

KLÁRA OSOLSOBĚ

HYPOKORISTIKA V KORPUSU SOUKROMÉ KORESPONDENCE KSK

Cílem článku není pojednání o hypokoristikách jako slovtvorné kategorii ani jejich sledování v rámci specifík jazyka mládeže. V malé materiálové sondě chceme pouze předložit přehled domáckých rodných jmen, která se vyskytla v elektronickém textovém korpusu soukromé korespondence (KSK) a jež je tudíž možno oprávněně pokládat za hypokoristika zastoupená a používaná v současném jazyce. V centru našeho zájmu stojí především problematika jejich zařazení k flektivním vzorům českých substantiv v souvislosti s potřebami automatické lemmatizace elektronických korpusů.

1. Počítačově čitelné jazykové korpusy

Počítačové korpusy představují v posledním desetiletí významný zdroj objektivních dat pro různé typy analýz přirozeného jazyka. Korpus KSK (více o něm Hladká v tomto sborníku) vznikl na FF MU v rámci projektu GAČR č. 405/03/0248 *Současná soukromá korespondence. Vytvoření databáze a zpracování vybraných jevů z pohledu lexikologicko-lexikografického a dialektologického*. Část, jíž se budeme zabývat, zahrnuje 2 000 dopisů soukromé korespondence, a to především dopisy mládeže. V elektronické podobě se jedná o necelý milion slovních tvarů (942 489 pozic, z čehož je asi 767 000 slovních tvarů, z toho 76 504 různých). KSK rozsahem zdaleka neodpovídá obecným korpusům, vzhledem ke specifičnosti materiálu však může poskytnout relevantní data pro lingvistické výzkumy už v stávající velikosti. Příkladem může být právě oblast hypokoristik.

2. Morfologické značkování KSK

Součástí projektu je lemmatizace a gramatické označkování korpusu. Pro tento účel byl vybrán program pro automatickou lemmatizaci a morfologické značkování korpusů *Ajka*, který vznikl na FI MU (Sedláček, 2004).

2.1 *Ajka*

Ajka je morfologicky analyzátor používaný na FI MU a FF MU pro lemmatizaci a následné morfologické značkování korpusů psaného jazyka. Z lingvistického hlediska vychází z algoritmického popisu české formální morfologie (Osolsobě 1996). Na základě dat získaných z korpusu BMK (Brněnský mluvený korpus je paralelní projekt k PMK – Pražskému mluvenému korpusu, oba jsou součástí projektu ČNK) byla navržena varianta popisu české morfologie rozšířená o popis koncovek typických pro mluvený jazyk – (Hlaváčková 1998, 2002).

Podrobnější informace o morfologickém analyzátoru lze rovněž získat na <http://nlp.fi.muni.cz/projekty/ajka/ajkacz.htm>.

2.1.1 Modifikace *Ajky* pro značkování korpusu KSK

Při přípravě automatického značkování korpusu byl nejdříve použit neupravený morfologický analyzátor *Ajka*. Výsledkem analýzy byl seznam slovních tvarů, které program „neumí“ analyzovat. Z celkového počtu 76 504 různých slovních tvarů *Ajka* „neznala“ cca 15 000 různých slovních tvarů, tedy asi 20 %. Díky tomuto materiálu bylo možné získat přehled o některých charakteristických lexikálních, morfologických i pravopisných rysech, jimiž se soukromá korespondence (mládeže) reprezentovaná korpusem KSK liší od standardních psaných projevů. Na základě zjištěných faktů byl program *Ajka*, vytvořený pro značkování spisovných textů, modifikován tak, aby analyzoval i nejčastější nespisovné varianty jazyka reprezentovaného korpusem KSK (Hlaváčková, Sedláček, 2004). V současné době probíhá ruční desambiguace korpusu KSK automaticky označovaného upravenou verzí programu *Ajka*.

3. Vlastní jména a jejich automatické rozpoznávání – značkování

Morfologický analyzátor *Ajka* pracuje se slovníkem, který zahrnuje kmeny vlastních jmen v jejich neutrálních podobách a vzory, podle nichž se tvoří jejich tvary. Vzhledem k vysokému počtu nejrůznějších domáckých podob křestních jmen v korpusu KSK se zdá účelné zařadit jejich kmeny do slovníku kmenů a přiřadit je k existujícím flektivním vzorům, popřípadě vytvořit pro ně odpovídající nové vzory.

