

Obsah / Content

Články / Articles

Lepší vrabec v hrsti nežli holub na střeše

Víceslovné lexikální jednotky v češtině: typologie a slovník

A Bird in the Hand is Worth Two in the Bush

Multiword expressions in Czech: typology and lexicon /3

*Milena Hnátková, Tomáš Jelínek, Marie Kopřivová, Vladimír Petkevič,
Alexandr Rosen, Hana Skoumalová, Pavel Vondříčka*

Slovní spojení *až tak* v současné mluvené češtině

Czech expression *až tak* in contemporary spoken Czech /23

Lucie Jílková

Kdo, kdy a proč skáče komu do řeči aneb překryvy ve spontánní mluvené češtině

Who, when and why is jumping into speech:

overlaps in spontaneous spoken Czech /41

Zuzana Komrsková, Petra Poukarová

Několik poznámek k synonymii a frekvenci příčinných spojek

Some notes on the synonymy and frequency of causal conjunctions /57

František Štícha

Recenze / Reviews

Korpuslinguistik

Rainer Perkuhn – Holger Keibel – Marc Kupietz: Korpuslinguistik. 2012 /66

Martin Šemelík

Strukturní rysy komunikace

Oldřich Uličný (ed.): Struktura v jazyce, jazyk v komunikaci. 2017 /70

Miloslav Vondráček

Zprávy / News

Co je v ČNK nového IX (Zprávy z Českého národního korpusu) /78

Hana Goláňová, Martina Waclawičová

Pokyny pro autory /83

Instructions for authors /84

Lepší vrabec v hrsti nežli holub na střeše Víceslovné lexikální jednotky v češtině: typologie a slovník

*Milena Hnátková, Tomáš Jelínek, Marie Kopřivová,
Vladimír Petkevič, Alexandr Rosen, Hana Skoumalová, Pavel Vondříčka*
Ústav teoretické a počítačové lingvistiky FFUK
*milena.hnatkova@ff.cuni.cz; tomas.jelinek@ff.cuni.cz; vladimir.petkevic@ff.cuni.cz;
alexandr.rosen@ff.cuni.cz; hana.skoumalova@ff.cuni.cz*
Ústav Českého národního korpusu FFUK
marie.koprivova@ff.cuni.cz; pavel.vondricka@ff.cuni.cz

A Bird in the Hand is Worth Two in the Bush. Multiword expressions in Czech: typology and lexicon

ABSTRACT: In this paper we present a multi-faceted typology of multiword expressions (MWE) in Czech, associated with a representative lexicon of MWEs (under development), whose entries reflect this typology. The typology and the lexicon are useful for many reasons, e.g. for lexicography, for teaching Czech as a foreign language, for dealing with important natural language processing tasks such as automatic morphological disambiguation and parsing, identification of MWEs in corpora of Czech, word sense disambiguation and semantic disambiguation. The typological classification has also a theoretical significance, since MWEs are interesting entities straddled between lexicon and grammar. In addition to various types of idiomaticity we record the variability of MWEs in morphology, syntax (including word order), and in the lexical setting of their components. We also take into account the usage of MWE (fragments) which still refer to the entire MWE.

KEYWORDS: multiword expressions, multiword lexicon, typology

KLÍČOVÁ SLOVA: víceslovné lexikální jednotky, slovník víceslovných jednotek, typologie

Slovní spojení *až tak* v současné mluvené češtině

Lucie Jílková

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

jilkova@ujc.cas.cz

Czech expression *až tak* in contemporary spoken Czech

ABSTRACT: The article answers the following research question: In what meanings/functions does the Czech expression *až tak* appear in spoken language? The subject of the analysis are the records of the Czech expression *až tak* in the corpora of the spoken Czech language DIALOG, ORAL and ORTOFON. The analysis reveals that the most prevalent usage of the expression is a measure expression in negative turns. One of the detected functions has not been reflected in linguistic works so far: the usage of the *až tak* expression as an independent dialogue turn. In these cases the expression enounces a) an extent of the speaker's degree of surprise/astonishment about what was said in the previous turn; b) an expression of surprise/astonishment itself, i.e. *až tak* in these cases bears the meaning of *Really? Are you kidding me? I can't believe it!* etc.

KEYWORDS: Czech expression *až tak*, DIALOG corpus, ORAL corpus, ORTOFON corpus, mass media dialogue, informal dialogue

KLÍČOVÁ SLOVA: slovní spojení *až tak*, korpus DIALOG, korpus ORAL, korpus ORTOFON, mediální dialog, neformální dialog

Kdo, kdy a proč skáče komu do řeči aneb překryvy ve spontánní mluvené češtině

Zuzana Komrsková, Petra Poukarová

*Ústav Českého národního korpusu, Filozofická fakulta UK Praha
zuzana.komrskova@ff.cuni.cz, petra.poukarova@ff.cuni.cz*

Who, when and why is jumping into speech: overlaps in spontaneous spoken Czech

ABSTRACT: Overlaps are very frequent in spontaneous communication. Intentional breaking in somebody speech can be motivated by a recipient's effort to take the floor and replace a speaker (competitive type) or to help the speaker when searching the right word for example (cooperative type). Unintentional overlaps can be found in the transition-relevant places most frequently. The specific kind of overlaps are continuers, mostly used by a recipient in order to demonstrate "I am still listening to you". In this paper, we ask the question what can influence the number and the types of overlaps in the speech where the same speaker speaks with two different people. The used data come from a corpus of spontaneous spoken Czech ORTOFON. We suppose that the number and types of overlaps can be influenced by the communication situation more likely and that the recipient breaks when understanding is threatened and, therefore, communication goal might not be achieved.

KEYWORDS: overlaps, spoken Czech, spoken language corpora

KLÍČOVÁ SLOVA: překryvy, mluvená čeština, korpusy mluveného jazyka

Několik poznámek k synonymii a frekvenci příčinných spojek

František Štícha

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

Some notes on the synonymy and frequency of causal conjunctions

ABSTRACT: In Czech, five synonymous causal conjunctions (*protože*, *neboť*, *jelikož*, *poněvadž*, *ježto*) – corresponding to the English *because* – are used. These five causal conjunctions differ in style and in frequency of use. Four of them (*neboť*, *jelikož*, *poněvadž*, *ježto*) are stylistically marked and only *protože* is stylistically unmarked and consequently used more frequently.

Until now, Czech grammarians have labeled *neboť* a coordinative conjunction and separated it from the other four ones in explaining causal relations in complex sentences. In fact, the five conjunctions have the same function to express a causal relation, except for *neboť* being positionally restricted to postposition. In this paper, I argue that whenever the conjunction *neboť* is used in a text it is synonymous with the other causal conjunctions. Therefore, this separation of *neboť* from *protože*, *jelikož*, *poněvadž*, *ježto* in grammar descriptions does not correspond to the function of these conjunctions. So far, *neboť* has been the second most frequent causal conjunction and together with *protože* used more frequently than the remaining three conjunctions. This article focuses on the idiolectal use of only one or two of the five causal conjunctions. Some Czech authors and translators into Czech mainly or solely use only one or two of all causal conjunctions and neglect the others. What is unpredictable and strange is that sometimes the less frequent conjunctions *poněvadž* and *jelikož* are strongly preferred.

