

# Obsah / Content

## Články / Articles

### **K slovotvorné analogii v češtině**

On the word-formative analogy in Czech /3

*Jana Bílková*

### **Tak v současných mediálních mluvených rozhovorech: ty zákony co se vztahují teďkon na nás soukromníky tak to asi budu muset zavřít (1. část)**

The word *tak* in contemporary spoken mass media dialogues (Part 1) /12

*Lucie Jílková*

### **K emocionálnosti interjekcí**

Towards the emotionality of interjections /34

*Josef Štěpán*

### **K synonymii primárních a sekundárních imperfektiv**

On synonyms of primary and secondary imperfective verbs in Czech /54

*František Štícha*

### **Neuhochdeutsche Lehnwörter im Tschechischen: Eine Frequenzuntersuchung zwecks der Tendenzermittlung**

New High German loanwords in Czech:

A frequency test to determine the general trends /67

*Aleksej Tikhonov*

## **Recenze / Reviews**

**Hana Bergerová, Lenka Vaňková et al. (eds.):  
Lexikalische Ausdrucksmittel der Emotionalität im Deutschen  
und im Tschechischen / 77**

*František Štícha*

## **Zprávy / News**

**Gramatika a korpusy po šesté / 82**

*Alexandr Rosen*

**Korpusová lingvistika Praha 2016 / 85**

*Michal Škrabal*

**Pokyny pro autory / 88**

Instructions for authors / 89

**Objednávka / 90**

Order form / 91

## Články / Articles

### K slovotvorné analogii v češtině

*Jana Bílková*

*Pedagogická fakulta, Univerzita Hradec Králové*

*jana.bilkova@uhk.cz*

#### **On the word-formative analogy in Czech**

**ABSTRACT:** The article looks into the basic word-formation relations – semantic motivation and formal foundation – and the tension between them. It presents three cases of word-formative analogy, i.e. imitating of a productive word-formative model by the copying of a morphemic structure of derivatives, regardless of the completeness or incompleteness of the derivative line. The article reflects experience gained during the work on Dictionary of affixes used in Czech.

**KLÍČOVÁ SLOVA:** slovotvorba, motivace, fundace, analogie, slovotvorný typ

**KEY WORDS:** word formation, motivation, foundation, analogy, word-formative type

# ***Tak v současných mediálních mluvených rozhovorech: ty zákony co se vztahují teď kon na nás soukromníky tak to asi budu muset zavřít*** (1. část)

Lucie Jílková

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

## **The word *tak* in contemporary spoken mass media dialogues**

**ABSTRACT:** The paper tried to answer the following research questions: What are meanings and functions of the Czech word *tak* [so in English] in spoken texts used in contemporary media? How is the word *tak* incorporated into a syntactic structure of a turn? Two TV programmes included in the Dialog Corpus were chosen to be analysed: *Otázky Václava Moravce* (a very formal discussion with politicians hosted by Václav Moravec) and *Uvolněte se, prosím* (a very informal, spontaneous talk show). The chosen programmes were intentionally quite different; they noticeably differed even in the total numbers of the *tak* occurrences caught in them: 74 occurrences in *Otázky Václava Moravce* vs. 149 occurrences in the talk show *Uvolněte se, prosím*. In the talk show, we identified an extremely high occurrence of the word *tak* as the preparative particle as it is called, and even a higher number of this word was found as a simple connector. The connector *tak* placed at the end of a line appeared only in the talk show. In both programmes, the word *tak* was commonly used to express various syntactic relations, most often expressing conditions and consequences. The examples often indicated that the word *tak* might be the only word expressing such relations.

**KEY WORDS:** word *tak*, DIALOG corpus, mass media dialogue, spoken language, meanings and functions of the Czech word *tak*

**KLÍČOVÁ SLOVA:** slovo *tak*, korpus DIALOG, mediální dialog, mluvený jazyk, významy a funkce slova *tak*

# K emocionálnosti interjekcí

Josef Štěpán

Peroutkova 1760/19, 150 00 Praha 5

stepan.josef@centrum.cz

## Towards the emotionality of interjections

**ABSTRACT:** The article delimits emotionality and presents a brief survey of research of emotional interjections. The methodology of the research (the issues of classification of the semantic field of emotionality, semantics of interjections, selection of the analysed texts) is outlined. The core of the paper consists in the characteristics of the five basic types of emotions, described on the basis of the material of interjections: (1) psychical states, e.g. pleasure-displeasure, joy, sadness, relief, surprise etc., (2) egoistic emotions (e.g. fear), (3) altruistic emotions (e.g. liking-antipathy, mistrust, admiration, wrath etc.), (4) esthetical emotions (beauty, disgust), (5) ethical emotions (e.g. pity). The sixth part of the article investigates the displays and consequences of emotions (for example, laughing and crying are expressed by onomatopoeic interjections. The general conclusions of the paper are represented by the findings that emotional interjections often have no inherent meaning, but these interjections are usually polysemic (as one interjection expresses more emotions, e.g. *ach, bože*), and that some emotions are expressed by a number of synonymous interjections (e.g. 'joy' *jupí, juchú, ach, joj, já, páni, bože* etc.).

