CELKOM ISTÉ SA PRÍDE NA TO, ŽE NIEKTORÉ SLOVÁ SA OPAKUJÚ DOSLOVNE: HORECKÉHO HYPERSYNTAX¹

DAN FALTÝNEK

Filozofická fakulta Univerzity Palackého, Olomouc, Česko

FALTÝNEK, Dan: It will certainly be found that some words are literally repeated: Horecký's hypersyntax. *Jazykovedný časopis (Journal of Linguistics)*, 2020, Vol 71, No 2, pp. 185 – 196.

Abstract: The article reflects the linguistic work of Ján Horecký in connection with hyper syntax and text linguistics. In his work *Základy jazykovedy* (Outline of linguistics) Ján Horecký remarks (1974, p. 90) that one of the principles of text construction is the literal repetition of certain words. We discuss this Horecký's assumption and describe its consequences for the *langue parole opposition* and the concepts of textual *isotopy* and textual *cohesion*. The main task of the article is to examine Horecký's assumption. For these purposes, we present an authorship attribution analysis of literary works by two Slovak authors: Svetozár Hurban Vajanský and Martin Kukučín. We focus on low-frequency lexicon, i.e. hapax legomena, which are supposed to be independent of the authorial style (e.g. Binongo, 2003) and should reflect random circumstances of communication (de Saussure, 1996, p. 50; Bloomfield, 1933, p. 170). This means that if the structure of the text were to be affected by the repetition of certain words, the low frequency layer of the lexicon should contain evidence of this repetition with a low degree of dependence on the content and style of the literary work (Baayen, 1996). The analysis and its presentation is based on separate processing of hapax legomena and their n-grams, cosine dissimilarity and multidimensional scaling (Torgerson, 1952). Contrary to the general notion of the text structure, we conclude that the authorial texts are based on the repetition of certain word forms and word forms combinations (by n-gram analysis), even in the level of low-frequency words.

Key words: textual cohesion, textual isotopy, langue parole opposition, hapax legomena, hapax legomena n-gram analysis, authorship attribution

ÚVOD

Jakkoliv u lingvisty málokdy vidíme tak široký záběr odbornosti, jako je tomu u Jána Horeckého (Těšitelová, 1980; Kořenský, 2007), můžeme přeci jen předmět jeho jazykovědného zájmu omezit, a to v oblasti slovní zásoby a gramatiky, s celkovým zapojením jazykových rovin od fonetické k větné. Toto omezení je poměrně globální, dovoluje nám ovšem konstatovat skutečnost, že do základního badatelské-

¹ Acknowledgement: Článek vznikl v rámci projektu Sinophone Borderlands, reg. No. CZ.02.1.01/0.0/0.0/16_019/0000791, Excellent research.

ho profilu Jána Horeckého nespadá hypersyntax. Přesto v jeho pracích najdeme místa, kde se tomuto tématu nevyhýbá a živě formuluje, jakou povahu má text z hlediska mezivětného navazování a jaká je struktura textu širokého rozsahu. Mluvíme zde o *Základech jazykovedy* (Horecký, 1978, s. 89 – 90), kde zmiňuje určité předpoklady o povaze textu, které v mnohém odrážejí přirozenou jazykovou intuici. Zároveň jsou ale v rozporu s některými teoretickými východisky lingvistiky, jako je *langue parolová opozice*, textová *izotopie* nebo textová *koheze*, což se dále pokusíme ukázat. Naším úkolem je ale především pojetí hypersyntaxe Jána Horeckého, jak se ukazuje v níže uvedených formulacích, prověřit.