KEYWORDS: causal conjunctions, synonyms, idiolectal use, style

KLÍČOVÁ SLOVA: příčinné spojky, synonymie, idiolekt, styl

Recenze / Reviews

Korpuslinguistik

**Rainer Perkuhn – Holger Keibel – Marc Kupietz: Korpuslinguistik
Paderborn: UTB, 2012, 144 S.**

Martin Šemelík

Institut für germanische Studien, Karls-Universität in Prag

martin.semelik@ff.cuni.cz

Ob von den einen „nur“ als Methodologie im Dienste der Forschung aufgefasst oder von den anderen als eigenständige linguistische Disziplin verstanden,¹ in beiden Fällen kann man zusammenfassen, dass die Korpuslinguistik seit mehreren Jahren als mehr oder weniger etabliert gilt. Pauschalisierende Einwände wie etwa „[a]ny natural corpus will be skewed. Some sentences won't occur because they are obvious, others because they are false, still others because they are impolite. The corpus, if natural, will be so wildly skewed, that the description would be no more than a mere list“ (Chomsky 1958, zit. nach Svartvik 2007: 15) sind in ihrem apodiktischen Ton zwar abzulehnen, zugleich muss man sich aber auch der Einschränkungen bewusst sein, die mit der Korpuslinguistik verbunden sind. So ist es beispielsweise in der lexikographischen Praxis gelungen, mit Hilfe der korpuslinguistischen Analysemethoden viele alte Probleme zu lösen (z.B. die Frage der Usualität der als lexikographische Beispiele verwendeten Wortverbindungen), andererseits kommen bei der Sichtung des immensen Korpusmaterials (*very large corpora*) neue Problemstellen zum Vorschein, die ältere Arbeitsmethoden (Introspektion etc.) nicht aufzudecken vermochten.²

Die Etablierung von korpuslinguistischen Analysemethoden in der Linguistik wie auch in den (Hochschul-)Curricula brachte auch unter den Lehrkräften wie Studierenden einen akuten Bedarf nach didaktisch durchdachten Einführungen in die Korpuslinguistik mit sich. Für den Bereich der Anglistik und Amerikanistik gibt es seit längerer Zeit eine Reihe von solchen Publikationen (vgl. McEnery – Wilson 1996, inter alia, von den neueren siehe z.B. Lindquist 2009), eine Tatsache, die wohl im Zusammenhang mit der Tradition des Firthschen Kontextualismus gesehen werden kann. Die erste Einführung für die Germanisten erschien jedoch erst viel später

¹ Auf die Frage, ob es sich im Falle der Korpuslinguistik um eine Methodologie oder eine Theorie handelt, kann hier aus Platzgründen nicht detaillierter eingegangen werden (dazu vgl. z.B. Gries 2010: 327ff., Mukherjee 2009: 30).

² Zur korpuslinguistischen Methodologie in der lexikographischen Praxis siehe z.B. Bergenholtz (2016).

(vgl. die verdienstvollen Publikationen von Lemnitzer – Zinsmeister 2006 und Scherer 2006), wobei die von Fandrych – Tschirner (2007: 195) formulierte Feststellung, dass „eine spezifische Einführung [in die Korpuslinguistik – M.Š.] für das Fach DaF [...] ein dringendes Desideratum [ist]“, bis heute gilt. Es fällt hierbei auch auf, dass es einen HSK-Band zur Korpuslinguistik erst seit 2008 gibt (Lüdeling – Kytö 2008/2009).

Umso erfreulicher ist es, dass den Germanistik-Studierenden (aber nicht nur ihnen) neben den oben genannten Einführungen seit 2012 eine weitere Publikation zur Verfügung steht. Der Band mit dem schlichten Titel *Korpuslinguistik* von Rainer Perkuhn, Holger Keibel und Marc Kupietz (alle Mitarbeiter bzw. ehemalige Mitarbeiter am IDS Mannheim), der in der Reihe LIBAC (*Linguistik für Bachelor* – herausgegeben von H. J. Heringer) des UTB-Verlags erschienen und Gegenstand der vorliegenden Rezension ist. Im Zusammenhang mit der LIBAC-Reihe ist festzuhalten, dass „[d]ie neuen Bachelor-Studiengänge [...] Lehrbücher [erfordern], die das exemplarische Lernen ermöglichen. Neben die theoretische Fundierung tritt in der Reihe [...] das Trainieren von Methoden. Mit den Bänden wird Studierenden speziell zu diesem Zweck entwickelte, didaktisch strukturierte Fachliteratur zu zentralen Themen der germanistischen Sprachwissenschaft zur Verfügung gestellt“ (http://www.utb-shop.de/utbseries/serie/index?utb_series=57, Zugriff am 19. 02. 2018).

Die in *Korpuslinguistik* vertretene Mannheimer Forschungsrichtung basiert auf einer engen interdisziplinären Verzahnung mit anderen Disziplinen, besonders mit der Psychologie und der Kognitionswissenschaft (S. 7). Dies ist positiv hervorzuheben, zumal Interdisziplinarität zu den wichtigsten Eigenschaften einer Methodologie zählt, die einen komplexen Einblick in ein facettenreiches Phänomen, wie etwa das der Sprache, zu gewähren sucht.

Primäres Anliegen der Publikation sei nach den Autoren weder eine genaue Definition des Begriffs „Korpuslinguistik“, noch ein erschöpfender Überblick über traditionelle Sichtweisen auf diesen Bereich (S. 9). Das Hauptziel bestehe darin, „ein Gefühl für eine eigene Sichtweise auf diesen Bereich zu vermitteln. Es wird weniger darum gehen, wie alte Fragen der Sprachwissenschaften beantwortet werden können, sondern eher darum, wie alte Fragen neu formuliert oder gar neue Fragen in den Mittelpunkt gerückt werden können“ (ebd.). Dieses Ziel wurde m.E. erreicht (vgl. auch weiter unten).

Das Buch umfasst insgesamt neun Kapitel, ein Literaturverzeichnis und ein Register. In Kap. 1 („Was ist Korpuslinguistik?“) geht es zuerst darum, sich Gedanken über die Sprache als Analysegegenstand zu machen. Thematisiert wird Sprache im Gebrauch bzw. in kommunikativer Funktion sowie Korpuslinguistik als Methodologie. Eingegangen wird auf Konzepte wie Introspektion, Kompetenz vs. Performanz, (nachgelagerte) Interpretation, Emergenz usw. Kap. 2 („Elementare Korpuseinheiten und einfache Recherche“) ist den Korpuseinheiten Text, Absatz, Satz etc. sowie praktischer Durchführung von Korpusrecherchen gewidmet (reguläre Ausdrücke u.Ä.). In Kap. 3 und 4 („Sprache sammeln“ und „Korpustexte anreichern“) wird auf Aufbereitung der Korpusdaten eingegangen (Repräsentativität, Designansätze, rechtliche Grundlagen, Typen von Korpusanreicherungen usw.), während in Kap. 5 und 6 („Was können Korpusbefunde aussagen?“ und „Korpusfrequenzen analysieren“)

an konkreten Beispielen allgemeine Grundsätze der Datenanalyse veranschaulicht werden (vgl. Ausführungen zu Häufigkeit, Dispersion etc.). In Kap. 7 („Assoziationsbeziehungen entdecken“) wird das Phänomen der statistischen Assoziation behandelt, mit deren Hilfe in Korpusdaten beispielsweise neue Lesarten bereits existierender Wörter bzw. Neulexeme identifiziert werden können. Kap. 8 („Sprachliche Strukturen aufspüren“) fokussiert die sprachliche Dimension „Kontext“ bzw. die am IDS Mannheim entwickelte Kookkurrenzanalyse. In Kap. 9 („Ähnliche Wörter – ähnlicher Gebrauch?“) wird u.a. illustriert, wie die Kookkurrenzanalyse zum Vergleich von Quasi-Synonymen verwendet werden kann.

Insgesamt liefert *Korpuslinguistik* viele Anregungen zum Nachdenken, von denen hier natürlich nur eine Auswahl punktuell genannt werden kann:

- Bei der Sprachanalyse kommt unserem subjektiven Gespür für Sprache eine bedeutende Rolle zu, aber nicht bei der Aufstellung von Anfangshypothesen, sondern erst in einem weiteren Schritt, und zwar bei der Interpretation der auf andere Weise als durch Intuition gewonnenen Befunde. Im Vordergrund steht hier also eine (nachgelagerte) Interpretation (vgl. S. 12), bei der versucht wird, „die subjektiven Analyseschritte so weit wie möglich nach hinten zu verschieben“ (S. 7).
- Damit hängt auch das Phänomen der Emergenz zusammen. Es handelt sich hierbei um die These, dass „das Systemhafte der Sprache aus ihr selbst heraus entgegentretritt“ (S. 13), d.h. dass die jeweiligen Gesetzmäßigkeiten des Sprachgebrauchs durch den menschlichen Geist nicht „erfunden“, sondern nur „gefunden“ werden müssen.
- Die am IDS Mannheim entwickelte Kookkurrenzanalyse bzw. die im Rahmen der Kookkurrenzdatenbank CCDB (<http://corpora.ids-mannheim.de/ccdb/>, Zugriff am 19. 02. 2018) verwendeten Visualisierungsmethoden (vgl. z. B. die Module SOM, CNS) ermöglichen Erschließung globaler Kontexte (thematischer Domänen), in denen ein Wort verwendet wird (Kriminalität, Verkehrsunfälle, Natur etc.), resp. Eruierung der Unterschiede zwischen sehr nahen Synonymen (*Gebüsch – Buschwerk*), und sind somit beispielsweise in der lexikographischen Praxis von besonderem Nutzen.³