**KEY WORDS:** derived (secondary) interjections, emotions, interjection, underived (primary) interjections

**KLÍČOVÁ SLOVA:** odvozené (sekundární) interjekce, emoce, interjekce, neodvozené (primární) interjekce

# K synonymii primárních a sekundárních imperfektiv

František Štícha

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

## On synonymy of primary and secondary imperfective verbs in Czech

**ABSTRACT:** Up to now, Czech linguistics generally assumed that the so called secondary imperfective verbs (SI) can only be used if the primary imperfective verb (PI) and the SI of the same root have different meaning. Thus, the SI *svažovat* (**S**-*vař-ova-t* (to **boil**; -*vař*- being the root, *s*- the prefix and *-ova-* and *-t* the suffixes) is used because the PI *vařit* (**vař-i-t** (to **cook**) has a different meaning. Contrary to that, *uvařovat* (\***U**-*vař-ova-t*) is not used because its meaning would be the same as in the basic PI *vařit*. Only a few exceptions to this rule were assumed to exist. However, the corpus findings show that this assumption was false. There are at least 160 SI derived from the basic PI in current standard written Czech language which are synonymous to the PI. This article concentrates on SI derivations with the prefix *VY-* and presents empirical data which enable us to postulate the synonymy of PI and SI with some 70 verb pairs.

**KEY WORDS:** aspect, primary and secondary imperfective verbs, prefixes, synonymy

**KLÍČOVÁ SLOVA:** vid, primární a sekundární imperfektiva, prefixy, synonymie

# Neuhochdeutsche Lehnwörter im Tschechischen: Eine Frequenzuntersuchung zwecks der Tendenzermittlung

*Aleksej Tikhonov*

*Institut Slavistiky, Humboldtova Univerzita Berlín*

*tikhonal@hu-berlin.de*

## **New High German loanwords in Czech: A frequency test to determine the general trends**

**ABSTRACT:** This article deals with germanisms in Czech. Frequencies of 26 different new High German loanwords were analyzed in the Czech National Corpus. These borrowed words were standing in competition with their Czech synonyms. This comparison is used to study the question of whether germanisms or their equivalents in Czech are more used by native speakers. For this analysis new High German loanwords were deliberately selected in order to verify the actuality of the topic. But the major part of the study was examined in a diachronic period. This shows not only the current situation but in most cases the frequency of the selected loanwords throughout their existence. The calculations of the average frequency are made for each century (since 1650), and also in the recent modern period (from 1947 to 2008).

**KEY WORDS:** germanisms, synchrony, diachrony, corpus, Czech, German, frequency, loanwords

## **Novohornoněmecké výpůjčky v češtině: Frekvenční výzkum ke stanovení obecných tendencí**

**ABSTRAKT:** Článek se zabývá germanizmy v češtině. Prostřednictvím Českého národního korpusu byly zjišťovány různé frekvence 26 novohornoněmeckých výpůjček a jim konkurujících českých synonym. Článek se na základě frekvenčních srovnání snaží odpovědět na otázku, zda čeští rodilí mluvčí preferují germanizmy či dávají přednost jejich českým ekvivalentům. Článek analyzuje nejen aktuální situaci, ale ve většině případů ukazuje frekvenci vybraných germanizmů z diachronního hlediska, po celou dobu jejich existence. Byla vypočtena průměrná frekvence za každé století (od roku 1650), včetně posledního moderního období (od roku 1947 do roku 2008).

**KLÍČOVÁ SLOVA:** germanizmy, synchronie, diachronie, korpus, čeština, němčina, frekvence, přejatá slova

## Recenze / Reviews

### Hana Bergerová, Lenka Vaňková et al. (eds.): **Lexikalische Ausdrucksmittel der Emotionalität im Deutschen und im Tschechischen**

Universität Ostrava, 2015

František Štícha

Ústav pro jazyk český AV ČR, v. v. i.

sticha@ujc.cas.cz

Tato kniha o lexikálních způsobech vyjadřování emocí, přesněji řečeno o jejich pojmenovávání, v němčině a v češtině je díky relativně velmi specifickému tématu jistě spíše monografií než tematickým sborníkem. Kromě úvodu, závěru, seznamu literatury a věcného rejstříku má pět kapitol pojednávajících o čtyřech základních emocích – radosti, smutku, zlosti, strachu a nádavkem ještě o lásce. Každá z těchto emocí má svého autora: Lenka Vaňková pojednává o radosti, Martin Mostýn o smutku, Hana Bergerová o zlosti, Eva Ciešlarová o strachu a Jiřina Malá o lásce.

Kniha je výstupem projektu *Ausdrucksmittel der Emotionalität im Deutsch-tschechischen Sprachvergleich* podporovaného Grantovou agenturou České republiky.

Autoři se poučili i z psychologické literatury, zejména z německého překladu francouzských autorů F. Lelorda a Ch. Andrého *Die Macht der Emotionen*. Tuto knížku, která v Německu vyšla v mnoha vydáních, by bylo dobré mít i v českém překladu.

Protože autoři při soustavném sledování jazykového výrazu oněch čtyř bazálních emocí a lásky z česko-německého kontrastivního hlediska soustavně pracovali s paralelním česko-německým a německo-českým korpusem textů obsaženým v ČNK, soudím, že náš časopis je pro zprávu o jejich monografii dobrým místem.

Budu se nyní věnovat jednotlivým vybraným výroky a zjištěním každé z pěti kapitol. Začínáme **radostí**.

Přestože všichni víme, kdy prožíváme radost a kdy nikoli, definovat „radost“ odborně tak, abychom se zavděčili všem, patrně není možné. To ostatně platí asi u každého abstraktního pojmu. Autorka této kapitoly, Lenka Vaňková, upozorňuje na to, že v odborné literatuře se pracuje i s „50 koncepty radosti“ (50 Konzepte der Freude). Na základě odborné literatury sestavuje autorka pro oba jazyky lexikální škálu odpovídající stupňům intenzity obecně pojaté „radosti“: *Erleichterung, Befriedigung, Zufriedenheit, Behagen, Gefallen, Vergnügen, Freude, Lust, Glück, Seligkeit, Entzücken, Begeisterung, Euphorie, Rausch, Ekstase* pro němčinu a *ulehčení, uspokojení, spokojenost, pohoda, zalíbení, potěšení, radost, rozkoš, štěstí, blaženost, zánícení, nadšení, euforie, opojení, extáze* pro češtinu. Samozřejmě bychom o všech



