

# Uživatelský manuál k morfologické anotaci korpusů CoCzeFLA

JOLANA TREICHELOVÁ, KLÁRA MATIASOVITSOVÁ, ANNA CHROMÁ,  
JAKUB SLÁMA, PETRA ČECHOVÁ & FILIP SMOLÍK

# Obsah

<b>1. VÝCHODISKA MORFOLOGICKÉ ANOTACE KORPUSŮ COCZEFLA .....</b>	<b>2</b>
1.1. MORFOLOGICKÁ ANOTACE A CHAT.....	11
1.1.1. <i>Zapojení kódů specifických pro formát CHAT do morfologické anotace .....</i>	<i>10</i>
1.2. MANUÁLNÍ KONTROLA .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
<b>2. OBECNÉ ZÁSADY URČOVÁNÍ LEMMATU, SLOVNÍCH DRUHŮ A MORFOLOGICKÝCH KATEGORIÍ . 3</b>	<b>3</b>
2.1. STRUKTURA AUTOMATICKÉ MORFOLOGICKÉ ANOTACE A LEMMATIZACE .....	3
2.2. SLOVNÍ DRUHY A JEJICH MORFOLOGICKÉ KATEGORIE .....	4
2.2.1. <i>Substantiva.....</i>	<i>4</i>
2.2.2. <i>Adjektiva.....</i>	<i>4</i>
2.2.3. <i>Zájmena.....</i>	<i>5</i>
2.2.4. <i>Číslovky.....</i>	<i>6</i>
2.2.5. <i>Slovesa .....</i>	<i>6</i>
2.2.6. <i>Neohebné slovní druhy.....</i>	<i>8</i>
2.3. PŘÍPADY, KDY NELZE URČIT SLOVNÍ DRUH/LEMMA/MORFOLOGICKÉ KATEGORIE.....	8
2.3.1. <i>Slovní druh a/nebo lemma nelze určit.....</i>	<i>Error! Bookmark not defined.</i>
2.3.2. <i>Slovní druh a/nebo lemma je třeba doplnit.....</i>	<i>Error! Bookmark not defined.</i>
<b>3 SPECIFICKÉ JEVY SPOJENÉ S DĚTSKÝM JAZYKEM/KORPUSEM .....</b>	<b>9</b>
3.1. ZAPOJENÍ KÓDŮ SPECIFICKÝCH PRO FORMÁT CHAT DO MORFOLOGICKÉ ANOTACE .....	<b>ERROR! BOOKMARK NOT DEFINED.</b>
3.2. ZNAČENÍ MORFOLOGICKÝCH CHYB .....	11
3.3. ZNAČENÍ NEOLOGISMŮ .....	11
3.4. ZNAČENÍ CIZÍCH SLOV.....	12

# 1. Východiska morfologické anotace korpusů CoCzeFLA<sup>1</sup>

V rámci morfologické anotace korpusů CoCzeFLA (Corpora of Czech as the First Language in Acquisition, <https://coczefla.ff.cuni.cz/>), přepisovaných podle **systému CHAT** (standardní formát přepisu mluvené komunikace vyvinutý speciálně pro potřeby přepisu dětské řeči v rámci databáze CHILDES<sup>2</sup>; <https://talkbank.org/manuals/CHAT.pdf>), je ke všem proneseným slovům přiřazeno lemma, slovní druh a hodnoty morfologických kategorií, které dané slovo vyjadřuje.

Pro toto morfologické značkování užíváme **nástroj** pro automatickou morfologickou anotaci **MorphoDiTa** (Morphological Dictionary and Tagger, <https://ufal.mff.cuni.cz/morphodita>)<sup>3</sup>. Během anotování používáme **značení MOR** (<https://talkbank.org/manuals/MOR.pdf>), které vychází z pravidel pro morfologickou anotaci dětských korpusů CHILDES. Pro převod ze značení MorphoDiTou na MOR formát, včetně některých úprav spojených se zásadami níže, využíváme skript vytvořený v Pythonu.