Das handliche Studienbuch *Korpuslinguistik* ist sehr benutzerfreundlich, wozu die Schlagwörter am Seitenrand beitragen, die sich auf den Inhalt der einzelnen (Unter)Kapitel und Absätze beziehen. Auch die sprachliche Gestaltung spielt hierbei eine Rolle; der Stil ist knapp, doch sehr gut lesbar, zumal die Ausführungen durch eine ganze Reihe von Schemata, Abbildungen u.Ä. begleitet werden. Die Publikation ist nicht nur zum Selbststudium geeignet, sondern auch als Kursbuch, umso eher, als sie am Ende jedes Kapitels Kontrollfragen und verschiedenste Forschungsaufgaben anbietet. Besonders motivierend wirkt, dass an jedem Kapitelanfang, dem dann ein „Warm Up“ folgt, jeweils die Ziele des Kapitels formuliert werden. Auf diese Weise

³ Eine minutiöse Analyse der Synonyme mit Hilfe des CNS-Moduls findet sich in Marková (2012); zur praktischen Umsetzung dieser Analysemethoden in der lexikographischen Praxis siehe v.a. Vachková (2007) und Šemelfk (2017).

wird die Leserschaft ungezwungen in das jeweilige Thema eingeführt. Zum Buch gibt es außerdem ein Online-Begleitmaterial (vgl. unter <http://corpora.ids-mannheim.de/libac/>, Zugriff am 19. 02. 2018), das vertiefende Erklärungen zu den in *Korpuslinguistik* behandelten Themen, diverse interaktive Lehrangebote zur Erforschung der im Buch vorgestellten Konzepte, sowie Empfehlungen für korpuslinguistische Hausarbeiten bietet.

Das Layout ist übersichtlich. Im Hinblick auf den Preis gibt es keinen Grund zur Unzufriedenheit, einzig die Schriftgröße könnte in der nächsten Auflage wohl vergrößert werden, zumal mit *Korpuslinguistik* ein Lehr- bzw. Kursbuch vorliegt, an das in dieser Hinsicht höhere Ansprüche gestellt werden können.

Den Autoren verdanken wir zwar nicht die allererste deutschsprachige Einführung in die Korpuslinguistik, nach meiner Meinung jedoch die inspirativste und benutzerfreundlichste. Den Autoren sei dafür an dieser Stelle herzlich gedankt.

LITERATUR

- BERGENHOLTZ, H. (2016): Eine Korpusanalyse ist eine überflüssige Zeremonie. In: S. J. Schierholz et al. (eds), *Wörterbuchforschung und Lexikographie*. Berlin: Walter de Gruyter, 155–168.
- CHOMSKY, N. (1958): Ein Beitrag an der *Third Texas Conference on Problems of Linguistic Analysis in English*. Austin, Texas.
- FANDRYCH, CH. – TSCHIRNER, E. (2007): Korpuslinguistik und Deutsch als Fremdsprache. Ein Perspektivenwechsel. *Deutsch als Fremdsprache* 3/44, 195–204.
- GRIES, S. T. (2010): Corpus Linguistics and Theoretical Linguistics. A Love–Hate Relationship? Not Necessarily... *International Journal of Corpus Linguistics* 3/15, 327–343.
- LEMNITZER, L. – ZINSMEISTER, H. (2006): *Korpuslinguistik. Eine Einführung*. Tübingen: Narr.
- LINDQUIST, H. (2009): *Corpus Linguistics and the Description of English*. Edinburgh: University Press.
- LÜDELING, A. – KYTÖ, M. (2008, 2009): *Corpus Linguistics* (Handbücher zur Sprach- und Kommunikationswissenschaft, Bd. 29.1., 29.2. Berlin – New York: Walter de Gruyter.
- MARKOVÁ, V. (2012): *Synonyme unter dem Mikroskop. Eine korpuslinguistische Fallstudie*. Tübingen: Gunter Narr.
- McENERY, T. – WILSON, A. (1996): *Corpus Linguistics* (Edinburgh Textbooks in Empirical Linguistics). Edinburgh: University Press.
- MUKHERJEE, J. (2009): *Anglistische Korpuslinguistik. Eine Einführung* (Grundlagen der Anglistik und Amerikanistik 33). Berlin: Erich Schmidt.
- SCHERER, C. (2006): *Korpuslinguistik*. Heidelberg: Winter.
- SVARTVIK, J. (2007): Corpus Linguistics 25+ Years on. In: R. Facchinetti (ed), *Corpus Linguistics 25 Years on*. Amsterdam – New York: Rodopi, 11–24.
- ŠEMELÍK, M. (2017): Zu neuen Möglichkeiten der lexikographischen Erfassung von Wortbildungskonkurrenzen. Ge- vs. -werk korpuslinguistisch betrachtet. *International Journal of Lexicography* 3/30, 350–381.
- VACHKOVÁ, M. (2007): Adjektive auf *-bar* in kontrastiver und korpuslinguistischer Sicht. Eine metalexikographische Betrachtung. *Linguistica Pragensia* 2/17, 57–75 Verfügbar unter http://dlib.lib.cas.cz/3139/1/1.Vachkova_s.57-74.pdf (19. 02. 2018)

Strukturní rysy komunikace

**Oldřich Uličný (ed.): Struktura v jazyce, jazyk v komunikaci
Liberec: Technická univerzita v Liberci, 2017**

Miloslav Vondráček

*Ústav pro jazyk český AV ČR, v. v. i., Praha
Pedagogická fakulta Univerzity Hradec Králové
vondracek@ujc.cas.cz, miloslav.vondracek@uhk.cz*

Editor sestavil vitráž z lingvistických studií zaměřených synchronně i diachronně, na jazyk i řeč, na češtinu spisovnou, běžně mluvenou i nářeční. Spojovacím prvkem, zřejmým už z názvu, je funkčně strukturní přístup k jazyku a komunikaci, odkazující k Pražskému lingvistickému kroužku, resp. Pražské škole. V metodologickém úvodu O. Uličný zdůrazňuje také nezanedbatelný vliv slovenské lingvistiky a rozšiřuje inspirační pole na československou lingvistiku vyvíjející se jako synergický celek – což autor recenze díky generační zkušenosti a současným kontaktům potvrzuje a vítá. Prostoru se dostává i přístupům kritickým či konkurenčním, které zvolenou centrální perspektivu nijak nezpochybňují. Kromě metodologie spojuje dílčí lingvistické exkurzy i materiálové východisko – korpusy současné psané i mluvené češtiny.

Termíny zvolené pro název publikace volně odkazují ještě hlouběji, ke strukturalismu a de Saussurově opozici langue : parole / jazyk : řeč. Tak čtenář dostává možnost ověřovat si ve dvanácti různorodých, ale vzájemně souvisejících a provázaných kapitolách, jak se systém a struktura jazyka promítají do řeči, nechávající v ní jako v autentickém komunikačním prostředku svědectví o své existenci; povaha systému a struktury se tak zvláště v diachronně pojatých studiích ukazuje jako proměnlivě dynamická, s variabilitou vázanou na hypotetické jazykové roviny a další faktory sociálního, pragmatického či kognitivního rázu.

V hutné precizní vstupní kapitole (I/1) nabízí **Ondřej Bláha** vzhled do sféry **syntaktických funkcí pádových forem**, o níž máme jako lingvisté své představy – jež se tváří v tvář korpusovému materiálu ukazují jako správné v globálním směřování, ne však přesné v odhadu masivnosti vývojového proudu; přesněji mohou se jevit jako předdimenzované vinou vstřícnosti ke generalizacím dílčích postřehů. Srovnáním rozsáhlého korpusového vzorku publicistického jazyka z roku 1915 a 2015, čítajícím v součtu kolem 30 tisíc slovních tvarů, autor dospívá ke zjištění, že k proměně syntaktických rolí pádů dochází, ale v míře menší, než jakou bychom čekali. Hovoří o pozvolné syntaktizaci pádu, pozbývajících některé sémantické rysy, jež bezprostředně neslouží k vyjadřování vztahů mezi větnými komponenty. Výraznější změny zaznamenává u genitivu posilujícího v roli neshodného atributu, dokumentuje spíše okrajové zvýrazňování objektové funkce akuzativu na úkor pádů sémantičtějších,

omezený ústup lokálu i v adverbialní platnosti a nepatrné prosazování nominativu v neprospěch instrumentálu v predikativní funkci, přičemž všechny změny se pohybují v řádu desetin až nízkých jednotek procent.