těchto slovech a jimi vyjadřovaných pojmech mohli diskutovat donekonečna. Podotknou tedy aspoň stručně, že všechna tato slova se pravděpodobně chovají sémanticky a syntakticky natolik různě, že budeme jen těžko hledat nějaký jejich společný jmenovatel. Například naleznou-li v něčem ulehčení, nemusí to znamenat, že mám radost. Např. v následující větě z ČNK jistě nejde o radost: *Nakonec jsme všichni s ulehčením souhlasili, že je škoda, že se právě Češi a sudetští Němci nechá-pou [...]*. Tyto věci ale nejsou v kapitole Lenky Vaňkové podstatné. Podstatné je to, co korpus objektivně ukazuje o užívání určitého slova označujícího nějakou emoci. Například je jistě jazykově i kognitivně signifikantní, jestliže němčina liší různé druhy radosti užíváním různé předložky: *Freude an* nebo *über etwas* – což čeština nečiní – v ní máme vždy radost pouze *z* něčeho. I v češtině bychom ale mohli usoudit, že při *radosti z přírody* (německy *an der Natur*) jde o poněkud jinou (např. méně intenzivní) radost než při *radosti z dárku* (německy *über das Geschenk*). Autorka hovoří o různých „spouštěcích radosti“, jejichž lexikální výraz se spojuje s různou předložkou, a dokládá to na poměrech korpusových nálezů; např. předložka *über* dominuje nejen u *dárku*, ale třeba i u *návštěvy a setkání*, zatímco u jiných spouštěčů radosti se užívá obou předložek téměř stejnou měrou. Autorka dochází k závěru, že rozdíl mezi oběma předložkami patrně souvisí s paragramy trvání a intenzity. Pokud jde o jedinečný a krátkodobý zážitek, užívá se předložky *über*, v opačném případě předložky *an*. Takto ovšem nevysvětlíme, proč se u spouštěče *Leistung* (výkon) užívá obou předložek téměř rovnoměrně (54 : 33; podstatně větší a/nebo jinak sestavený korpus by ovšem pravděpodobně přinesl jiný poměr). Z hlediska češtiny se ale také zdá, že německému *Freude an etwas haben* by mohlo odpovídat české *mít z něčeho potěšení*, kdežto *Freude über etwas haben* by odpovídalo českému *mít z něčeho radost*. Šlo by pak primárně o rozdíl v intenzitě prožitku.

2500 německých kompozit obsahujících jako bázi substantivum *Freude*, o nichž se autorka zmiňuje, svědčí nejen o obrovské slovotvorné produktivitě tohoto obecného německého slovotvorného modelu, ale také o tom, že „spouštěčů“ radosti není málo.

Soustavná práce s paralelními korpusy snadno odhalí překladatelskou vynalézavost (nebo nevynalézavost) a její míru, zejména tam, kde pro určité slovo jednoho jazyka schází ve druhém jazyce přímý a bezprostřední lexikální ekvivalent. Autorka kapitoly o lexikálních způsobech vyjadřování radosti to ukazuje na řadě případů, z nichž nejnápadnější se mi zdá německé substantivum *Wucht* a jeho překladové ekvivalenty. Půl století starý Siebenscheinův dvousvazkový *Německo-český slovník* (větší a lepší zatím bohužel nemáme) tomuto německému slovu přiřazuje jako české lexikální ekvivalenty substantiva *váha, síla, tíha*. Ani náznak něčeho, co by mohlo připomínat radost. Navíc *váha* je německy především *Gewicht, síla* je *Kraft* nebo *Stärke* a *tíha* je *Schwere*. Už jen z tohoto faktu této asymetrie samé nutně plyne, že *Wucht* vlastně není ani *váha*, ani *síla* a ani *tíha*. Že je to pojem, který má od všech těchto pojmů něco, ale celkově jde o pojem svérázný a jazykově specifický, který česky nelze jedním slovem vyjádřit. Pochopitelně se zde dotýkám jevu notoricky známého, nicméně soudím, že neuškodí v detailech si připomínat to, co velmi dobře známe v obecné rovině. Německý *Duden. Deutsches*

*Universalwörterbuch* (v 5. vydání z roku 2003) jako 4. význam substantiva *Wucht* uvádí: **eine Wucht sein** (salopp; *beeindruckend, großartig sein*) a jako příklad tohoto významu uvádí *das Hotel ist eine Wucht*. Vzhledem k základnímu významu substantiva *Wucht* by patrně nejbližším ekvivalentem v češtině bylo slovo *bomba – ten hotel je bomba*. Autorka však v paralelním korpusu našla tyto ekvivalenty: *jednička, k sežrání, borec, třída, sexbomba, kus*. Jde tu pravda o samé lidi, nikoli o věci – u nich by musely být ekvivalenty jiné. Jedním z autorčiných závěrů je povzdech nad celkem známou věcí, že ve slovnících často nalézáme něco, co v praxi nelze použít.

Příznám se, že teprve z této knihy o vyjadřování emocí, konkrétně z její kapitoly o **smutku**, jsem se dozvěděl, že německé substantivum *Trauer* označuje něco jiného než z něj derivované *Traurigkeit*. Konkrétně mi to napověděla citace z Freuda, v níž se (v mém překladu) praví, že „Trauer“ je „reakce na ztrátu milované osoby nebo na její místo povýšené abstrakce, jako je vlast, svoboda, ideál atd.“ To však o českém substantivu *smutek* rozhodně neplatí. Ukazuje to dosti zřetelně např. tato věta z ČNK: *Jako novinář sleduji se smutkem, kterak hrubne slovník lidí ve veřejném prostoru*. Novinářův smutek nevyvolává ani ztráta milované osoby, ani ztráta nějakých vznešených ideálů, ale pouze to, jak hrubne něčí slovník. Je navíc dosti pravděpodobné, že konkrétní smutek novinářův z tohoto hrubnutí není velký. V němčině by tu tedy nešlo o *Trauer*, ale o *Traurigkeit*. Autor na tento rozdíl bohužel nepoukazuje.

