

Pražské přednášky Davida Gila

Jan Chromý

Ústav českého jazyka a teorie komunikace, FF UK, Praha
jan.chromy@ff.cuni.cz

Eva Lehečková

Ústav českého jazyka a teorie komunikace, FF UK, Praha
eva.leheckova@ff.cuni.cz

Ve dnech 18.–20. listopadu 2008 se na Filozofické fakultě Univerzity Karlovy v Praze pod záštitou Ústavu lingvistiky a ugrofinistiky uskutečnila série čtyř přednášek typologa Davida Gila pod souhrnným názvem *How different can languages be? Compositional semantics in typological perspective*.

David Gil je zaměstnán v lingvistickém oddělení Institutu Maxe Plancka pro evoluční antropologii v Lipsku, větší část roku však působí v terénním pracovišti tohoto institutu v indonéské Jakartě. Hlavním předmětem jeho zájmu je jazyková typologie (spolu s Bernardem Comriem, Martinem Haspelmathem a Matthew Dryerem je jedním ze spoluautorů známého Atlasu jazykových struktur světa¹). V posledních letech svou pozornost zaměřuje na běžně mluvenou indonéštinu na ostrově Riau.

Hlavním pilířem Gilova odborného díla však není pouhý popis exotických jazyků. Toto materiálové bádání se v Gilově vědecké perspektivě stává nutným předpokladem pro poznání jazyka jako takového, pro výstavbu adekvátní teorie jazyka. Určujícím principem je důsledná deeurocentrizace lingvistického popisu – Gil požaduje, aby teorie jazyka byla s to popsat i jazyky typologicky odlišné od jazyků indoevropských, zvláště když jimi, jako v případě jazyků austronéských, hovoří mnoho set milionů mluvčích.

Výše uvedený název série českých přednášek naznačuje, že jednou ze základních otázek, k nimž se David Gil vyslovuje, je problematika jazykových univerzálií. Gil se svými postoji dogmaticky nepřiklání ani k jednomu z vyostřených názorů (ať už jde o generativní lingvisty předpokládající existenci univerzálií nebo o křídlo typologů, kteří existenci jakýchkoliv univerzálií zpochybňují²), nýbrž volí „střední cestu“: Tvrdí, že rozdíly mezi jazyky jsou četnější a častější, než si obvykle jazykovědci myslí, ale že zároveň shody mezi jazyky existují, a tak je možné i vhodné o určitých jazykových univerzáliích uvažovat. Pro jejich popis Gil vytváří (na pozadí kritiky generativní gramatiky) vlastní elegantní aparát kategoriální gramatiky (Gil 2000).

Svůj cyklus pražských přednášek a ústřední otázku „How different can languages be“ David Gil otevřel kontrastem mezi naší osvojenou představou o struktuře jazyka a typickou indonéskou větou „Ayam makan“, jež obsahuje indonéská slova „ayam“ (kuře) a „makan“ (jíst). Ukázal, že při rozboru této věty není přesné užívat tradiční kategorie podmět a přísudek. Slova „ayam“ a „makan“ fungují jako obecné pojmy, věta může znamenat nejen „kuře jí“,

1 World atlas of linguistic structures, <http://wals.info/>

2 Srov. např. Dryer (1997) a Haspelmath (2007), pro které nekonečné variace ve struktuře jazyků světa vedou k hypotézám o neexistenci univerzálních mezijazykových (kroslingvistických) kategorií.

ale i „kuřata byla sněžena“, „důvod, proč kuře jí“, „někdo jí kvůli kuřeti“, „někdo jí spolu s kuřetem“ a mnohé další.

Věta „Ayam makan“ funguje podle Davida Gila na základě asociace, její význam je „entita asociovaná s kuřetem a jídlem“. Jako řešení potíží s analýzou daných jazykových struktur pomocí klasických syntaktických kategorií představil David Gil koncept „IMA language“ (izolační, monokategoriální a asociativní jazyk), tj. izolační jazyk, v němž je výhodné chápat všechny segmenty jako náležející k jediné kategorii a který disponuje pouze malou, uzavřenou skupinou morfologických markerů a v němž je význam struktury tvořen pomocí (konvenčních i situačních?) asociací. Gil ukázal, že takovým IMA jazykem jsou např. neverbální sémiotické systémy, jazyk malých dětí nebo komunikace u některých druhů opic. Gil konstatuje, že mezi současnými jazyky světa neexistuje žádný čistý IMA jazyk, ale lze nalézt tzv. relativní IMA jazyky, např. právě riauskou indonéštinu, které se čistému IMA jazyku blíží (například v nich lze nalézt řetězce gramatických struktur vykazujících vlastnosti IMA jazyka, vysoký stupeň sémantické asociativnosti, pouze početně omezenou skupinu funkčních slov – morfologických příznaků).³

V posledních dvou přednáškách Gil podrobně prezentoval svůj nejnovější výzkum, v němž se věnuje právě zkoumání míry asociativnosti v různých izolačních jazycích. V době přednášek tento výzkum ještě nebyl uzavřen, už průběžné poznatky však přinášely pozoruhodná zjištění. Z několika závěrů uveďme například to, že míra asociativnosti určitého jazyka byla statisticky vyšší u jazyka užívaného v menších komunitách (do 100 000 mluvčích), nebo to, že jazyk méně vzdělaných mluvčích vykazuje vyšší míru asociativnosti.

