

Posudek diplomové práce *Slovesný systém italštiny a jeho úskalí z perspektivy českých studentů italianistiky*, kterou předložila Bc. KAROLÍNA PRANEVSKÁ, FF UK, Ústav románských studií, ak. r. 2022/2023.

Diplomová práce Karolíny Pranevské je velmi originálním textem, jenž je výsledkem dobře promyšlené empirické práce vycházející z celé řady zajímavých hypotéz. Mohu proto hned zkraje říci, že její četba a nyníjší posouzení pro mě není jen povinností, ale i osobním potěšením. Jádrem práce je totiž typologie chyb, její kvalitativní i kvantitativní analýza u studentů a studentek italštiny od 1. po 3. ročník na základě dostatečně velkého a anonymizovaného vzorku různých testů a cvičení. Data, která diplomantka shromáždila, jsou nesmírně cenná a zajímavá, ale také velice různorodá, nehomogenní a nevyvážená. Jejich korektní analýza tak představuje určitou výzvu.

Diplomová práce je přehledně rozdělena do sedmi hlavních kapitol. Po úvodu se autorka věnuje různým teoriím osvojování jazyka s důrazem na klíčový rozdíl mezi učením se cizímu jazyku a jeho osvojováním v bilingvním prostředí; postupně definuje základní pojmy, jako je mj. i jazykový transfer. Ve třetí kapitole stručně nastiňuje typologické charakteristiky češtiny a italštiny, aby bylo patrné, jak se slovesný systém v obou jazycích liší. Čtvrtá kapitola je věnována popisu slovesné flexe v italštině. Byť je tato kapitola jen shrnutím toho, co obvykle přináší gramatiky, působí souhrn kompaktně a přehledně. Pátá kapitola přináší důležitou typologii chyb, v šesté jsou vyloženy cíle a hypotézy a sedmá představuje podrobnou kvantitativní a kvalitativní analýzu sebraného materiálu. Po obligátním résumé v češtině a v italštině následuje literatura a seznam obrázků a tabulek. Lze tedy jednoznačně říci, že i po formální stránce nelze diplomové práce nic vytknout, ba dokonce lze jistě vyzdvihnout pečlivě a pozorně zpracování každé části (kterému na dokonalosti ubírá jen sporadický překlep či nějaká ta chyba v interpunkci – nepovažuji za nutné zde žádný takový příklad uvádět).

Jak už jsem zmínil v úvodu, rozhodně platí, že jde o mimořádně zajímavou práci, z níž je na každém kroku patrný osobní, velmi samostatný postup, který by však také v jedné konkrétní fázi vyžadoval konzultaci např. se statistikem, aby byla data správně analyzována a korektně vyhodnocena. Na tento nedostatek, který se projevuje v několika ohledech, se v následujících bodech zaměřím, ale zmíním také pár dalších – spíše zanedbatelných – aspektů. Těmi začnu.

Oceňuji, jak se diplomantka pouští do obtížné sekundární literatury, kterou skutečně samostatně studuje, dobře interpretuje a vhodně využívá k tomu, co chce říci. Přesto bych řekl, že by jistě mohla např. některé tabulky převzaté i s anglickou terminologií (např. č. 1, s. 15, č. 2, s. 18) směle převést do češtiny. Určitě by se tím odstranily i chybičky, které ztěžují četbu (např. v tab. č. 1, s. 15 „Complecity Theory“ místo „Complexity Theory“, v tab. č. 2 na s. 18 je „fewer inhabitions usually“ místo „inhibitions“). Také některé reference k sekundární literatuře by mohly být uváděny přesněji (viz např. souborný odkaz k „Bahnikové, Hamplové a Salviho a Venallio“, když jde o tři různé texty, navíc ten poslední je od Salviho a **Vanelliové**).

Nyní pár otázek k aplikační části, která je opravdu pozoruhodná. Zaujal mě značný počet hypotéz (a jedna zajímavější než druhá...) vyložený na s. 76-78. Byť se pak každá může vyhodnocovat zcela samostatně na základě kvantitativních dat, určitě by stálo za to je nějak rozdělit podle typologie chyb. Je asi zřejmé, že některé chyby jsou formální, spočívající jen v nesprávném tvarosloví (např. Hypotéza č. 1, č. 9), některé jsou obsahové, konceptuální. Tento rozdíl mezi formální komplexitou (složitá alomorfie, nepravidelná slovesa atd.) a konceptuální komplexitou je často zmiňován i u studií zabývajících se osvojováním mateřského jazyka. V seznamu hypotéz je však někdy obtížné toto oddělit (např. podmínkové věty třetího typu, tzv. kontrafaktuály, vyžadují pokročilé uvažování a zároveň jsou morfologicky i syntakticky složité). Stejně tak by bylo dobré oddělit konceptuální chyby tohoto typu (tj. kontrafaktuály) od konceptuálních chyb zapříčiněných absencí dané kategorie ve výchozím jazyce (např. chyba v *trapassatu prossimo*, protože čeština ho nemá). Dále je pak sporná hypotéza č. 3 – nejde přece o valenci, v té nejspíš chyby nikdo nedělá, neboť je dána opravdu konceptuálně (rozdíl mezi intranzitivním „dormire“ a tranzitivním „ammazzare“ je jasný), ale o to, čemu se běžně říká slovesná rekce, a to je již formální vyjádření (a z definice této hypotézy je zřejmé, že jde právě o rekci – chybné užití předložky, kterou sloveso vyžaduje, popř. její absence apod., srov. též s. 100).

