Productivity of Verbal Nouns. A corpus-based approach on Czech and Polish Verbal Nouns

key words: productivity; verbal nouns; Polish; Czech; language corpus

Czech and Polish verbal nouns (hereinafter VN) (Czech *čtení*, Polish *czytanie* 'reading') represent a highly productive and regularly formed group of nouns (Daneš 1987; Пчелинцева 2016; Łaziński 2020; Kocková 2022; Shcherbii et al. 2024). Since the derivational model «verb stem + –ije» is of Proto-Slavic origin, this type of deverbal nouns is present in all Slavic languages, including dialects. However, a comparative perspective reveals significant differences between VNs in terms of productivity and retention of verbal properties (cf., lexical meaning of the verb, aspectual characteristics). Studies of Slavic verbal nouns demonstrate that the productivity of VNs, as well as their retention of verbal properties, increases considerably "from east to west" (Dickey 2000; Пчелинцева 2016; Kocková 2022: 144). This paper presents a comparative corpus-based study of the productivity of VNs and explores the possibility of using a semi-automatic method to compare nominal and verbal lemmas obtained from language corpora.

In Czech and Polish, VNs exhibit a very high productivity and regularity in the formation; it is possible to form a VN from almost every verb and from both aspects of the verb. At the same time, some discrepancies can be observed between the frequency of verbs and verbal nouns (cf., cs *nutkání* 'compulsion' – *nutkat* 'to compel*', bání se* 'fearing' – *bát se* 'to fear'). Many of VNs exist only as potential, or occasional units, and it is preferred to use another deverbal noun instead (cf. in Czech, low-frequent VN *pracování* 'working': 8 occurrences in the corpus syn2020 CNC; high-frequent deverbal noun *práce* 'work'*:* 84 959 occurrences*,* cf. also Karlík 2004). Some authors suggest that the causative, modal, frequentative and state verbs in Czech (Havránek/Jedlička 1930; Křížková 1968: 134; Karlík 2004: 77), and verbs with relational or modal meaning, or impersonal meaning of the verb in Polish (Puzynina 1969, Marzniakowa 1993 etc.) do not form the VNs. However, the corpus data do not confirm these restrictions on productivity (for Czech cf., Giger/Kocková 2024).

In order to compare the frequency of verbs and VNs, we have developed a procedure for semi-automatic comparison of verb stems and stems of verbal nouns, which allows larger amounts of data to be processed. In this way, we have obtained a list of verbs that do not have a corresponding verbal noun and vice versa. The present paper relies on data from the Czech balanced corpus syn2020 CNC and on the Polish corpus NKJP, from which a list of the most frequent verbal lemmas and lemmas of verbal nouns was selected. The stem was then extracted from both lists by automatically removing the suffixes. Based on the identification of regular changes in the formation of VNs (e.g., change in vocalic quantity and quality) automatic stem changes were made in each relevant group of VNs: e.g. in Czech, a noun stem ending in -á (očekává-ní 'expectation') was changed to -a (očekáva-). Subsequent comparison of the lists of lemma stems (e.g. očekáva-t 'to expect') demonstrates a high degree of agreement, yet also reveals a group of verbs that do not show corresponding verbal nouns (pl *poszumieć* 'to make noise'- *?,* cs *připadat* 'to seem' – ?) or occur with low frequency, cf., (pl *potrzymać* 'to hold'– *potrzymanie* 'holding'*,* cs *smát se* 'to laught' – *smání* 'laughing') and a group of nouns that synchronically are not associated with a corresponding verb (cs *ponětí,* pl *pojęcie* 'concept, notion'). The lemmas obtained will also be examined in terms of aspect, semantics (incl. lexicalization) and other factors.

References:

Daneš, František (1987): *Mluvnice češtiny 3. Syntax*. Praha: Academia.

Dickey, Stephen M. (2000): *Parameters of Slavic aspect: a cognitive approach*. Stanford, Calif.: CSLI Publ (Dissertations in linguistics).

Giger, Markus/Kocková, Jana (2024): Grenzüberschreitungen an der Peripherie: Aspektuelle Funktionen von Aktivpartizipien und Verbalsubstantiven im Tschechischen. In *Zeitschrift für Slawistik*, 69(1), 1–26.

Havránek, Bohuslav/Jedlička, Alois (1930): Česká mluvnice. Praha: SPN.

Karlík, Petr (2004): Mikrosyntax českých deverbálních jmen. In *Sborník prací Filosofické fakulty brněnské university. A, Řada jazykovědná*, 52(1), 71–81.

Kocková, Jana (2022): *Neurčité tvary slovesné v češtině, ruštině a němčině a jejich vzájemná ekvivalence*. Praha: Academia, Slovanský ústav AV ČR.

Křížková, Helena (1968): Substantiva s dějovým významem v ruštině a v češtině. In *Kapitoly ze srovnávací mluvnice ruské a české III. O ruském slovese*. Praha: Academia, 81–152.

Łaziński, Marek (2020): *Wykłady o aspekcie polskiego czasownika*. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.

Marzniakowa, Irena (1993): *Gramatyka konfrontatywna rosyjsko-polska*. Warszawa: PWN.

NKJP: National Corpus of Polish (2008): https://nkjp.pl/.

Pchelintseva, Elena (2016): *Ot glagola k imeni: aspektualnost‘ v russkyx, ukraynskyx i pol’skyx imenax dejstvija*. St. Petersburg: Nauka.

Puzynina, Jadwiga (1969): *Nazwy czynności we współczesnym języku polskim (słowotwółrstwo, semantyka, składnia)*. Warszawa: PWN SA.

Shcherbii, Nataliia/Vorobets, Oleksii/Mytsan, Dariia/Korpalo, Oksana/Kuravska, Nataliia (2024): The debate on the nature of verbal nouns in Slavic languages: nominal or verbal? In *Eduweb*, 18(2), 238–250.

SYN v9 (2021): Corpus SYN, version 9 from 5. 12. 2021. Ústav Českého národního korpusu FF UK, Praha. https://www.korpus.cz.

SYN2020 (2020): representative corpus of written Czech: Ústav Českého národního korpusu FF UK, Praha. http://www.korpus.cz.