V této práci představujeme jeden z možných modelů zpracovaní rozšířené textové koreference a asociační anafory na velkém korpusu textů, který dále používáme pro anotaci daných vztahů na textech Pražského závislostního korpusu.
Na základě literatury z oblastí teorie reference, diskurzu a některých dalších poznatků teoretické lingvistiky na jedné straně a s použitím existujících anotačních metodik na straně druhé jsme vytvořili detailní klasifikaci textově koreferenčních vztahů a typů vztahů asociační anafory.
V rámci textové koreference rozlišujeme dva typy textově koreferenčních vztahů - koreferenční vztah mezi jmennými frázemi se specifickou referencí a koreferenční vztah mezi jmennými frázemi s nespecifickou, především generickou referencí.
Pro asociační anaforu jsme stanovili šest typů vztahů: vztah PART mezi částí a celkem, vztah SUBSET mezi množinou a podmnožinou/prvkem množiny, vztah FUNCT mezi entitou a unikátní funkcí na této entitě, vztah CONTRAST sémantického a kontextového protikladu, vztah ANAF anaforického odkazování mezi nekoreferenčními entitami a vztah REST pro jiné případy asociační anafory.
Jedním z úkolů výzkumu bylo vytvořit systém teoretických principů, které je nutno dodržovat při anotaci koreferenčních vztahů a asociační anafory. V rámci tohoto systému byl zaveden například princip důslednosti anotace, pnnctp dodržování maximálního koreferenčního řetězce, princip kooperace se syntaktickou strukturou tektogramatické roviny, princip preference koreferenčního vztahu před asociační anaforou a další.
Vypracovanou klasifikaci jsme aplikovali na koreferenční a anaforické vztahy v Pražském závislostním korpusu (Prague Oependency Treebank, POT). Byla provedena anotace těchto vztahů na polovině korpusu POT (cca 25 tis. vět).