Výsledky automatické morfologické anotace **dětských řádků** (nikoli řádků dospělých mluvčích) následně manuálně kontrolujeme, a to následujícím způsobem. V případech, kdy automatická morfologická anotace lemma, slovní druh (a podkategorii) či hodnotu morfologické kategorie neurčí, doplňujeme chybějící informaci. V rámci ohebných slovních druhů potom především opravujeme MorphoDiTou nesprávně vyhodnocené pádové formy a jmenné rody, rozlišujeme tázací a vztažná zájmena a doplňujeme hodnoty slovesného vidu, případnou modalitu slovesa a dále povahu slovesa *být* (pomocné/sponové/plnovýznamové). Naopak příliš nezasahujeme v případech neohebných slovních druhů, u kterých převážně ponecháváme výsledek automatické anotace. Realizované manuální zásahy jsou dále podrobněji popsány v kapitole 3. Slovní druhy a jejich morfologické kategorie.

---

<sup>1</sup> Vznik morfologické anotace a tohoto Uživatelského manuálu byl podpořen programem *Grantová schémata na UK*, reg. č. CZ.02.2.69/0.0/0.0/19\_073/0016935.

<sup>2</sup> MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Lawrence Erlbaum Associates.

<sup>3</sup> Spoustová, D., Hajič, J., Raab, J., & Spousta, M. (2009). *Semi-Supervised Training for the Averaged Perceptron POS Tagger*. In A. Lascarides, C. Gardent, & J. Nivre (Eds.), *Proceedings of the 12th Conference of the European Chapter of the ACL* (pp. 763–771). Association for Computational Linguistics. Straková, Jana, Milan Straka & Jan Hajič. 2014. *Open-source tools for morphology, lemmatization, POS tagging and named entity recognition*. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 13-18. Baltimore, Maryland: Association for Computational Linguistics.

## 2. Obecné zásady určování lemmatu, slovních druhů a morfologických kategorií

### 2.1. Struktura automatické morfologické anotace a lemmatizace

Morfologický řádek se nachází hned pod hlavním řádkem. U každého slova jsou určeny:

- (Detailní) slovní druh (odděleno svislicí).
- Lemma (odděleno spojovníkem).
- Hodnoty morfologických kategorií (spojeny ampersandem).

Př. 1

\*CHI: to je hroch.

%mor: pro:dem|ten-1&SG&N v:cop|být-3&SG&ind&pres&akt&impf n|hroch-1&SG&MA .

U jmen a sloves jsou vždy hodnoty morfologických kategorií označovány následujícím způsobem:

morfologická kategorie	hodnota	Způsob značení
pád	1-7	1-7
číslo	jednotné	SG
	množné	PL
jmenný rod	ženský	F
	střední	N
	mužský životný	MA
	mužský neživotný	MI
	mužský (u jiných sl. druhů než subst.)	M
osoba	1-3	1-3
způsob	oznamovací	ind
	rozkazovací	imp
	podmiňovací	cond
čas	přítomný	pres
	budoucí	futur
	minulý	past
slovesný rod	činný	akt
	trpný	pas
vid	dokonavý	pf
	nedokonavý	impf

Seznam všech zkratk pro značení jednotlivých slovních druhů, příp. jejich podrobnější specifikace, a příklady lemmatizace a určování morfologických kategorií a dalších jevů naleznete v tabulkách s příklady níže.

## 2.2. Slovní druhy a jejich morfologické kategorie

### 2.2.1. Substantiva

Následující tabulka obsahuje přehled (detailnější) kategorizace substantiv a příklady určování hodnot morfologických kategorií u substantiv.