Pokud jde o sběr dat, jejich povahu a vyhodnocení, je určitě obdivuhodné, jak obrovskou práci diplomantka vynaložila, aby je získala a zpracovala. Je zároveň škoda, že některé metodologické otázky, které si sama klade a které zcela intuitivně považuje za zásadní, nejsou zodpovězeny tak, aby výsledné tabulky – či vůbec celkové zpracování v Excelu – bylo úplně v pořádku a aby bylo kdykoli a kýmkoli

znovu použitelné. Jsou zde totiž dva problémy, které ale není vůbec těžké odstranit (a jejich odstranění diplomantka v práci zvažuje, navrhuje, vyhledává – srov. např. na s. 89).

Za prvé jde o to, že není-li pro daný typ chyby v datech žádná hodnota prostě proto, že k tomu nejsou data (např. daný typ cvičení v 1. roč. není, viz např. tab. č. 32, cvičení „cloze“ 1. ročník), pak nemůžeme tuto absenci do datových souborů zapisovat jako nulu, ale jako NA („not applicable“). Nula by přece znamenala, že zrovna v daném cvičení studenti chyby nedělají (a v celkovém kvantitativním hodnocení se to projeví ve srovnání s počtem chyb v jiných cvičeních). V tomto smyslu by bylo nutné předělat všechny tabulky s hodnotou „0“ tam, kde tato „nula“ opravdu neoznačuje hodnotu proměnné, ale absenci dat.

Za druhé je struktura dat nevyvážená a heterogenní právě proto, že některá cvičení v daném ročníku jsou, některá ne. Takže srovnávat počty chyb napříč ročníky je pak velice zkršené (ne-li přímo chybné). Karolína Pranevská si je obou těchto překážek velmi dobře vědomá (sama dokonce říká, že by bylo nutné mít identický typ cvičení pro každý ročník zvlášť, abychom mohli nějaký pokrok posoudit).

V neposlední řadě je zde určitá konfúze mezi tím, co lze očekávat od kvantitativních dat, a tím, jak je pak interpretujeme. To je patrné na s. 103, kde v první větě máme tezi, kterou by bylo možné – na základě čistých dat – statisticky ověřit (např. kdyby byl počet chyb ve stejném cvičení pro 1. roč. 100, pro druhý 50, pro třetí pak 25, měli bychom jasný trend). Ale hned v druhé větě si diplomantka klade otázku po příčinách chyb – to je ale až druhotné. Získáme-li perfektní data, z nichž bude statisticky prokazatelné, že je zde vyšší chybovost u tvarů *passata remota* než u indikativu prézentu, pak můžeme tu příčinu hledat (jako odpověď na nějakou z uvedených hypotéz). Proto je např. tab. č. 33 na s. 104 jen částečně užitečná. Ano, v absolutních číslech to vypadá, jako by už v 3. roč. těch chyb bylo celkově méně, ale nemůžeme si tím být jisti, když jsou ta data tak nevyvážená (a 2. roč. má vysoká čísla proto, že pro něj máme více dat z vícero cvičení, a ne proto, že zrovna 2. ročník chybí nejvíce...). Aby tohle bylo možné ošetřit bez perfektních dat, musely by se použít pokročilé statistické techniky, které takto „děravá“ data umí vyhodnotit (nějaké smíšené modely apod.). Navíc je tab. č. 33 zajímavá v tom, že ukazuje víceméně stejný poměr mezi obecnými a konkrétními cvičeními, který se mění jen mírně (a zde by byl opět nutný statistický test ne-proporcionality, který by nám řekl, že jsou ty poměry v podstatě stejné).

Celkově je tedy potřeba ke kvantitativnímu zpracování dat říci, že je opravdu cenné, že je máme, ale že by bylo nutné je teď zpracovat tak, aby se s nimi dalo zacházet zcela správně, např. podle zásad „tidy data“ (srov. např. zde: https://fu.ff.cuni.cz/STAT/21_tidy_data.html), kde bude na jedné straně (v řádcích) každý jednotlivý test (T1 – T138), na straně druhé (ve sloupcích) každá proměnná (a v řádcích pak u každého testu daná hodnota, je-li k dispozici, jinak NA) bez ohledu na typologii (tu lze pak z dat dodatečně vyčíst na základě kontingenčních tabulek apod.), tj. každý typ chyby a hodnota, pak počet (např. T1 – chyba_tvar_remoto – 3 atd.), včetně dalších kategoriálních proměnných (typ testu, ročník, atd.). Mám za to, že takovou transformaci by šlo udělat celkem snadno a do budoucna by stála za pokus – mj. i proto, že částečně je tento formát už započat (viz např. Obr. 6 na s. 82).

Tolik ke kvantitativnímu zpracování, které je sice nedokonalé, ale zdaleka ne tak bezcenné, aby se nemohlo přistoupit ke kvalitativní analýze. Ta je v této práci opravdu pečlivá, detailní a v mnoha ohledech velmi přínosná, jen je potřeba se ujistit, zda závěrečné vyhodnocení hypotéz na s. 107 a dále je možné opravdu jen na základě kvalitativní analýzy, anebo zda k tomu potřebujeme statisticky signifikantní data.

Jak je zřejmé z výše uvedeného, připomínek k této jinak velmi zdařilé diplomové práci je možné vznést hned několik a všechny se tak či onak týkají kvantitativního zpracování a vyhodnocení sesbíraných dat, ale vzhledem k tomu, že studijní plán mgr. studia nezahrnuje povinnou přípravu do základní ani pokročilé statistiky, považoval bych za nespravedlivé tyto výhrady promítnout do hodnocení (a také proto, že ani já sám v tom nejsem dostatečně trénovaný).

Proto diplomovou práci Karolíny Pranevské vřele doporučuji k obhajobě a hodnotím ji jako **výbornou**.

V Praze dne 30. května 2023

doc. Mgr. Pavel Štichauer, Ph.D.,
oponent diplomové práce