Kromě těchto základních vývojových trendů má podle O. Bláhy nominativ tendenci stát se výhradním subjektivním pádem (autor recenze tento vývoj ilustruje i teoretickými argumenty), konkurenci mu netvoří ani predikativní instrumentál, ve většině konstrukcí nahraditelný univerzálnějším nominativem. Genitiv i díky četným sekundárním předložkám opouští větněčlenská pozice objektu, stává se pádem nominální skupiny – pádem přívlastkovým. U dativu dominuje funkce druhého objektu na úkor funkcí vedlejších. Tendence akuzativu být prvořadě pádem objektu zvláště vyniká ve srovnání s dalšími slovanskými jazyky; na poli češtiny se navíc akuzativ jednoznačně prosazuje ve valenčním poli slovesných neologismů s obligatorní objektovou vazbou. Uplatnění lokálu v roli objektu mírně ustupuje, dominuje jeho platnost adverbialní (byť i ta slábne vlivem adverbí) s postupnou další diferenciací funkcí. O instrumentálu ve verbo-nominálním predikátu již zmínka padla – ani přes jistou módnost se nestává v této roli primárním pádem.

O. Bláha potvrzuje směřování k přirozenému souladu pádových významů a funkcí pádu, ačkoli je vývoj systému jako celku subtilnější, než by bylo možno očekávat pod vlivem nápadných periferních jevů. Předpokládá, že v méně konzervativní mluvené komunikaci může být zvýraznění syntaktické povahy pádu silnější. Frekvenční zastoupení pádů ve staré češtině mohlo být rovnoměrnější dík tomu, že pády plnily rozmanité sémanticko-syntaktické funkce bez těsné vazby k rolím větněčlenským.

Na **instrumentál předložkový**, přesněji na funkci předložky *s* v instrumentálových vazbách a na dynamiku těchto konstrukcí soustředila pozornost **Hana Prokšová** (1/2). Téma nahlíží v kontextu evropského jazykového areálu se zvláštním zřetelem k jazykům západoslovanským, pro něž je/bylo dosud příznačné formální odlišování instrumentální a komitativní platnosti instrumentálu (na komitativní roli váže autorka i typologii zmíněných jazyků). Ve sledovaném vývoji shledává projev unifikáčnické tendence patrné do jisté míry i v okolních jazycích. Autorka vhodně upozorňuje na skutečnost, že větněčlensky je předložka vnímána jako součást pádové formy, zatímco v závislostní syntaxi se jeví jako dominanta nominální skupiny a může být vnímána jako jednotka samostatná; hovoří proto o jisté významotvorné autonomii prepozic projevujících se v dynamice předložkových a bezpředložkových pádů obecně.

H. Prokšová připomíná předložkově instrumentálově hypotaktické vyjádření koordinace (*otec s matkou*) volně přecházející v platnost atributivní (*koláč s povidly*), zmiňuje specifickou analytickou sekundární předložku *spolu s* a obdobné deverbální skupiny *v souladu s*, *v souvislosti s*, udržující si obligatorní pádovou vazbu původních sloves, především se ale zaměřuje na povahu instrumentálu při predikátorových slovesech. Neustálenost předložkové vazby ukazuje na slovesech typu *šetřit (s)*, *chlubit se (s²)*, *hýbat (s)* s různě hodnocenými odlišnostmi významovými a stylovými. Uvádí příklady vazeb tradičně bezpředložkových, z nichž je ovšem patrné, že ani ty nebudou – se stejnými dopady – předložkově kompletace ušetřeny (pozn. MV).

Mezi pragmatickými motivy zdůrazňuje autorka především výhodnost redukováného souboru výrazových prostředků pádu, projevující se synergicky zejména ve spojení se zájmeny, resp. jejich *n*-formami (*nalepit s ní, práť s ní* – ‚páskou‘, ‚pračkou‘). Podstatný je i argument kognitivní – posilování konceptu sounáležitosti na úkor nástrojovosti objektu (*jet s autem*).

Svatava Škodová se dlouhodobě zabývá mj. slovesy pohybu v nezákladních sémanticko-syntaktických rolích. Tak je tomu i ve studii soustředěné na **sloveso *jít* tvořící složku analytického predikátu** (I/3). Sleduje sémantické vyprazdňování slovesa – ústup významu translokačního či obecně kynetického/funkčního ve prospěch vyjádření iniciální fáze progresivního děje označeného infinitivem slovesa (*jdeme spát*), nebo možnosti nastání děje, tj. funkce modální (*nejde odolat*). Hovoří-li autorka o gramatikalizaci slovesa *jít* ve spojení s infinitivem jiného slovesa, odkazuje tím implicitně ke stavu, který dokumentuje dále A. M. Černá obligatorním spojením supina jako historického neurčitého tvaru plnovýznamového slovesa označujícího cílový děj/stav se slovesem pohybu, zpravidla *jít* (II/5, s. 158) – pozn. MV.

Autorka konfrontuje svá zjištění s dřívějšími příklady L. Saicové Římalové, přičemž u konstrukcí typu *jdeme koupit* připouští dvojí možnou interpretaci ve smyslu ‚chůzí se přemístit za účelem nákupu‘ nebo ‚v blízkém okamžiku koupit‘. Výtečně postihuje i sémantickou mezihodnotu metaforického pohybu vyjádřenou konstrukcemi *jít dělat jako dělník, jít luštit název*. Upozorňuje tak na neprototypické užívání slovesa *jít* a na proces ustalování jeho gramatických funkcí, získávaných v analytickém predikátu.

Vojtěch Veselý s Barborou Štěpánkovou předkládají zevrubnou strukturalistickou (lexikálně-syntaktickou) **analýzu výrazu *jediný*** (I/5). Na tomto lexému demonstrovují zásadní kvalitativní změnu ve zpracování nového lexikografického díla, oživující tak – více či méně explicitně – i otázku relevance kritérií slovnědruhové klasifikace a pojetí větých členů. Současně ověřují, že některé z významů, stereotypně po-/přepisovaných dosavadními slovníky, jsou již neživé.

Lexém *jediný*, hodnocený tradičně jako adjektivum, řadí vzhledem k základní složce jeho sémantiky – významu čísla 1 – k číslovkám, pro invariantní hodnotu ‚jen jeden‘ k číslovkám solitérním (*má jediné dítě*). Konstatují přitom, že vztah výrazu k číslu 1 může být zvláštní, pro číslovky netypický. Autor recenze s tímto přístupem sympatizuje, byť s numerálií jako slovním druhem nepočítá a výraz by hodnotil jako (solitérně) numerální adjektivum. Ve skupině dalších druhů slov však považuje sémantickou specifikaci uplatněnou V. Veselým a B. Štěpánkovou, jež dá vyniknout dalším významům variantním, za pregnantní.

Na pozadí stávajících popisů dokumentují autoři dosud víceméně nereflexitovanou aktualizace částicovou lexii s významem ‚jen‘, ‚pouze‘ (*jediná já jsem nebyla chráněná*) v těsném postavení před restrigovaným výrazem, popř. v postpozici (*cara chrání on jediný*). Autor recenze dodává, že pro částicovou platnost může hovořit omezeně přípustná alternace s formálním adverbium (*jedině já jsem nebyla chráněná*), vylučujícím změnu slovosledu, ev. užití komponentu *jako* (**já jediné jsem nebyla*

*chráněná/*já jako jediné jsem nebyla...).* S výrazem *jediný* porovnávají autoři význam vytykáací částice *samotný, sám*. Výraz *jediný* řadí ve srovnatelných kontextech k částicím omezovacím zdůrazňovacím, zatímco výrazy *sám, samotný* k částicím zahrnovacím s příznakem překvapivosti.

V odlučeném postavení, jestliže slovosledně náleží k přísudku, má výraz *jediný* syntaktickou platnost doplňku určovacího, jakkoli odlišného od doplňků prototypických (*bratr se jako jediné nestyděl*).