SUBSTANTIVA	příklad	morfologická anotace
substantiva	(o) lidech	n člověk-6&PL&MA
nesklonná substantiva	(k smaženému) filé	n filé-3&SG&N
propria*	(v) Brně	<b>n:prop</b>  Brno-6&SG&N
pluralia tantum	(s) kalhotami	<b>n:pt</b>  kalhoty-7&PL&MI
pomnožná propria	(v) Kozojedech	<b>n:prop:pt</b>  Kozojedy-6&PL&MI

#### \* Propria

V případě prvních slov názvů pohádek, písní apod. neoznačujeme první slovo psané velkým písmenem jako proprium, tj. *n:prop*, ale jako apelativum (*n*). Jako *n:prop* značíme pouze vlastní jména osob, zvířat, postav a míst.

Př. 2

\*CHI: táto, "**Kočka** leze dírou".

%mor: n|táta-5&SG&MA, **n|kočka-1&SG&F** v|lézt-3&SG&ind&pres&akt&impf  
n|díra-7&SG&F.

Př. 3

\*CHI: **Bublina** (jméno psa) není.

%mor: **n:prop|Bublina-1&SG&F** v|být-neg-3&SG&ind&pres&akt&impf.

### 2.2.2. Adjektiva

Následující tabulka obsahuje příklady určování hodnot morfologických kategorií u adjektiv, včetně příkladu morfologické anotace slovnědruhově podkategorie jmenná adjektiva a značení negace a komparativu/superlativu u přídavných jmen.

ADJEKTIVA	příklad	morfologická anotace
pozitiv	červený (kabát)	adj červený-1&SG&M
adjektiva s negací	nepěkná (věc)	adj pěkný- <b>neg</b> -1&SG&F
komparativ	(mám) hezčí (šaty)	adj hezký- <b>CP</b> -4&PL&M
superlativ	nejoblíbenějšímu (plyšákovi)	adj oblíbený- <b>SP</b> -3&SG&M
nesklonná adjektiva	(v) khaki (bundách)	adj khaki-6&PL&F
přivlastňovací adjektiva*	(pro) matčiny (psy)	adj khaki-4&PL&M
jmenná adjektiva	(princezna je) ráda	<b>adj:short</b>  rád-1&SG&F

## \* Přivlastňovací adjektiva

Nepisovná použití posesivních adjektiv značíme jako posesiva.

Př. 4

\*MOT: do **Lukášového** dárku ?

%mor: prep|do **adj:poss|Lukášův-2&SG&M** n|dárek-2&SG&MI ?

### 2.2.3. Zájmena

Následující tabulka obsahuje informace o detailnější kategorizaci zájmen a příklady určování hodnot morfologických kategorií u zájmen.

ZÁJMENA	příklad	morfologická anotace
zájmeno ukazovací	ty (kluci si hrajou)	<b>pro:dem</b>  ten-1&PL&M
zájmeno vztažné*	(kluci), kteří (si hrajou)	<b>pro:rel</b>  který-1&PL&M
zájmeno tázací*	čí (je to hračka)?	<b>pro:int</b>  čí-1&SG&F
zájmeno osobní**	ona (si hraje)	<b>pro:pers</b>  on-1&SG&F
zájmeno přivlastňovací	jejich (maminka)	<b>pro:poss</b>  jejich-1&SG&F
zájmeno záporné	žádná (hračka tam není)	<b>pro:neg</b>  žádný-1&SG&F
zájmeno neurčité	nějaký pes (tam je)	<b>pro:indef</b>  nějaký-1&SG&M
zájmeno reflexivní**	(házejí) sebou	<b>pro:refl</b>  se-7&SG
zájmeno refl. přivl.	(máš) svůj (koláč)	<b>pro:refl:poss</b>  svůj-4&SG&M
čí	(to je) čí (hračka)?	<b>pro:int:poss</b>  čí-1&SG&F
jejichž	(kamarádi), jejichž (psa hlídáme)	<b>pro:rel:poss</b>  jejichž-4&SG&M

## \* Tázací a vztažná zájmena

Mezi tázacím (*pro:int*) a vztažným (*pro:rel*) zájmenem se vždy rozhodujeme během manuální kontroly. V případě nejistoty na morfologickém řádku zůstává značení *pro:rel/int*.

