

Posudek disertační práce

Mgr. Anny Nedoluzhko: Rozšířená textová koreference a asociační anafora

Předmětem předložené disertační práce je studium textové koreference a asociačních vztahů v textu, klasifikace jejich typů a návrh, jak postupovat při jejich anotaci ve velkém korpusu českých textů. Tento návrh autorka ověřila na části Pražského závislostního korpusu (cca 25 tisíc vět), přičemž se snažila také vyhodnotit vztahy mezi anotacemi různých anotátorů, a tím mimo jiné ověřit, zda anotace koreferenčních vztahů a asociační anafory je reálným úkolem.

Autorka ve své práci přesvědčivě doložila, že se velmi dobře orientuje v dostupné světové literatuře o daném tématu. Přitom nejde o téma nikterak snadné: především sám pojem koreference a anafory je obtížné definovat, protože jde o vztahy překračující hranice jazykové formy a zahrnující odkazy k mimojazykové skutečnosti. Za druhé, většina prací dotýkajících se problematiky zachycení koreferenčních a anaforických vztahů v textu na základě korpusů přihlíží k jazykovému materiálu odlišnému od češtiny (kromě velké studie Bergerovy, v níž ovšem jde o úzce vymezený typ koreferujících výrazů), takže nebylo možné klasifikaci v nich navrženou jednoduše převzít a pouze ověřit.

Pokud jde o materiálovou stránku práce, autorka se mohla opřít o již provedenou anotaci morfologických a hloubkových syntaktických vztahů v Pražském závislostním korpusu a využít i existující zachycení gramatické koreference v tomto korpusu a jistého „zárodku“ anotace koreference textové.

Je pochopitelné, že při zdolávání tak široce vymezeného úkolu se autorka nevyhnula některým nedostatkům, podle mého názoru ve značné míře způsobeným skutečností, že nejde o rodilou mluvčí češtiny. Tato skutečnost neovlivnila vlastní lingvistický rozbor, ale formu, ve které je její výzkum předkládán. Nevyhnula se často neobratným doslovným překladům ze svého mateřského jazyka – ruštiny, nebo i pro češtinu nezvyklým či dokonce nesprávným překladům z jazyka originálu práce, kterou popisuje (převážně angličtiny); rovněž po stránce terminologické by bylo třeba jistých úprav (generování, nikoli generace; je třeba rozlišovat „jména“ a „substantiva“, popř. jmenná fráze; „využití zájmen“ místo „existence zájmen“; odkaz k prvku spíše než reference na prvek; parafráze a nikoli perifráze; „stav výzkumu“ místo „state of the art“ apod.). Domnívám se ale, že bychom neměli být na autory dizertací, jejichž mateřským jazykem není čeština, přísnější než na české doktorandy, kteří předkládají své práce v angličtině, a proto jazykové nedostatky ve svém posudku zmiňuji jen takto obecně a návrhy jejich oprav vpisuji do práce v její tištěné podobě, aby jich autorka mohla využít, pokud by práci nebo její části hodlala publikovat v češtině. Zde uvádím jen některé dílčí poznámky či dotazy týkající se obsahu práce:

str. 6: asociační anafora vnáší do modelu textu informaci novou jen z hlediska textu, ale z hlediska posluchače jde o informaci identifikovatelnou

str. 10: co se míní „pokleslým referentem“?

str. 23: u Mendozové nejde o referenční nástroje, ale o přesný popisný aparát

str. 28-29: ve větě (55) je snad možné užít v subjektové pozici zájmena „On“

str. 53, předposlední odstavec: tato instrukce platí v případě, že není možné určit typ víceznačnosti?

str. 61: i v případě gramatické koreference je třeba počítat s víceznačností!

str. 63: v tektogramatické rovině nejsou předložky samostatným uzlem (první věta odst. III.1.2)

str. 66: zasloužilo by vysvětlení, proč u atomických uzlů nevede žádný vztah ke kořenu tektogramatického stromu (nebo jde jen o příliš zhuštěnou formulaci?)

str. 68: co se míní „diskurzivními jednotkami“ (odst. 2)

str. 139, př. (97): které interpretaci dáváte přednost – specifické?

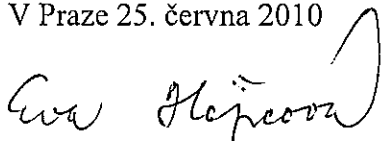
str. 197, odst. III.5.1.6.6: argument, že asociační vztah typu událost – argument je již v Pražském závislostním korpusu vyznačen jako závislostní vztah mezi slovesem a jeho argumenty, neplatí např. pro (104): jde o užití ve dvou různých větách

V textu není na některých místech jasné rozlišování dvou různých pojmů: anafora (femininum) jako vztah mezi dvěma jednotkami textu a anafor (maskulinum) jako člen, od kterého vede anaforický vztah k antecedentu.

Shrnutí: Úkol, před kterým doktorandka stála, je nesmírně rozsáhlý, obtížný a nesnadno uchopitelný. Domnívám se, že autorka předložené dizertace potvrdila, že je velmi dobrou lingvistkou, že umí ve své práci využívat četných pramenů odborných i materiálových a v neposlední řadě že si je vědoma jisté otevřenosti problémů či prozatímnosti některých řešení, ke kterým dospěla (to se odráží v celé části práce, kde dokládá své poznatky na širokém empirickém materiálu, i ve vlastním závěru přiléhavě nazvaném Další otázky a výhledy). Rozborem příkladů z korpusu (který tvoří více než polovinu celé práce) dokládá, že jí je vlastní práce s jazykovým materiálem i s počítačovým korpusem; vhodně využívá i toho, že v korpusu, s nímž pracuje, jsou také označeny morfologické charakteristiky jednotlivých jazykových jednotek i závislostní vztahy mezi nimi. Poučná je i krátká kapitola o sledování mezianotátorské shody. Obdivuhodné je také to, jak se dokázala vyrovnat i s počítačovými aspekty své práce. Po jazykové stránce trpí práce četnými nedostatky, ty však jsou vyrovnány jak vlastním obsahem, tak i formální stránkou (obrázky, přehledné grafy).

Z uvedených důvodů považuji předloženou doktorskou dizertaci za plně vyhovující a **doporučuji ji jako podklad pro získání titulu Ph.D.**

V Praze 25. června 2010



Prof. PhDr. Eva Hajičová, DrSc
Ústav formální a aplikované lingvistiky
MFF UK
Malostranské nám. 25
118 00 Praha 1