

## Automatická slovnědruhov<sup>á</sup> desambiguace slova „to“ v ustálených větných výrazech

**ABSTRACT:** This paper deals with an automatic part-of-speech disambiguation of Czech texts containing the word *to* (E. *it*) in fixed collocations used especially in spoken Czech, and, moreover, with case identification of the pronominal reading of this word. The word *to* is ambiguous: the result of automatic morphological analysis of this word is either the pronominal lemma *ten (it)* as a nominative/accusative singular neuter, or the particle lemma *to*.

It is very difficult to automatically distinguish the nonprepositional nominative and accusative case in Czech texts. Therefore, the paper primarily focuses on *to* as a particle.

The software module performing automatic identification of collocations in Czech corpus texts is part of the automatic morphological rule-based disambiguation used for tagging texts of synchronic Czech in the corpora of the SYN series: it deals mainly with the disambiguation of nongrammatical collocations and phrases. The paper focuses on fixed expressions listed in the Dictionary of Czech Phraseology and Idiomatics and is based on the description of automatic identification and classification of collocations comprising the word *to* in the SYN2010 corpus. Also, examples (primarily idioms) are presented where automatic disambiguation using general grammatical rules yields unreliable results.

**KEYWORDS:** corpus, automatic morphological disambiguation, automatic identification of collocations, sentential phrases, word form *to*

**KLÍČOVÁ SLOVA:** korpus, automatická morfologická analýza, vyhledávání ustálených slovních spojení, větné frazémy, slovní tvar *to*

### ÚVOD

Příspěvek se zabývá automatickou slovnědruhovou desambiguací poměrně frekventovaného užití slovního tvaru „to“ (v textu dále označováno jako TO) v českých ustálených větných výrazech (zejména v mluvené češtině), a následně i problémem určení pádu v případě užití zájmena TO ve větných frazémeh. Forma TO je homonymní a rozpoznat slovní druh a pád v českém textu počítačovým programem v případě homonymního tvaru TO je obtížné. Zaměříme se především na desambiguaci slova TO jako částice.

Vyhledávací procedura ustálených slovních spojení FRANTA (**FR**azémová **AN**otace a **T**extová **A**nalýza) je program automatického vyhledávání kolokací v korpusových datech, pracuje na jednoznačném morfologicky označovaném textu. Seznamy frazémů jsou uvedeny v tabulce, kterou program vyhledávání využívá. Tento počítačový program automatického vyhledávání ustálených slovních spojení FRANTA byl doplněn o vyhledávání frekventovaných větných frazémů ze slovníku Slovníku české frazeologie a idiomatiky – část Výrazy větné (SČFI4), srov. Čermák et al., 2009, a o některé varianty větných frazémů vyskytující se v korpusových datech (Hnátková et al., 2012). Program je také součástí automatické morfologické desambiguace, kdy se řeší zejména slovnědruhov<sup>á</sup> desambiguace v negramatických spojeníh a frazémeh.

Konkrétně se tedy zaměříme na ustálená spojení uvedená ve Slovníku české frazeologie a idiomatiky (SČFI4) a na základě automatického vyhledávání větných ustálených spojení v českých textech v korpusu SYN2010 představíme typy větných výrazů obsahujících slovo TO a podle úspěšnosti dosavadního automatického značkování ukážeme příklady, kdy automatická desambiguace pomocí obecných gramatických pravidel není spolehlivá a kdy jde převážně o užití frazému.

### 1. Automatická identifikace frazémů a ustálených slovních spojení v automaticky desambigovaném textu

Zobecnit desambiguaci slova TO na všechny výskyty tohoto slova v českém jazyce (nejen ve frazémeh a nejen jako částice) je velice náročný úkol, protože použití tohoto „zástupného“ slova je velice rozmanité a složité v závislosti na významu celé výpovědi.

Automatická desambiguace tohoto spojení je velice komplikovaná, kompletní studie pro všechny použití tvaru TO není doposud provedena, ve svém příspěvku se tedy soustředíme pouze na formulaci některých pravidel pro automatickou pravidlovou desambiguaci, zejména ustálených slovních spojení s částicí TO.