Z nejrůznějších překladatelských ekvivalentů, na něž autor upozorňuje a jež prezentuje – jako všichni autoři knihy – v tabulkách a rámečcích, mne zaujala například lexikální neekvivalence frazeologizovaného užití substantiva *Jammer* a jeho českého překladového textového ekvivalentu *pláč*. Siebenscheinův slovník překládá různé významy substantiva *Jammer* jako *bída, nouze, tíseň; útrapy, trápení; hoře, žal; smutek, zármutek, naříkání, bědování; škoda* – tento význam je uplatněn v kontextu *Es ist ein Jammer, daß ...* Avšak ani všechny tyto lexikální ekvivalenty někdy překladateli, jak se zdá, nestačí: Hanuš Karlach totiž ve známé próze G. Grasse *Die Rätin* přeložil výrok *Ein Jammer, dieser humane Pfusch!* takto (tučně FŠ): *Bylo to k pláči, tohleto lidské fušerství*.

Jako spouštěče emoce německy nazvané *Trauer* korpus podle autora dokládá např. (v mém výběru; uvedená substantiva jsou jádrem postponovaného předložkového atributu, např. *Trauer über*) *Ableben, Abschied, Ermordung, Leid, Tod, Trennung, Untergang, Verlust, Zerstörung*. V souhrnném korpusu syn v3 ČNK ale po syntagmatu *smutek* z nejběžnější následují substantiva *porážka, prohra a vyrazení* (celkem 173 výskytů). To svědčí jednak o tom, že našemu sociálnímu světu dominuje sport (nikoli umění nebo věda), jednak o tom, že v češtině *smutek* nemusí znamenat nějakou zvlášť velkou ztrátu.

Ze všech poznatků této kapitoly upozorním již jen na jediný detail. Když autor porovnává sémantické příznaky německého slovesa *greinen* v německé originální větě *greinte wie ein Kind* s českým slovesem *rozbřečet se* v překladatelském ekvivalentu *rozbřečel se jako dítě*, sledává tu sice rozdíly v sémantických příznacích, bohužel si ale netroufá konstatovat, že toto je opravdu nepovedený a špatný překlad. V tomto kontextu už totiž bylo jednou řečeno, že *„Staršina se rozplakal.“*

Následují věty *Ruce si přitiskl na svůj široký obličej sibiřského sedláka. Motal se a rozbrečel se jako dítě.* Sotva tu má jít o to, že staršina se nejprve rozplakal, pak plakát přestal a vzápětí se rozbrečel znovu. Německé *greinen* podle charakteristik slovníku nejvíce odpovídá jednak našemu *fňukat*, jednak našemu *bědovat*. Protože však o dětech obvykle neříkáme, že bědují, odpovídalo by rozhodně lépe originálu sloveso *fňukat*. Navíc německé *sich herumwerfen*, které stojí v originále, je spíše *házet sebou* než *motat se*. Místo *Motal se a rozbrečel se jako dítě* bychom tedy měli spíše napsat *Házel sebou a fňukal jako dítě*.

Také v případě *hněvu*, *zlosti* a *vzteku* (v němčině *Zorn*, *Ärger* a *Wut* – nejde ale o přesné lexikální ekvivalenty) umožňují „adjectivní partneři“ těchto substantiv, jak je shledáváme v korpusech, objektivněji identifikovat shody a rozdíly ve významu těchto kandidátů synonymie. Např. jak v němčině, tak v češtině se s adjektivem *boží* nebo *spravedlivý* spojuje pouze jedno z uvedených třech substantiv: *Zorn* a *hněv*. V češtině se ještě participiální adjektivum *potlačovaný* spojuje pouze se substantivem *hněv*, zatímco německé adjektivum *unterdrückt*, lexikálně odpovídající českému adjektivu *potlačovaný*, se spojuje se všemi třemi substantivy. Autorka této kapitoly konstatuje, že překladatelé zacházejí s výběrem lexikálního ekvivalentu pro německá substantiva *Zorn*, *Ärger* a *Wut* dosti svévolně, neboť ke každému z nich je v dosti hojném počtu doloženo kterékoli z české trojice substantiv *hněv*, *zlost* a *vztek*. K tomu podotýkám, že nic jiného se nedalo očekávat. Navíc podotýkám, že k německému *Wut* je možné přiřadit i *zuřivost*, *běs* či *běsnění*. Na s. 118 autorka píše, že důkazem toho, že uživatelé jazyka tyto výrazy nepokládají za synonyma, je to, že tato slova kladou vedle sebe v koordinovaných spojech, např. *Zorn und Wut*. Tak tomu ovšem není. Lze totiž prokázat, že v koordinativních spojech se vyskytují i prokazatelná synonyma, a sice ze stylizačních, např. rytmických důvodů.

Pokud jde o **strach** a jeho „spouštěče“, pak je poněkud překvapivé zjištění autorky této kapitoly, že zatímco mezi běžnými českými kolokáty tohoto substantiva jsou předložkové výrazy *z výšek*, *z reakce* a *z neznámého*, v němčině mezi prvními 50 kookurencemi substantiva *Angst* odpovídající výrazy scházejí. Naopak mezi častými německými kookurenty jsou výrazy *vor Repressalien*, *vor Zinserhöhung* (*z represálií*, *ze zvýšení úroků*), které zase nejsou doloženy mezi běžnějšími kookurenty českými. Jako možný důvod těchto rozdílů autorka zmiňuje žánrově nestejně složení českého a německého korpusu. O tom, zda je tomu opravdu tak, bychom se poměrně snadno přesvědčili, kdybychom hledali například pouze v novinových korpusech.