3.1 Zdrobnělé a domácké podoby vlastních jmen

Informace o domácích podobách rodných jmen jsou ve slovnících zpracovány nejednotně a poněkud nepřehledně. SSJČ zahrnuje hesla zachycující rodná jména a u některých odkazuje i k jejich domáckým podobám. SSČ uvádí samostatný seznam rodných a křestních jmen v jejich neutrálních (matrikových) podobách. Podrobněji se zabývá danou problematikou specializovaná literatura (např. Knappová, 1989). Materiál, představený v následujících odstavcích, přináší sondu do oblasti hypokoristik založenou na korpusových datech. Jak jsme již zmíni-

li výše, korpus KSK není příliš rozsáhlý. Nepokoušíme se tudíž o frekvenční analýzu, která by podle našeho mínění nemohla být nikterak reprezentativní. Nicméně se zdá, že shromážděná data mohou poskytnout podklad pro stanovení hypotéz ověřitelných na rozsáhlejších korpusech.

3.2 Hypokoristika v korpusu KSK

Vcelku nepřekvapí, že se v seznamu slovních tvarů nerozpoznaných nemodifikovaným programem *Ajka* objevilo poměrně velké množství hypokoristik. Z cca 15 000 různých nerozpoznaných tvarů jsme ručně vyřídili cca 800 různých tvarů hypokoristik, což představuje zhruba 5,7 %. Je zřejmé, že počet hypokoristik je v tomto případě podmíněn žánrem textů a pravděpodobně i věkovým složením pisatelů a adresátů dopisů.

V korpusu zcela převažují jména osob ženského pohlaví. Lze předpokládat, že je tomu tak proto, že něžný vztah k něžnému pohlaví se v jazyce projeví větším počtem hypokoristik ženských jmen a větší vynalézavostí při jejich tvoření. Důvod větší pestrosti a frekvence feminin může však být vzhledem ke způsobu získávání dat pro korpus KSK prozaičtější. Pisatelkami i adresátkami dopisů jsou totiž většinou ženy, v případě KSK z velké části studentky FF MU.

3.2.1 Přehled hypokoristik v korpusu KSK podle derivačních sufixů

Dříve než uvedeme jednotlivé skupiny, upozorníme na některé zvláštnosti podmíněné materiálem, z něhož vycházíme. Ve výčtu nalezených tvarů zachycujeme mj. případy krácení vokálů uvnitř kmene. Nelze však vždy zjistit, zda se jedná o skutečné krácení vokalické délky, anebo o vynechání diakritiky. Podobně uvádíme varianty *-i/-y(-í/-ý)*, u nichž také nelze spolehlivě dokázat, zda se jedná o záměr, či zda jde o pravopisnou chybu (překlep); totéž platí o variantách (překlepech) *s/z, d/d', n/ň, t/t' a e/ě* a o vynechaných písmenech. Krácení vokalické délky v sufixech uvádíme zvlášť, ačkoliv ani v těchto případech nelze spolehlivě dokázat, jde-li o záměr, či o pravopisnou chybu.

Materiál vytěžený z korpusu soukromé korespondence KSK vcelku potvrdil přehledové údaje o derivačních sufixech hypokoristik tvořených z křestních jmen, které uvádějí gramatiky (Čechová, 1996). Početně nejvíce zastoupena jsou deminutiva. Jedná se především o hypokoristika tvořená sufixem *-ka* a jeho variantami. Tento typ je bohatě reprezentován femininy. U maskulin mu četností nejvíce odpovídá typ se sufixem *-ek* s variantami a sufixem *-ík*. Ojedinělý je tvar *Haňátko*, formální neutrum svědčící o tom, že se využití napětí mezi přirozeným a gramatickým rodem při tvorbě hypokoristik netýká jen záměny femininum-maskulinum (*Evík, Léňas*) či případů, kdy podle nom. jména nelze spolehlivě určit jeho rod (*Kájulka, Peťulka, Peťuša, Pavluša*). Nicméně materiál korpusu přináší pouze jediný příklad formálního neutra tohoto typu. Z toho, že se v tomto případě jedná o modifikaci jména ženského, nelze vyvozovat žádné závěry (potencionálně můžeme vytvořit obdobné hypokoristikum i od jména mužského, např. *Jiřátko, Mišátko*).

Podívejme se nyní podrobněji na jednotlivé doložené příklady. Především, že u většiny nalezených tvarů nebyl problém určit neutrální tvar jména, vytvořit lemma a určit rod analyzovaného tvaru. V některých případech však bylo třeba nahlédnout do korpusového kontextu. U několika málo dokladů zůstalo vytvoření lemmatu a přiřazení ke vzoru sporné.