Konkrétní použití tvaru slova TO v textu je buď homonymní forma zájmena *ten* v nominativu (subjekt, neutrum singulár nominativ), například ve spojení: *To se teprve ukáže. To je ale překvapení.*, nebo v akuzativu (objekt, neutrum singulár akuzativ), například ve spojení: *Kdo to má vydržet. Já to říkám pořád.*, nebo částice TO, například ve spojení: *To pěkně děkuju. To budeš koukat.*

Pro přehlednost jsou jednotlivé výskyty tvaru TO v příkladech barevně rozlišeny: zájmeno v nominativu **modře**, zájmeno v akuzativu **zeleně** a částice **červeně**.

Z automaticky získaného seznamu všech variant větných frazému ze Slovníku české frazeologie a idiomatiky, části Větné výrazy (SČFI4), jsme automaticky vytvořili seznam VF1 všech větných výrazů obsahující tvar TO z SČFI. Z celkového počtu variant 17 462 větných výrazů v SČFI4 seznam VF1 obsahoval celkem 4803 větných výrazů se slovem TO, tedy 27,5 % všech variant uvedených ve slovníku.

V morfologicky označovaném korpusu SYN2010, který obsahuje celkem 121 666 511 pozic, je celkem 969 496 výskytů tvaru slova TO, z toho jako zájmeno *ten* v nominativu (PDNS1) je značkováno celkem 603 013 výskytů, jako zájmeno *ten* v akuzativu (PDNS4) je označeno 362 790 výskytů a jenom 3 693 výskytů slova TO je automaticky lemmatizováno jako částice TO.

Program automatického vyhledávání ustálených víceslovných spojení FRANTA označil v korpusu SYN2010 celkem 43 332 výskytů tvarů TO jako součást víceslovného frazému, z toho 28 378 výskytů ve větných výrazech, 14 304 výskytů ve slovesných ustálených spojeních a pouze 121 výskytů tvaru TO v nalezených příslovích. Obecně se tvar slova TO málo vyskytuje v příslovích nebo v přirovnání, dokonce můžeme tvrdit, že se v příslovích nevyskytuje nikdy jako částice. Nejčastější identifikované přísloví se slovem TO v korpusu SYN2010 je: *Co na srdci, to na jazyku.* (celkem 15 nálezů).

## 2. Automatická desambiguace slova „to“ v ustálených větných výrazech

Hlavním cílem tohoto příspěvku je zmapovat slovo to ve frazémeh a ustálených slovních spojeních, hlavně jeho užití jako částice. Neuvažujeme tedy případy výskytu tvaru TO:

a) po předložce, včetně frazémového spojení „být s to“, kdy se jedná vždy o akuzativ zájmena *ten* a automatická desambiguace tohoto spojení je zřejmá.

Příklady:

(1) *Nebyl s to přijít včas. Jen se na to třese. Na to ho moc neužije. Já se na to vyto!  
O to běží! Na to nejsem zvědavej! Na to se neumírá. Na to ti kašlu. Na to zapomeň!*

b) na pozici rozvíjejícího přívlastku, kdy následuje substantivum v nominativu nebo v akuzativu, kdy se jedná o tvar zájmena *ten* a automatická desambiguace pádu probíhá na základě shody s

příslušným substantivem.

Příklady:

(2) *Tady to lidové rčení sedí. To slovo jí nešlo z úst. S kým mám to potěšení?*

c) ve spojení: *to celé/poslední/pravé/nejlepší, to všechno*, kdy se jedná opět o tvar zájmena *ten* a automatická desambiguace pádu probíhá na základě shody s příslušným adjektivem nebo zájmenem *všechno*.

Příklady:

(3) *To je to samý v bledě modrým. To nejlepší nakonec!*  
*To všechno má svůj čas. Každý chválí to svoje.*

Po odstranění případů typu a) až c) ze seznamu VF1 (větné frazémy ze slovníku SČFI obsahující tvar TO), jsme získali celkem 4 502 variant větných frazémů s tvarem slova TO.