V záporných větách nabývá výraz významu ‚žádný‘, podle autorů tedy i platnosti zájmena záporného, s nímž je částečně synonymní – v závislosti na počitatelnosti denotátu řídicího substantiva (*nesvítilo jediné/žádné okno* vs. *neměl žádný/*jediný vliv*). Autor recenze by jej hodnotil relativně shodně – jen s jiným akcentem – jako pronominální numerální adjektivum.

Cenná část kapitoly se zabývá podmínkami jedno-/víceznačnosti vět s výrazem *jediný* (*povodni odolal jediné tábor; jediné dva tábory odolaly*). V dalším oddílu autoři sledují limity výše naznačené synonymie (*neměl žádný/jediný/ani jeden důkaz : žádný vliv*).

Studie **Ludmily Veselovské** se týká podobného tématu – výrazu ***málo*** v **kvantifikovaných jmenných frázích** (1/4), označeného zde proklamativně za číslovku – přičemž autorka současně zdůrazňuje jeho obecnou kategoriální ambiguitu. Autorka se snaží nástroji generativní morfo-syntaktické analýzy odhalit/eliminovat úskalí tradiční slovnědruhové taxonomie, jež podle jejího soudu není schopna jednoduchým a zobecnitelným způsobem vystihnout specifika podobných modelů. Vychází z obecné (strukturalistické) definice číslovek, podle možnosti i z deskripce chování konkrétního výrazu, který je spárován a sledován paralelně s výrazem *mnoho*.

Autorka pozoruje chování obou výrazů ve strukturních (přímých) pádech a v pádech nepřímých (lexikálních), jakož i mimo nominální skupinu – v platnosti adverbia. Morfém *mál-/mal-* ve funkci lexikálního prefixu eviduje ve složených zájmenech a kompozitech typu *maloměsto, malorážka, malomyslný*, kde mu připisuje plnou gramatikalizaci, navrhuje sledovat rozdíly mezi lexikální a gramatikalizovanou variantou.

Kategorie číslovek je podle L. Veselovské definována nikoli jako jednotka, ale jako soubor charakteristik často na sobě nezávislých, jejichž spojování je pro hlubší jazykovou analýzu nepřínosné. Lexém *málo* vytváří dvě paradigma: stupňovatelností ukazuje k adjektivům/adverbiím, dílem je schopno i substantivní flexe; syntaktické argumenty však adekvátní hodnocení dále ztěžují. Autorka považuje za metodologicky nežádoucí klasifikovat touž slovní formu jako substantivum ve strukturním pádu, jako adjektivum/adverbium v pádu lexikálním; poukazuje k původně (hloubkově) adjektivní povaze lexémů *málo/mnoho*.

Na korpusových datech autorka ověřuje, do jaké míry naplňuje číselně užitý výraz *málo* zmíněná paradigma – a vedle nich třetí, tzv. číslovkové. Dochází k závěru, že směrodatné charakteristiky adjektivní/adverbiální a substantivní, zmiňované normativními gramatikami, splňuje jako příznakové substantivum (singulárové neutrum), přičemž se v lexikálních pádech vyhýbá kombinacím s nepočítatelnými substancemi,

a v kombinaci s jiným číselným výrazem jako jeho ustrnulý pandán s podpůrnou funkcí (*několik málo*). Jako číslovka základní s heterogenním paradigmatickým je nedoložen; paradigmatická nedostatečnost je kompenzována dalším lexémem. Autorka soudí, že disociace slovnědruhových charakteristik (pád, shoda, distribuce) nesvědčí o dostatečnosti tradičních zhruba deseti slovních druhů jako univerzální taxonomie jednotek jazykového systému. Jinými slovy – v tradiční slovnědruhové klasifikaci nalézá zobecnění shluku samostatně existujících nezávislých charakteristik, jež je třeba samostatně zkoumat a vysvětlovat jejich typologicky podmíněné kombinace.

K saussurovskému systému/langue směřuje i kapitola **Štěpána Šimka**, zaměřená na **rekonstrukci formální a významové stránky lexémů přiekora, přiekořizna, přiekořizna (I/6)** z „parolových“ podkladů pro tvorbu Staročeského slovníku. Tyto jednotky hodnotí jako synonymní, přičemž úsilím o slovtvornou analýzu (přecházející pochopitelně v rozklad etymologický) naznačuje, že se jedná o slovtvorné varianty. Odkrývá předčeský původ těchto dějových jmen s významem ‚příkoří, pohana, potupa, urážka‘, resp. ‚křivda, nespravedlnost‘.

V editorově intenci byl i poukaz k projevům strukturních rysů jazyka v komunikaci. Kapitoly věnované řeči/parole začínají hodnotnými návraty k pramenům pojmu jazyková kultura a postupnému formování axiologických kritérií, mezi nimiž se systematická prostředků objevuje vedle noremnosti a adekvátnosti s nástupem strukturalistické lingvistiky, tj. Pražské školy. První tři oddíly se více či méně dotýkají i kodifikace spisovného jazyka.

Martin Prošek se zkušenostním základem daným dlouholetou poradenskou praxí rekonstruuje a kriticky hodnotí **současný stav jazykové kultury (II/1)** včetně nezbytného posouzení jejích cílů, metod, nástrojů, pojmoslovného aparátu. Rekapituluje odbornou diskusi vyvolanou v minulosti F. Čermákem, P. Sgallem a P. Vybíralem, následovanou později alternativním konceptem minimální intervence V. Cvrčka. V souvislosti s jistým rozporem mezi očekávanými a reálnými dopady kodifikace hovoří o stigmatizaci oboru ústící v úpadek zájmu o obor mezi mladými lingvisty, resp. o potřebě zbavit jazykovou kulturu ideologického nánosu, demotivujících přístupů označujících jazykově kultivační činnost za nežádoucí a nelegitimní.

Autor se vrací k cílům klasické teorie jazykové kultury, spocívajícím v pěstění spisovného jazyka. Nezbytně se tak dostává ke kritériím spisovnosti, případně k alternativnímu pojetí prestižní variety jazyka (standardní češtině) a opodstatnění útvorového členění vůbec. Oprávněně dospívá k dílčímu skeptickému závěru, že i prestižní útvar ustavený výhradně na základě empirických dat samotnými lingvisty by byl neproblematický nanejvýš pro ně. Autor recenze dodává, že s dopady většinové volby různého typu má tato společnost zkušenost; sféra jazyka jich i v proměnlivých společenských podmínkách zůstala ušetřena patrně proto, že při obecné znalosti původního významu slova *kultura* dokážeme domyslet důsledky rezignace na každodenní péči o prostředí, v němž žijeme, včetně klimatu jazykového.

V souvislosti s možným přeceňováním reálné schopnosti kodifikace ovlivnit stav jazyka zvažuje M. Prošek efekt tištěných a elektronických zdrojů poučení o jazyce,

mezi něž nepochybně patří Internetová jazyková příručka. Na zřetel bere i další nástroje zpracování přirozeného jazyka – morfologické analyzátoři, korpusy ap., mezi něž lze řadit nejspíš i korekturní a slovníkové služby textových editorů (pozn. MV). Mezi místy možného poučení eviduje i svépomocně vznikající diskusní fóra, tvořená přirozeně decentralizovanými skupinami mluvčích projevujících zájem o jazyk. Je zřejmé, že ze strany nelingvistické veřejnosti je poptávka po jazykovém poradenství vysoká a jazyková kultura je považována za závažné společenské téma.

Za zmínku stojí i autorův postřeh, že mimopražští lingvisté neodmítají jazykovou kulturu jako disciplínu, ani její metody a způsoby uplatňování, ba co víc, aktivně naplňují tradiční jazykovědné postuláty Pražské školy, uplatňující i hodnotové postoje k jazyku a jeho kultivaci. Právě v těchto intuicích a axiologických přístupech vidí lecky jedinou argumentační bázi nezbytnou pro posouzení empirických dat.

