

Články / Articles

Monokolokabilní slova v češtině: jejich hlavní aspekty

Monocollocable Words in Czech. Their Main Aspects /3

František Čermák

Zum adverbialen Argument in den reflexiven Dispositionen des Tschechischen II

The adverbial argument in Czech dispositional reflexives II /13

Aleš Půda

Koreference a elipsa v Pražském závislostním korpusu

Coreference and Ellipsis in Prague Dependency Treebank /35

Kateřina Rysová, Magdaléna Rysová

Slovotvorný přehled slov s číselným významem II: číslovky neurčité

A word-formation oriented overview of words with numeral meaning II: indefinite numerals /48

Josef Šimandl

Opakování a elipsa reflexivního morfému *se* v koordinovaném spojení zvrtných sloves

Repetition and omission of the reflexive morpheme *se* in coordination of reflexive verbs /61

František Štícha

Recenze / Reviews

Schneiderová Soňa: Analýza diskurzu a mediální text. Praha: Karolinum, 2015. /84

Lucie Jílková

Koupil Ondřej: Grammatyka česka – mluvnice češtiny v 16. až 19. století (katalog výstavy). Filip Tomáš – Akropolis, 2015. /88

František Štícha

Zprávy / News

Co je v ČNK nového VII (Zprávy z Českého národního korpusu) /91

Michal Křen

Pokyny pro autory /93

Instructions for authors /95

Objednávka /98

Order form /99

Monokolokabilní slova v češtině: jejich hlavní aspekty

František Čermák

Ústav Českého národního korpusu Filozofické fakulty Univerzity Karlovy, Praha

frantisek.cermak@ff.cuni.cz

Monocollocable Words in Czech. Their Main Aspects

ABSTRACT: Monocollocable words are such words and word forms that occur in a single lexical combination only or in very few, whose number is severely restricted and set. Practically, they are found as parts of set idioms and multi-word terms. They are found in many other languages, cf. English *tenterhooks* or Russian *bakluši*. Czech examples *dát/dostat najevo, na viděnou, je mi líto, říct/mluvit/hrát nahlas, je známo, je zapotřebí, být třešničkou na dortu, není divu, jít/chodit pěšky, dát/dostat zadarmo* illustrate this in more detail, showing, at the same time, that there might be a limited variation found, too, but, above all, that these are, in fact, no full-fledged words, lacking most of their basic characteristics, such as meaning, word-class membership, etc. In the sense of their severely limited combinatorial capacity, these words, less known under such alternative labels as cranberry words, form a substantial and irregular periphery of language and its lexicon. The contribution briefly comments on some of their aspects and suggests that broadly some classes or types can be recognized.

KEY WORDS: monocollocable words, language periphery, combination

KLÍČOVÁ SLOVA: monokolokabilní slova, periferie jazyka, kombinace

Zum adverbialen Argument in den reflexiven Dispositionen des Tschechischen II

Aleš Půda

Slavisches Institut, Ruprecht-Karls-Universität, Heidelberg

ales.puda@slav.uni-heidelberg.de

The adverbial argument in Czech dispositional reflexives

ABSTRACT: The study presents some corpus-based results about dispositional reflexive constructions in Czech. One type of Czech dispositional reflexives expresses the evaluation of a mental state with regard to an event associated with the predicate (*Snad se mu žije lehčeji*). The evaluation of experiencing an underlying event in a particular way is usually encoded lexically by an adverbial phrase, the licensing conditions of which have not yet been investigated both within Western formal approaches (“manner adverb”, “some adverbial modification”) and within the Slavic grammatical tradition. Based on the analysis of corpus data, the formal (lexical, syntactic and prosodic) properties of this (explicit or implicit) adverbial argument will be examined. Then the frequency distribution of the adverbial argument (initial, preverbal, postverbal, final) within the clause will be established. Attention will also be given to the frequency of specific adverbs in both affirmative and negative clauses. Furthermore, an explanatory account of the adverbial selection in dispositional reflexives will be developed. When a reflexive verbal form is combined with an adverb (θ -identification), different main types of manner adverbs need to be distinguished, according to the semantic feature of ‘activation’ of the adverb [eval] and the θ -grid ($\langle\theta$ -Exp \rangle , $\langle\theta$ - x \rangle): obligatory evaluative (*nerado* ‘unwillingly’), facultative evaluative (*výborně* ‘excellently’) pragmatically evaluative (*přirozeně* ‘naturally’) and non-evaluative adverbs (*rozhořčeně* ‘indignantly’, *rychle* ‘quickly’). Within the class of evaluative adverbs, one can further differentiate between primary evaluative adverbs (*dobře* ‘well’) and secondary evaluative adverbs (*vesele* ‘cheerfully’), according to their semantic structure. The behaviour of adverbs that do not define their pole as either positive or negative (*normálně* ‘normally’, *nějak* ‘somehow’), or those that can be tied contextually to both poles (*zvláště* ‘oddly’) will be outlined briefly. Finally, the proposed paper will discuss three distinct functions of the modifier *sám*, *-a*, *-o* (‘alone’) in syntactic reflexives: autocausative, autoagentive and – last but not least – dispositional.