-ka(-ačka/-čka/-inečka/-uška/-ouška/-ulka/-ušenka/-inka/-enka/-inečka/-linka/-ulinka/-unka/-ušenka/-ičečinečička/-ulililinka)

-feminina

Adelka, Agátka, Alenka, Alička, Alinka, Alžbětka, Andrejka, Aňka, Barulinka, Belinka, Bertička, Betka, Bobinka, Broňka, Dagmarka, Dájinka, Dájka, Darinka, Deniska, Ditulka, Domika, Domka, Donika, Dorinka, Endička, Evěnka, Evulka, Gábuška, Gabrielka, Gabruška, Ilonka, Irénka, Irunka, Iruška, Ivanka, Iváňuška, Ivka, Ivetka, Ivonka, Ivunka, Ivuška, Jájinka, Janka, Janinečka, Jaňulililinka, Jaňulka, Januška, Jěňulka, Jiřulka, Jituličečinečička, Jitušenka, Jojka, Kájinka, Kájulka, Kájůška, Karolinka, Katulka, Klárinka, Klárka, Klaudinka, Ladka, Lenilinka, Lenuška, Lidečka, Lidka, Lindička, Lindinka, Lucička, Lucinečka, Lucinka, Luisička, Majduška, Mariánka, Marka, Martínka, Marťulka, Marušenka, Miřulka, Mišička, Miška, Moňulka, Naďulka, Nastěnka, Nelinka, Nikolka, Nikouška, Olínka, Pajunka, Pavluška, Peťulka, Peťka, Silvička, Silvinečka, Raďulka, Raduška, Renatka, Renátka, Renka, Romanka, Silvička, Silvinečka, Simonka, Staňka, Táňuška, Tereska, Terezinka, Teruška, Terryka, Terynka, Tutinka, Verka, Věrka, Verunečka, Věruška, Věrulinka, Věrunka, Věruška, Vladěnka, Vladislávka, Zdeňička, Zdenulka, Zitka, Zituška

-maskulina

Iváňuška, Jěňulka, Jirka, Joška, Lubka, Mišička, Miška, Miřulka, Peťka, Peťulka, Ráďulka, Renka, Saška, Staňka, Vítulka, Zdendulka

-átka

-formální neutrum (ženské jméno)

Haňátka

-ák/-ek/-eček/-iček/-ineček/-enek/-ínek/-išek/-ešek/-ošek/-ásek

-maskulina

Alešek, Benáček, Beníček, Bobišek, Daneček, Danek, Davidek, Denošek, Fanošek, Fredeček, Honzíček, Iváňošek, Irísek, Ivošek, Janek, Janíček, Jencek, Jeníček, Jindříšek, Jindřišek, Jiřánek, Jířek, Josefek, Joženek, Kájíček, Ladásek, Láďánek, Licek, Lubineček, Marešek, Matěk, Mišánek, Mišák, Mišutka, Nikoušek, Olínek, Ološek, Ondrášek, Peťánek, Peťulínek, Radeček, Radimek, Reneček, Renek, Tedánek, Víteček, Vitek, Vladanek, Vladánek, Zdeneček, Zděnek

-formální maskulina (ženská jména)

Ivinek, Marťasek, Peťulínek, Zuzíček

-ík/-ik**-maskulina**

Bobík, Bertík, Daník, Gustík, Honzík, Janík, Kájík, Lexík, Lubík, Mišík, Ondřík, Staník, Vojtík, Zdeník,

Božík, Frantik, Honzik, Janik, Ladik, Ládik, Maxik, Pepik, Pétik, Ondřik, Vlastik

-formální maskulina (ženská jména)

Elík, Evík, Janík, Lucík

Evik, Janik

Druhou velkou skupinou jsou hypokoristika využívající formant *-a*, jež narmnoze provází krácení základní podoby jména (*Blažena – Blaža*), popř. jeho hláskové obměny (*Karel – Kája*). V některých případech dochází rovněž ke konsonantické alternaci souhlásky před formantem *-a*. Tato změna však není, jak by se na první pohled mohlo zdát, vázaná na některé skupiny konsonantů (*-d*, *-t*, *-n*). Z analyzovaného materiálu vyplývá, že k němu nedochází v případě, když jméno není pocíťováno jako plně domácí, např. *Tína, Týna (Kristína, Kristýna), Gina (Regina/Regína)*.

I v této skupině, podobně jako u hypokoristik popsaných výše, je charakteristické užívání formantů s nenulovou koncovkou jak pro ženská, tak pro mužská jména (první se pak skloňují podle vzoru *žena/Bláža*, druhá podle *předseda/bača*). Problematické je ovšem při nedostatku dokladů zjistit, o který z přirozených rodů se jedná (*Kája, Pavča, Peťka*, ...), protože tuto informaci nelze z kontextu vždy stoprocentně určit (disambiguovat). V případech, že ji zjistit nelze, řadíme k oběma rodům. Lemmatizace a přiřazení ke vzorům nečinilo v této poměrně bohatě tvarově zastoupené skupině žádné problémy.