David Gil se české odborné veřejnosti představil coby poutavý přednášející a rádi konstatujeme, že se tak na všech přednáškách dalo za hojné účasti posluchačů, ať už z řad pedagogů nebo studentů, a že po přednáškách následovala plodná diskuse, motivovaná zájmem o problematiku, nikoli potřebou být na přednášce viděn.

Více informací o Davidu Gilovi se lze dozvědět na jeho osobních stránkách: <http://email.eva.mpg.de/~gil/>. Oba jeho zmíněné články jsou ke stažení na stránkách http://lingvistika.cz/?2008_gil.

Literatura:

DRYER, Matthew (1997): Are grammatical relations universal? In: Joan Bybee – John Haiman – Sandra A. Thompson (eds.), *Essays on Language Function and Language Type: Dedicated to T. Givón*, s. 115–143.

GIL, David (2000): Syntactic categories, cross-linguistic variation and universal grammar. In: Petra M. Vogel – Bernard Comrie (eds.), *Approaches to the Typology of Word Classes (Empirical Approaches to Language Typology)*, s. 173–216.

GIL, David (2008): How much grammar does it take to sail a boat? In: Andrew D. M. Smith – Kenny Smith – Ramon Ferrer-i-Cancho (eds.), *Proceedings of the 7th International Conference on the Evolution of Language*, s. 123–130.

HASPELMATH, Martin (2007): Pre-established categories don't exist. *Linguistic typology*, 11, s. 119–132.

3 Podrobněji viz Gil (2008).

Workshop „Žákovské korpusy a možnosti jejich využití ve výuce C2J“

Zuzanna Bedřichová

Ústav českého jazyka a teorie komunikace, FF UK, Praha

Z.Bedrichova@seznam.cz

Není samozřejmostí, aby lingvistické projekty měly za cíl bezprostřední využití, které by mohlo přinést konkrétní pomoc. A už vůbec není běžné, aby taková pomoc byla zaměřena na příslušníky menšin, které v současnosti v České republice žijí.

Místem, kde se o takových možnostech lingvistického výzkumu mluvílo, se stala Technická univerzita v Liberci. Na její půdě proběhl 9. září 2009 workshop „Žákovské korpusy a možnosti jejich využití ve výuce C2J“, uskutečněný v rámci projektu Inovace vzdělávání v oboru čeština jako druhý jazyk (projekt CZ.1.07/2.2.00/07.0259 je podpořen ESF a ze státního rozpočtu České republiky z OP Vzdělávání pro konkurenceschopnost). V rámci tohoto workshopu byly představeny žákovské korpusy jako prostředek inovací oboru čeština jako cizí jazyk, a zvláště projekt žákovského korpusu češtiny jako druhého jazyka, jehož cílem je shromáždit jak psané, tak mluvené jazykové projevy zejména vietnamských a ruský mluvících minorit žijících v České republice. Součástí tohoto korpusu je také co do rozsahu významný subkorpus jazykového materiálu romských uživatelů češtiny, u kterých se sice nejedná o češtinu jakožto druhý jazyk (předpokládá se, že romština není jejich prvním jazykem), ale kde by výzkum přinesl ucelenější soubor jazykového materiálu romského etnolektu češtiny.

Žákovský korpus češtiny jako druhého jazyka vzniká v rámci spolupráce Ústavu českého jazyka a teorie komunikace FF UK, Ústavu teoretické a počítačové lingvistiky FF UK a Katedry českého jazyka a literatury Fakulty přírodovědně-humanitní a pedagogické Technické univerzity v Liberci. Hlavním koordinátorem projektu je Karel Šebesta, řešitelský tým dále tvoří Vladimír Petkevič, Svatava Škodová, Barbora Štindlová, Kateřina Váňová, Václav Lábus, Kateřina Šormová a Zuzanna Bedřichová. Korpus je zajímavý již svým plánovaným rozsahem – cca 2 000 000 slov – i tím, že shromažďuje jak písemné projevy, tak projevy ústní (předpokládaný rozsah orálního korpusu je 600 000 slov). Oba typy dat jsou získávány položením; většinou se jedná o texty v rozsahu 500 slov, které vznikají v rámci běžné výuky (např. slohové práce, písemná cvičení apod.), či o nahrávky v časovém rozsahu 10–20 minut. Unikátní ve světovém měřítku je nejen fakt, že se jedná o žákovský korpus flektivního jazyka, ale i samotná koncepce anotace a snaha postihnout jazykové projevy od začátečnických úrovní (tj. od úrovně A1 podle SERRJ) až po pokročilé, a tak pokrýt celý rozsah osvojování češtiny jako druhého jazyka. Sami autoři projektu ovšem hovořili o tom, že anotace jazykových projevů začátečnické úrovně bude značně složitá.