Jak bude ukázáno na jednotlivých příkladech v tomto příspěvku, obecně se automatická disambiguace tvaru TO může řídit jednak (i) pravidlem shody nebo neshody řídicího slovesa se zájmenem TO v rodě a čísle – v případě shody v rodě a čísle se jedná o potenciální subjekt v nominativu, v případě neshody je tvar TO pouze dvojznačný, tedy jedná se o objekt v akuzativu nebo o částici, případně (ii) neakuzativní nebo akuzativní nepředložkovou valenci řídicího slovesa – v případě „neakuzativního“ slovesa (například *koukat, pršet, přijít*) je slovo TO dvojznačné, buď je TO subjekt v nominativu, nebo částice, a dále (iii) na základě toho, zda je pozice subjektu nebo objektu ve větě již zaplněna jiným jednoznačným nominativem nebo akuzativem – v případě výskytu subjektu nebo objektu ve větě se jedná pouze o homonymii částice a zájmena. Například v případě neshody tvaru TO s tvarem netranzitivního (neakuzativního) slovesa jde ve větě o částici TO, například: *To koukáš, co?*. V případě neshody tvaru TO s tvarem slovesa s nepředložkovou akuzativní valencí a výskytu substantiva v akuzativu ve větě se jedná také o částici TO, například: *To mi říkáš novinu*.

Vyhledávací procedura FRANTA je součástí automatické pravidlové morfologické disambiguace, která provádí disambiguaci právě takových slovních spojení, u kterých se výše uvedená pravidla nedají aplikovat (například chybí sloveso). V těchto případech je tvar TO identifikován převážně jako částice.

Příklady:

(4) *To jo. To určitě! To zrovna! To zrovna ne. To vůbec ne. To asi sotva.*  
*To teda/tedy jo/ne. To teda ani náhodou! To snad ne.*  
*To určitě ano/ne. Cokoli/proboha/všechno/ne, jen to ne.*

V některých případech frazeologických větných výrazů, které neobsahují plnovýznamové sloveso, není určení slovního druhu u tvaru TO zřejmé a jednoznačné. Například ve větě: *To nemůžu, paní Nováková*. můžeme předpokládat, že se jedná o objekt v akuzativu ve smyslu: *To vám nemůžu udělat*. V tomto užití se ale asi jedná o zkrácení frazému: *To nemůžu sloužit.*, kdy se jedná o částici TO.

## 2. 1. Příklady automatické slovnědruhovú desambiguace

Jedním z příkladů složité desambiguace, kdy automatická statistická desambiguace selhává a k určení významu je nutné zapojení frazémů do desambiguace, je v případě tázací věty: *Kde to žiješ/zijeme/žijete?*, kde tvar TO je částice, na rozdíl od věty: *Je to střed města, kde to žije.*, kde TO je zájmeno v nominativu.

Obecně otázka: *Kde to žije?* je dvojnásobná a jednoznačná automatická desambiguace není možná, protože například v následujícím kontextu (5) jde o zájmeno v nominativu, v případě kontextu (6) jde o částici TO.

(5) *Co to je? Jak to vypadá? Kde to žije? Má to srst?*

(6) *Takhle vidí pan Steigerwald ženy? Kde to žije? Problém je trochu jinde.*

Zvláštní skupinu tvoří specifické slovesné spojení typu: *hasit si to někam*, kde TO můžeme desambiguovat jako nadbytečný akuzativ zájmena *ten*, potom ale ve spojení *drandit si to někam* by sloveso *drandit (si)* mělo akuzativní valenci. V případě, že TO v tomto spojení budeme považovat za částici, je otázka, zda TO je součástí lemmatu takového slovesa, nebo jeho valenční doplněním. V korpusech ČNK nalezneme další slovesa tohoto typu: *šinou si to, mastit si to, valit si to, mířit si to, namířit si to, zamířit si to, odhasit si to, řídit si to, hnát si to, šinou si to, štrádat si to*.

Příklady:

(7) *A odhasil si to. Mířil si to domů. Mastil si to honem domů. Namířil si to do lesa.*

*Zamířil si to přímo nad náš člun. Přihasil si to profesor z Ameriky. Auto si to hasilo ...*

*Řítil si to sto šedesátkou. Hnali si to přes pole pryč. Šinul si to jako šnek. Ten si to štráduje!*

*Drandíme si to po Panamerikáně.*

V užití sloves tohoto typu nalezneme i případy se změnou slovosledu a se slovy vloženými mezi jednotlivé části slovesného spojení (podobně jako u slovesných tvarů slovesa, například *naplánoval jsem si (to)*)

(8) *Kdo jiný než ona si to metl po Landing Road.*

*Sledujete, jak si to lod' mete přímo do nich.*

*Po záplatované silničce si to proti nim šinul stařík na kole.*

*Prázdňá bárka si to šinula proti proudu.*

Při automatické desambiguaci je nutné rozlišovat následující případy dvou odlišných sloves: *mazat si to někam* (příklad 9a) a *mazat si něco* (příklad 9b).