KEY WORDS: adverbial argument, autoagentive, autocausative, dispositional reflexive constructions, disposition, evaluation, evaluative adverb, experiencer, Czech syntactic reflexives, middle construction

Koreference a elipsa v Pražském závislostním korpusu

Kateřina Rysová, Magdaléna Rysová

*Ústav formální a aplikované lingvistiky, Matematicko-fyzikální fakulta, UK Praha
rysova@ufal.mff.cuni.cz, magdalena.rysova@ufal.mff.cuni.cz*

Coreference and Ellipsis in Prague Dependency Treebank

ABSTRACT: The paper focuses on which role is given to elided sentence participants in coreference chains, i.e. whether (and to which degree) the participants that are present only implicitly in the surface layer are involved in relations of textual and grammatical co-reference. Generally, the paper introduces the methods how it is possible to examine the interplays of different language phenomena in corpus data of the Prague Dependency Tree-bank containing multilayer annotation.

KEY WORDS: ellipsis, coreference, corpus linguistics, Prague Dependency Treebank, explicitness, implicitness

KLÍČOVÁ SLOVA: elipsa, koreference, korpusová lingvistika, Pražský závislostní korpus, explicitnost, implicitnost

Slovotvorný přehled slov s číselným významem II: číslovky neurčité

Josef Šimandl

Ústav pro jazyk český AV ČR, v.v.i., Praha
simandl@ujc.cas.cz

A word-formation oriented overview of words with numeral meaning II: indefinite numerals

ABSTRACT: This article is a counterpart to the overview I: definite numerals in the previous number. Three kinds of indefiniteness are introduced: (1) the problematic subclass of traditional indefinite numerals in the strict sense (lexical means for giving quantities on the scale „few–much“), the most representants of it being rather adverbs of measure; (2) the subclass of algebraic numerals and (3) the subclass of deictic numerals containing demonstrative, interrogative/relative and indefinite numerals in the sense of indexical indefiniteness. One group of indefinite deictics, *nevímkolik* and suchlike, is presented as one pole of continuum with expressions in the sentence form on the other pole. Combinations of indefinite numerals with other numerals, with other parts of speech etc., used when expressing an indefinite quantity, are mentioned as well.

KEY WORDS: building of Czech numerals, indefinite numerals, indefinite numerals in the strict sense, algebraic numerals, deictic numerals

KLÍČOVÁ SLOVA: tvoření českých číslovek, neurčité číslovky, číslovky neurčité v užším smyslu, číslovky algebraické, číslovky deiktické

Opakování a elipsa reflexivního morfému *se* v koordinovaném spojení zvratných sloves

František Štícha

Ústav pro jazyk český AV ČR, v.v.i., Praha

sticha@ujc.cas.cz

Repetition and omission of the reflexive morpheme *se* in coordination of reflexive verbs

ABSTRACT: It is only the aid of large corpora (several billions of words) that enables us to discover some intuitively and spontaneously followed rules of grammar. Different kinds of ellipsis and non-ellipsis (repetition of a word or a nominal phrase, which – under some conditions – can be omitted) can also be governed by such rules. The corpus findings of sentence structures as (1) *Zastavila se a podívala se na hodinky* or (2) *Zastavila se a podí-vala na hodinky* (*She stopped and looked at her watch*) have clearly shown that ellipsis as well as repetition is a (strict) rule under specific semantic and syntactic conditions.

KEY WORDS: ellipsis, repetition, reflexive verbs, reflexive morpheme, coordination

KLÍČOVÁ SLOVA: elipsa, opakování, reflexivní slovesa, reflexivní morfém, koordinace

Soňa Schneiderová: Analýza diskurzu a mediální text

Schneiderová, S.: *Analýza diskurzu a mediální text*. Praha: Karolinum, 2015, 164 stran.