Př. 5

\*CHI: **co** je tady?

%mor: **pro:int|co-1&SG&N** v|být-3&SG&ind&pres&akt&impf adv:pro|tady?

V morfologické anotaci je zároveň zohledněno, že tázací/vztažné zájmeno *kdo* je vždy rodu mužského a *co* rodu středního a vyjadřují jednotné číslo.

kdo (a odvozeniny)	kdo si hraje?	<b>pro:int</b>  kdo-1&SG&M
co (a odvozeniny)	(kluci), co (si hrají)	<b>pro:rel</b>  co-1&SG&N

## \*\* Osobní a reflexivní zájmena

V morfologické anotaci je zohledněno, že personální zájmena *já/ty/my/vy* a zvrtné *se* nevyjadřují rod.

já, ty	(vidím) tě	pro:pers ty-4&SG
my, vy	(bez) nás	pro:pers my-2&PL
se	(staví) se / (snaží) se	pro:refl se-4&SG
	(hrají) si	pro:refl se-3&SG

## 2.2.4. Číslovky

Následující tabulka obsahuje příklady určování hodnot u číslovek a detailnější kategorizaci slovního druhu.

ČÍSLOVKY	příklad	morfologická anotace
číslovka základní	(o) jedné princezně	num:card jeden-6&SG&F
číslovka řadová	(přijíždí) první (autíčko)	num:ord první-1&SG&N
číslovka násobná	čtyřikrát	num:mult čtyřikrát
další druhy číslovek	(kupuju) čtyřvery (kalhoty)	num čtyřvery-4&PL&M

### Specifické případy

V morfologické anotaci je zohledněno, že číslovky s dvoutvarou flexí nabývají singuláru v 1. a 4. pádě a plurálu v ostatních pádech. Neohebné řadové číslovky jsou značeny jako neohebné slovní druhy.

číslovky s dvoutvarou flexí (i kolik, tolik, mnoho...)	(dám tam) pět (hrníčků)	num:card pět-4&SG
	(do) pěti (hrníčků)	num:card pět-2&PL
dva, oba, tři, čtyři, souborové č.	(vidím) dvě (kočičky)	num:card dva-4&PL&F
neohebné řadové číslovky	poprvé	num:ord poprvé

## 2.2.5. Slovesa

Následující tabulka obsahuje příklady určování hodnot morfologických kategorií u sloves a příklady jednotlivých podkategorií sloves.

SLOVESA	příklad	morfologická anotace
verbum (dokonavé), jednoduchý tvar	přečtu	v přečíst-1&SG&ind&pres&akt&pf
verbum (nedokonavé), jednoduchý tvar	letíš	v letět-2&SG&ind&pres&akt&impf
participium (s auxiliárem)	(jsem) říkal	v říkat-SG&past&akt&M&impf
participium (bez auxiliáru)	(ona) vyhrála	v vyhrát-SG&past&akt&F&pf
pasivum	(je) rozsvíceno	v rozsvítit-SG&pas&N&pf
imperativ	zazpívej	v zazpívat-2&SG&imp&akt&pf
infinitiv	(chci) pomoci	v pomoci- <b>inf</b> &pf
negace u sloves	nevím	v vědět- <b>neg</b> -1&SG&ind&pres&akt&impf
modální sloveso*	chci (pomoci)	<b>v:mod</b>  chtít-1&SG&ind&pres&akt&impf
kopula**	buď (hodná)	<b>v:cop</b>  být-2&SG&imp&akt&impf
auxiliár** (participium minulého času)	jsem (říkal)	<b>v:aux</b>  být-1&SG&ind&pres&akt&impf
auxiliár (kondicionál***)	bychom (říkali)	<b>v:aux</b>  být-1&PL&cond&akt&impf

auxiliár (budoucí čas)	budu (číst)	<b>v:aux být-1&amp;SG&amp;ind&amp;futur&amp;akt&amp;impf</b>
idiosynkratické slovo/morfologická inovace	ozbílám	v ozbílám-1&SG&ind&pres&akt&pf- <b>neo</b>

### \* Modální slovesa

Jako modální značíme v modálním použití následující slovesa: *mít, chtít, smět, muset, moci*.