(9a) *ale on na vás plivne kapičky zloby, vycení polozvířecí chrup a maže si to dál*

(9b) *Mamka jí poradila, aby si to mazala krémem.*

Samostatným případem složité automatické desambiguace je slovní spojení „a to“, kdy se jedná o spojku. Slovník spisovného jazyka českého popisuje toto slovní spojení jako spojku takto:

spojka **a**

**3.** ve spojení **a to** (část. než *a sice*)

vysvětluje, zpřesňuje n. omezuje platnost toho, co předchází (před *a* se píše čárka):

*v září, a to na počátku; v knize, a to až v závěru; kromě mne přebýval v tom domě ještě jeden nájemník, a sice v přízemí*

Pokud by ve spojce *a to* bylo slovo TO zájmeno *ten*, jako první se nabízí tento výskyt tvaru TO ve spojce *a to* desambiguovat jako nominativ, ale potom by desambiguace vět v příkladu (10) byla následující:

(10) *Přiběhl pes-NNMSI, a to-PDNSI malý.*

*Viděl jsem psa-NNMS4, a to-PDNSI malého.*

*Prohráli jsme, a to-PDNSI na celé čáře.*

Je otázkou, zda tento tvar TO ve spojení neoznačovat jako částici, případně přímo jako spojku. Přesně identifikovat případy, kdy se jedná opravdu o spojku *a to*, je pro automatickou desambiguaci obtížné. V případě spojky musí předcházet čárka (nebo tečka místo čárky), ale v některých příkladech z korpusu čárka chybí, kdy jde o zřejmou chybu, například: *budou podle něj trénovat už pouze doma a to jednofázově*. Automaticky rozlišit případ spojky *a to* nebo částice TO není jednoduché i ve větě: *Česká společnost má prý navíc jeden hloupý zvyk, a to koukat druhým do kapsy*. Ve SČFI4 jsou uvedeny větné výrazy se spojením *a to* například: *A to ještě není všechno. A to jo! A to tedy jo! A to zase ne! A to zas prr!* V korpusu jsme našli následující příklady s částicí TO: *A to vážně nepřeháním. A to budete koukat. A to večer zdaleka ještě nekončil.*

Větný výraz: *A to jsem ještě neslyšel*. je dvojnásobný. TO je buď objekt v akuzativu – *a toto jsem ještě neslyšel* – nebo TO je částice – *v případě, že jsem byl hluchý a stalo se to v době, kdy jsem ještě neslyšel...* Stejně tak je dvojnásobná věta: *A to vážně nepřeháním*. V tomto případě můžeme předpokládat, že jde o částici.

Příkladem, kdy v současné automatické desambiguaci selhává, je spojení se slovesem *bolet*. Pravidlová desambiguace využívá k desambiguaci objektu v akuzativu soupis sloves s obligatorní akuzativní valencí, kam je zařazeno i sloveso *bolet*, a to vzhledem k užití typu: *Maminku-NNFS4 bolí hlava*. V příkladech (11) větných výrazů nalezených v korpusu je slovo TO v nominativu.

(11) *Ten je blbej, až to bolí! Bolí to jako čert. To (ne)bolí. (Ne)Bolí to. To to bolí. To bude bolet. To bolelo.*

Při automatické desambiguaci větných členů se slovesem *bolet* bude nutné doplnit u homonymních slov informaci, zda se jedná o pojmenování věci nebo osoby, kdy v případě osoby jde o substantivum v akuzativu, v případě věci o nominativ, například: *Soudce-NNMS4 bolí kosti-NNFP1*. Ve spojení slov *srdce bolí* bude tedy substantivum desambiguováno jako nominativ, ale až na jedinou výjimku v přísloví: *Co oči nevidí, to srdce nebolí.*, kde se jedná o akuzativ slova *srdce*. Sloveso *bolet* je také zajímavé tím, že tvary v 1. nebo 2. osobě se skoro vůbec nevyskytují, v korpusu nalezneme jenom pár výskytů v poezii:

(12) *Bolíš mě, lásko. A bolíš mě také vším. Strašlivě mne bolíš.*

Podobná slovesa jsou snad jenom *brnět* a *svědit*.

