III.2.2	Kvazikomplexní uzel v pozici anaforu	86
III.2.3	Kořeny souřadných struktur v pozici anaforu	86
III.2.4	Kořeny seznamových struktur v pozici anaforu	87
III.3	Gramatická koreference	89
III.4	Textová koreference	93
III.4.1	Pronominální textová koreference	93
III.4.2	Rozšířená textová koreference	95
III.4.2.1	Typologie textově koreferenčních vztahů	97
III.4.2.1.1	Koreferenční vztah mezi výrazy se specifickou referencí (coref_text, typ = SPEC)	104
III.4.2.1.2	Koreference generických jmenných frází (coref_text, typ = GEN) .	108
III.4.2.1.3	Koreferenční řetězce s prolínající se specifickou a nespecifickou referencí	112
III.4.2.2	Textová koreference z hlediska lexikálních skupin	113
III.4.2.2.1	Koreference abstraktních jmen	113
III.4.2.2.2	Koreference deverbativ	117

III.4.2.2.3	Koreference pojmenovaných entit	119
III.4.2.3	Problematické případy označování textové koreference	125
III.4.2.3.1	Hraniční případy mezi [coref_text, typ = SPEC] a [coref_text, typ = GEN]	125
III.4.2.3.2	Hraniční případy mezi [coref_text, typ = GEN] a vztahy, které lze chápat jako nekoreferenční	129
III.4.2.3.3	Spojení s výrazy s významem „kontejneru“	131
III.4.2.3.4	Dvě místní určení vedle sebe (<i>tady v Praze, u nich doma</i> apod.)	135
III.4.2.3.5	Technické problémy	136
III.4.2.4	Nejednoznačný výběr antecedentu	140
III.4.2.4.1	K otázce výběru antecedentu v případě apoziční skupiny	140
III.4.2.4.2	K otázce výběru antecedentu v případě koordinační skupiny	142
III.4.2.4.3	K otázce více možností odkazování (identická koreference)	143
III.5	Asociační anafora (bridging vztah)	145
III.5.1	Typologie vztahů asociační anafory	147
III.5.1.1	Vztah PART mezi částí a celkem (PART: PART_WHOLE a WHOLE_PART)	151
III.5.1.1.1	Vztah PART uvnitř jedné věty	154
III.5.1.1.2	Vztah PART u generických NP a deverbativ	154
III.5.1.1.3	Hraniční případy u asociační anafory typu PART	155
III.5.1.2	Vztah SUBSET mezi množinou a podmnožinou/prvkem množiny (SUB_SET a SET_SUB)	156
III.5.1.2.1	Vztah SUBSET uvnitř jedné věty	157
III.5.1.2.2	Asociační anafora typu SUBSET u generických NP a deverbativ	158
III.5.1.2.3	Hraniční případy u asociační anafory typu SUBSET	160
III.5.1.3	Vztah FUNCT mezi entitou a určitým objektem, který má vzhledem k této entitě jedinečnou funkci (P_FUNCT a FUNCT_P)	161
III.5.1.3.1	Označování vztahu FUNCT v párech typu <i>prezident Klaus - ČR</i>	163
III.5.1.3.2	K otázce hloubky „vloženosti“ funkce ve vztazích typu <i>ministr zemědělství - vláda - stát</i>	164
III.5.1.3.3	Hraniční případy u asociační anafory typu FUNCT	165
III.5.1.4	Vztah CONTRAST sémantického a kontextového protikladu	166
III.5.1.4.1	Hraniční případy u asociační anafory typu CONTRAST	170
III.5.1.5	Vztah ANAF anaforického odkazování mezi nekoreferenčními entitami	170

III.5.1.5.1	Hraniční případy u asociační anafory typu ANAF	172
III.5.1.5.2	Anaforické odkazování na nevyjádřený antecedent	173
III.5.1.6	Vztah REST pro jiné případy asociační anafory	174
III.5.1.6.1	Vztah „rodinná příslušnost“	174
III.5.1.6.2	Vztah „místo – obyvatel“	174
III.5.1.6.3	Vztah typu „autor -- dílo“	175
III.5.1.6.4	Vztah „věc -- majitel“	175
III.5.1.6.5	Vztah mezi stejně vyjádřenými nebo synonymními nekoreferenčními NP	176
III.5.1.6.6	Vztah „událost – argument“	177
III.5.1.6.7	Vztah „objekt – velmi typický instrument“	177
III.5.1.6.8	Jiné možné vztahy, o kterých jsme uvažovali	178
III.5.2	K omezení počtu vztahů asociační anafory	178
III.5.2.1	Preference koreference	179
III.5.2.2	Ne více než jeden vztah od jednoho uzlu	179
III.5.2.3	Kooperace s TGS – omezení na anotace asociační anafory u závislých uzlů s některými funktoři	181
III.5.3	Nejednoznačný výběr antecedentů	182
III.5.3.1	K otázce výběru antecedentu v případě apoziční konstrukce	183
III.5.3.2	K otázce výběru antecedentu v případě koordinační skupiny	183
III.5.3.3	Spojení se slovy s funkcí „kontejneru“	185
III.6	Textová koreference nebo asociační anafora. Problematické případy	187
III.6.1	Dlouhé vzájemně propojené řetězce s textovou koreferencí, asociační anaforou a koordinačními konstrukcemi	187
III.6.2	Specifická konstrukce – typ „faktory – jeden z faktorů“	191
III.6.3	Případ „zaměstnanci – každý ze zaměstnanců“	192
III.6.4	Propojení koreferenčních řetězců jediným vztahem asociační anafory	194
III.7	Speciální typy reference (coref_special)	195
III.7.1	Exoforické odkazování	195
III.7.2	Odkazy na segmenty textu	196
III.7.2.1	Hraniční případy mezi typem coref_special, typ = segm a asociační anaforou typu SUBSET	198

III.8 Zásah do anotace původní zájmenné koreference	201
IV Aplikace a evaluace probíhající anotace	205
IV.1 Technické provedení	209
IV.1.1 Formát dat	209
IV.1.2 Pomoc anotátorům	209
IV.1.2.1 Předanotace dat	209
IV.1.2.2 Automatická pomoc v průběhu anotace	210
IV.2 Aplikace anotace rozšířené textové koreference a asociační anafory	215
IV.3 Měření mezianotátorské shody	219
IV.4 K rozdílům v mezianotátorské shodě	221
V Závěrem	233
V.1 Další otázky a výhledy	239
Summary	241
Seznam zkratk a značek	245
Seznam obrázků	249
Seznam tabulek	253
Seznam grafů	255
Literatura	257
Internetové odkazy	269