HORECKÝ O HYPERSYNTAXI

V *Základech jazykovedy* se Ján Horecký věnuje oblasti hypersyntaxe v návaznosti na výklady o syntaxi, které jsou do značné míry svázány s generativním přístupem. Jeho *Úvod do matematickej jazykovedy* (1969) prozrazuje, že autor má velmi rozsáhle zazdrojovány algebraické přístupy k popisu jazyka (Revzin, Chomsky, Sgall atd.) a staví na nich svůj pohled na jazyk a především na strukturu pojmenování, která je jedním z jeho základních zájmů (viz 1969, s. 55 – 60). Naopak přístupy k popisu jazyka spadající do kvantitativní lingvistiky (entropie, jazykové zákony apod.) reflektuje úsporně. Až v návaznosti na syntax a hypersyntax se Horecký ve svých *Základech* věnuje popisu zbylých jazykových rovin. Tradiční úvody do studia jazyka či obecné jazykovedy se většinou přidržují postupného výkladu od hlásky k větě či textu (viz např. Čermák, 2011; Černý, 1998; Palek, 1989) a jednotlivé roviny představují na základě specifik jejich popisu. Horecký tuto posloupnost ve výkladu obrací. To je v úvodové publikaci ambiciózní, ale zároveň koncepčně průhledné a poctivé, vzhledem k tomu, že autor vzhledem k výchozím teoretickým inspiracím tímto směrem skutečně postupuje, viz např. výklad spojený s generativním modelem slova (Horecký, 1978, s. 108). Dodejme, že Horeckého úvod se zároveň staví po boku úvodu Helbigova (1991) nebo Dolníková (2009), které čtenáři nepředkládají konceptuálně jasně dané a neproblematizované lingvistické instrumentarium, ale široce reflektují zdroje a tlumočí čtenáři výchozí literaturu, čímž mu otevírají přístup do lingvistického diskurzu na základě obeznámenosti se způsoby vedení diskuse nad danými tématy (viz Faltýnek, 2015).

Ve výkladech věnovaných hypersyntaxi a teorii textu Horecký popisuje strukturu textu jako systematizovanou souslednost vět zajišťovanou konektivitou, textovou a situační deixí a lexikálně sémantickými relacemi (Horecký, 1978, s. 86). Vysvětluje dále, že mezivětné relace jsou zprostředkovány tematickými posloupnostmi. Tyto výklady ze základních kompendií známe (Hirschová, 1989). Horecký ovšem na závěr kapitoly předkládá také vizi hloubkové struktury věty dekomponované na sémantické role, které tvoří strukturu v širší perspektivě textu. V tu chvíli, kdy připouští, že jednotlivé komponenty věty – mluví o funktorech a argumentech – můžeme reprezentovat jejich slovními formami, formuluje následující:

„Takisto možno zisťovať zásady výberu z radu synonym, ktoré má daný jazyk k dispozícii pre jednotlivé funkctory a argumenty. (...) Celkom iste sa pri takejto analýze príde na to, že v súvislých textoch sa niektoré (najmä opakujúce sa) argumenty (ale i funkctory) vyjadrujú zámenami, ďalej, že niektoré slová sa opakujú doslovne, iné v istých variáciách a pod.“ (Horecký, 1978, s. 90).

Dovoľme si do Horeckého výkladu doplniť autora. Horecký hovorí v reflexívnom pasívu („sa vyjadrujú“) o tom, že je text komponovaný na základe výberu medzi synonymnými jazykovými prostriedkami (ve vypuštenej časti citácie uvádza príklad pojmenovania parlamentu NSR), ďalej uvažuje o tom, že je text konštruovaný na základe opakovania konkrétnych slovných foriem. Všetchny tieto konštrukčné postupy majú autorský pôvod. Uvažovať bychom také mohli o voľbách stylových, situačných apod. (Kelih a kol., 2005), autor je ale vždy pôvodcom voľby v jazykovom systéme, pouze se nám do strukturního nebo generativního rámce špatně vkládá (jak výstižně ukazuje interakční stylistika; Orgoňová – Bohunická, 2018). Pokusíme se dále parafrázovat výše uvedenou Horeckého citaci, abychom byli schopni dohlédnout jejich souvislosti: Horecký v závěrečné části výkladu o hypersyntaxi a teorii textu říká, že text má určitou *myšlenkovou osnovu* (1978, s. 89), kterou si můžeme představit jako logicky organizovanou síť sémantických rolí ve výpovědi a jejich přesazích. Na povrchové struktuře textu registrujeme realizaci těchto rolí formou autorských voleb synonymních vyjádření a autorským opakováním určitých slov v textu. Jak jsme poznamenali v předcházejícím výkladu, takový pohled na strukturu textu konvenuje přirozené jazykové intuici, jsme zvyklí hodnotit jazykové projevy tak, že říkáme, že mají podobu, *jako by je řekl Karel* nebo *Benedikt*, a to právě na základě výběru slovních forem a jejich kombinací. V čem by představa toho, že autorský slovník strukturuje text, měla odporovat východiskům lingvistické teorie, jak naznačujeme výš?