## 2. 2. Rozdělení použití tvaru TO podle pozice ve větě

Zvláštními případy složité automatické desambiguace jsou i dva výskyty tvaru TO na začátku věty, z nichž ani jedno není po předložce nebo v konstrukci se sponou. V SČFI4 jsou uvedeny tyto příklady větných výrazů: *To to dneska frčí! To to dopadlo! To to frčí! To to trvá! To to uteklo!*, kdy TO na začátku je částice a druhé TO je zájmeno v nominativu. Jednotlivé případy dvou tvarů TO na začátku věty nalezené v korpusu můžeme rozdělit na případy:

- (i) TO(částice) TO(subjekt, zájmeno v nominativu) ...  
*To to bouchlo/trvalo/uteklo/dopadlo/uběhlo/začalo.*  
*To to letí/zebe/studí/bolí/smrdí/začíná/utíká/trvá/vypadá/spěchá/zní.*  
*To to bude vypadat. To to dopadne. To to uteklo.*
- (ii) TO(částice) TO(objekt, zájmeno v akuzativu) ...  
*To to vyvedl. To to nepoznají? To to nevidíš?*



(iii) TO(subjekt) TO(objekt) ...  
*To to vysvětluje. A to to trošku zachraňuje.*

(iv) TO(částice) TO(částice) ...  
*To se to směje cizímu neštěstí ! To se to utrácí. To se to řekne.*

Speciální jsou případy výskytů dvou tvarů TO ve větě typu: je/bylo/bude/není/nebylo TO TO <nominální skupina>, kdy se jedná o dvě zájmena v nominativu. V SČFI4 je uveden frazém: *Není to to pravý vořechový. Věděl, že to nebude to pravé ořechové.* V korpusu nalezená spojení jsou tohoto typu: *Není to to, co bývalo. Je to to jediné, co můžeme udělat. Je to to samé. Je to to nejdůležitější. Není to to správné slovo.*

Nyní uvedeme několik příkladů výskytů slova TO rozdělených podle pozice tohoto tvaru ve větě.

### 2. 2. 1. TO na počátku výpovědi

Uvažujeme-li tvary slova TO na první pozici ve výpovědi, zahrnujeme v to i případy výskytu TO „po spojce a“, „po čárce a spojce“ nebo „po adverbium“(právě to, přesně to), případně „po částici no“ bezprostředně nebo oddělené čárkou, (ne „po předložce“). V SČFI4 jsou uvedeny větné výrazy: *To nepadá v úvalu! To nestojí za řeč. To nevěstí nic dobrého. To pláčeš na špatným hrobě. To mrkáš na drát.*

Podrobněji můžeme rozdělit jednotlivé případy podle následující slovního tvaru:

#### A) TO <tvar slovesa>

Jedná se o případ, kdy tvar slova TO je na začátku výpovědi následovaný finitním tvarem slovesa, současně ve větě není další větný člen v nominativu, přitom se může jednat o sloveso v různém tvaru, například: *To se mohu rovnou jít oběsit. To se mohu rovnou oběsit. To se půjdu rovnou oběsit. To se mám rovnou oběsit? To se jdu oběsit.*

Slovosled zájmena a slovesa může být i opačný, desambiguace slova TO je v konkrétních případech různá podle shody a tranzitivity slovesa. Případy různých slovních druhů ve spojení „to má“ jsou uvedeny v příkladech (14).