Materiál je získáván ve spolupráci s nekomerčními jazykovými školami, pedagogy základních škol (často se jedná o školy zaměřené na výuku jinojazyčných žáků, tj. o školy s multietnickými třídami), praktickými a speciálními školami, různými volnočasovými organizacemi, romskými sdruženími, romskými asistenty, ale také s Ústavem jazykové a odborné přípravy UK.

Podstatnou částí práce na korpusu v rámci projektu je v současné době probíhající vytvoření chybové anotace žákovského korpusu, která bude analyzovat a označovat sebraný jazykový materiál z různých hledisek (morfologie, syntax, slovní zásoba a řada dalších).

Právě nad možnostmi anotace, využitím různých anotačních nástrojů a vůbec nad problémy spojenými s chybovou analýzou textů jinojazyčných mluvčích se v závěru workshopu po příspěvcích Vladimíra Petkeviče, Mileny Hnátkové, Tomáše Jelínka, Alexandra Rosena a Hany Skoumalové, kteří představili různé možnosti počítačových anotačních nástrojů, rozpoutala velice zajímavá a přínosná diskuze. Oproti řadě jiných korpusů má tento totiž velkou výhodu v tom, že se už od počátku jeho vzniku předpokládá konkrétní využití pro výuku češtiny jako druhého/cizího jazyka. Měl by např. pomoci osvětlit, jakým způsobem se utváří mezijazyk mluvčích zkoumaných skupin, co je pro nerodilého mluvčího češtiny problematické a co nikoli (v diskusi několikrát zaznělo, že rodilí mluvčí považují za obtížné často úplně jiné jevy než ti, kteří se češtinu učí jako cizí jazyk), bude možné jej využívat jako referenční zdroj při tvorbě výukových materiálů včetně testů, dále k vyhledávání častých chyb spojených s konkrétním jevem (např. problematické vazby s konkrétními slovesy apod.), může sloužit jako databáze příkladů apod. Tato jasně představitelná využití usnadňují koncepci anotace jak po uživatelské stránce (jak by mohl vypadat anotační nástroj, jak složité značkování volit apod.), tak po stránce teoretické (jaké chyby má smysl značit a jaké ne).

V případě subkorpusu romských mluvčích češtiny bude získaný jazykový materiál možné využít k tvorbě učebních materiálů pro romské žáky (včetně internetových aplikací využitelných např. v rámci doučování ve volnočasových organizacích i individuálně), k vytvoření nových koncepcí výuky češtiny ve třídách, v nichž je zastoupena romská minorita, či jako podklad pro školení pedagogů a budoucích pedagogů.

Velkým přínosem pro chystanou anotaci je propojení probíhajícího projektu s výukou na řešitelském pracovišti – Katedře českého jazyka a literatury Fakulty přírodovědně-humanitní a pedagogické TUL. Už v tomto akademickém roce se zde vyučují předměty čeština jako cizí jazyk 1, čeština jako cizí jazyk 2, úvod do počítačové lingvistiky, čeština v kontrastivním pohledu, čeština jako cizí jazyk – praktická cvičení a analýza učebnic, které mohou kvalitně připravit studenty-anotátory pro práci na korpusu. V rámci workshopu se hovořilo také o možnosti specializovaného semináře na této katedře, který by byl přímo zaměřen na školení anotátorů. Z diskuze jasně vyplývalo, že jsou si autoři projektu vědomi náročnosti chystané anotace, zejména v oblasti přepisu ústního projevu, při němž budou muset být anotátoři připraveni na specifické problémy vlastní té které jazykové menšině (především v oblasti výslovnosti, ale i syntaktické stavby apod.). Pro každou jazykovou menšinu je plánována supervize rodilého mluvčího daného jazyka, který bude kontrolovat přepis. Právě na problémy spojené s přepisem orálního materiálu upozorňovala Hana Goláňová, která má s touto problematikou zkušenosti z práce na projektu mluveného korpusu Schola a Databáze Pedagog, který na workshopu ve svém příspěvku představila a který bude v blízké době zpřístupněn veřejnosti.

Diskuze i příspěvky byly nesené kompetentností všech zúčastněných, a to jak hlavních autorů lingvistické koncepce projektu, kteří mají sami zkušenosti s výukou češtiny pro cizince, podílejí se na vytváření její metodiky (Karel Šebesta, Svatava Škodová, Barbora Štindlová a další), tak počítačových odborníků, kteří vyvíjejí adekvátní anotační nástroj. Bylo proto velmi zajímavé a pozitivní být svědkem toho, jak se z takového prostředí rodí nový projekt, a bude jistě neméně zajímavé sledovat i jeho průběh a konečná využití.