-a/-‘a**-feminina**

Ada, Alína, Baja, Bája, Bára, Bára, Blaža, Daja, Dáša, Gina, Hája, Jéna, Jiřa, Jíta, Kája, Lára, Lúca, Máca, Madla, Mára, Máša, Miša, Míša, Mýša, Pája, Saša, Síma, Tína, Týna, Zúza

Aďa, Aňa, Berťa, Bláňa, Bobaňa, Háňa, Jáňa, Jíťa, Leňa, Léňa, Moňa, Marťa, Páťa, Ráďa, Reňa, Taňa, Zdeňa

-maskulina

Daja, Dava, Deja, Fana, Feja, Honza, Jéna, Kája, Mara, Pája,

Miša, Míša, Mýša, Pavluša, Rýša, Saša, Váša, Zbyša,

Kost’a, Laďa, Luďa, Lúďa, Maťa, Old’a, Ond’a, Peťa, Péťa, Ráďa, Rost’a, Sáňa, Staňa, Viťa, Vít’a, Vlad’a, Vlast’a, Vojř’a, Voloďa, Vraťa, Zbyňa, Zdeňa, Zdeňa

(Do této skupiny jsme zařadili jména zakončená na *-ďa*, která by mohla být zařazena též níže. Učinili jsme tak z toho důvodu, že v uvedených případech kmen neutrální podoby jména obsahuje *-d* - srv. *Radka/Radek – Ráďa, Ladislav – Laďa, Luděk – L(u/ů)ďa, Oldřich – Old’a, Ondřej – Ond’a*.)

Také další skupina nepřináší z hlediska přiřazení k flektivním vzorům žádné problémy.

-ča

-feminina

Alča, Berča, Danča, Darča, Domča, Elča, Evča, Gabča, Ilča, Irča, Ivča, Kača, Kamča, Lidča, Lucča, Marča, Monča, Nelča, Olča, Pavča, Renča, Romča, Simča, Verča, Zdenča

-maskulina

Lubča, Pavča, Zdenča

Varianta *-ula* byla stejně jako varianta *-ule* zastoupena vok. tvary (*Peřulo, Jaňulo, Leňulo, ...; Barule, ...*). Tvary s vok. zakončeným *-ulo* se tvoří podle vzoru *žena*, lemma, je pak např. *Jaňula*, tvary na *-ule* se tvoří podle vzoru *růže*, lemma je pak *Jaňule*. Mužské tvary se skloňují podle vzoru *předseda*.

-ula/(-ule)

-feminina

Jaňula, Leňula, Peřula, Barule, Evule, Haňule, Jaňule, Jitule, Katule, Kářule, Martule, Nařule, Peřule

-maskulina

Fandula, Honzula, Vojřula

Pro následující skupinu jsme v korpusu našli také převážně tvary vokativu. U feminin není tedy zcela jednoduché vytvořit odpovídající lemma. Končí-li nom. na *-a*, popř. vok. na *-o*, lemma vytvoříme podle vzoru *žena* (s klasicky popsanou variantou *Bláža*), končí-li nom., ak. a vok. na *-e*, vytvoříme lemma podle vzoru *růže*. Má-li ovšem tvar nom., ak., i vok. (jiné pády nejsou doloženy) konc. *-o*, tvar lematizujeme jako nesklonné substantivum. U maskulin řadíme bez problémů odpovídajícím způsobem ke vzorům *muž, předseda (bača)*.

-uše/-uša/-áš/-oš/-uš/-iš/-ouš

-feminina

Evuše, Ivuše, Katuše, Linduše, Petruše

Kláruša, Petruša

Daďuš, Danuš, Fanuš, Ivuš, Jaruš, Jituš, Kristuš, Madluš, Maruš, Miluš, Miruš, Petruš, Veruš, Věruš

Blaniš, Leniš

Marcoš

-maskulina

Ondruša, Peřuša, Pavluša

Ondráš

Ivoš, Jenoš, Luboš, Marcoš, Tedoš
Fanuš, Ituš, Petruš,
Bohouš, Jarouš

Následující poměrně málo zastoupené skupiny nepředstavují pro lemmatizaci žádný problém.

-ina/-ena

-feminina

Gabina, Gábrina, Lucina, Sašena

-maskulina

Palina

-an/-án/-in/-ín/-yn/-ino/-íno

-maskulina

Jiřan, Ríšan, Rýšan, Vítan

Bobin, Honzin, Honzyn, Jarin, Jožin, Ladín, Lukin, Robin

Palino, Pepíno

-formální maskulina (ženská jména)

Lucián

-ďa/-da

-feminina

Bohďa, Handďa, Janďa, Venďa

Alda, Dada, Dáda, Denda, Majda

-maskulina

Bohďa, Rendďa, Venďa, Zbyňďa, Zbyňďa

Fanda, Jenda, Milda, Renda, Tanda, Zbyňďa

-íta

-feminina

Anita, Evita

- áč/-oč

-maskulina

Honziňáč, Luboč

-ál

-maskulina

Zdenál

-our

-formální maskulina (ženská jména)

Lucour

Následující skupinku tvoří maskulina skloňující se podle vzoru *muž*. Výjimku tvoří formální maskulina *Léñas, Marías*, u nichž jsme kromě pravidelně tvořeného vokativu na *-i* (*Martási*) našli i doklad vokativ/nominativ (*Léñas*). V ostatních pádech se patrně skloňují podle vzoru *muž*.