Řekli jsme, že Horeckého výklad by mohl být dán do kontrastu s *langue parolovou hypotézou*, konceptem textové *izotopie* a textové *koheze*. Pokud by byl text strukturován na základě autorské voľby synonym a opakování konkrétních slov, jak se Horecký zmiňuje ve výše uvedené citaci ze *Základů jazykovědy*, pak bychom vzhledem k těmto konceptům museli uvažovat o následujícím:

a/ Z hlediska *langue parolové hypotézy* máme za to, že *langue* představuje naindividuální soubor jazykových pravidel, která se používají v konkrétní promluvě zasažené v jedinečných komunikačních podmínkách, *parole*. Ferdinand de Saussure předpokládá, že do oblasti *langue* spadají prostředky strukturující roviny fonologickou, morfologickou, lexikální. Větu chápe jako volní akt mluvčího, protože její výsledná podoba závisí na potřebách vyjádření mluvčího a z toho důvodu také evidujeme nespočetné množství vět v různých podobách, kterým nenacházíme společného jmenovatele v analogii k fonému nebo lexému. Z toho důvodu větu řadí do *parole* (Saussure, 1996, s. 132 – 133). Např. zmiňovaný Horeckého *Úvod do matematické jazykovědy* (1969) společně s vývojem lingvistiky v posledních sto letech jasně ukazují, že

věta má striktní konstrukční pravidla. Určitá pravidla nacházíme i v rovině mezivětného navazování, Horecký je nakonec v *Základech* představuje. O rovině textu širšího rozsahu už ale nepředpokládáme, že je formována nějakými pravidly, nepředpokládáme zde žádnou jednotku analogickou větě, slovu atd. (viz Bloomfield, 1933, s. 170). Použijeme-li v parafrázi znovu de Saussura (1996, s. 50), můžeme konstatovat, že struktura textu širšího rozsahu je „náhodná“, tzn. že je ovlivněna aktuálními potřebami mluvčích a specifickými okolnostmi komunikace, kontextem. Pokud by měla být tato struktura ovlivněna určitými slovními tvary, které mluvčí opakovaně používá, jak naznačuje Horecký, museli bychom z oblasti parole po větě vydělit další velmi podstatný díl, protože bychom předpokládali, že aktuální kontext (*parole*) se přizpůsobuje způsobům vyjádření autora – jeho zvyklostem v užívání aktivní slovní zásoby.

b/ Pod textovou *izotopii* rozumíme sémantickou souvislost částí textu. Texty jsou soudržnou veličinou na základě toho, že dávají smysl, a to vzhledem k tomu, že k sobě můžeme vztáhnout referované jevy v nějaké očekávatelné nebo myslitelné konstelaci. Zároveň koncept izotopie můžeme formulovat tak, že části textu mají takové sémantické vlastnosti, že alespoň některé z nich k sobě můžeme vztáhnout jako související. Algirdas J. Greimas (1966) to vyjadřuje následujícím způsobem:

„The isotopy between the terms of the elementary structure ensures and is, in a way, the foundation of the microuniverse seen as a unit of meaning.“ (Greimas, 1987, s. 66).²

Jak říká A. Greimas, ze sémantické struktury slov zapojených v určitém textu jsou vybírány spojitelné obsahové aspekty, které vyjadřují aktuální kontext. Co když je tomu ale tak, že slova použitelná pro výstavbu textu jsou předem dána – nejsou použita vzhledem k jejich optimální vlastnostem vzhledem k potřebám vyjádření kontextu, ale jsou autorsky preferovaným (opakovaným) způsobem vyjádření. V takovém případě se na Greimasově konceptu izotopie nic nemění, pouze se obrací vzhůru nohama: slova se nepřizpůsobují kontextu a aktuálním podmínkám sémantické struktury vyjádření, ale kontext se přizpůsobuje zadané množině lexika, z nějž jsou vybírány potřebné sémantické rysy k jeho vyjádření a zároveň výstavby soudržné struktury textu.

c/ Textová koheze je definována jako soudržnost textu daná formálními prostředky jazyka – tzn. nikoliv založená na sémantických souvislostech lexika (viz b/). Takovými prostředky jsou: pronominalizace (nahrazování předmětů řeči vyjádřenými v textu substantivy a dalšími slovními druhy zájmenem: **Karel hraje pexeso**; **On je velmi hravý člověk**), elipsa (**Karel hraje pexeso**; **Je velmi hravý**), kongruence (**Karel hrával pexeso**; **Býval**, nikoliv **bývala** – **velmi hravý**, nikoliv **hravá**), konektivita (**Karel hrál pexeso a kouřil dýmku**; **Karel hrál pexeso nebo kouřil dýmku**), opakováním slovních tvarů (**Hrej Karle pexeso!**; **Nemůžeš Karle pořádně jen kouřit dýmku**), formálně sémantickými vztahy lexikálních forem (**Karel se bavil hrou v pexeso**; **Nikdy se nenudil**). Povšimněme si, že příkladové vztahy textové soudržnosti jsou realizovány v těsné blízkosti vět.