Př. 6

\*CHI: **nechce** to jíst .

%mor: **v:mod|chtít-neg-3&SG&ind&pres&akt&impf** pro:dem|ten-4&SG&N v|jíst-inf&impf .

Př. 7

\*CHI: já **chci** maminku.

%mor: pro:pers|já-1&SG **v|chtít-1&SG&ind&pres&akt&impf** n|maminka-4&SG&F .

### \*\* Sloveso „být“

Mezi auxiliárem (*v:aux*), pomocným slovesem (*v:cop*) a lexikálním, plnovýznamovým slovesem (*v*) vždy rozhodujeme během manuální kontroly. V případě nejistoty na morfologickém řádku zůstává značení *v:aux/cop*.

Př. 8

\*CHI: já **budu** rytíř.

%mor: pro:pers|já-1&SG **v:cop|být-1&SG&ind&futur&akt&impf** n|rytíř-1&SG&MA .

Př. 9

\*CHI: já **jsem** našel hůlku.

%mor: pro:pers|já-1&SG **v:aux|být-1&SG&ind&pres&akt&impf** v|najít-SG&past&akt&M&pf n|hůlka-4&SG&F .

Př. 10

\*CHI: ale tady **nejsou** bubáci.

%mor: past|ale adv:pro|tady **v|být-neg-3&PL&ind&pres&akt&impf** n|bubák-1&PL&MA .

### \*\*\* Kondicionál

Kondicionál (*cond*) je značený pouze na slovese “být”, který je součástí kondicionálového tvaru, u l-ového participia je značený minulý čas. U kondicionálových tvarů vyjádřených kontrakcí spojek *aby/kdyby* se slovesem *být* určujeme morfologické kategorie stejně jako u slovesa *být* (včetně osoby a čísla, a to i u 3. os. sg. a pl.). Výjimkou je kondicionál 2. os. sg. u zvratných sloves, kdy je sloveso *být* vyjádřeno kontrakcí se zvratným zájmenem „ses“ (př. 22), v tom případě osobu a číslo neurčujeme na „aby“, ale pouze na „ses“.



Př. 11

\*CHI: ten **by** ti nechutnal .

%mor: pro:dem|ten-1&SG&M **v:aux|být-3&SG&cond&akt&impf** pro:pers|ty-3&SG v|chutnat-neg-SG&past&akt&M&impf .

Př. 12

\*CHI: **aby** ses napila .

%mor: **conj:sub\_v:aux|aby\_být-cond&akt&impf** pro:refl\_v:aux|se\_být-4&SG\_2&SG&ind&pres&akt&impf v|napít-SG&past&akt&F&pf .

Př. 13

\*CHI: **aby** se mohla koupat .

%mor: **conj:sub\_v:aux|aby\_být-3&SG&cond&akt&impf** pro:refl|se-4&SG v:mod|moci-SG&past&akt&F&impf v|koupat-inf&impf .

## 2.2.6. Neohebné slovní druhy

Následující tabulka obsahuje přehled neohebných slovních druhů a jejich podkategorií.

NEOHEBNÝ SLOVNÍ DRUH	morfologická anotace
adverbium	adv
zájmenná příslovce	adv:pro
zájmenná záporná příslovce	adv:pro:neg
předložka	prep
spojka podřadící	conj:sub
spojka souřadící	conj:coord
částice	part
citoslovce*	int

### \* Citoslovce

Pro zachování jednotnosti všechny jednoslovné výpovědi obsahující lexémy *ano, ne, jo* značíme jako citoslovce.

## 2.3 Případy, kdy nelze určit slovní druh/lemma/morfologické kategorie

Během ruční kontroly automatického určování lemmatu, slovního druhu a morfologických kategorií bereme vždy v potaz kontext celé věty a v případě potřeby také širší kontext. V případě, že z širšího kontextu není určení některého údaje možné, aplikujeme následující pravidla.