V poslední kapitole věnované **lásce** jsem na nic – z mého pohledu – zvláštního, na co bych chtěl čtenáře této recenze speciálně upozornit, nenarazil. Tím ale nechci říci ani naznačovat, že tato poslední kapitola nestojí za přečtení. Naopak. Účelem této stručné zprávy opravdu není, aby potenciálnímu čtenáři nahrazovala čtení celé knihy. Teoretik v ní možná mnoho nového neobjeví (to ale nemohu příliš kompetentně posoudit), zato jazykový praktik si tu přijde na své. Knižka o lexikálních prostředcích emcionality stojí za pozorné čtení. Mohu jen osobně zalitovat, že okruh potenciálního českého čtenářstva je omezen na germanisty a ostatní znalce němčiny.

# Gramatika a korpusy po šesté

Alexandr Rosen

Ústav teoretické a počítačnické lingvistiky, Filozofická fakulta UK, Praha  
alexandr.rosen@gmail.com

Šestý ročník mezinárodní konference *Gramatika a korpusy* uspořádal ve dnech 9. až 11. listopadu 2016 mannheimský Ústav německého jazyka.<sup>1</sup> Konference se konala poprvé v roce 2005 v Praze, pak 2007 v Liblicích, 2009 v Mannheimu, 2012 v Praze a 2014 ve Varšavě. Deklarovaným tématem je obecně využití korpusů při popisu a teoretické analýze gramatiky, a to se zaměřením na germánské, románské a slovanské jazyky, mimo jiné i z perspektivy kontrastivní, synchronních variant i vývoje jazyka. Může přitom jít o hledání empirických podkladů teorií a gramatik, ověřování hypotéz a nejrůznější metodologické i teoretické otázky, jakými jsou např. anotační schémata, statistické modely, hranice mezi gramatikou a diskursem a ostatně i vztah mezi korpusovou a teoretickou lingvistikou. Rozdílnost teoretických východisek a empirických cílů se přitom považuje za vítanou šanci k diskusi.

Spojitosť mezi gramatikami a korpusy není samozřejmá. Teoretická lingvistika někdy preferuje introspekci před doklady z korpusu, na druhé straně korpusová a počítačová lingvistika často pochybuje o tom, zda hledat zobecnění v podobě gramatických principů a pravidel má vůbec nějaký smysl. Příležitostí pro setkávání s důrazem právě na spojení teoretické a empirické povahy lingvistiky a převážně lingvistický pohled tak není mnoho. Právě v tom se tato konference liší od jí podobných, a to i od podobně nazvané *Treebanks and Linguistic Theories*, jejímž hlavním tématem jsou syntakticky anotované korpusy.

Konference je pestrá svým záběrem i při srovnání jednotlivých ročníků. Každý má výrazně odlišné složení účastníků. Tento rok jich bylo 110, ze 40 akademických institucí a 18 zemí. Nejvíce (85) z Německa, z toho jen z hostitelské instituce 48. S velkým odstupem následovaly Švýcarsko, Belgie, Francie, Nizozemsko, Rusko, Česká republika (4), Chorvatsko, Rakousko, Polsko, Francie, Finsko, Kazachstán, USA, Gruzie, Velká Británie, Itálie a Rumunsko. Ke 30 příspěvkům ve dvou paralelních sekcích a 5 plenárním přednáškám je třeba přičíst 15 posterových prezentací. Témata se týkala 15 různých jazyků, jednacím jazykem byly tradičně angličtina a jazyk hostitelské země. Prezentace se dočkají i písemné podoby v plánovaném sborníku.

---

<sup>1</sup> <http://gac2016.ids-mannheim.de>

Den před začátkem vlastní konference byly na programu 4 tutoriály, které byly zaměřené převážně na uživatelské aspekty práce s korpusem a statistiku ve službách lingvistiky, konkrétně na webové korpusem, korpusový vyhledávač *KorAP*, vizualizaci jazykových dat pomocí programu *R* a možnosti využití paralelního korpusu *InterCorp*.

Vlastní konference měla kromě plenárních přednášek a posterového bloku dvě paralelní sekce. Nejvíc příspěvků se týkalo výzkumu jednotlivých jevů současného jazyka, ale část se věnovala i jeho historickým nebo nářečním podobám (obojí s výrazným zastoupením němčiny) a textové lingvistice. I když relativně málo příspěvků se soustředilo pouze na problematiku počítačových metod a nástrojů nebo budování a anotace korpusů, byla tato témata přítomna jako součást většiny referátů s převážně lingvistickým obsahem.

Výslovně češtiny se týkaly čtyři příspěvky. Vývoji českého lexému *ovšem* od 16. století do současnosti se věnuje František Martínek z Univerzity Karlovy v příspěvku *Vom Adverb über die Partikel zur Konjunktion: Der Fall des tschechischen ovšem*, a to na základě mnoha zdrojů, včetně korpusu *DIAKORP*. Variantnost koncovek v české deklinaci zkoumají Matías Guzmán Naranjo z Lipska a Olivier Bonami z Paříže v příspěvku *Overabundance as hybrid inflection. Quantitative evidence from Czech*. Další dvě prezentace byly posterové. Elżbieta Kaczmarek z Varšavy se zabývá některými českými lexémy vyjadřujícími duševní stav a navrhuje metodu pro zjištění jejich významu a volbu polského překladu v daném kontextu, částečně na základě paralelního korpusu *InterCorp*. Stefan Heck z Tybinek porovnává užívání aspektu v českém, polském a ruském imperativu na základě ČNK a webových korpusů *Araneum*. Z hlediska bohemistického byly inspirativní i některé další prezentace, např. *From dictionary to morphological analyzer. The case of Baroque Polish* Witolda Kieraše z Varšavy, *Unintegrated co relative clauses in conversational Polish* od Wojciecha Guza z Lublinu nebo *Clitic climbing and infinitive clusters in Bosnian, Croatian and Serbian – A corpus-driven study* od Björna Hansena, Zrinky Kolaković a Edyty Jurkiewicz-Rohrbacher z Řezna.