² „Izotopie mezi pojmy elementární výpovědní struktury je původcem a svým způsobem základem mikrosvěta v podobě určité významové veličiny.“ (prekl. D. F.)

Tyto a další prostředky textové koheze (viz např. výše zmíněné tematické posloupnosti) jsou realizovány s ohledem na tzv. koreferenci, tzn. že se mezivětné navazování vztahuje k určitému předmětu řeči, v našem případě konkrétní osobě *Karla*, *pexesu* atd. Pokud bychom uvažovali o tom, že autor při výstavbě textu používá uzavřenou množinu lexika, doslovně opakuje konkrétní slovní formy, jak říká Horecký, znamenalo by to, že jsou tyto slovní formy umístěny v libovolné vzdálenosti v textu a vztahují se k odlišným předmětům řeči nebo odlišným kontextům, kdy je k nim referováno. Tzn. že takové opakování slov nevykazuje koreferenci. Můžeme potom hovořit o tom, že jsou tyto lexikální formy prostředkem textové koheze? Pokud by se autorský styl vyznačoval rozsáhlým opakováním konkrétních slovních forem, znamenalo by to, že jsou tyto slovní formy výchozím prostředkem strukturace textu. Každá z těchto forem by totiž měla své specifické gramatické vlastnosti, které by vyžadovaly specifickou výstavbu výpovědi, vyskytovala by se např. v určitých vazbách (*Karel se bavil hraním pexesa*, *Karel se bavil hrou v kostky*). Tyto slovní formy by tedy vyžadovaly specifické obsazení částí výpovědi, včetně predikátu, tzn. že by jejich umístění ve výpovědi určovalo např. slovesnou valenci prostřednictvím spojitelnosti s určitým tvarem slovesa (*Karel si včera dal špagety*, *Karel si včera dal buček* – *Karel šel včera na buček*; korpusové doklady všeobecně ukazují, že poslední varianta je zcela ojedinělá a ve spojení s bučkem je preferována pouze forma se slovesem *dát si*). Můžeme to shrnout následujícím způsobem: pokud by text obsahoval určité množství opakovaných autorských slovních forem, pak by tyto slovní formy určovaly výstavbu konkrétních výpovědí, jejich formální a sémantickou návaznost a stávaly se určujícím faktorem ovlivňujícím výše zmíněné prostředky textové koheze vázané v koreferenci. Vzhledem k formálně sémantickým vlastnostem textu to Greimas vyjadřuje následujícím způsobem:

„Any discourse, once it has posited its own semantic isotopy, is nothing more than a very partial exploitation of the considerable virtualities the lexematic thesaurus makes available to it.“ (Greimas, 1987, s. 114).³

ANALÝZA

Rozhodli jsme se prověřit Horeckého formulaci o doslovném opakování slov v autorských textech na dvou slovenských autorech, prvním je Svetozár Hurban Vajanský a druhým Martin Kukučín (texty byly převzaty ze serveru <https://eknizky.sk>, datum přístupu: 16. 9. 2020). Autoři jsou současníci, doba vzniku jejich autorských děl se od 80. let 19. stol. překrývá. V obou případech se jedná o autory píšící prózu dobového realistického směru. Tím je alespoň částečně dána homogenita vzorku, jeho stylová a obsahová podobnost. Níže analyzujeme od každého z autorů vždy sedm literárních děl (viz popis Obrázku č. 1).

³ „Jakýkoli diskurs, jakmile si vytvoří vlastní sémantickou izotopii, není ničím jiným než omezeným využitím virtualit, které k tomu nabízí lexikální systém.“ (prekl. D. F.)

Specifikujeme-li otázku směřující k opakování slov v autorských textech, můžeme pro následující analýzu formulovat předpoklad, že se budou autoři lišit v oblasti nízko frekventovaného lexika v určitých skupinách slov, která se sice s malou frekvencí, ale pravidelně, opakují. Lexikum s touto frekvencí odráží aktuální kontext referovaných předmětů řeči, jedná se o lexikum přímo nezávislé na obsahu (styl, žánr a tematika jsou u obou autorů obdobné). Připomeňme, že základním předpokladem určování autorství je, že tato vrstva lexika vykazuje nejhorší vlastnosti při identifikaci autora (tzn. nepředpokládá se, že se nízko frekventované lexikum opakuje v závislosti na autorovi, viz Binongo, 2003; Boukhaled – Ganascia, 2015; Burrows, 2003; Juola, 2008; Juola – Baayen, 2003; Koppel – Schler – Argamon, 2009; Zhao – Zobel, 2005).