Pokud nelze určit, o jaký slovní druh se u slova jedná, značíme jako **x|lemma**. V příkladu 9 z kontextu věty není možné určit, zda je „a“ spojka či částice.

Př. 14

\*CHI: a xxx .

%mor: x|a .

Pokud nelze určit, o jakou podkategorii se u slovního druhu jedná, značíme hlavní slovní druh a za dvojtečkou používáme x. V příkladu 10 z kontextu věty není možné určit, zda je „jsem“ auxiliár, kopula či plnovýznamové sloveso.

Př. 15

\*CHI: já jsem panenku [\*] .

%mor: pro:pers|já-1&SG v:x|být-1&SG&ind&pres&akt&impf n|panenka-4&SG&F-err .

Pokud nelze určit o jaký slovní druh (podkategorii) a lemma se jedná, značíme jako x|x. V příkladu 11 z kontextu věty není možné určit, zda je „je“ sloveso být či zájmeno.

Př. 16

\*CHI: xxx je .

%mor: x|x .

### 3. Morfologická anotace a CHAT

V systému CHAT jsou využívány specifické znaky pro zachycování verbálních i neverbálních jevů a jiných řečových, popř. zvukových realizací. Pro snazší orientaci ve značení v korpusech rané češtiny CoCzeFLA je dostupný Manuál pro přepisovatele, verze 3.0 zde: [https://coczeffa.ff.cuni.cz/wp-content/uploads/2022/10/manual\\_prepis\\_22-05-30\\_v3.0.pdf](https://coczeffa.ff.cuni.cz/wp-content/uploads/2022/10/manual_prepis_22-05-30_v3.0.pdf).

Mezi jevy a situace, které **nejsou morfologicky anotovány**, patří kromě případů, jako jsou např. zvukové tělesné projevy (&=kašel), doprovodné komentáře ([?]), fragmenty, nesrozumitelné části výpovědí a anonymizované části (xxx, zzz), také části označené jako **básnička/zpěv** (<> [=! zpěv]), **falešné začátky/sebeopravy** (<> [/] a <> [//]), **obsahy hranatých závorek** zachycující alternativní interpretaci – [=? ahoj@i]), **opakování, zvuky ne/souhlasu, prodlevy** či **hezitační zvuky** (&em, &to, &eee, hm, mhm, emem).

Př. 17

\*CHI: +< <co> [/] co dělá ?

%mor: pro:int|co-4&SG&N v|dělat-3&SG&ind&pres&akt&impf ?

Př. 18

\*CHI: hele@i [x 2] .

%mor: int|hele .

### 3.1. Zapojení kódů specifických pro formát CHAT do morfologické anotace

V prepisech rané češtiny jsou speciálně značeny některé jevy, které jsou zachyceny i v morfologické anotaci, jako je **užití interjekcí, idiosynkratických forem, morfologických inovací, cizích** (jinojazyčných) **slov a chyb**.

Jako interjekce jsou automaticky anotována slova značená pomocí @i, příp. @z:ip (interjekce ve funkci predikátu), @z:in (interjekce v nominální pozici) či @z:ia (interjekce v atributivní pozici).

Př. 19

\*CHI: dělám bubli@z:ip .

%mor: v|dělat-1&SG&ind&pres&akt&impf int|bubli .

V případě podtržitek u vícečetných interjekcí je tento znak opomíjen a interjekci morfologicky značíme jako celek (tj., *pane\_jo@i* je anotováno jako „panejo“ > int|panejo; *bim\_bam@i* je anotováno jako „bimbam“ > int|bimbam apod.).

Morfologické inovace (@n) a idiosynkratická slova (@c) jsou automaticky značeny pomocí **-neo** na konci morfologické anotace slova.

Př. 20

\*CHI: to jsou puchlíci@c .