(14) *To má někdo štěstí/nervy. Má to někdo štěstí.*  
*To má něco do sebe. Má to něco do sebe.*

Nejčastější výskyty tohoto typu v korpusu s částicí TO jsou například:

(15) *To jsem se vás lekla. A to se podívejme. To se pleteš! To se těš! To se mě moc ptáš.*

#### B) TO jsem/jsi/jsme/jste ...

Jedná se o homonymní tvar slova TO (částice, zájmeno v akuzativu nebo v nominativu) následovaný finitním tvarem slovesa v 1. nebo 2. osobě. V SČFI4 jsou uvedeny větné výrazy: *To jsi uhodl! To jsem si oddechl! To jsem tedy rád. To jsme to dopracovali.*

Příklady (16) představují nejčastější výskyty v korpusových textech s částicí TO, příklady (17) výskyty se subjektem TO (v nominativu) a (18) příklady s TO jako objektem (v akuzativu)

(16) *To jsem dopadl.*  
*(Ale) to jsem odbočil. To jsme se nasmáli.*  
*To jsem úplně mimo. To jsi úplně vedle.*  
*To jsi teda na omylu.*

(17) *To jsem celá já.*  
*To jsi celá/ý ty.*  
*To jsou věci!*

(18) *To jste neviděli!*  
*To jsem nechtěl!*  
*To jsem si mohl myslet!*

#### C) TO se/si <imperativ>

Jedná se o tvar slova TO na začátku věty následovaný slovesem v imperativu, jde o případy, kdy tvar TO je ve většině případů částice. V SČFI4 jsou uvedeny větné výrazy tohoto typu: *To si zapiš za uši. To se opovaž! To se podrž! To se podívejme! To se těš!*

Nejčastější spojení tohoto typu v korpusových datech je: *(Tak) TO SI PIŠ* (, že), přičemž současná desambiguace je v některých případech chybná. Další nejčastější výskyty s částicí TO jsou následující případy (19), tvar TO je desambiguován jako akuzativ v příkladech (20).

(19) *To se vsad. To se neptej. To se neboj.*  
*To si buď jistý. To se nediv. To se podrž.*  
*To se podívejme.*

(20) *To si neber. To si nemysli! To si pamatuj!*  
*To si nechte zarámovat. To si nedovoluj!*

#### D) TO <infinitiv>

Případy, kdy tvar TO na začátku věty je následován infinitivem slovesa, můžeme rozdělit na následující podtypy:

D1) TO se/si mám <infinitiv slovesa> – jedná se většinou o otázku, při neshodě v osobě s tvarem slovesa *mít se* vždy jedná o částici. Automatická desambiguace v korpusových datech v případech shody slovesa *mít se* zájmenem TO je většinou chybná.

(21) *To se mám rozkrájet? To se mám zbláznit?*  
*To se máme na co těšit. To se má čím chlubit.*

Obtížná je také desambiguace tvaru TO ve větách z příkladu (22), kdy jde o věty dvojznačné: má se změnit *on* nebo *ono*?

(22) *To se má brzy změnit. To se má změnit?*

D2) TO bude/budeš/budete/budeme/budou <infinitiv> – jedná se o tvar TO na začátku věty následované futurem slovesa, tvořeným pomocným slovesem *být* a infinitivem. Poměrně časté spojení: *To bude koukat!* je dvojznačné (*ono bude koukat/přečínvat* nebo ve významu *on se bude divit*), ale výskyty v korpusu jsou převážně frazémové použití s částicí TO. Na rozdíl od věty: *To budete koukat.*, kdy je TO jednoznačně částice.

Při vyhledávání tohoto typu větného výrazu v korpusu byly nalezeny všechny infinitivy v tomto slovním spojení. Značnou část tvoří slovesa vnímání.

Příklady:

(23) *To budeš/budete civět/koukat/mrkat/čumět/čučet/zírat/čubrnět/vidět.*  
*To budou na nás výřit!*  
*To budeš vidět ten rachot!*

E) **TO** když/jako ... – po tvaru TO na začátku věty následuje spojka. TO je vždy ve významu částice, současná automatická desambiguace je v některých případech chybná. Příklady:

- (24) *To když je zima jak v ruském filmu, ...*  
*To když mu dojde, že už je doma.*  
*To když padám na ústa, ...*  
*To když potká doktora, ... .*  
*To když se profláklo, že ...*  
*To jako proč?*  
*To jako kdo?*  
*To jako říká kdo?*

V případech desambiguace vět: *A to jako ředitel nemohu připustit. Formy mezd se dělí do dvou skupin, a to jako základní a doplňkové.* se nejedná o výskyt částice TO, ale o přirovnání nebo výše zmiňovanou spojku *a to*.