Z hlediska nízko frekventovaného lexika zvolíme pro analýzu její krajní formu, zaměříme se na slovní formy vyskytující se v autorských dílech pouze jednou, tzn. hapax legomena. Texty srovnáme na základě excerptce souborů slovních forem s frekvencí 1, pro vyjádření blízkosti analyzovaných textů použijeme kosinovou nepodobnost, metodou zobrazení je vícerozměrné škálování (MDS; Torgerson, 1952).

Obrázek 1: Analýza hapax legomen z autorských děl Svetozára Hurbána Vajanského a Martina Kukučína. Seznam analyzovaných děl: Svetozár Hurbán Vajanský: *Babie leto*, *Kandidát*, *Mier duše*, *Na Bašnárovom kopci*, *Podrost*, *Tri kabanosky*, *V jasku*; Martin Kukučín: *Dies irae (Dni hnevu)*, *Keď báčik z Chocholova umrie*, *Neprebudený*, *Regrúti*, *Rysavá jalovica*, *Tichá voda*, *Veľkou lyžicou*. Srovnání hapax legomen z vybraných děl je provedeno na základě kosinové nepodobnosti, pro zobrazení je použito vícerozměrné škálování.

Analýza hapax legomen 14 literárních děl Svetožára Hurbana Vajanského a Martina Kukučína ukazuje, že texty se shlukují ve skupinách na základě autorství. To znamená, že v jednotlivých dílech obou autorů se opakovaně objevují určité slovní formy, které u druhého autora nevidujeme. Zároveň to znamená, že se tyto slovní formy objevují s nízkou frekvenční charakteristikou, nikoliv vyšší. Tím jsou popřeny předpoklady disciplíny určování autorství (viz reference uvedené ve výkladu nad obrázkem) a všeobecný lingvistický předpoklad o náhodné struktuře textu širšího rozsahu. Připusťme ale, že výběr hapax legomen je do jisté míry umělý, jedná se o slova použitá v textu v nejmenší možné míře, pouze jednou. My se ovšem v ověřování Horeckého hypotézy o opakování slov v autorských textech chceme zaměřit na oblast nízko frekventovaného lexika v širším rozsahu. V následující analýze proto provedeme náhodný výběr slov z vybraných textů a z něj vybereme opět hapax legomena pro stejný postup analýzy jako v případě Obrázku č. 1.

Obrázek 2: Analýza hapax legomen z náhodného výběru 2500 slov z autorských děl Svetožára Hurbana Vajanského a Martina Kukučína, viz Obrázek č. 1.

Výsledek analýzy hapax legomen z náhodného výběru slov z vybraných děl obou autorů ukazuje, že se autoři odlišují na základě širšího rozsahu nízko frekventovaných slov. Zjišťujeme tedy, že se auttribuční vlastnosti prokázané v analýze

z Obrázku č. 1 netýkají pouze nejnižše položených položek ve frekvenčním seznamu (hapax legomen), ale podobné autorské rysy prokazuje širší rozsah nízko frekventovaného lexika použitého v textu. Opět se tak potvrzuje Horeckého formulace o pravidelném doslovném opakování slov v textu. Vzhledem k diskusi uvedené výše v bodech a/ *langue parolová hypotéza*, b/ *textová izotopie* a c/ *textová koheze* je dále nutné prověřit, zda se autorské hapaxy opakují samostatně, nebo v určitém okolí. V následujícím Obrázku č. 3 proto předkládáme analýzu totožného vzorku textu jako v předchozích analýzách. Pro ověření toho, zda se autorské hapaxy objevují v textu v určitých frázích, slovních spojeních či gramatických vazbách, využijeme n-gramovou analýzu. Použity jsou bi-gramy slov, zbylé vlastnosti analýzy zůstávají stejné jako v předchozích případech.

Obrázek 3: Analýza hapax legomen bi-gramů z autorských děl Svitozára Hurbana Vajanského a Martina Kukučína, viz Obrázek č. 1.

