

Posudek doktorské disertační práce (vyjádření školitele)

Mgr. Otakar Smrž: Functional Arabic Morphology, Formal System and Implementation

Hlavním výsledkem práce Mgr. Otakara Smrže je originálně navržený a implementovaný formalismus pro popis arabské morfologie (ElixirFM), který vychází ze solidní lingvistické teorie. S tímto výsledkem úzce souvisí výsledek ve formě modulu MorphoTrees zapojeného do procesu anotování Pražského arabského závislostního korpusu. Posledním (dle pořadí prezentace v disertaci) výsledkem je modul pro počítačovou reprezentaci arabštiny.

Práce je přehledně rozdělena do třech tématických celků – jejich vzájemná provázanost je jasně popsána. Celkům předcházejí dvě kapitoly, svým charakterem úvodní. První kapitola je průvodcem po dosavadních přístupech k morfologické analýze v rámci počítačového zpracování arabštiny. Ota vysvětluje, proč se rozhodl navrhnout a implementovat další přístup. Četné citace jasně dokládají, že se velmi dobře orientuje v dané tématické oblasti. V této kapitole je rovněž uveden seznam všech softwarových nástrojů, jejichž prostředků Ota využívá k dosažení svých cílů. Kapitola vrcholí desaterem hlavních výsledků, kterých Ota ve své práci dosáhl. Druhá kapitola je úvodem k třetímu tématickému celku, tedy k řešení počítačové reprezentace arabštiny.

První tématický celek (kapitoly 3-6) se věnuje arabské morfologické analýze. Třetí kapitola je rešerší dosavadních přístupů. Čtvrtá kapitola je krátkým úvodem k funkčnímu programovacímu jazyku Haskell, který si Ota vybral pro implementaci vlastního přístupu, pojmenovanou ElixirFM. ElixirFM je předmětem kapitoly páté, ve které je přirozeně největší pozornost věnována popisu slovníku, nejdůležitějšího prvku každé morfologické analýzy. Kapitola 6 je vhodné nalistovat při počátečním seznamování se s ElixirFM.

Druhý tématický celek (kapitoly 7-8) zasazuje morfologickou analýzu do kontextu anotování korpusu. Mezi pražské závislostní korpusy patří i Pražský arabský závislostní korpus, jehož anotaci Ota koordinoval (což se bohužel z práce samotné nedozvíme). Po vzoru Pražského závislostního korpusu byly anotace arabského korpusu zahájeny morfologickými anotacemi, tj. zjednodušením výstupu morfologické analýzy. Kapitola 7 popisuje modul MorphoTrees, který usnadňuje anotátorům výběr správné analýzy vzhledem ke kontextu. Kapitola 8 propojuje arabskou morfologickou analýzu s Pražským arabským závislostním korpusem (který by si zasloužil, aby se jeho jméno objevilo v názvu kapitoly).

Třetí tématický celek (kapitola 9) popisuje implementaci modulu Encode Arabic v jazyce Haskell.

Posláním krátkého závěru je především to, že všechny moduly jsou volně ke stažení. Bohatý seznam literatury dokládá široký Otův počítačově-lingvistický přehled.

Otova disertace působí uceleným dojmem. Je psána anglicky a po formální stránce je naprosto v pořádku.

Po celou dobu doktorského studia Ota pracoval velmi intenzivně na tématu arabské morfologie a prokázal samostatnost, i ve vedení pracovního týmu anotátorů arabského korpusu. Vzhledem ke všem uvedeným skutečnostem doporučuji uznat předkládanou práci jako doktorandskou a Otovi přiznat titul Ph.D.

V Praze, 16. srpna 2007