%mor: pro:dem|ten-1&SG&N v:cop|být-3&PL&ind&pres&akt&impf n|puchlíci-1&PL&MA-**neo** .

Slova značená jako @z:c jsou značena pomocí **-ciz** na konci morfologické anotace slova.

Př. 21

\*CHI: lorry@z:c .

%mor: n|lorry-x&SG&x-**ciz** .

Slova obsahující morfologickou chybu jsou značena pomocí **-err** na konci morfologické anotace slova.

Př. 22

\*CHI: mám velikou hlad [\*] .

%mor: v|mít-1&SG&ind&pres&akt&impf adj|veliký-4&SG&F-**err** n|hlad-4&SG&MI

Morfologicky rovněž **značíme** obsahy závorek v případech kontrakce plnovýznamového slovesa a slovesa *být* (např. *přišels = přišel jsi*).

Př. 23

\*CHI: +< **tys [: ty jsi]** mě vyhodila.

%mor: **pro:pers|ty-1&SG v:aux|být-2&SG&ind&pres&akt&impf**  
pro:pers|já-4&SG v|vyhodit-SG&past&akt&F&pf .

## 4. Specifické jevy spojené s dětským jazykem/korpusem

Hodnoty morfologických kategorií určujeme rovněž u jevů specifických pro korpusy rané češtiny – u morfologických chyb, neologismů a cizojazyčných vyjádření.

### 3.1 Značení morfologických chyb

Na morfologickém řádku značíme pomocí **-err chyby v hodnotách morfologické kategorie**. Zachycujeme přitom hodnoty morfologických kategorií formy slova, kterou dítě použilo, nikoli té, kterou pravděpodobně mělo použít.

Př. 24

\*CHI: ona nechtěla toho **manžel** [\*].

%mor: pro:pers|on-1&SG&F v|chtít-neg-SG&past&akt&F&impf pro:dem|ten-4&SG&M n|**manžel-1&SG&MA-err** .

%err: manžel = manžela.

Př. 25

\*CHI: a ty **budeš jít** [\*].

%mor: part|a pro:pers|ty-1&SG v:aux|být-2&SG&ind&futur&akt&impf v|jít-**inf&impf-err** .

Pokud není možné pomocí kontextu odhadnout, o jaké hodnoty morfologických kategorií u daného, nesprávně použitého, lexému jde, používáme místo hodnoty dané kategorie **x**.

### 3.2 Značení neologismů

U slov značených jako **@c** a **@n** (s **-neo** na konci morfologické anotace slova) určujeme, pokud je to možné, slovní druh i hodnoty morfologických kategorií na základě stavby slova a kontextu věty. Pokud ne, značíme pomocí **x**.

Př. 19

\*MOT: hele@i, co tady je ?

\*CHI: **tá@c** .

%mor: x|**tá-neo** .

%com: tá = letadlo (opakovaně používaná slabika pro letadlo).

Př. 20

\*CHI: **ozbílám@c** kladívkem .

%mor: v|**ozbílám-1&SG&ind&pres&akt&impf-neo** n|kladívko-7&SG&N .

Př. 21

\*CHI: to je **ticejapoty@c** .

%mor: pro:dem|ten-1&SG&N v:cop|být-3&SG&ind&pres&akt&impf n|**ticejapoty-1&SG&x-neo** .

### 3.3 Značení cizích slov

Celé výpovědi pronesené v cizím jazyce morfologicky neznačíme a morfologický řádek v těchto případech odstraňujeme. Naopak jednotlivá cizí slova v rámci promluvy v češtině morfologicky značíme, pokud je to na základě kontextu možné (pokud ne, doplňujeme x).

Př. 22

\*MOT: kde je **head@z:c** ?

%mor: adv:pro|kde v: cop|být-3&SG&ind&pres&akt&impf **n|head-1&SG&x-ciz** ?

Př. 23

\*CHI: **bunny@z:c** .

%mor: **n|bunny-x&SG&x-ciz** .