N-gramová analýza vybraného vzorku textů ukazuje, že ve srovnání s analýzou hapaxů literárních děl (absolutních hapaxů) a hapaxů z náhodných výběrů slov z textu (lokálních hapaxů) vykazují dvojkombinace slovních forem těsnější shlukování a tedy jasnější výsledek atribuční analýzy (položky vykazující shodu u daného autora jsou redukovány, ale zároveň nevykazují více shod s druhým autorem). Tzn. že

pokud použijeme hapaxy reprezentující jednotlivé texty ve vazbách, v nichž se vyskytují v textu, jasněji vyznačíme jejich autorský původ. Autorská hapax legomena proto můžeme považovat za jednotky, které se v textu objevují včetně svého textového okolí a určují tak strukturu textu v širším rozsahu, jak jsme to diskutovali výše v bodech a/ *langue parolová hypotéza*, b/ *textová izotopie* c/ *textová koheze*. Toto zjištění doposud nebylo publikováno a ukazuje na mnohem rozsáhlejší vázanost struktury textu na autorský slovník, než jak ukazuje zjištění o pravidelnosti samostatných hapaxů (viz Faltýnek – Matlach – Owsianková, 2020; Faltýnek – Matlach, 2020). Zároveň je tak posílena evidence o vnitrojazykovém výběru autora mezi synonymními způsoby vyjádření a způsobech zapojení těchto autorských voleb do struktury textu, kterou naznačuje Horecký.

DISKUSE

Můžeme konstatovat, že předpoklad Jána Horeckého ze *Základů jazykovedy* (1974, s. 90), že texty jsou do určité míry strukturovány na základě doslovného opakování slov, je možné vzhledem k předloženým analýzám považovat za jednoznačně správný. Prověřili jsme na 14 literárních dílech dvou autorů, respektive na nízko frekventovaném lexiku těchto literárních děl. Nízko frekventované lexikum bylo vybráno proto, že u něj nepředpokládáme opakování slovních forem v souvislosti s obsahem literárního díla nebo jeho stylem. Můžeme konstatovat, že na základě samostatného zpracování hapax legomen a hapax legomen z náhodného výběru slov z textu, jejich srovnání na základě kosinové nepodobnosti a zobrazení formou vícerozměrného škálování (viz Faltýnek – Matlach – Owsianková, 2020; Faltýnek – Matlach, 2020) se ukazuje, že texty vykazují autorské rysy v oblasti nízko frekventovaného lexika. Tento fakt doposud nebyl znám (Koppel a kol., 2009). V oblasti lingvistiky se předpokládá, že jsou nízko frekventované slovní formy závislé na kontextu a proto jsou dílem náhody a nevykazují žádnou pravidelnost či strukturu. V oblasti určování autorství se tento lingvistický předpoklad promítl tím způsobem, že nízko frekventované slovní formy nebyly analyzovány nikdy samostatně (případně se tak činilo v součinnosti s dalšími vlastnostmi textu, viz Baayen, 1996) a jako určující faktor autorského stylu byly považovány především vysoce frekventované položky. Nástroje, které byly pro analýzu využívány, pak nízko frekventované slovní formy odstraňovaly jako nepotřebný znak (Boukhaled – Ganascia, 2015; Zhao – Zobel, 2005).

Na závěr diskuse si dovolueme doplnit citát z Jána Horeckého, jehož hypotézu jsme po celou dobu prověřovali:

„Formou jazykového vedomia je banka dát, uložená v jazykovom vedomí. Báza dát – ako lexikálna zložka banky dát – je zložitý systém jazykových znakov, akýsi tezaurus, ktorý možno znázorniť ako sieť pomenovaní (jednoslovných i via-

cslovných). Každá substančná jednotka v tejto sieti tvorí uzol, charakterizovaný dvoma typmi vzťahov. Sú to predovšetkým vzťahy priradenosti, podradenosti a nadradenosti. Ale veľmi dôležité sú aj vzťahy atribúcie, statickej i dynamickej. (...) Je tu zrejma paralela k neuronovým sieťam v mozgu.“ (Horecký, 1991, s. 81 – 82).