Lucie Jílková

Ústav pro jazyk český AV ČR, v.v.i., Praha

jilkova@ujc.cas.cz

Spojení jména autorky a titulu recenzované publikace zřejmě nikoho, kdo se danou oblastí zabývá, nepřekvapí, neboť Soňa Schneiderová se analýze českých mediálních textů věnuje dlouhodobě. Její monografii *Analýza diskurzu a mediální text* lze rozdělit na část teoretickou a praktickou. V první, teoretické části autorka čtenáře uvádí do prostředí nesnadno vymežitelné vědní disciplíny, do prostředí analýzy diskurzu. Věnuje se zahraničním i českým nejvýznamnějším badatelům, jejichž jméno je s analýzou diskurzu, respektive kritickou analýzou diskurzu tradičně spojováno. Upozorňuje, že dnes široce užívaný termín *diskurz* není snadné jednoznačně definovat, že tento výraz zjevně nelze jednoznačně užívat jako synonymum ke slovu *text*, k čemuž však u některých autorů dochází. Za užitečné autorka považuje vymezení Foucaultovo, tedy pojetí diskurzu jako určitý princip a pravidla podílející se na zobrazování, resp. vytváření skutečnosti. Mnoho prostoru je logicky věnováno zásadní osobnosti N. Fairclougha, který mimo jiné tvrdí, že při diskurzivně zaměřené analýze textu je třeba znát společenský kontext, ve kterém text vznikl, a že taková analýza by měla o jazykových datech říci více než formální jazykověda. Autorka se dále pokouší Faircloughovo pojetí analýzy diskurzu propojovat s postřehy dalších významných badatelů, a to M. Foucaulta, M. M. Bachtina, M. A. K. Hallidaye, T. van Dijka a J. Habermase. Je třeba říci, že četba těchto teoretických pasáží věnovaných zahraničním autorům vyžaduje čtenáře v dané problematice již dosti zblhlého. Čtenáři méně připravenému jako úvod do analýzy diskurzu snad lépe než recenzovaná publikace poslouží třeba texty J. Hoffmannové (1997), J. Nekvapila (2006), J. Homoláče (2006), K. Mrázkové (1998) nebo E. Lehečkové (2015), což jsou – s výjimkou textu posledně jmenovaného – zároveň práce, na které se S. Schneiderová odvolává. Domácí autoři, kteří se věnují analýze diskurzu, rovněž nezůstávají S. Schneiderovou opomenuti. J. Hoffmannová je zde představena jako autorka mnoha publikací věnovaných analýze dialogu, konkrétněji třeba tematické výstavbě dialogu, reprodukci cizí a vlastní řeči, neverbálním projevům, prostředkům souladu a konfliktu aj., v nichž je využíváno postupů konverzační analýzy. Poněkud zarážející je, že prakticky stejnou měrou jsou citovány i práce O. Müllerové, S. Čmejrkové a J. Zemana, jejichž jména se však do názvu kapitoly nedostala. Podobná je situace u oddílu věnovaného pracím J. Nekvapila, v němž jsou citovány práce, které tento autor napsal společně s I. Leudarem, v titulku oddílu opomenutým. J. Nekvapil a I. Leudar, spjatí s členskou kategorizační analýzou, jsou mimo jiné autory konceptu mediální dialogické sítě, na který se autorka odvolává i v další, praktické části. Další domácí osobností, která je v recenzované knize představena, je J. Homoláč, jenž ve svých analýzách propojuje postupy etnometodologie, členské kategorizační analýzy a diskurzivní psychologie. U J. Krause autorka připomíná jeho tradiční disciplínu – rétoriku, přičemž rétorické strategie mohou mimo jiné objasnit negociační a komunikační funkci diskurzu.

Po představení prací domácích badatelů následují dva opět čtenářsky náročné oddíly nazvané *Analýza diskurzu v různých vědních oborech* (tj. v literární vědě, v sociálních vědách, v psychologii, ve filozofii a v edukaci) a *Možné přístupy k analýze diskurzu*. Také pro tyto části platí, že jejich čtenář už musí mít nastudovány nějaké (např. výše zmíněné) texty jiné, aby těmto pasážím porozuměl. S. Schneiderová v nich ukazuje svůj obdivuhodný rozhled nejen v literatuře

jazykovědné, ale právě i v literatuře oborů spřízněných. Jako poněkud problematické se mi jeví, že některé zde uváděné poznatky byly již zmíněny dříve. Například o diskurzivní psychologii byla již řeč dříve v souvislosti s pracemi J. Homoláče, zde se o této disciplíně mluví znova, byť jsou zde zmíněni další (domácí i zahraniční) badatelé, kteří se jí zabývali.