Kritériím jazykové správnosti se věnuje i **Martin Beneš** (II/2). Jako východisko volí konceptualizaci jazyka z perspektivy tzv. ontologického socialismu. Z ní, resp. z pravidla tzv. sdílených znalostí odvozuje kritérium jazykové správnosti jako míru shody s touto regulí (lze připomenout podobnou teorii oprávnění a závazků, pozn. MV). Na tomto teoretickém pozadí představuje další, tradičně užívaná, v čase proměnlivá kritéria jazykové správnosti jako dříve zohledňovanou historičnost, doloženost v jazyce lidovém nebo jazyce dobrého autora, estetičnost, logičnost, účelnost, srozumitelnost, významovost, nezbytnost, ev. soulad s jazykovým citem; nověji pak systémovost, adekvátnost funkcím, noremnost, popř. kritérium přijatelnosti z hlediska regionálního chápání spisovnosti, resp. územní akceptovatelnosti jazykového prostředí (O. Uličný) a kritérium většinové volby mluvčími (V. Cvrček).

Některá ze zmíněných kritérií mohou být prizmatem zvoleného teoretického základu shledána odborníky, zvláště tehdy, jestliže se jazykový prostředek navzdory neshodě s nimi, avšak ve shodě se sdíleným pravidlem dostane do užívání a je mluvčími považován za správný. Autor skepticky konstatuje, že odbornická kritéria jsou interakčním partnerům většinou zcela cizí. Situaci ilustruje poznámkou, že případné problémy autentických uživatelů v autentických situacích nejsou řešitelné spekulací lingvistů, zda konkrétní prostředek vyhovuje jejich kritériím. Současně hovoří o tom, že otázkou není, zda jazykový prostředek (ne)může narušit strukturu jazyka, ale zda (ne)ruší jeho uživatele.

Tradiční kritéria jazykové správnosti nemají stejnou váhu, se zřetelem ke komunikační praxi jsou hierarchizována. I přesto nejspíš mohou jen podporovat či tlumit výše zmíněná sdílená pravidla.

Žádoucí detail nášší do komplexní tematiky jazykové kultury exkurz na periferii jazykového systému – do oblasti **proprií a jejich kodifikace** (II/3) – zprostředkovaný **Kamilou Smejkalovou**. Téma je přitažlivé už tím, že propriální sféra se od apelativní v mnohém liší, především v rovině morfologické, v důsledku pak i zvláštnostmi kodifikačního procesu, který není jen záznamem jazykové normy; ta se vzhledem k nízké frekvenci jazykových jevů ani nemůže vytvořit. Z praktických potřeb vzniká proto apriorní kodifikace založená na kritériu systémovosti a vývojové progresivnosti.

Autorka nabízí vhled do problémů spojených s tvaroslovnou a slovtvornou stránkou vlastních jmen, posléze s pravopisem velkých písmen (tedy zejm. s antroponymy, toponymy a chrématonymy, jejich skloňováním, tvorbou slovtvorných derivátů, v případě antroponymy zejm. s přechylováním). Právě pro výlučnost propriální oblasti mohou být kodifikační přístupy podnětem i pro přístup k jevům centrálním.

Typem jazykového materiálu – antroponymy a toponymy – na K. Smejkalovou částečně navazuje **Alena M. Černá** kapitolou o **češtině nejstarší knihy tábořské** (II/5). Jde o cenný diachronní doklad dynamiky jazykového systému odrážející se v konkrétních parolových realizacích – autentických datovaných zápisech z první poloviny 15. století, pořázených v průběhu zhruba dvaceti let devíti písaři.

Autorka dokumentuje vývoj jazykových rovin v jejich vzájemné provázanosti, tj. roviny hláskové, tvaroslovné, syntaktické, zvláštní pozornost věnuje rovině lexikální (se složkou apelativní a propriální), především ustavující se právní terminologii. Stejně závažné jsou ovšem poznatky o konstituování normy úřední komunikace, základů pozdějšího administrativního stylu.

Jiný typ faktorů ovlivňujících vývoj jazykového systému při stejné šíři jazykových jevů reflektuje **Sergej Skorvid** v kapitole zaměřené na **současný stav** češtiny, resp. **českých přistěhovaleckých nářečí** (II/6), udržujících se v malých enklávách Krasnodarského kraje a Omské oblasti Ruské federace.

Tyto dialekty jsou důsledkem migrace rodin českých rolníků od konce 60. let 19. století na černomořské pobřeží severního Kavkazu a druhotně pak počátkem 20. století z ukrajinského Čechohradu na Sibiř. Autor zaznamenává podstatné znaky fonetické, morfologické a lexikální, z nichž vyvozuje jihozápadočeský a severovýchodočeský původ dialektů. Ty souběžně odrážejí dlouhodobý vliv ruštiny, zčásti i ukrajinštiny; bez výraznějších jazykových dopadů na česká nářečí zůstala koexistence s polskými mluvcími ve smíšených obcích. Mimo dialektologicky relevantní informace o vývoji jazykového systému pod tlakem systémů jinojazyčných a jinonářečních přináší text i metainformace důležité z hlediska sociolingvistického.

Není lepšího důkazu vzájemné provázanosti langue a parole než poukaz ke **koherenci mluvených textů** (II/4), který předkládá s využitím korpusu mluvené češtiny **Petra Poukarová**; neboť komunikát, má-li být jako komunikát vnímán, má systémovou oporu v textu a textotvorných (strukturativních) principech. Téma nebylo zvoleno náhodně; vychází ze zkušenosti, že mluveným komunikátům a spontánním nepřipraveným dialogům zvláště je často přisuzována nižší/nízká míra koherence, popř. i koheze, aniž by byly dále zkoumány prostředky, jakými se jí – odlišně, nikoli však nepřítomně – dosahuje.

Autorka dokumentuje prostředky soudržnosti komunikátu na situacích, kdy mluvčí sám utváří svou promluvu, mluvčí je v promluvě přerušen komunikačním partnerem a následně v komunikaci pokračuje, nebo je formování komunikátu věcí všech účastníků hovoru.

Mezi základní prostředky textovosti (v míře, v jaké spolu slova text a textil, textura etymologicky souvisejí – pozn. MV) patří tematické posloupnosti, na lexikální a gra-

matické úrovni paradigmatické vztahy ve slovní zásobě, pronominální a deiktické prostředky v širokém smyslu včetně časoprostorových orientátorů, spojovací prostředky, opakování, valence, kongruence, tedy i dokončování syntaktických struktur, intonace, uplatnění částic apod.

Publikace je dobrým obrazem současné lingvistické bohemistiky, která je kompetentní pojednat témata procházející napříč hypotetickými rovinami jazyka, napříč hypotetickými útvary jazyka, napříč časovými i geografickými souřadnicemi, reflektovat konstrukty typu jazyk i řeč, aniž by ztrácela jistotu v předmětu bádání, smyslu tohoto bádání a jeho cílech, nacházejíc oporu ve struktuře poznávaného i poznávajícího. K tomu má dostatek nadhledu a síly, aby byla schopna nerezignovat na hodnotová kritéria, dobírat se jich, formulovat je a uplatnit, vyjádřit osobnostní postoje, protože na nich stojí nejen jazyková kultura, jak o ní hovoří M. Prošek, ale kultura obecně.

Co je v ČNK nového IX (Zprávy z Českého národního korpusu)

Hana Goláňová, Martina Waclawičová

*Ústav Českého národního korpusu Filozofické fakulty UK, Praha
golanova@ujc.cas.cz; martina.waclawicova@ff.cuni.cz*

V rámci projektu **Český národní korpus** byla na začátku června 2017 zveřejněna trojice korpusů mluvené češtiny,¹ a to dva obecné, **ORTOFON** a sjednocení korpusů řady **ORAL**, a jeden specializovaný, nářeční korpus **DIALEKT**. Podrobné pojednání o korpusu ORAL bylo v KGA zveřejněno,² v této zprávě se tedy zaměříme na zbývající dva nové korpusy – ORTOFON a DIALEKT.

Korpus **ORTOFON**³ navazuje svou koncepcí sběru dat na korpusy neformální mluvené češtiny řady ORAL. Stejně jako ony se i ORTOFON zaměřuje na spontánní mluvený jazyk užívaný v neformálních situacích mezi mluvčími, kteří se znají. Na rozdíl od korpusů řady ORAL však ORTOFON disponuje víceúrovňovým prepisem. Prezентuje jazyková data na rovině ortografické a fonetické, přičemž je prepis propojen se zvukovou stopou. Korpus je lemmatizován a morfologicky značkován stejně jako korpusy ORAL a DIALEKT.