Dvě plenární přednášky se týkaly analýzy konkrétních jevů (Anke Holler: *Tracking the implicit causality bias in corpora – A case study* a Anne Abeillé: *Grammar of extraction and ellipsis: An empirical view*), další pak problematiky popisu a zpracování nestandardních jazykových variet (Alexandr Rosen: *Coping with unruly language: non-standard usage and phraseology in a corpus*), využití korpusové lingvistiky při jazykové přípravě expertů (Susan Conrad: *Beyond Grammar Description: Applying Corpus Analysis to Disciplinary Education*) a konečně vývoji teoretické, korpusové a počítačové lingvistiky (John Nerbonne: *Vaulting ambition*). Pro obecnější dosah zasluhují poslední dvě podrobnější zmínku.

Susan Conrad prezentovala své zkušenosti z jazykové přípravy stavebních inženýrů. Gramatická analýza odborných textů studentů, čerstvých absolventů a zkušených expertů metodami korpusové lingvistiky pomohla zaměřit se na problematické jevy a posloužila k optimalizaci výuky. Výsledkem bylo výrazně lepší hodnocení absolventů zaměstnavateli. Autorka zdůraznila, že přes „nevědeckost“ tématu jde o velmi užitečnou aplikaci jazykovědy, která vede k maximálně srozumitelnému a jednoznačnému odbornému vyjadřování.

John Nerbonne ukázal, že teoretická, korpusová i počítačová lingvistika prošla za poslední desetiletí podobným vývojem. Od odhodlání analyzovat, zpracovat, popsat nebo vysvětlit daný jazyk v jeho úplnosti s úspěšností blížící se bezchybnosti si dnes všechny tři disciplíny kladou cíle skromnější: přes stále bohatší repertoár analýz nejrůznějších jevů v mnoha jazycích jde dnes teoretické lingvistice o to, analyzovat jen některé výrazy, konstrukce či jevy; přes stále sofistikovanější a výpočetně náročnější metody jde korpusové i počítačové lingvistice o to, vytvářet modely jen některých výrazů a konstrukcí, případně rezignovat na překonání bariéry úspěšnosti v analýze standardních textů a soustředit se na jazyky handicapované nedostatkem zdrojů dat nebo na nestandardní jazykové variety. Obecným řešením může být program bližší korpusové lingvistice, která si vždy kladla o poznání skromnější cíle.

Z uvedeného by mělo být zřejmé, že konference Gramatika a korpusy se v evropské lize lingvistických událostí úspěšně prosazuje a že má díky organizátorům i účastníkům nakročeno k dalším úspěšným ročníkům. Je příjemné si uvědomit, že za tím vším stál od počátku zejména český tým pod vedením Františka Štíchy.

# Korpusová lingvistika Praha 2016

17. září 2016

Michal Škrabal

Ústav Českého národního korpusu FF UK, Praha

michal.skrabal@ff.cuni.cz

V polovině září se na pražské filozofické fakultě konala v pořadí již pátá ze série konferencí Korpusová lingvistika Praha. Oproti předchozím ročníkům (např. o KLP 2011 vyšla zpráva v KGA 4/2011, s. 83–87; k KLP 2014 viz ČMF 1/2015, s. 5–6) byla pojata skromněji a komorněji – spíše jako kolokvium – a její program byl vtěsnán do pouhého jednoho dne. Tak trochu to předjímá už podtitul kolokvia „Od mluvené češtiny k psané“, který výrazně omezil dřívější tematickou šíři příspěvků. Dlužno však dodat, že to nebylo nikterak na újmu: akcentovala se především problematika korpusově založeného popisu mluveného jazyka, postižení jeho jednotlivých variet a registrů, šířeji též kontrastivní pohled na různé komunikační módy – a i takto vymezené tematické pole nabízí široké badatelské angažmá, což ostatně dokázal i pestrý program (je k dispozici též online na <http://ucnk.ff.cuni.cz/kl2016/files/kl2016-program-20160913.pdf>; resp. abstrakty k jednotlivým příspěvkům zde: <http://ucnk.ff.cuni.cz/kl2016/abstracts.php>; ověřováno 2. 2. 2017), rozčleněný do šesti tematických sekcí.

V úvodní plenární přednášce, nazvané „Výstavba mluvených dialogů a jejich výzkum (od pauz a ‚malých slov‘ k výpovědím, replikám, sekvencím, promluvám)“, promluvila Jana Hoffmannová. Východisko jejích úvah bylo třístupňové: nejprve shrnula současně penzum poznatků o specifičnosti mluvené češtiny oproti té psané, vzápětí ukázala, jak k rozšíření těchto znalostí přispívají jazykové korpusy, a zamyslela se též nad tím, co bychom se o mluvených projevech potřebovali dozvědět do budoucna. Připomínám jen, že Hoffmannová je hlavní řešitelkou grantu GA ČR „Syntax mluvené češtiny“, běžícího do r. 2017; poté tedy lze očekávat první, jakkoliv nutně jen dílčí, výsledky tohoto pionýrského výzkumu (dosavadní ucelené popisy a výklady české skladby jsou založeny vesměs na psaných datech).