Budeme-li uvažovať o dôsledkoch toho, že jsou autorské texty jednoznačne podobné na základě souborů nízko frekventovaných lexikálních položek, mění se nejen představa o volnosti použití jazyka v oblasti *parole*, o sémantické určenosti výběru lexika *kontextem* nebo o soudržných mechanismech textu založených především na prostředích využívajících *koreference* slovních forem – protože opakované slovní formy v širokém rozsahu textu referují k různým předmětům řeči napříč celým textem. V návaznosti na Horeckého citáty o mentálním charakteru jazyka uvedené výš můžeme konstatovat, že se mění naše celková představa o podobě mentálního slovníku. Tato představa dostává mnohem více než doposud podobu sítě (De Deyne a kol., 2016). V této síti jsou v určitých vrstvách autorského lexika určité položky závaznější, než položky jiné, a to bez ohledu na jejich absolutní frekvenci v daném textu nebo pomyslném jazykovém korpusu. Tyto položky určují zapojení ostatních položek do textu, ať již se jedná o položky stejně, níže nebo výše frekventované, protože jsou to právě ony, které musejí být ve výpovědi zapojeny, což je v přímém rozporu s všeobecnými předpoklady o rovnocenné roli frekvence napříč analyzovanými texty a jazykovými referenčními zdroji jako jsou korpusy (viz Forster – Chambers, 1973). Tuto disproporci mezi všeobecně jazykovým a autorským lexikem formuluje J. Dolník následujícím způsobem:

„A sme nastavení aj na zachovanie nášho behaviorálno-akčného poriadku. Ide nám aj o istotu nášho jazykovo-komunikačného bytia, k čomu patrí zachovanie poriadku prirodzenej jazykovej asimilačnej bázy.“ (Dolník, 2018a, s. 87).

Pokud si pripustíme rozsah autorské idiosynkrazie – „jistotu jazykového bytí“ – vzhľadom k rozsáhlému opakování určitých položek lexika napříč celým spektrem frekvence našich textů, jak budeme chápat identičnost prožívání každého aktuálního kontextu, k němuž se právě vztahují naše slova?

„Vyslovujem hypotézu, že pri tejto aktivite (používaní slov) ľudia prežívajú ekvivalentnosť používaných slov s tým, čo prežívajú, keď sú v intencionálnom stave, t. j. keď sú v zrkovom, sluchovom, hmatovom, čuchovom, chuťovom, emocionálnom, vôľovom, racionálnom aj iracionálnom styku so svetom, do ktorého patria aj oni sami.“ (Dolník, 2018b, s. 306).

Odpověď na Dolníkovu hypotézu a způsob „jazykového“ prožívání aktuálního světa zůstává na introspekci každého z nás. Slova a s nimi také „intencionální stav“ se v našich textech opakují mnohem pravidelněji, než jsme čekali. Zbývají nám spory o to, nakolik je polysémie každého slova, autorem stále opakovaného, vzhľadom k zapojení do aktuálního kontextu jedinečná. Každopádně: autor se vrátil (Barthes, 2006) a popřel jemu neznámé předpoklady spojené s absolutní frekvencí a s abstraktní povahou systému projevenou ve struktuře jeho textu.

Bibliografia

BAAYEN, Harald – VAN HALTEREN, Hans – TWEEDIE, Fiona: Outside the cave of shadows: using syntactic annotation to enhance authorship attribution. In: *Literary and Linguistic Computing*, 1996, roč. 11, č. 3, s. 121 – 132.

BARTHES, Roland: Smrt autora. In: *Aluze*, 2006, roč. 10, č. 3, s. 75 – 77.

BINONGO, José, Nilo, G.: Who wrote the 15th book of Oz? An application of multivariate analysis to authorship attribution. In: *Chance*, 2003, roč. 16, č. 2, s. 9 – 17.

BLOOMFIELD, Leonard: *Language*. New York: Henry Holt & Co 1933. 566 s.

BOUKHALED, Mohamed, Amine – GANASCIA, Jean-Gabriel: Using Function Words for Authorship Attribution: Bag-of-Words vs. Sequential Rules. In: *The 11th International Workshop on Natural Language Processing and Cognitive Science*, Oct 2014, Venice, Italy. de Gruyter, *Natural Language Processing and Cognitive Science Proceedings 2014, 2015*, s. 115 – 122.

BURROWS, John: Questions of Authorship: Attribution and Beyond. In: *Computers and the Humanities*, 2003, roč. 37, č. 1, s. 5 – 32.

ČERMÁK, František: *Jazyk a jazykověda. Přehled a slovníky*. Praha: Karolinum 2011. 382 s.

ČERNÝ, Jiří: *Úvod do obecné jazykovědy*. Olomouc: Rubico 1998. 248 s.

DE DEYNE, Simon – KENETT, Yoed, N. – ANAKI, David – FAUST, Miriam – NAVARRO, Daniel: Large-scale network representations of semantics in the mental lexicon. In: *Frontiers of cognitive psychology. Big Data in Cognitive Science*. Ed. N. M. Jones. Routledge/Taylor & Francis Group 2016, s. 174 – 202.

DOLNÍK, Juraj: *Všeobecná jazykověda: opis a vysvetľovanie jazyka*. Bratislava: Veda, vydavateľstvo SAV 2009. 376 s.