-os/-us/-as

-maskulina

Lukas, Pavlos, Peás, Zbynkus

-formální maskulina (ženská jména)

Léñas, Marías

Poslední typ je doložen jen jedním tvarem, je tedy těžké říci něco opodstatněného o jeho flexi (zda se skloňuje podle vzoru *muž*, zda se vypouští vkladné *-e*-*Vlastece* x *Vlastce*, zda jde o nesklonné substantivum).

-ec

-maskulina (formální?)

Vlastec

3.3 Morfologické zvláštnosti

V seznamu domácích podob křestních jmen se vyskytly i tvary, u nichž bylo obtížné jednoznačně vytvořit jejich základní tvar a zároveň množinu tvarů od potencionálního lemmatu odvozených (tedy vlastně flektivní paradigma, které by mohlo být vzorem).

Na základě zkoumaného materiálu jsme vyčlenili následující skupiny:

Blan-i/Mon-y

-feminina

Blani, Domi, Hani, Heli, Iri, Jani, Jiři, Kači, Kristi, Leni, Leňuli, Lindi, Luci, Madli, Magdi, Mari, Marti, Miši, Moni, Pavli, Romi, Silvi, Simi, Soňi, Stani, Susi, Viki, Zuzi,

Fanyny, Mony, Tery

-maskulina

Domi, Honzi, Jiři, Káji, Lubi, Mili, Miši, Oli, Pali, Radi, Zdeni, Luky, Rendy

Iv-í

-feminina

Alí, Iví

-maskulina

Daví, Iví, Oldí, Pavlí, Tomí, Vojtí

Ver-u

-feminina

Haňuňu, Veru

V naprosté většině uvedených případů se jednalo o vok., nicméně se objevil i nom. (*Alí, Fanyny, Jani, Jiři, Luci, Susi, Viki, Zuzi, Daví, Luky, Oli, Oldí, Pali, Rendy, Tomí, Vojtí*), gen. (*Mišiho*), dat. (*Lukymu*), lok. (po *Lukym*) a instr. (s *Jani, Lukym*). Při začlenění k ohýbacím vzorům jsme narazili na určité nesnáze. Přijali jsme hypotézu, že se jedná o analogii k hypokoristickým apelativům typu *mami/í, babi/í, teti/í, tati/í*. Na základě porovnání slovníku a ČNK jsme zjistili, že odpovídající lemmata ke tvarům nalezeným v KSK lze tvořit dvěma způsoby:

1. tvar=„mami“ – lemma=„mami“ – nesklonné femininum (ženská jména), tvar=„tati“ – lemma=„tati“, gen. tati-ho (mužská jména). Tak uvádí SSJČ, ale pouze u hesel *mami, tati/í, babi, teti*, případy *mami, babi, tetí* neuvádí vůbec. SSČ uvádí pouze hesla *mami, tati* se stejnými charakteristikami jako SSJČ. Morfologický analyzátor *Ajka* řadí všechny jmenované případy jako nesklonná substantiva, tomu však odporují doklady *Mišiho, Lukymu, Lukym*. Řešení 1. (podle SSJČ/SSČ) se tedy zdá být adekvátní, a to i s ohledem na možnost obdobně zacházet se zaznamenanými tvary domácích podob cizích křestních jmen (*Dejvy, Jerry, Jimmí/y, Johnny, Miki/y*). Jejich popis ovšem zůstává mimo rámec naší studie.

Sondou do korpusu SYN2000 jsme zjišťovali lemmatizaci i značkování slovních tvarů typu *mami/tati*, která se tu řeší poněkud nejednotně. Tvary *tati, mami* se řadí k lemmatu *táta, máma* a je jim přiřazena odpovídající značka. Tvarům *teti, tatí, mamí i babi* se přiřazuje lemma odpovídající tvaru samotnému, první pozice značky tvaru *babi* je ale rozdílná od značky u tvarů *teti, tatí, mamí*. U všech čtyř tvarů se shoduje druhá pozice, která nabývá hodnoty X, tj. „slovní tvar, který byl rozpoznán, ale značka ve slovníku chybí“, ostatní pozice jsou ponechány nerozhodnuty (srv. <http://ucnk.ff.cuni.cz/bonito/>). Stejnou značku jako tvary *teti, tatí, mamí* má i tvar *tatim*, a tvar *tatím*, kde se ovšem zřetelně jedná o překlep (*zatím*). Tvar *babí* se vyskytuje pouze jako měkké adjektivum, takže se o něm jakožto o tvaru substantiva z korpusu nelze nic dozvědět.