Následuje praktická část, *Mediální diskurz*, v níž autorka analyzuje konkrétní texty z prostředí českých médií, a to výhradně ze zpravodajství. Ukázky pocházejí především ze zpráv televize Nova, méně často z České televize a televize Prima. Hned poznamenejme, že tato prakticky zaměřená část publikace se čte daleko snáze, než tomu bylo u předcházejících teoretických částí, ovšem čtenářskou přístupností není nijak snížena odborná úroveň textu. Naopak, například Faircloughův (a také třeba Wodakové a dalších významných autorů) tzv. triangulační či trojfázový princip uplatňovaný při analýze jazykového materiálu autorka na vybraných příkladech názorně předvádí. Triangulačním principem se rozumí oddělení a) analýzy jazykových prostředků, b) analýzy postihující provázanost textů a situačního kontextu a konečně c) analýzy společenského dopadu textů; autorka však poznamenává, že těžištěm její analýzy jsou především části a) a b).

Analýzou jazykových prostředků užívaných v současných zpravodajských textech a s přihlédnutím ke kontextu, ve kterém tyto texty vznikají, lze podle autorky sledovat některé obecné postupy uplatňované při vytváření zpravodajských relací.

K nim patří, že média divákům poskytují tzv. interpretativní a analytické zpravodajství, tj. k popisované skutečnosti (tématu, události) zauímají nějaký postoj a vedou diváka k tomu, aby také on k dané skutečnosti zaujal týž postoj.

Dále se ukazuje, že zpravodajství vytvářejí do nějaké míry sami diváci. Jejich názory se stávají součástí informování o nějakém tématu, případně diváci sami vnášejí do zpravodajství svá témata. Informace jsou často podávány zábavným až bulvárním způsobem, zpravodajství se proměňuje ve společenský magazín. S tím souvisí i výběr témat, mezi nimiž se objevují také třeba informace ze soukromí známých osobností. Snaha po zábavném podávání informací může vést až k nesouladu mezi sdělovaným obsahem a užitými jazykovými prostředky; autorka k tomu uvádí příklad, kdy jsou záběry loupežného přepadení uvedeny slovy *nabízíme sestřih toho nejlepšího*.

Nedílnou součástí zpravodajství jsou upoutávky na další pořady vysílané touž televizí. Prostřednictvím upoutávek se tak do zpravodajství dostávají strategie diskurzu reklamního. Extrémním případem bulvarizace, který autorka uvádí, je třeba propojení informace o úmrtí herce s upoutávkou na seriál, který televize právě vysílá a v němž zesnulý herec hraje.

Ve zpravodajství je užíván jazyk, který poutá pozor sám o sobě, jazyk, který vyvolává emoce, jazyk, který přitáhne pozornost diváka k médiu. S. Schneiderová ukazuje, že ve zpravodajství jsou velmi běžně užívány výrazy (původně) expresivní, jejichž volba však nemusí být vždy dobře promyšlená, například ve větě *S nadměrným používáním igelitových tašek chtějí zatočit úředníci Evropské unie* není bez dalšího komentáře jasné, zda se výrazem *zatočit* míní, že úředníci EU chtějí používání igelitových tašek *zakázat* nebo *omezit*. Velmi běžné je začleňování frazémů, třeba *Ministr zdravotnictví má další zářez na pažbě*, nebo aluzí: *Civilizovaný svět, řečeno slovy klasika, lehce pozvedává obočí*. Autorka upozorňuje, že při nadměrném užívání expresivního lexika se expresivní síla výrazů časem vytrácí, že původně neotřelé frazémy se stávají klišé.

Novým, dnes však již plně etablovaným mediálním žánrem je tzv. příběh, též story či storka. S. Schneiderová v souvislosti s ním upozorňuje na interní materiál Mladé fronty DNES, ve kterém jsou novináři tohoto periodika vyzýváni k tomu, aby *zkoušeli dostat prvky story, kam to jde, i do běžných zpráv*. Dovolují si k tomu poznamenat, že takovéto návody pro pisatele mediálních textů mohou být pro jazykovědce nepochybně zdrojem zásadního poučení o tom, jak výsledná podoba mediálního textu vznikala. Je ovšem zjevné, že přístup k těmto materiálům je zpravidla velmi omezen. S. Schneiderová konkrétními ukázkami dokládá, že hlavním aktérem příběhu (story) ve