Korpus ORTOFON zpřístupňuje nejnovější nahrávky současného mluveného jazyka. Skládá se z 332 nahrávek z let 2012–2017 a obsahuje 1 014 786 textových slov, tj. 1 236 508 pozic. V sondách je zachyceno celkem 624 různých mluvčích. Nahrávky byly pořizovány v Čechách, na Moravě i ve Slezsku, tedy na území celé České republiky, a jejich celková délka je téměř 103 hodin. Pokrytí jednotlivých oblastí je možné zjistit i na mapě sběru.⁴

Sběr materiálu pro korpus byl prováděn v souladu s kritérii platnými pro korpusy řady ORAL. Byl zaměřen výhradně na prototypický spontánní mluvený jazyk, který se používá při bezprostřední interakci mluvčích v neformálních komunikačních situacích. Hlavními kritérii pro získávání nahrávek byly: synchronní komunikace mluvčích, dialogičnost promluv, vzájemný blízký vztah mluvčích, nepřipravenost,

¹ V souvislosti se zveřejněním korpusů byly přidány i nové funkce do korpusového rozhraní KonText.

² Kopřivová et al. 2017.

³ Podrobné informace o korpusu jsou zveřejněny na <http://wiki.korpus.cz/doku.php/cnk:ortofon>.

⁴ https://wiki.korpus.cz/doku.php/cnk:ortofon#slozeni_korpusu_a_sber_dat

spontánnost, neveřejná a neoficiální komunikační situace. Vzhledem k fonetické vrstvě přepisu byl kladen vyšší důraz na kvalitu nahrávek. Maximální možné míry autenticity jednotlivých nahrávek bylo dosaženo tak, že mluvčí většinou nebyli o nahrávání informováni předem, ale až po jeho skončení. Všichni nahraní mluvčí souhlasili s použitím nahrávek pro účely ČNK. V tomto sběru se bude i nadále v ÚČNK pokračovat.

Korpus ORTOFON je první korpus mezi korpusy mluvené češtiny ČNK, který je plně **vyvážený** v rámci všech základních **sociolingvistických kategorií mluvčích** (pohlaví, věk, nejvyšší dosažené vzdělání a nářeční oblast, v níž mluvčí strávil většinu života do svých 15 let). Již při sběru dat bylo dbáno na maximální možnou variabilitu mluvčích v rámci všech nářečních oblastí a materiál byl průběžně vyvažován. První tři kategorie, tj. pohlaví, věk, vzdělání, byly rozděleny binárně, čtvrtá kategorie byla rozvržena do deseti skupin, tj. deseti nářečních oblastí. Toto rozdělení navazuje na koncepci korpusů řady ORAL. S ohledem na cílenou velikost korpusu (1 000 000 slov) byla pro každou kategorii prezentovanou kombinací čtyř proměnných – pohlaví (2) × věk (2) × vzdělání (2) × nářeční oblast pobytu do 15 let (10) – stanovena cílová hranice 12 500 slov. Ve snaze o co největší pestrost mluvčích v rámci každé kategorie byl stanoven minimální limit pět různých mluvčích. Cílem tohoto opatření bylo omezit vliv idiolektu. Sociolingvistické údaje o mluvčích byly rozšířeny, např. byla doplněna informace o genderovém a generačním složení mluvčích v nahrávce a o velikosti sídla, odkud mluvčí pochází. Podle těchto údajů lze vytvářet subkorpusy.

Pro **přepis** sond určených pro korpus ORTOFON (stejně jako pro korpus DIALEKT) byl použit transkripční program ELAN, jenž umožňuje propojení víceúrovňového přepisu se zvukem. Přepis byl realizován na dvou rovinách: **ortografické** a **fonetické**. Zápis na **ortografické** rovině slouží primárně k porozumění nahranému rozhovoru a orientaci v něm, a snaží se proto co nejvíce přiblížit tradičnímu zápisu slov podle pravidel českého pravopisu. Na rozdíl od korpusů řady ORAL se zde tedy nezachycuje zjednodušená či redukováná výslovnost, pro její zaznamenání slouží fonetická rovina. V korpusu ORTOFON je použita **pauzová interpunkce**. Rozlišují se tři druhy pauz: velmi krátká pauza nebo spíše intonační předěl (méně než 120 ms), slyšitelné ticho nebo neverbální zvuk (např. smích, odkašlání) (120 ms – 2 s) a odmlčení (více než 2 s). **Fonetická** rovina zachycuje skutečnou realizaci vyřčeného pomocí upraveného fonetického přepisu. V korpusovém rozhraní KonText si ji lze zobrazit jako poziční atribut. Zápis využívá běžnou českou abecedu s několika speciálními znaky (např. ŋ, ŋ̥, ɣ) a nepředstavuje ani detailní fonetickou transkripci, ani transkripci dialektologickou. Požadavek transkribovat to, co bylo skutečně vysloveno, se odráží jak na rovině segmentální (asimilace, zjednodušená výslovnost souhláskových skupin, kvantita samohlásek, sdílení hlásky, tj. případy, kdy první slovo končí a druhé začíná stejnou hláskou, ale v realizaci mluvčího je tato hláska vyslovena jen jednou), tak na rovině suprasegmentální (přízvukové takty). Tyto dvě roviny doprovází ještě **metajazyková**, která zachycuje doprovodné zvuky produkované mluvčími (např. smích, kašel) nebo přítomné v okolí s možností ovlivnit rozhovor (např. zvonění telefonu, které může vést k přerušení dialogu).

Korpus ORTOFON umožňuje zkoumat různé aspekty mluveného jazyka jako např. lexikum, morfologii, syntax, pragmatiku nebo výstavbu dialogu. Korpus sice není primárně určen pro dialektologický či fonetický výzkum, přesto ale umožňuje ověřit existenci výslovnostních či regionálních variant nebo jevů souvisejících s výslovností. Může také poskytnout data pro doplňkový nářeční výzkum, neboť obsahuje oblastní varianty v rovině hláskoslovné, morfologické nebo regionální či nářeční výrazivo. Zveřejněním korpusu ORTOFON a korpusů řady ORAL se uživatelům otevírá možnost zkoumání neformální mluvené češtiny v dosud nejrozsáhlejším komplexu dat (v součtu 6,4 mil. slov) z patnáctiletého období (2002–2017).

Korpus **DIALEKT**⁵ má za cíl prezentovat tradiční teritoriální dialekty zachycené na území celé České republiky. Je složen z přepisů nahrávek nářečních mluvčích, přičemž každý přepis je realizován na dvou rovinách, dialektologické a ortografické, a je spojen se zvukovou stopou. Rozsah korpusu je v jeho první verzi přibližně 100 000 slov⁶ a počítá se s tím, že se bude dále rozšiřovat a budou zveřejňovány jeho další verze. (Každá jeho verze bude referenční, tj. neměnná.) V korpusu je na 324 nahrávkách, čítajících téměř 13 hodin, zachyceno 178 mluvčích. Větší počet mluvčích je dán tím, že nářeční sondy jsou zaměřeny vždy na jedno konkrétní téma, a tudíž poměrně krátké (cca od 1 min. do 6 min.). Korpus obsahuje jazykový materiál teritoriálně různorodý, jsou v něm zastoupeny všechny nářeční oblasti České republiky. Pouze data z oblasti českého, moravského a slezského pohraničí se zatím nepodařilo získat, do dalších verzí se však jejich začlenění plánuje. Navíc bylo ale zařazeno několik nahrávek pořízených v českých jazykových ostrovech na území Polska. V současné fázi budování korpusu je hlavním cílem především shromáždění dostatku jazykových dat, proto korpus v jeho prvních verzích nebudeme vyvažovat.

Korpus **DIALEKT** je podobně jako korpusy ORAL a ORTOFON lemmatizovaný a morfologicky označovaný. Vzhledem k velké variabilitě nářečního materiálu a nedostatku trénovacích dat byl ale proces značkování a lemmatizace značně komplikovaný a s vědomím toho je také třeba k výsledku přistupovat.