F) **TO** by(ch(om))/bys(te) ... – tvar slova TO je následovaný kondicionálem slovesa.

V SČFI4 se vyskytují větné výrazy tohoto typu: *To by bodlo. To by byla věčná škoda. To by sis dal! To bych si dal líbit! To bych si vyprosil!* Jedná se většinou o spojení s částicí TO – příklady (25) nebo s akuzativem zájmena *ten* – příklady (26).

- (25) *To by tak hrálo, aby ...*  
*To by tak chybělo, aby ...*  
*To by tak scházelo, aby ...*  
*A to by v tom byl čert, aby ...*  
*Ale to bych předbíhal. To bych se nenadál.*  
*No to bych prosila. To bych byl špatný politik.*

- (26) *To bych nerad.*  
*To bych do tebe neřekl.*  
*To bych také rád věděl.*

G) **TO** aby(ch(om))/abys(te) – částice TO je následovaná spojkou *aby*. V SČFI se vyskytují větné výrazy: *To aby měl člověk pořád ruku v kapse. To abych se šel pak rovnou voběsit. To abych už šel.* Příklady nejčastějších výskytů tohoto spojení v korpusu jsou:

- (27) *To aby se neřeklo/nezapomnělo*  
*To aby nedošlo k mýlce.*  
*To abychom nedělali ostudu.*  
*To abych běžel/šel.*  
*To aby řeč nestála.*

### 2.2.2. **TO** na druhé pozici ve výpovědi

Podrobněji můžeme rozdělit jednotlivé případy výskytu tvaru TO na druhé pozici ve výpovědi podle předcházejícího slovního tvaru:

A) KDE (jsme/se/jsme se/jsi se/...) **TO** – jedná se o otázku nebo vedlejší větu (*Nevím, kde se to stalo.*) V SČFI jsou uvedeny otázky tohoto typu: *Kde jsme to přestali? Kde se to tu bere? Kde to vážně? Kde to žiješ?*



V případě slovního spojení „*kde to žije*“ se v otázce: *Kde to žije/žiješ/žijete/žijí?* jedná o částici TO, na rozdíl od věty: *Kde to žije bohatým životem/dechovkou.* je tvar TO v nominativu. V příkladech (28) jsou uvedena nejčastější spojení v korpusu s částicí TO, příklady (29) ukazují případy, kdy TO je zájmeno v akuzativu (objekt), nejčastější výskyty na pozici v nominativu (subjektu) jsou uvedeny v příkladech (30).

(28) *Kde to (sakra/proboha/vlastně) jsme/jsem/jsi/jsou?*  
*Kde to vězíš/vězíte? Kde to vězí? Kde to lítáš? Kde to lítá?*

(29) *Kde to, co ...*  
*Kde to bere?*

(30) *Kde to bylo? Kde to bude?*  
*Kde se to stalo?*  
*Kde to bolí/bouchlo?*

B) KAM (jsme/se/jsme se/jsi se/...) TO – jedná se o otázku nebo vedlejší větu (*Nevím, kam se to ženeme. Nechápu, kam se to vše podělo/ztratilo.*). V SČFI4 jsou uvedeny otázky: *Kam by to vedlo? Kam se to podělo?* Frekventované výskyty s částicí TO v korpusu jsou uvedeny v příkladech (31).

(31) *Kam to brejliš? Kam to koukáš? Kam se to, k čertu, hrneš?*  
*Kam se to ten vrtáček vrtl? Kam jsme se to dostali?*  
*Kam se to řítím? Kam se to ženeš? Kam to ten svět spěje?*

Zajímavé je použití ustáleného větného výrazu: *Kam to bude, šéfe/slečno/kamaráde?* Zde by měl být tvar TO identifikován jako částice.

V tomto případě je slovo *kdo* v nominativu (subjekt), proto tvar TO bude buď akuzativ (32), nebo částice (33) nebo v případě slovesa *být* nominativ (34). V SČFI4 jsou uvedeny větné výrazy: *Kdo by to do něj řekl! Kdo to má vydržet? Kdo to má čuchat? Kdo to ukradne, tomu ruka upadne.*

(32) *Kdo mi to dokáže? Kdo mu to udělal? Kdo to kdy slyšel?*  
*Kdo to jakživ viděl, aby .... Kdo to má čuchat?*

(33) *Kdo mi to loupe perníček? Kdo to mluví? Kdo to přišel?*

(34) *Kdo to je? Kdo to byl?*

D) CO TO ...