zpravodajství nemusí být nutně jen člověk: *Volby – to jsou i příběhy voličů*, ale i zvířata: *příběh šestiměsíční srnky Milušky* nebo neživé objekty: *příběh mostu*, *příběh systému vyplácení dávek* atd. Autorka vysvětluje, že v některých případech příběh může dobře splnit svou primární úlohu, tedy prostřednictvím jednotlivého osvětlit obecné, jindy však od informace, která je pomocí příběhu divákům sdělována, spíše odvede pozornost (příběh cestujícího, jemuž propadne letenka kvůli stávce zaměstnanců letiště, ovšem tématem zprávy byla právě stávka, její důvody atd.).

Zpravodajství je běžně vytvářeno prostřednictvím citátů osobností, o nichž se ve zprávě hovoří. Vysoký podíl citátů ve zpravodajství autorka vysvětluje snahou autorů mediálních textů být autentický, zároveň se autoři prostřednictvím citátů preventivně brání tomu, že budou nařčeni z nepřesnosti, a konečně může jít i o signál pohodlnosti (jednodušší je citovat někoho jiného než vytvářet vlastní text). Zásadní samozřejmě je, zda je citát přesný, nezkrácený, do jakého kontextu je vsazen atd.

Závěrem lze konstatovat, že v knize S. Schneiderové je pohledem analýzy diskurzu podrobně analyzován dobře vybraný, značně rozsáhlý jazykový materiál. Na textech ze současného televizního zpravodajství autorka přehledně ukázala, prostřednictvím jakých jazykových prostředků je současné zpravodajství utvářeno, jaké normy a tendence jsou pro ně platné. Zajímavou výzvou pro učitele (a jistě nejen pro ně) je pokusit se o takový rozbor společně se žáky/studenty v rámci vyučování, k čemuž autorka na konci své publikace vybízí.

LITERATURA:

HOFFMANNOVÁ, J. (1997): *Stylistika a...*, Praha: Trizonia

HOMOLÁČ, J. (2006): Diskurz o migraci Romů na příkladu internetových diskusí, *Sociolo-gický časopis / Czech Sociological Review* 42, č. 2, 329–351.

NEKVAPIL, J. (2006): Úvodem k monotematickému číslu: Analýza promluv a textů, analýza diskurzu, *Sociologický časopis / Czech Sociological Review* 42, č. 2, 263–267.

LEHEČKOVÁ, E. (2015): Analýza diskurzu a kritická analýza diskurzu in: *Aplikovaná lin-gvistika. Příručka pro studenty Bc. Studia ČJL* (Šebesta, K. – Lehečková, E. – Pierścieniak, P. P. – Šormová, K.), Praha: Karolinum, 78–95.

MRÁZKOVÁ (=Karhanová), K. (1998): Norman Fairclough: Discourse and Social Change. *Slovo a slovesnost*, 59, č. 3, 225–228.

Ondřej Koupil: Grammatyka czeska

Koupil, O.: Grammatyka czeska – mluvnice češtiny v 16. až 19. století (katalog výstavy). Vyd. Filip Tomáš – Akropolis, 2015, 174s.

František Štícha

Ústav pro jazyk český AV ČR, v.v.i., Praha
sticha@ujc.cas.cz

Kniha, o které zde podáváme zprávu, je katalogem k výstavě gramatik a jim podobných jazykových (zejména pravopisných) příruček českého jazyka vydaných knižně od 16. do konce 19. století. Výstava probíhala od října 2015 do konce ledna 2016 v Národním muzeu v Praze.

Tato krásná kniha není ovšem pouhým „katalogem“ v základním smyslu toho slova. Lze ji vnímat spíše jako stručné dějiny českého mluvnictví (gramatikografie). Představeno a charakterizováno je 65 příruček, některé z nich jsou však jen různá vydání (v podstatě) téhož díla.

Kniha má 9 oddílů: I. Úvod, II. Syntetický pohled, III. Katalog, IV. Galerie autorů, V. English Summary, VI. Zkratky a literatura, VII. Seznam vyobrazení, VIII. Rejstřík osob, IX. Příloha: *Rozepře nebolizto Controversiae českých gramatiků*.