DOLNÍK, Juraj: Jazykové znalosti a ovládanie jazyka. In: *Jazykovedný časopis*, 2018a, roč. 69, č. 1, s. 77 – 89.

DOLNÍK, Juraj: Lexikálny význam v recepcii. In: *Jazykovedný časopis*, 2018b, roč. 69, č. 3, s. 302 – 315.

FALTÝNEK, Dan: Má tedy Kroužek sedmou publikaci. In: *Bohemica Olomucensia*, 2013, roč. 4, s. 318 – 328.

FALTÝNEK, Dan – MATLACH, Vladimír: Hapax Remains: authorial features of textual cohesion in authorship attribution. Preprint, 2020, DOI: 10.13140/RG.2.2.19865.24164.

FALTÝNEK, Dan – MATLACH, Vladimír – OWSIANKOVÁ, Hana: Hapax legomena jako indikátor autorského stylu a formální znak koheze textu. Preprint, 2020, DOI: 10.13140/RG.2.2.16509.79847.

FORSTER, Kenneth, I. – CHAMBERS, Susan, M.: Lexical access and naming time. In: *Journal of Verbal Learning & Verbal Behavior*, 1973, roč. 12, č. 6, s. 627 – 635.

GREIMAS, Algirdas, Julien: *On Meaning Selected Writings in Semiotic Theory*. Minneapolis: University of Minnesota Press 1987. 251 s.

HELBIG, Gerhard: *Vývoj jazykovědy po roce 1970*. Prekl. J. Nekvapil – J. Holšánková. Praha: Academia 1991. 303 s.

HIRSCHOVÁ, Milada: *Úvod do teorie textu*. Olomouc: Univerzita Palackého v Olomouci 1989. 243 s.

HORECKÝ, Ján: Úvod do matematickej jazykovedy. Bratislava: Univerzita Komenského 1969. 92 s.

HORECKÝ, Ján: A Model of Slovak Conjugation. In: Prague Studies in Mathematical Linguistics, 1970, č. 3, s. 259 – 271.

HORECKÝ, Ján: Základy jazykovedy. Bratislava: Slovenské pedagogické nakladateľstvo 1978. 176 s.

HORECKÝ, Ján: Jazykové vedomie. In: Jazykovedný časopis, 1991, roč. 42, č. 2, s. 81 – 88.

JUOLA, Patrick: Authorship attribution. In: Foundations and Trends in Information Retrieval, 2008, roč. 1, č. 3, s. 233 – 334.

JUOLA, Patrick – BAAYEN, Harald: A controlled-corpus experiment in authorship identification by cross-entropy. In: Literary and Linguistic Computing, 2003, roč. 20, č. Suppl. Issue, s. 159 – 167.

KELIH, Emmerich – ANTIĆ, Gordana – GRZYBEK, Peter – STADLOBER, Ernst: Classification of author and/or genre? The impact of word length. In: Classification: The Ubiquitous Challenge. Eds. C. Weihs – W. Gaul. Heidelberg: Springer, 2005, s. 498 – 505.

KOPPEL, Moshe – SCHLER, Jonathan – ARGAMON, Shlomo: Computational methods in authorship attribution. In: Journal of the American Society for Information Sciences and Technology, 2009, roč. 60, č. 1, s. 9 – 26.

KOŘENSKÝ, Jan: Za Jánem Horeckým. In: Slovo a slovesnosť, 2007, roč. 68, č. 1, s. 78.

ORGOŇOVÁ, Oľga – BOHUNICKÁ, Alena: Interakčná štylistika. Bratislava: Univerzita Komenského 2018. 208 s.

PALEK, Bohumil: Základy obecné jazykovědy. Praha: SPN 1989. 288 s.

TĚŠITELOVÁ, Marie: Profesor Ján Horecký šedesátníkem. In: Slovo a slovesnosť, 1980, roč. 41, č. 2, s. 157 – 160.

TORGERSON, Warren S.: Multidimensional scaling: I. Theory and method. In: Psychometrika, 1952, roč. 17, č. 4, s. 401 – 419.

ZHAO, Ying – ZOBEL, Justin: Effective and scalable authorship attribution using function words. In: Proceedings of 2nd Asian Information Retrieval Symposium, 2005, s. 174 – 189.

ZUMRÍK, Miroslav: Ján Horecký's Approach to Language and Thinking. In: Jazykovedný časopis, 2017, roč. 68, č. 2, s. 426 – 431.