S ohledem na data nalezená v korpusu SYN2000, jsme tedy zkoumali druhou variantu:

2. tvar=„mami“ – lemma=„máma“ (ženská jména), tvar=„tati“ – lemma=„táta“ (mužská jména). (Uvádíme v pořadí tvar – potencionální lemma. V závorce a s otazníkem uvádíme potencionální lemma, které je v rozporu s tvary nalezenými v KSK.)

Alí – (*Ala?*), *Blani* – *Bla(n/ň)a*, *Domi* – *Doma*, *Fanyny* – (*Fanyna?*), *Hani* – *Ha(n/ň)a*, *Haňuňu* – *Haňu(n/ň)a*, *Heli* – *Hela*, *Iri* – *Ira*, *Ivi* – *Iva*, *Jani* – (*Ja(n/ň)a?*), *Jiři* – (*Jiřa?*), *Kači* – *Kača*, *Kristi* – *Krista*, *Leni* – *Le(n/ň)a*, *Leňuli* – *Leňula*, *Lindi* – *Lin(d/ď)a*, *Luci* – (*Luca?*), *Magdi* – *Mag(d/ď)a*, *Madli* – *Madla*, *Mari* – *Mara*, *Marti* – *Mar(t/ť)a*, *Miši* – *Miša*, *Mony* – *Mo(n/ň)a*, *Pavli* – *Pavla*, *Romi* – *Roma*, *Silvi* – *Silva*, *Simi* – *Sima*, *Soni* – *Soňa*, *Stani* – *Sta(n/ň)a*, *Susi* – (*Susa?*), *Tery* – *Tera*, *Veru* – *Vera*, *Viki* – (*Vika?*), *Zuzi* – (*Zuza?*)

Daví – (*Dava?*), *Domi* – *Doma*, *Honzi* – *Honza*, *Iví* – *Ivo!*, *Jiři* – *Jiřa*, *Káji* – *Kája*, *Lubi* – *Luba*, *Luky* – (*Luka?*), *Mili* – *Mila*, *Miši* – (*Miša?*), *Oli* – (*Ola?*), *Oldí* – (*Ol(d/ď)a?*), *Pali* – (*Pala?*), *Radi* – *Ra(d/ď)a*, *Rendy* – (*Ren(d/ď)a?*), *Tomí* – (*Toma?*), *Vojtí* – (*Voj(t/ť)a?*), *Zdeni* – *Zde(n/ň)a*.

Kromě rozporů, které přináší tvary *Mišiho*, *Lukymu*, *Lukym* a další (nom. na *i/í/y/u*), je zřejmé, že v některých případech dostáváme samostatný příznakový tvar k neutrálnímu lemmatu (*Hana*, *Jana*, *Linda*, *Marta*, *Pavla*, *Silva*, ...). To ovšem není metodicky čisté, protože nejde o obdobu substandardní (nespisovné, popř. nářeční) varianty koncovek u apelativ, nýbrž o variantu slovotvornou (hypokoristikum). Přiklonili jsme se tudíž k řešení 1.

Další morfologickou zvláštnost představují jednotlivé substandardní (nespisovné, nářeční) koncovky kombinované jednak s neutrálními podobami kmenů (*Lukáša*, *Mikuláša*, *Tomáša*, *Ingridu*, ...), jednak s kmeny hypokoristik (*Honzoj*, *Jenošu*, *Jendem*, *Jirkem*, *Jirkoj*, *Laďem*, *Péťoj*). Zmiňujeme se o nich pouze pro doplnění, protože se jedná o problémy řešené komplexně při úpravě té části analyzátoru *Ajka*, která obsahuje množiny koncovek a zachycuje vztahy mezi spisovnými koncovkami a jejich nespisovnými variantami (srov. Hlaváčková, 2004).

3.4 Základní podoby vlastních jmen a jejich výskyt v korpusu KSK

Na základě hypokoristik nalezených v korpusu KSK jsme vytvořili seznam vlastních jmen, od nichž jsou potencionálně příslušné domácí formy odvozeny. Jak jsme již poznamenali výše, nebylo je vždy možno s jistotou určit, nicméně u naprosté většiny případů jsme se pokusili nalézt přijatelné řešení.