V *Syntetickém pohledu* Ondřej Koupil zajímavě poukazuje na fakt, že „Pěstování mluvnic se ani v českých zemích nikdy neobešlo bez emocí; i tady platí slova Briana Cummingse o tom, že studium gramatiky nebylo nikde 'unaffected' a vždycky souviselo s literárními, duchovními, kulturními a politickými polemikami.“ K tomu mám jednu zásadní poznámku (která ovšem nijak bezprostředně nesouvisí s recenzovanou knihou): Je přirozené, že emocí je tím více, čím méně objektivních dat se veřejnosti předestírá a čím subjektivnější jsou tvrzení autora. Dnes, díky možnostem, které lingvistice poskytují velké elektronické korpusy, se případné emoce veřejnosti mohou oprávněně dotýkat jen hrstky jednotlivostí, pro které ve standardním psaném úzu neexistuje dostatečná opora. Zcela jinak tomu ovšem bylo v dobách minulých.

V knize se na s. 15 dozvídáme, že v roce 1533 vydaná první mluvnice češtiny, takzvaná „náměšťská“, byla dokonce první mluvnicí živého slovanského jazyka vůbec. Koupil dále upozorňuje na to, že její autoři nebyli jazykoví teoretici, ale praktici, jejichž hlavním cílem bylo přeložit do češtiny Nový zákon, a že první česká mluvnice má charakter jakéhosi jejich pracovního deníku.

Po uplynutí několika staletí i našeho obrození v 1. pol. 19. století se s druhou polovinou tohoto století „dostáváme do terénu málo známého“, píše Koupil. „Gramatika češtiny se stahuje, obrazně řečeno, z literárního výsluní a z kolbiště, kde se bojuje o prestiž jazykově české kultury, do škol a na bohemistické univerzitní katedry. Adresáty mluvnic se postupně stávají hlavně středoškolské studenty a praktici jazykového provozu – redaktoři, spisovatelé, žurnalisti.“ K tomu bych si opět dovolil podotknout, že v průběhu 20. století se spisovatelům podařilo zcela se vymanit z jakýchkoli ohledů na jakoukoli knižní gramatiku a lingvistiku vůbec. Posledním velkým českým spisovatelem, který ještě něco věděl o lingvistice, byl patrně Karel Čapek.

Hlavní část knihy, prostě nazvaná Katalog, je rozčleněna do oddílů podle jednotlivých století. V jejich rámci jsou pak jednotlivé gramatiky a jim podobné příručky (průběžně číslované, opatřené zkratkou názvu, rokem vydání a dalšími bibliografickými údaji) stručně charakterizovány co do obsahu. Obvykle se dozvíme, do kolika jak nazvaných oddílů je daná příručka rozvržena – bývá to

většinou jen pravopis a tvarosloví (zahrnující v podstatě pouze deklinaci a konjugaci), v tehdejších gramatikách ovšem zvané „etymologie“.

Jestliže první tištěná česká „gramatika“, takzvaná „náměšťská“, má dvě části, z nichž první se věnuje pravopisu a druhá „etymologii“, v níž se probírají „tvaroslovné a překladatelské otázky“, pak z dnešního hlediska bychom k této příručce měli poznamenat, že jde o „gramatiku“ jen podle názvu a z dobového, nikoli dnešního hlediska.

Pokud se některé z „gramatik“ 17. a 18. století věnovaly kromě skloňování a časování také syntaxi, pak šlo opět jen o její „kostru“ – kongruenci a rekcii. V příkrém kontrastu k tomu jsou dnešní velké gramatiky angličtiny, které jsou velkou většinou popisem a výkladem jevů syntaktických. Z prostého důvodu: angličtina totiž, jak známo, ve srovnání se slovanskými jazyky téměř žádnou morfologii nemá. O velké většině syntaktických jevů popisovaných a vykládaných gramatikami 2. pol. 20. století a století 21. neměl ovšem žádnou představu nejen Dobrovský, ale ani Gebauer – čímž chci říct pouze to, že lingvistika prošla během 20. století, zejména v jeho druhé polovině, vývojem podobným tomu, jaký přírodní vědy, zejména biologie, prodělaly už ve století osmnáctém a devatenáctém.

Mezi české gramatiky 17. století je zařazena i mluvnice latiny od „francouzského filozofa a gramatika Petra Rama“, která je do katalogu zahrnuta proto, že je „příkladem souvislosti domácího mluvnictví s gramatikografií světových ja-zyků.“

Z charakteristiky česko-latinské mluvnice jezuitského překladatele Jiřího Konstance se dozvídáme, že tato gramatika, vydaná v roce 1667, je prvním knižním projevem „brusičského“ postoje k užívání jazyka.

