



Ústav germánských studií FF UK

POSUDEK NA DIPLOMOVOU PRÁCI

Bc. Jitky Křesťanové

s názvem

**Vergleich der Ergebnisse der Kookkurrenzdatenbank (CCDB)
und der Kookkurrenzanalyse**

ÚGS FF UK, Praha 2017, 88 stran

Tématem diplomové práce je porovnání dvou aplikací, které jsou používány pro corpus-driven zkoumání korpusu německého jazyka DeReKo. Jedná se o statickou aplikaci Kookkurrenzdatenbank (dále jen CCDB) a Kookkurenční analýzu (dále jen KA). Cílem práce je analýza a následný popis výhod a nevýhod obou aplikací, které se projevují jak v rozdílné velikosti korpusu, který zkoumají, tak v možnosti nastavení parametrů a také v porovnání výsledků a jejich využitelnosti v lingvistické praxi.

Hlavním cílem bylo vyzkoumat rozdíly ve výsledcích vázané na rozdílnou velikost analyzovaného korpusu a sledovat míru využitelnosti těchto nástrojů pro účely studia ustálených slovních spojení (kolokací i frazémů).

Práce je členěna na teoretický úvod, v němž se autorka vypořádává s potřebnými pojmy a uvádí čtenáře do problematiky. Součástí teoretického úvodu je i velmi detailní popis fungování obou aplikací, který má návodný charakter a bylo by možné ho tedy využít jako „manuál“ pro práci s oběma aplikacemi. Ve vlastní empirické části (některé důležité aspekty a postřehy lze nicméně nalézt i v části teoretické) autorka podrobuje analýze deset slov (jedno od každého slovního druhu), a to vždy v obou aplikacích, popisuje shody a rozdíly ve výsledcích a vyvozuje závěry směřované k cíli, který je uveden výše. V závěru svá zjištění shrnuje a potvrzuje svou úvodní hypotézu, že velikost zkoumaného korpusu zásadním způsobem ovlivňuje získané výsledky (což ovšem bylo očekávatelné). Práce také disponuje německým a českým shrnutím a seznamem sekundární literatury.

Práce je vcelku přehledně členěna a psána slušnou němčinou, jen s několika málo překlepy či typografickými chybami (např. str. 11 –*Passagen*, str. 68 chybí tečka na konci prvního odstavce). Na některých místech lze nalézt jisté stylistické nedostatky, kostrbaté konstrukce a nadměrné opakování téže informace (např. str. 11 „*Der andere Unterschied ist auch das, ob oder wie ...*“; časté opakování informace, že CCDB je statická, anebo že nastavení granularity *sehr grob* znamená analýzu jen primárních kolokátorů apod.). Autorka správně cituje a odkazuje na sekundární literaturu, z níž čerpá. Přestože se zdá, že je v práci vloženo mnoho screenshotů (které jsou ale nutné pro pochopení a lepší orientaci v problematice), splňuje práce bez problému rozsahový limit 60 normostran.



Ústav germánských studií FF UK

Po obsahové stránce lze zhodnotit, že teoretická část nepřináší nic zásadně nového a čerpá převážně ze sekundární literatury, což se ovšem dalo očekávat vzhledem k obtížnosti a charakteru zvoleného tématu. Na tomto místě je třeba vyzdvihnout odvalu, se kterou se autorka zvoleného tématu ujala. V empirické části a všude tam, kde dochází k vyhodnocení výsledků, je patrná absence lehkého nadhledu nad problematikou, který může být způsoben nezkušeností práce v tomto oboru a menším rozhledem v problematice jako celku. Stejně tak dochází k občasnému přeceňování kvantifikačních údajů (typicky u procentuálního podílu daných syntagmatických vzorů apod.). Jedná se však o práci poctivou, na níž je vidět množství času stráveného analýzami, nicméně je třeba říci, že z daného materiálu bylo možné vytěžit i více. Metodologicky se nabízí např. možnost provést analýzy všech slov v obou aplikacích v „továrním“ nastavení, tyto vyhodnotit a v návaznosti na toto vyhodnocení pokračovat v analýzách s jiným nastavením, a to vždy po důkladném uvážení cíle výzkumu.

Z pozice vedoucí práce také pozitivně hodnotím poctivý přístup autorky k psaní práce, pravidelné konzultování a vzorný time-management.

Po přečtení práce vyplývá několik témat a otázek vhodných k diskusi u obhajoby:

1. Jmenujte několik situací, v nichž byste doporučila uživatelům korpusu užít spíše KA a kdy spíše CCDB.
2. Na str. 11 se věnujete reprezentativitě korpusů a jako příklad uvádíte rozložení DWDS-Kernkorpus. Proč zmiňujete právě tento korpus, který jinak v práci nepoužíváte, a jak jsou na tom s reprezentativitou ostatní korpusy, např. studovaný DeReKo nebo ČNK?
3. Na str. 11nn. rozebíráte dělení korpusů dle různých kritérií. Vysvětlíte rozdíl mezi korpusem reprezentativním a referenčním.
4. V textu na str. 29 popisujete nabídku *als Kookkurrenzpartner Treffer ausschließen*. Zdá se mi, že na tomto místě není text zcela srozumitelný a čtenář by nemusel pochopit, o jaké omezení nastavení se jedná. Můžete uvést konkrétní příklad tohoto parametrického nastavení a jeho důsledky pro výslednou analýzu?
5. Na str. 42 tvrdíte, že kolokátorem adjektiva *schwarz* je slovo *weiß*, ale neuvádíte ho zde jako kolokaci/frazém. Opravdu tomu tak není/nemůže být? Co je pro posouzení tohoto a jemu podobných jevů nanejvýš důležité (při práci s dotčenými korpusovými aplikacemi)?
6. Mezi kolokátory v CCDB je kromě popsaných ještě *schreiben* (které nekomentujete) a pak ještě zmiňovaná substantiva *Humor* a *Liste*, která KA na prvních 15 postech



Ústav germánských studií FF UK

nezmiňuje (str. 54). Nemůže se tedy jednat o to, že CCDB, jejíž výsledky vznikly kombinací několika různých analýz s různými parametry nakonec opravdu podává o ustálených spojeních relevantnější informaci než KA, která je závislá na zadání konkrétních parametrů? Podobně i u *der* je u CCDB sice prvním kolokáterem *in*, ale syntagmatické mustry přinášejí lepší informaci o kolokacích v CCDB (*in der Nacht*, *in der Nähe von*, *in der Regel*), zatímco KA toto nepřináší.

7. Jsou obě aplikace řízené čistě přístupem corpus-driven nebo lze v některé z nich/v obou nalézt prvky corpus-based?

Vzhledem k výše řečenému a s přihlédnutím k obtížnosti zadaného tématu a k úrovni jeho zpracování **doporučuji** předloženou diplomovou práci Bc. Jitky Křest'ánové **k obhajobě** a předběžně **navrhuji hodnocení „velmi dobře (2)“**.

Mgr. Věra Hejhalová, Ph.D.
vedoucí práce