



FILOZOFICKÁ FAKULTA Univerzita Karlova

Ústav Asijských studií

Posudek oponenta bakalářské práce

Student: Kristýna Trmalová

Specializace: koreanistika

Název práce: **Koronavirová krize v tisku KR a KLDK: korpusová analýza lexika**

Vedoucí práce: Mgr. Vladislava Mazaná

Oponent: Tomáš Horák, Ph.D.

Řešitelka si vybrala zajímavé téma korpusové analýzy publicistických textů věnovaných koronavirové pandemii. Spojuje tak obsahově aktuální globální problematiku s perspektivním metodologickým přístupem zkoumání jazyka. Korpusová lingvistika je totiž pro nerodilého mluvčího, postrádajícího jazykovou intuici mluvčího rodilého, jednou z mála oblastí jazykovědy, v níž může provádět plnohodnotný výzkum. Kvalitní výzkum samozřejmě předpokládá použitelný korpus, jehož vytvoření vyžaduje velké materiální i lidské zdroje. Nejbezpečnější metodou je využít stávajících institucionálních korpusů a subkorpusů, které bývají nejen reprezentativní co do obsahu, ale i kvalitně lemmatizované a tagované a navíc u nich bývají k dispozici i korpusové vyhledávače a nástroje, viz např. Český národní korpus (korpus.cz).

Řešitelka zvolila poněkud riskantnější a pracnější cestu tvorby korpusu vlastního, což mohlo být dáno specifičností sledovaného tématu, nebo obtížnou veřejnou dostupností profesionálně budovaných korpusů. Na jednu stranu je nutné ocenit odvahu i vložené úsilí, na druhou stranu nelze přehlédnout některé problémy, které jsou s tím spojené, jak bude poukázáno níže.

Práce jako celek je koncipována přehledně a systematicky. Cíle práce jsou vhodně formulovány, řešitelka čerpá z relevantní literatury a prokazuje základní znalost klíčových pojmů korpusové lingvistiky. Požadavky kladené na bakalářskou práci splňuje i formální zpracování a jazyková úroveň. Hlavní nedostatky práce řešitelka zmiňuje sama v závěru: „Pro přesnější výsledky by nejspíš bylo třeba zvolit jiný referenční korpus a věnovat čas lemmatizaci typů“ (str. 46).

Pokud jde o referenční korpus pro potřeby analýzy klíčových slov, tak ten byl zjevně zvolen naprosto nevhodně, v podstatě proti logice věci. Chápeme-li klíčové slovo jako jednotku, jejíž frekvence ve zkoumaném textu je výrazně vyšší než je obvyklé v běžných textech, pak referenční korpus má odrážet právě tento běžný úzus, vytvářet



FILOZOFICKÁ FAKULTA Univerzita Karlova

Ústav Asijských studií

neutrální jazykové pozadí, v kontrastu k němuž vyvstává prominence klíčových slov. Použijeme-li jako referenci korpus složený ze stejně specializovaných textů, tak to v zásadě postrádá smysl. V tomto případě je to ještě umocněno tím, že se používá technicky vzato „jiný jazyk“ (severokorejská kulturní řeč x jihokorejská řeč standardní). Při současném stupni jazykové divergence je téměř nezbytné porovnávat severokorejské texty se severokorejským územ a jihokorejské s jihokorejským. Pokud bychom například porovnali úvodník z Rodong sinmun s jihokorejským územ, vyšly by nám pravděpodobně jako vysoce prominentní klíčová slova „soudruh“, „vůdce“, „straníky“, „revoluce“ apod., ovšem na pozadí severokorejského úzu by šlo o běžnou slovní zásobu. Samotná volba referenčního korpusu tedy do značné míry devalvovala výsledky hledání klíčových slov i práci s nimi.

Podobným problémem je i lemmatizace, resp. její absence. V seznamu nejfrekventovanějších „slov“ se objevují samostatné i nesamostatné výrazy a jejich spojení bez zjevného systému. Platí to zejména pro partikule a koncovky, které se objevují jak ve spojení se jménem, resp. se slovesem, tak i samostatně (navíc každý alomorf jako samostatný typ, viz např. 을/를). Výsledná duplicita slov (stejně slovo s různými partikulami a koncovkami) v reálu snižuje počet dohledaných nejfrekventovanějších slov, kde pak chybí očekávaná slova jako „test“ (antigenní, PCR), „odběrné, testovací místo“ apod. Přiznání slovního statusu gramatickým prostředkům (partikulám a koncovkám) zase výrazně posouvá a zkresluje výsledky ve prospěch korejského lexika. Ještě problematičtější je absence lemmatizace u analýzy článků Rodong sinmun, kde se nám mezi tzv. „slovy“ objevují jmenné a adnominální fráze typu 신형코로나비루스감염증피해, 신형코로나비루스감염자수가, 확대되고있는 atd. Za prověření by stálo, jestli tyhle slepence nejsou spíš výsledkem chyb při kopírování souborů. K další nejednoznačnosti pak přispívá nerozlišení homonymie u uváděných výrazů, kdy např. u šestého výrazu 수 není zřejmé, jestli se opravdu všech více než 8 tisíc výskytů týká korejského pomocného výrazu s významem „možnost“, nebo jde o sinokorejské slovo s významem „počet“; 27. výraz může být sinokorejská číslovka, nebo korejská partikule apod. Vzhledem k proklamovanému cíli práce by byla důsledná lemmatizace mnohem vhodnější, než práce s typy, navíc nejednoznačného rozsahu a obsahu.

Navzdory výše uvedeným výtkám práci jako celek oceňuji a doporučoval bych v řešitelce v této oblasti pokračovat. V případě zájmu by bylo užitečné obrátit se např. na Národní ústav korejského jazyka a požádat o přístup k profesionálním korpusům. Pro



FILOZOFICKÁ FAKULTA Univerzita Karlova

Ústav Asijských studií

předkládanou práci mohl být relevantní např. novinový korpus „Sinmun malmungčchi 2020 (신문 말뭉치 2020), který by ušetřil spoustu mechanické manuální činnosti a umožnil by soustředit se na zajímavější aspekty práce s korpusem. I když procentní čísla, která jsou součástí výsledků nemají příliš vypovídací hodnotu, práce tohoto typu má dle mého názoru velice dobrý potenciál praktického využití, např. při jazykové výuce středně a více pokročilých, kde mohou být užitečné kolokace generované z korpusu, nebo lze vytvářet specializované tematické slovníčky frekventovaných slov.

Práci doporučuji k obhajobě s navrhovaným hodnocením **velmi dobře**.

V Zelenči, dne 24.08. 2020