O mluvnici Františka Jana Tomsy Koupil říká: „V mluvnici projevil Tomsa kromě znalosti předchozí gramatikografické tradice i pozoruhodný vhled do běžně mluveného jazyka své přítomnosti, uvádí hláskově rozlišená paradigmata (např. typu *pěkný/pěknej*) včetně stylistického, resp. sociokomunikačního odlišení [...]“

Soustavnější výklady o tvoření slov se patrně poprvé objevují až v gramatice Jana Nejedlého z roku 1804, neboť zde se o tomto gramatickém jevu Koupil zmiňuje poprvé.

Pokud jde o jevy slovosledné, věnoval se jim zřejmě, soudě podle charakteristik Koupilových, poprvé Vincenc Pavel Žák v gramatice nazvané *Böhmische Sprachlehre für Deutsche*. Toto dílo se však prý „nesetkalo s příznivým ohlasem kruhů napojených na pražské centrum“ – Koupil cituje z jednoho dopisu: „Škoda, překoda, že to není gramatika českého spisovného jazyka, alebrž jakéhosi vy-myšleného moravsko-českého nářečí.“

Katalog českých gramatik Koupil uzavírá známým dílem Gebauerovým na-zvaným *Příruční mluvnice jazyka českého pro učitele a studium soukromé*. Jím se začínají dějiny gramatikografie 20. století.

Kniha O. Koupila prezentovaná především jako „katalog“ k výstavě může dobře posloužit jako „úvod do dějin české gramatikografie“ nejen diachronně orientovanému bohemistovi či slavistovi, ale i gramatikovi zaměřenému syn-chronně. Zkoumáme-li např. obrovskou variantnost v dnešní deklinaci substantiv, může být pro nás poučné, jak ten či onen jev z oblasti deklinace popisovali a hodnotili autoři gramatik minulých století. Více nám přitom o tehdejších spisovném úzu často napoví spíše to, jaké postoje tito vzdělanci zaujíмали k určitým jevům, zda je doporučovali, nebo odmítali, než sama jimi poskytovaná fakta.

Co je v ČNK nového VII

(Zprávy z Českého národního korpusu)

Michal Křen

Ústav Českého národního korpusu Filozofické fakulty UK, Praha

micHAL.kren@ff.cuni.cz

Korpus SYN2015

Jednou z nejdůležitějších událostí roku 2015 v ČNK bylo bezesporu zveřejnění nového stomilionového reprezentativního korpusu SYN2015. Tento korpus je pokračováním řady započaté korpusy SYN2000, SYN2005 a SYN2010, jejímž cílem je mapovat v pětiletých intervalech současnou psanou češtinu. Korpus SYN2015 však přináší celou řadu změn a vylepšení, z nichž ty nejpodstatnější shrnují následující body:

1. Pojetí psanosti bylo pro korpusy řady SYN zúženo pouze na tištěné a veřejně publikované texty; do korpusu SYN2015 proto nebyly zařazeny žádné internetové zdroje.

2. V korpusech řady SYN budou nadále uplatňována aktualizovaná kritéria pro určení hranice synchronie. Pro beletrii bude za synchronní považováno dílo vydané v posledních 25 letech, pokud doba od prvního vydání díla nepřesahuje 75 let. U odborných textů platí požadavek prvního vydání v posledních 25 letech, pro publicistiku musí být text vydán v období mapovaném daným korpusem (v případě SYN2015 se jedná o roky 2010–2014).

3. Základním principem výstavby korpusů se počínaje SYN2015 stává aktualizovaná klasifikace textů. Stejně jako klasifikace původní, používaná dosud v korpusech řady SYN, je založena na externích mimotextových kritériích, tj. především komunikační funkci textu. Při aktualizaci původní klasifikace textů jsme postupovali se zřetelem k zachování kontinuity, k úpravám jsme tedy přistoupili jenom tam, kde byla původní kategorizace problematická a změny nezbytné:

Publicistika byla na úrovni typu textu rozdělena na *tradiční* (typicky zpravodajské deníky) a *volnočasovou* (typicky časopisy pro volný čas); tradiční publicistika je dále rozdělena na *celostátní* a *regionální*, zatímco volnočasová je členěna podle převažujících témat.

V oborové literatuře (dříve odborná) byl vyčleněn nový textový typ *profesní literatura* charakterizovaný převahou praktických informací určených především profesionálům v oboru.

Žánry (*genre*) v oborové literatuře jsou převzaty z předmětové kategorizace používané v Národní knihovně, z níž vychází také klasifikace konkrétních textů.