Korpus je budován ve dvou vrstvách nářečního materiálu. **Starší nářeční vrstva** obsahuje nahrávky, které byly pořízeny v období od konce 50. let až do 80. let 20. století. Do **nové vrstvy** jsou zařazeny sondy pocházející z období od 90. let 20. století až do současnosti. V případě obou vrstev máme k dispozici jazyková data, v nichž jsou zachyceny archaické nářeční prvky, které se už dnes v běžném úzu většinou nevyskytují. Část starší vrstvy tvoří jazykový materiál získaný sběrem dialektologického oddělení Ústavu pro jazyk český AV ČR, v. v. i., publikovaný v *Dodacích Českého jazykového atlasu* (Balhar 2011), z něhož pocházejí také nahrávky pořízené na polském území. Zbývající část starší vrstvy tvoří soukromé sběry jednotlivců, které byly také z velké

⁵ Podrobné informace o korpusu jsou zveřejněny na <http://wiki.korpus.cz/doku.php/cnk:dialekt>.

⁶ Na dialektologické rovině (v KonTextu pod názvem dialekt v1 – dial) obsahuje 128 289 pozic, resp. 99 552 textových slov, na ortografické rovině (v KonTextu dialekt v1 – ort) obsahuje 126 131 pozic, resp. 99 581 textových slov.

části publikovány. V nové vrstvě korpusu jsou zařazeny sondy z vlastního sběru institucí, pocházejících zejména z fakult různých univerzit, soukromého nahrávání jednotlivců a v neposlední řadě také ze sběru nářečních sond ÚČNK.

Co se týká **metodiky sběru dat**, praktikují se zásady, které jsou běžné v oblasti české dialektologie. V této úvodní fázi, zaměřené na získávání nářečního materiálu, se přednostně soustředíme na zachycení nejstaršího stavu tradičního teritoriálního dialektu a zatím neregistrujeme generační rozdíly v nářečí. V rámci obou korpusových vrstev, starší i nové, se tedy terénní nářeční výzkum zaměřuje výhradně na příslušníky nejstarší generace, aby byly pokud možno zachyceny původní nářeční rysy. V promluvách nářečních mluvčích jsou dochovány dialektismy ve všech jazykových rovinách (fonetické a fonologické, morfologické, syntaktické i lexikální). Mluvčími jsou především místní rodáci zpravidla z venkovských lokalit, kteří patřili ke starousedlé vrstvě obyvatelstva, většinou se nestěhovali a byli spjati se zemědělským způsobem života nebo s řemeslem. Vybírání byli hlavně nářeční mluvčí ve věkové kategorii nad 60 let, tedy ti, kteří se narodili v období od konce 19. století až do poloviny 20. století.

Nahrávky mají spíše neformální ráz, přestože velká řada z nich zachycuje situaci, kdy explorátoři (nahrávající) prováděli s informátory (nářečními mluvčími) výzkum formou řízeného rozhovoru – metodou užívanou v dialektologii. Výsledkem jsou nepřipravené promluvy převážně monologického charakteru realizované v soukromém domácím prostředí. Témata promluv souvisejí s tradičním venkovským životem a tehdejším světem, jsou tedy spojena se zemědělstvím, řemesly, místními zvyky a tradicemi, folklorem, dobovými událostmi atp. Škála témat je poměrně široká, nalezneme mezi nimi třeba tkalcovství, začátek II. světové války nebo vyprávění o zakletém hadovi.

Přestože korpus není sociolingvisticky vyvážený, obsahuje velmi pestrý jazykový materiál co do typů mluvčích, míst a situací při nahrávání. To vše je zachyceno v bohatém **sociolingvistickém značkování**. U každého mluvčího je zachycena řada údajů od pohlaví, věku, vzdělání, pobytu mluvčího do 15 let, místa nejdelšího pobytu a jeho podrobného nářečního začlenění až po nejdéle vykonávanou profesi. Ve vztahu k místu nahrávání lze pracovat např. s údaji o typu lokality (městská, venkovská) a její velikosti, příslušnosti do státu, regionu (Čechy, Morava, Slezsko), nářeční skupiny, podskupiny, případně i nářečního úseku a typu. Informace o nahrávce zachycují kromě jejího zařazení do systému nářečního členění např. i zdroj pořízení nahrávky (instituce apod.), rok, téma, počet mluvčích a přítomnost explorátora. Všechny tyto údaje si lze zobrazit a využít pro tvorbu subkorpusů nebo zjišťování frekvenčních charakteristik.

Nářeční materiál je v korpusu **DIALEKT**, stejně jako v korpusu ORTOFON, zpracováván v prepisovacím programu ELAN. Jazyková data jsou zachycována na **dvou úrovních prepisu**, dialektologické a ortografické. Ty jsou propojeny se zvukovou stopou a v další úrovni jsou zachyceny i metajazykové údaje. V korpusovém rozhraní KonText je možné si obě úrovně prepisu zobrazit vedle sebe v podobě paralelního korpusu. Jako základní je brána **dialektologická** rovina prepisu. Vychází z *Pravidel*

pro vědecký přepis dialektických zápisů českých a slovenských (1951) a dále z Českých nářečních textů (Lamprecht, A. – Michálková, V. ad. 1976) a Dodatků Českého jazykového atlasu (Balhar 2011). Kromě běžných znaků jsou v ní využívány i znaky tradičně užívané v dialektologických přepisech (např. *w* pro obouretně *v*, *ä* pro výrazně široké *a* apod.). **Větná interpunkce** je používána ve shodě s pravidly českého pravopisu, přičemž zvláštní rysy syntaktické stránky mluvených projevů (opakování slov, přiřazování vět a větných úseků apod.) jsou zachyceny zjednodušeně, většinou pouze pomocí interpunkčních znamének, zvláště čárky. Druhou úroveň transkriptu představuje **ortografický** přepis, jenž je srovnatelný s obecnými pravidly stanovenými pro mluvené korpusy v ČNK. Blíží se běžnému ortografickému zápisu, přičemž zůstávají zachyceny některé rysy mluveného jazyka a regionální prvky. Kvantita je převáděna na spisovnou (i v případě slezského systémového krácení). **Hláskové rozdíly** v kořenech slov se kvůli potřebám značkování a lemmatizace sjednocují do spisovné podoby (*bili*, *buli*, *boli*, *beli*, *byli* na dialektologické rovině je sjednoceno pod *byli* na rovině ortografické). V původní podobě zůstávají **morfologické jevy**, např. koncovky všech typů skloňování (*synoj*, *perletěj*) a časování (*nosijó*). Nářeční nebo regionální **lexikum** (*rulík*, *trávnica*, *ostat*) se zachovává, a pokud nemá paralelu ve spisovném jazyce, ponechává se i původní hlásková podoba (*kútky*). Na ortografické rovině je používána **interpunkce pauzová**, v rámci které se rozlišuje několik druhů pauz podobně jako v korpusu ORTOFON (popis pauz viz výše). Pauzová interpunkce nám kupříkladu umožňuje zpětně u dialektologické roviny zjistit, proč není realizována předpokládaná mezislovní asimilace znělosti.

Předpokládáme, že korpus DIALEKT, který je ve svém složení i způsobu zpracování unikátní, bude sloužit nejen odborné veřejnosti (dialektologům, ostatním lingvistům a badatelům z příbuzných oborů), ale například i jako praktická pomůcka při výuce na středních nebo vysokých školách. V budoucnu by měl být nejen rozšiřován, ale i doplněn o interaktivní mapy s nářečními rysy z jednotlivých nářečních oblastí, o ukázky přepisů a nahrávek z vybraných lokalit a o další užitečné doplňky.

GRANTOVÁ PODPORA

Tento text vznikl při realizaci projektu Český národní korpus (LM2015044) financovaného Ministerstvem školství, mládeže a tělovýchovy v rámci aktivity Projekty velkých infrastruktur pro VaVal.

LITERATURA

BALHAR, J. ad. (2011): *Český jazykový atlas Dodatky*. Praha: Academia.

Dialektologická komise České akademie věd a umění (1951): *Pravidla pro vědecký přepis dialektických zápisů českých a slovenských*. Praha: Česká akademie věd a umění.

KOPŘIVOVÁ, M. – KOMRSKOVÁ, Z. – LUKEŠ, D. – POUKAROVÁ, P. (2017): Korpus ORAL: sestavení, lemmatizace a morfologické značkování. *Korpus – gramatika – axiologie*, 15, 47–67.

LAMPRECHT, A. – MICHÁLKOVÁ, V. et al. (1976): *České nářeční texty*. Praha: Státní pedagogické nakladatelství.