Po plenární přednášce následovaly ve třech blocích dvě paralelní sekce, vždy po třech či dvou příspěvcích. Převažovala témata gramatická, byť se, jak už to prve naznačila Hoffmannová, kolem gramatiky mluveného jazyka spíše opatrně našlapovalo. Pro představu o rozmanitosti nastíním syžety několika přednášek: Jan Kolář, Veronika Kolářová a Marie Mikulová hledali v psaných i mluvených korpusech způsoby vyjádření propozice pomocí verbálních substantiv; Magdalena Žiková, Jan Křivan a Eva Lehečková se zaměřili na jev českými syntaktiky dosud blíže neprozkoumaný: na pravou dislokaci v mluvené češtině; užití pasivního participia v mluvené češtině se věnovala Kamila Mrázková; Lucie Jílková analyzovala výraz *tak* ve funkci členicích partikule. Techničtěji laděn byl příspěvek Radka Čecha,

Miroslava Kubáta a Jána Mačutka, nahlížející negaci v mluvené češtině z perspektivy Menzerathova-Altmanova zákona.

Nechyběla ani témata kontrastivní: Anna Nedoluzhko[vá] srovnala osobní a reflexivní posesiva v češtině a ruštině, zatímco Nicole Richter[ová] a Dagna Zinkhahn-Rhobodes[ová] se věnovaly česko-polské analýze ukazovacích zájmen. Díky příspěvku Františka Martínka a Marty Rybové „Modální částice v korespondenci Karla Havlíčka“ získalo kolokvium též aspekt lingvo-literární.

Každá konference potřebuje své „koření“, tedy příspěvky originální a rázovité, ať už svým obsahem či formou. Jiří Milička nám – způsobem sobě vlastním, tj. nahony vzdáleným jakékoliv škrobenosti a suchopáru – představil modely zašuměných kanálů, s nimiž se musí potýkat mluvený jazyk, včetně svých vlastních a porovnával je s reálným šumem. A tak publikum tajilo dech a napjatě naslouchalo, až bude vysílané ticho bezcítěně protnuto tu ojedinělým zařinčením, tu více či méně soustavným bříkotem. Půvabný kriminalistický nádech měl referát Radka Skarnitzla a Jitky Vaňkové „Databáze obecné češtiny pro forenzní účely“; ta má sloužit k srovnávání hlasů ve forenzní fonetické praxi. Pan rada Vacátko by asi jen uznale pokýval hlavou a pochvalně zamručel.

Je potěšitelné, že vytěžovány jsou nejen velké, „vlajkové“ korpusy projektu ČNK, ale též ty menší a specializované. Katja Brankačec a František Martínek např. zkoumali valenci českých a lužickosrbských ekvivalentů německého (*sich*) *an etw. gewöhnen* v korpusech DIAKORP, HOTKO a DOTKO (korpusy horní a dolní lužické srbštiny, jež ČNK hostuje; oba byly vytvořeny Lužickosrbským institutem v Budyšině). Další osobité korpusy vznikají: trojice tvůrkyň Karolína Vyskočilová, Zuzana Komrsková a Klára Dvořáková představila mluvený korpus BANÁT, zachycující specifika mluvy banátských Čechů, a také první poznatky získané jejím srovnáním s češtinou na území ČR. Mluvený žákovský korpus začátečníků (A1–A2) pro češtinu jako cizí jazyk pak prezentovala Barbora Kočařová.

Tolik, ovšemže jen výběrově, o dopoledním a popoledním programu. Na začátku posledního bloku představili pracovníci mluvené sekce Ústavu Českého národního korpusu chystané novinky v oblasti mluvených korpusů. Jednak přibudou samotné korpusy: korpus neformální dialogické češtiny ORTOFON a nářeční korpus DIALEKT, oba nově se dvěma úrovněmi anotace, ortografickou a fonetickou (DIALEKT s dialektickou). Dalším vítaným přírůstkem je možnost zobrazit si konkordanci, aktuálně pouze v lineární podobě, z mluvených korpusů ČNK v uživatelsky přívětivějším formátu, ať už pomocí ad hoc vytvořeného nástroje MluvKonk (autor: David Lukeš), či převzatého nástroje ANNIS. V neposlední řadě by u nových korpusů měla být implementována dlouho očekávaná lemmatizace, nově tak bude možno mluvená data prohledávat pomocí lemmatu a tagu.

Celé kolokvium završila diskuse pětice pozvaných účastníků kulatého stolu na téma „Korpusová rozhraní pro práci s mluvenými daty – jejich možnosti a limity“. Role moderátorky se zhostila vedoucí mluvené sekce Marie Kopřivová, vyzvavší jednotlivé aktéry, aby stručně představili, jak s mluvenými daty pracují, co při práci oceňují a co naopak postrádají. Vpravdě pestrý vzorek prominentních uživatelů tak měl možnost poskytnout zpracovatelům mluvených dat zpětnou



vazbu, ale zejména též podněty k budoucím vylepšením, které rozhodně nebývají oslyšeny (jak o tom svědčí např. nová funkce Výběr řádků, přidaná do rozhraní KonText „na objednávku“ autorského kolektivu vznikajícího *Akademického slovníku spisovné češtiny*).

Jako první promluvila Jana Hoffmannová z ÚJČ AV ČR: poukázala na nevhodný způsob zobrazení dat v rozhraní KonText a od toho se i odvíjející potíže při vyhledávání v korpusech řady ORAL. Že lineární zobrazení mluvených dat, kde se jen stěží rozlišují jednotliví mluvčí (neřkuli překryvy jejich promluv), není rozhodně ideální, si jsou dobře vědomi sami tvůrci mluvených korpusů, a chystají proto nový, výše zmíněný, způsob přepisu, který by – soudě podle uveřejněných slajdů – měl být nesrovnatelně přehlednější.