Vznikla nová rovina *genre_group*, která je rozlišována pouze pro oborovou literaturu a která vznikla seskupením jednotlivých oborových disciplín do větších skupin (humanitní vědy, sociální vědy, přírodní vědy, formální a technické vědy).

Explicitní subklasifikace beletrie není dále zveřejňována, protože je značně interpretativní povahy; byla však rozšířena a interně zohledněna při výběru textů do korpusu SYN2015, což výrazně napomohlo dosažení pestrosti složení rozsáhlejších kategorií.

Bylo zavedeno podrobnější vnitřní členění významných publicistických titulů na úrovni jednotlivých článků; ty byly nově rozděleny do 13 tematických oblastí, což umožňuje jejich cílený výzkum.

4. SYN2015 je založen na nové koncepci reprezentativnosti, jejímž cílem je zachytit co nejširší spektrum různých typů veřejných psaných komunikátů. Vytváření korpusu proto sledovalo cíl rovnoměrného zastoupení hlavních typů textu na základě stanovené klasifikace tak, aby bylo dosaženo co největší pestrosti nejenom z hlediska typů textu a žánrů, ale – v rámci možností daných dostupností textů – také autora, překladatele, původního jazyka, nakladatelství apod. Přestože si tato koncepce nečiní nárok na vyváženost odpovídající proporcím, v jakých by texty měly být recipovány v celé populaci textů, korpus SYN2015 se svým složením od svých předchůdců nijak zásadně neliší; podstatnější změnou je podrobná klasifikace textů a její precizní aplikace.

5. Při přípravě korpusu SYN2015 byla použita kompletně přepracovaná sada nástrojů pro zpracování textů, která umožnila mj. zachování jejich bohatší struktury. Kromě přejmenování některých strukturních jednotek (<doc>, <text>) je nyní k dispozici také členění do odstavců (<p>); mimoto jsou v korpusu vedle hranic vět (<s>) zaznamenány také struktury <hi> (zvýraznění a řezy písma) a <lb> (označení hranice verše v poezii).

6. Pro zpracování SYN2015 byly použity – a často přímo pro tento korpus dále upraveny – nejnovější verze dostupných anotačních nástrojů (zejména pro morfologickou analýzu a desambiguaci), nově přibyl atribut *proc* informující o nástroji, který provedl konečnou desambiguaci.

7. Změnil se také způsob zpracování interpunkčních znamének: zatímco v předchozích korpusech řady SYN byly sjednoceny různé způsoby zápisu uvozovek, apostrofů, a byl také setřen rozdíl mezi spojovníkem a pomlčkou, v SYN2015 je již původní podoba těchto znaků v maximální možné míře zachována, a to v atributu *word*; v atributu *lemma* je pak uvedena sjednocená podoba, kterou je výhodné používat pro vyhledávání.

8. Vedle morfologické anotace se v SYN2015 poprvé objevuje automatická syn-taktická anotace. Anotace vychází z analytické roviny Pražského závislostního korpusu, která je svým pojetím blízká skladbě V. Šmilauera. Syntaktická anotace byla provedena stochastickým parserem a její chybovost je pochopitelně vyšší než u anotace morfologické, což je při jejím využívání potřeba brát v úvahu. Přesto věříme, že může v řadě situací práci s korpusem výrazně usnadnit a že jde o důležitý krok směrem k dalším úrovním jazykového popisu v ČNK.

Úplný a podrobný popis korpusu SYN2015 včetně výsledné podoby syntaktické anotace v rozhraní KonText je k dispozici na wiki ČNK na adrese <https://wiki.korpus.cz/doku.php/cnk:syn2015>

Korpus SYN verze 4

V době vzniku tohoto textu vrcholily práce na korpusu SYN verze 4, který je stejně jako jeho předchůdci sjednocením všech dosud vydaných korpusů této řady. Kromě nich však obsahuje také dosud nezveřejněnou publicistiku převážně z let 2010–2014 v ročních objemech přesahujících 200 mil. slov, takže celková velikost korpusu dosahuje přibližně 3,6 mld. slov. Korpus SYN verze 4 odpovídá korpusu SYN2015 jak způsobem zpracování a struktury textů, tak i jejich klasifikací,

kteřá byla zpětně aplikována na všechny v něm obsažené texty; výjimkou je pouze syntaktická anotace, která byla nahrazena pilotní verzí anotace frazémů.

Nové verze korpusu SYN budou dále vycházet pravidelně každý rok s přírůstkem v podobě aktuálních publicistických dat a budou referenční, tj. zpětně dostupné.