Seznam neutrálních podob jmen, k nimž jsme našli domácí podoby, zahrnuje ženská jména:

Adéla, *Agáta*, (*Alberta*?), *Alena*, *Alexandra*, *Alice*, *Alžběta*, *Andrea*, *Anna*, *Barbora*, *Běla*, *Blanka*, *Blažena*, *Bohumila*, *Bohuslava*, *Bronislava*, *Daniela*, *Dagmar*, *Darja*, *Denisa*, *Dita*, *Domínika*, *Dora*, *Ela*, *Eva*, *Gabriela*, *Hana*, *Ilona*, *Irena*, *Iva*, *Iveta*, *Jana*, *Jiřina*, *Jitka*, *Kamila*, *Karla*, *Karol(i/í)na*, *Kateřina*, *Klára*, *Klaudie*, *Krist(i/ý)na*, *Ladislava*, *Lenka*, *Lucie*, *Ludmila*, *Luisa*, *Marie*, *Marta*, *Martina*, *Michaela*, *Monika*, *Naďa*, *Nataša*, *Nela*, *Nikola*, *Olga*, *Patricie*, *Pavla*, *Pavčina*, *Petra*, *Radka*, *Reg(i/í)na*, *Renata*, *Romana*, *Simona*, *S(i/y)lva*, *Táňa*, *Tereza*, *Veronika*, *Věra*, *Vladislava*, *Vlasta*, *Zdena*, *Zita*, *Zuzana*
a mužská jména:

Augustin, *Albert*, *Alexandr*, *Aleš*, *Benedikt*, *Bohumil*, *Bohuslav*, *Daniel*, *David*, *Eduard*, *Fred*, *František*, *Gustav*, *Ivo*, *Jan*, *Jaroslav*, *Jindřich* *Jiří*, *Josef*, *Karel*, *Ladislav*, *Lubor*, *Luděk*, *Lukáš*, *Marek*, *Mari(á/a)n*, *Matěj*, *Max*, *Maxmilián*, *Micha(e)l*, *Mikuláš*, *Miloš*, *Nikola*, *Oldřich*, *Ondřej*, *Patrik*, *Pavel*, *Petr*, *Radek*, *Radim*, *René*, *Richard*, *Robert*, *Rostislav*, *Stanislav*, *Tomáš*, *Vít*, *Vladislav*, *Vladimír*, *Vlastimil*, *Vojtěch*, *Vratislav*, *Zbyněk*, *Zden(ě/e)k*

Z uvedených neutrálních podob nebyla v korpusu žádným tvarem doložena ženská jména *Agáta*, *Blažena*, *Bohumila*, *Bronislava*, *Darja*, *Karla*, *Ladislava*, *Ludmila*, *Vladislava* a mužská jména *Benedikt*, *Bohumil*, *Eduard*, *Matěj*, *Miloš*, *René*, *Rostislav*, *Vladislav* (jméno *Augustin* se vyskytlo pouze v kontextu sv. Augustin, tvar jména *Albert/Alberta* ve významu obchodní dům, *Vratislav* jako jméno místní, a tedy femininum). Všechny ostatní neutrální podoby křestních jmen se v korpusu KSK vyskytly alespoň jednou.

Pro zajímavost uvedeme přehled výskytu slovních tvarů jména *Jan/Jana* a jejich hypokoristik

(v pořadí tvar jména – počet výskytů podle nalezených kmenů *Jan-/J(e/é)n-/Honz-*): Jana -162, Jani - 47, Janou - 45, Jano - 44, Janu - 28, Janě - 26, Janičko - 22, Jan - 21, Jany - 17, Janička - 16, Janča - 15, Janinko - 7, Janičce - 5, Janina - 5, Jaňulko - 5, Jaňulo - 4, Jančo - 4, Janičku - 3, Janičkou - 3, Jaňule - 3, Jaňuli - 2, Janik - 2, Jand'ou - 2, Janka - 2, Janči - 2, Janinka - 2, Januška - 1, Janíku - 1, Janinou - 1, Janička - 1, Janče - 1, Janušky - 1, Janikem - 1, Janek - 1, Jand'a - 1, Jančou - 1, Janovi - 1, Janču - 1, Janinečko - 1, Janičku - 1, Janík - 1, Jankou - 1, Janulkou - 1, Jaňďu -1, Jaňulilinka - 1, Jaňulky - 1, Jaňulka - 1, Jaňuly - 1, Jaňula - 1, Jenda - 6, Jendo - 2, Jencek - 2, Jendem - 1, Jendovi - 1, Jendou - 1, Jenóšu - 1, Jeničku - 1, Jenička - 1, Jeníček - 1, Jéna - 1, Jéňulka - 1, Honza - 99, Honzou - 33, Honzovi - 23, Honzík - 17, Honzu - 17, Honzy - 8, Honzíkem - 7, Honzíku - 4, Honzíka - 4, Honzo - 4, Honzíček - 2, Honzíčka - 2, Honzulo - 1, Honzoi - 1, Honzíkovi - 1, Honzi - 1, Honzynem - 1, Honzíčkovi - 1, Honziňáč - 1, Honzik - 1, Honzin - 1.

