

Posudek disertační práce

Dan Faltýnek

autor: Ing. Michaela Nohejlová Zemková

název: Lingvistické přístupy v genomice a lingvistická metafora v biologii

Předložená disertační práce se zaměřuje na lingvistickou analýzu genetických textů. Jednou z jejích klíčových otázek je, zda je tento typ analýzy vhodný pro rozlišení genomu parazitů a volně žijících organismů. Metoda, kterou práce užívá, je založena na n-gramové analýze textu. Lingvistické snažení v oblasti teorie informace prokázalo, že n-gramová analýza je dobrý nástroj k výzkumu charakteru textu, jeho příslušnosti k jazyku, případně jazykovým stylům, autorům apod. Využití těchto metod pro analýzu genetického textu se tedy nabízí a řada výzkumů prokazuje, že je funkční (autorka zmiňuje např. reprezentativní monografii: Bolshoy, A. et al.: *Genome Clustering from Linguistic Models to Classification of Genetic Texts*). Dalším tématem textu je využití n-gramové analýzy k rekonstrukci evoluce genomu z původně jednoduchých motivů či repetit. K tomu se těsně váže následné téma, kde je nejvíce přítomen lingvistický obsah, a tím je analogie evoluce genomu a vývoj dětské řeči, což je nepochybně zajímavá konkretizace zmiňované analogie.

Autorka se podrobně věnuje komplexitě textu ve smyslu informace a redundance, navazuje na iniciační text v této oblasti Popovové a Trifonova, nezapomíná na lingvistické pozadí teorie informace. Vysvětluje pojetí slovníku studovaného genetického textu, kde nalézá minimum poměru aktuálního a maximálního slovníku při délce 5 (bez nějakého nátlaku na jazykovou analogii je zajímavé podotknout, že průměrná délka českého slova se pohybuje /v závislosti na studovaném korpusu/ okolo 4,7 grafémů). Autorka v práci detailně představuje sběr a zpracování analyzovaných dat. Hlavní analytické hledisko představuje bohatost slovníku. Pro vyjádření závislostí byla velmi vhodně využita shluková analýza (PCA). Autorka následně interpretuje výsledky kvantitativní analýzy a ukazuje její spolehlivost v klastrování studovaných vzorků. V další části je pozornost věnována sekundárním strukturám proteinu v komparaci se strukturou textu přirozených jazyků. Analogie je zde vedena mezi kombinatorikou hydrofobních a hydrofilních aminokyselin a kombinatorikou hlásek, vokálů a konsonantů. Autorka upozorňuje, že analogie obou textů je pomocná a centrální je v přístupu podobnost analytického postupu – s tím, že lze komparovat rysy obou textů (délka, velikost abecedy) na výsledky analýzy, což je důležitý metodologický postřeh, na který nezapomíná. Projevy kombinatoriky aminokyselin autorka znovu vztahuje k parazitům a interpretuje ji jako možný důsledek únikové strategie parazita vůči hostiteli. Práce dále navazuje na přitažlivou Trifonovovu hypotézu evoluce genomu na základě repetitivních struktur. Ty jsou znovu analogizovány s přirozeným jazykem – s opakujícími se slabikami v dětské řeči. Tato analogie je opět pomocná. Lingvistická teorie o možnosti rozvoje slovních tvarů či gramatických struktur na základě ablautu v repetitivních strukturách neuvažuje (autorka na to upozorňuje). Toto téma je na druhou stranu mimo lingvistický zájem, témata evoluce řeči v těchto souvislostech byla na pořadu dne před více než sto lety, text tedy přináší motivaci k návratu do diskuse.

Text disertační práce odpovídá standardům, místy má nejednotnou typografii a obsahuje malé množství typografických chyb a velmi malé množství ortografických a stylistických prohrěšků. Na základě práce navrhuji udělit doktorandce titul Ph.D.

Témata k diskusi:

1. Bylo by možné uvažovat o využití Zipfova zákona jakožto další vlastnosti vstupující do shlukové analýzy (PCA)?
2. Jak autorka hodnotí komparaci n-gramových analýz s analýzami náhodně generovaných textů? Podpořila by taková komparace tvrzení z konce textu, že život a jazyk jsou sebeutvářející se systémy (s. 49)? Neplatilo by to pak pro jakékoliv takové systémy?
3. Je možné analogizovat vývoj dětské řeči, jazykovou ontogenezi, a jazykovou fylogenezi? Lingvistika takovou možnost dříve potvrzovala, vývoj jazykových prostředků v historii a u dítěte postupu analogicky. Dnes je ale tento pohled zapovězený.