Daně Hlaváčkové (a zdaleka nejen jí) z brněnské filozofické fakulty leží na srdci především lemmatizace mluveného jazyka, problematika to veskrze palčivá, zvláště se zřetelem k nářečním jevům. K současné pracovní verzi lemmatizace měla řadu cenných připomínek, nabídla též zkušenosti ze značkování Brněnského mluveného korpusu.

Dialektologickou obec zastupovala Martina Ireinová z ÚJČ AV ČR; z její poněkud vyhyčivé odpovědi usuzuji, že dialektologové se s možnostmi korpusů řady ORAL a jejich využití pro svou práci teprve seznamují.

Radek Skarnitzl z Fonetického ústavu FF UK uvedl, že obecně pojaté a budované mluvené korpusy jsou pro specifické potřeby fonetiků víceméně nepoužitelné, především vinou nízké kvality nahrávek. (To nicméně nebránilo, aby fonetici výrazně pomohli při tvorbě pravidel pro fonetickou transkripci.) Ovšemže jde o nepřekonatelnou překážku: v dokonale odhlučněné nahrávací místnosti sice nebudou rušit zvuky ulice ani (u přepisovačů velice neoblíbené) chrchlání kávovaru, zároveň z ní však zcela vymizí jakákoliv živost a spontaneita přirozeného mluveného dialogu. Tak či tak, byla podnětná i opačná perspektiva někoho, kdo pracuje s korpusy výrazně menšími – třebaže mluvené korpusy vznikají mnohem klopotněji než ty psané, oproti těm pro fonetické účely se budují relativně rychle a snadno.

Konečně Mária Šimková z Jazykovedného ústavu L. Štúra SAV přidala neméně cennou zkušenost slovenskou, zejména s nářečním korpusem. Podle ní je potřeba zachytit svéráz pozvolna ustupujících dialektů, byť by nahrávky zůstaly prozatím nezveřejněné – aby bylo později vůbec co zveřejňovat. Při této příležitosti apelovala na všechny, kdo takovými daty disponují, aby spojili své síly.

Osobně letošní zacílení na jeden konkrétní aspekt korpusové lingvistiky vítám, podobně orientované kolokvium proběhlo, rovněž na půdě FF UK, před pár týdny – KOLT 2016: „Korpusy v kontrastivní lingvistice a translátologii“. Snad by bylo do budoucna dobré tento dynamický model zachovat, tj. střídát větší, všeobecněji zaměřené konference s menšími, úžeji profilovanými kolokvii. To si dovoluji nadnět jako podnět organizátorům Korpusové lingvistiky Praha, kteří i tentokrát zdárně dostali své úloze.

*Za cenné připomínky a doplnění děkuji Marii Kopřivové.*

# Objednávka

Objednávka časopisu <i>Korpus – gramatika – axiologie</i>				
Odběratel	Název instituce:			
	Dodací adresa:			
	Fakturační adresa:			
	IČO / DIČ / NIP / VAT number:			
	IBAN:			
	BIC (SWIFT):			
Objednávám	zasílání ..... kusu/kusů počínaje číslem .....			
	jednorázové zaslání čísla ..... v počtu ..... kusu/kusů			
	s úhradou v měně:	Kč	€	USD
Další informace pro dodavatele:				
Dodavatel	Název instituce: Univerzita Hradec Králové, Pedagogická fakulta			
	Adresa: Rokitanského 62, 500 03 Hradec Králové, ČR			
	IČO: 62690094, DIČ: CZ62690094			
	Bankovní spojení: Česká spořitelna, a.s., číslo účtu: 2733582/0800			
	IBAN: CZ65 0800 0000 0000 0273 3582			
	BIC (SWIFT): GIBACZPX			
Datum:	Za odběratele:			
<p>Objednávku odešlete na adresu redakce:            Korpus – gramatika – axiologie, Univerzita Hradec Králové, Pedagogická fakulta,            KČJL, Rokitanského 62, 500 03 Hradec Králové, Česká republika,            nebo elektronickou formou na adresu: <a href="mailto:kga@ujc.cas.cz">kga@ujc.cas.cz</a>.</p> <p>Cena bude objednateli fakturována ve zvolené měně:            60 Kč/ks + poštovné a balné (vč. DPH), nebo            11 €/ks včetně poštovného, balného a bankovního poplatku (bez DPH), nebo            15 USD/ks včetně poštovného, balného a bankovního poplatku (bez DPH).</p>				

# Order form

Order form for <i>Korpus – gramatika – axiologie</i>				
Subscriber	Institution:			
	Delivery address:			
	Billing address:			
	IČO / DIČ / NIP / VAT number:			
	IBAN:			
	BIC (SWIFT):			
Ordering information	regular delivery of ..... copy/copies, starting with issue No. ....			
	single issue No. ...., number of copies .....			
	currency of payment:	Kč	€	USD
Additional information for distributor:				
Distributor	Institution: Univerzita Hradec Králové, Pedagogická fakulta			
	Address: Rokitanského 62, 500 03 Hradec Králové, Česká republika			
	IČO: 62690094, DIČ: CZ62690094			
	Bank: Česká spořitelna, a.s.; Account no.: 2733582/0800			
	IBAN: CZ65 0800 0000 0000 0273 3582			
	BIC (SWIFT): GIBACZPX			
Date:	Subscriber signature:			
<p>Order form should be mailed to:            Korpus – gramatika – axiologie, Univerzita Hradec Králové, Pedagogická fakulta,            KČJL, Rokitanského 62, 500 03 Hradec Králové, Česká republika,            or electronically to the following address: <a href="mailto:kga@ujc.cas.cz">kga@ujc.cas.cz</a>.</p> <p>The price on the invoice will be quoted in the currency selected by the subscriber:            60 Kč/copy + shipping and handling (incl. DPH), or            11 €/copy, including shipping, handling, and a bank transfer fee (without DPH), or            15 USD/copy, including shipping, handling, and a bank transfer fee (without DPH).</p>				