Neutrálních podob bylo nalezeno 344 od 2 lemmat (*Jan/Jana*), tvarů nejrůznějších hypokoristik 423 od 31 lemmat.

4. Slovtvorné vzory pro derivaci hypokoristik

Nástroje vyvinuté v Laboratoři zpracování přirozeného jazyka FI MU, morfologický analyzátor *Ajka* a program *I_par* (Veber, 2003) pro tvorbu a údržbu morfologické databáze umožňují efektivní využití materiálu popsaného v odd. 3. Propojením neutrálních lemmat a lemmat odpovídajících hypokoristik lze snadno vytvořit jádro slovníkové databáze pro integrovaný popis formální a derivační morfologie vlastních jmen (existuje již část pro příjmení) na základě empirických dat, kterou bude možno nadále rozšiřovat a udržovat.

5. Závěr

Rodná jména a odpovídající hypokoristika tvoří nepřehlédnutelnou složku korpusu KSK. Tento podíl je dán žánrovou specifičností korpusu, který se tak stává empirickým zdrojem dat pro lingvistický výzkum domácích podob křestních jmen. Zařazení kmenů do slovníku *Ajka* je prvním krokem pro jednotné automatické označování a následnou desambiguaci tvarů rodných jmen v jejich různých variantách. Ve chvíli, kdy bude k dispozici lemmatizovaný, označovaný desambiguovaný korpus, bude možné vyhodnotit poměr neutrálních apelativ a odvozených domácích forem v nejrůznějších vzájemných vztazích.

LITERATURA

- ČECHOVÁ, M.: *Čeština – řeč a jazyk*, 1996.
- HLAVÁČKOVÁ, D.: *Korpus mluvené češtiny*, dipl. práce FF MU, (rkp.), Brno 1998.
- HLAVÁČKOVÁ, D.: *Brněnský mluvený korpus a jeho morfologická analýza*. In: 3. mezinárodní setkání mladých lingvistů, Olomouc. 2002, s. 167 – 173.
- HLAVÁČKOVÁ, D., SEDLÁČEK, R.: *Morfologické značkování korpusu soukromé korespondence*, XIV. kolokvium mladých jazykovedců, 8. – 10. 12. 2004, Šintava pri Seredi, Slovenská republika (v tisku).
- KARLÍK, P., NEKULA, M., RUSÍNOVÁ, Z.: *Přůruční mluvnice češtiny*, 1996.
- KNAPPOVÁ, M.: *Rodné jméno v jazyce a společnosti*, 1989.
- OSOLSOBĚ, K.: *Algoritmický popis české formální morfologie a strojový slovník češtiny*, disert. práce, (rkp.), 1996.
- PALA, K., SEDLÁČEK, R., VEBER, M.: *Vztah mezi tvarotvornými a slovotvornými vzory v češtině*. In: Čeština – univerzália a specifika. Praha : Nakladatelství Lidové noviny, 2004, s. 151–162.
- PALA, K., VEBER, M.: *CED – A Program for Editing Corpora*. In *Proceedings of Telri Seminar 2002*. 1. vyd. Birmingham, 2002.
- SEDLÁČEK, R.: *Morfologický analyzátor češtiny*, dipl. práce FI MU, (rkp.), Brno 1999.
- SEDLÁČEK, R.: *Morphematic analyser for Czech*, disert. práce FI MU, (rkp.), Brno 2004.
- VEBER, M., SEDLÁČEK, R., PALA, K., OSOLSOBĚ, K.: *A Procedure for Word Derivational Processes Concerning Lexicon Extension in Highly Inflected Languages*. In *Proceedings of the Third International Conference on Language Resources and Evaluation, LREC*. Las Palmas de Gran Canaria: ELRA, 2002. s. 1254–1259.
- VEBER, M.: *Nástroje pro textové korpusy a morfologické databáze*, disert. práce FI MU, (rkp.), Brno 2003.
- <http://nlp.fi.muni.cz/projekty/ajka/ajkacz.htm>.
- <http://ucnk.ff.cui.cz/bonito/>

HYPOCORISM IN THE CORPUS OF PRIVATE CORRESPONDENCE (KSK)

The goal of this paper does not consist in describing a derivation of hypocorism concerning the language of the young generation in general. We report on an exploration of the first names and their homely (family) forms extracted from the KSK, i.e. the corpus of private correspondence. We set it as our task to create a complete list of the hypocorism occurring in the KSK and to associate them with their respective inflectional paradigms. This description was integrated into the new version of AJKA (morphological analyser and tagger of Czech) with respect to the need of automatic morphological tagging of the corpus mentioned above.

Klára Osolsobě
Ústav českého jazyka FF MU
Arna Nováka 1
602 00 Brno
e-mail: osolsobe@phil.muni.cz