Jedná se o případy, kdy tvar TO je na druhé pozici v otázce po tázacím CO. V SČFI jsou uvedeny větné výrazy: *Co to furt meleš? Co to koštuje? Co to má znamenat? Co to na mě zkoušíš? Co to obnáší? Co to slyším? Co to vidím? Co to? Ať to stojí, co to stojí.* Tyto případy můžeme dále rozšířit o dva typy frekventovaných spojení, vyskytujících se v korpusových datech:

DI) CO (se/jsme se/jste se/jsi se/ses ...) TO ...

Nejčastější větný výraz se slovním spojením *Co to ...* v korpusových datech jsou otázky: *Co se to děje/dělo/stalo/stane?* Při vyhledávání tohoto spojení jsme zjistili, že na další pozici ve větě po tvaru TO se vyskytují často výrazy klení:

(35) *Co se to*  
*proboha/propána/krucci/krucinál/sakra/k čertu/ksakru/*  
*pro všechno na světě/u všech čertů/ke všem čertům/*  
*u všech hromů/prokristapána/u všech všudy/kurva/kurva fix/*

*hergot/do hajzlu/do prdele/u všech fotonů  
děje/dělo/stalo/stane?*

Další frekventované otázky s TO na druhé pozici v nominativu jsou například:

(36) *Co to (vlastně) bylo/je? Co to (vlastně) znamená?*

D2) CO (si/jsme si/jste si/jsi si/sis ...) **TO** ...

Příklady nejčastějších otázek tohoto typu v korpusu:

(37) *Co si to dovoluje(š)? Co si to namlouvá? Co si to vlastně naháváš?  
Co sis to zamlouval/usmyslil? Co sis to vzal do hlavy? Na co si to hraje(š/te/me)?*

### 2. 2. 3. **TO** na konci věty

Tvar TO také může stát na poslední pozici ve větě, a to na konci výpovědi (za TO následuje tečka), například (38) ve větných výrazech z SČFI4, nebo na konci hlavní věty (za TO následuje čárka), například (39) ve větných výrazech z SČFI4. Na této pozici není tvar TO částice.

(38) *Bere ho to. Jde to. Klaplo to. Vleče se to. Vodsejpá to. (38)*

(39) *Ber to, jak to leží a běží. Stane se to, než se naděješ. (39)*

V korpusových datech jsme našli dva následující typy frekventovaných větných výrazů se slovem zájmenem TO na konci.

(i) ... <tvar slovesa> (ani/aspoň) **TO**.

V těchto případech je TO pravděpodobně zájmeno v nominativu, když se tvar slovesa shoduje s TO v rodě a čísle (příklady (40)), v opačném případě je slovo TO zájmeno v akuzativu (příklady (41)).

(40) *Bolí to. Fungovalo to. Nešlo to. Neprošlo to. Zabralo to.*

(41) *Dokázal to. Neudělá to. Pochopí to. Neví to.*

(ii) ... <pocitové sloveso> (by) mě/ho **TO**.

V případě tohoto typu se zájmeno TO shoduje s tvarem slovesa, jedná se tedy o subjekt (zájmeno v nominativu). Příklady:

(42) *Žere mě to. Štve mě to. Mrzí mě to. Dojalo mě to. Dopaluje mě to.*

*Deptalo mě to. Děsilo mě to. Fascinuje mě to. Flustrovalo mě to. Polekalo ho to.*

Při vyhledávání větných výrazů typu (ii) v korpusu SYN2010 jsme při zadání slovního spojení „mě/ho TO“ jsme našli převážně spojení s pocitovými slovesy – přehled (43) a (44).

(43) ... *mě/ho to mátló, nadchlo, nakoplo, nadzvedlo, naštválo, dojalo, drásalo, dráždilo, překvapilo, rozhodilo, rozčílilo, potěšilo, napružilo, nadzvedlo, nudilo, pobavilo, popletlo, ...*

(44) ... *mě/ho to baví, drásá, děsí, nudí, potěší, trápí, ...*

Mezi pocitová slovesa můžeme zahrnout i frazémová verbonominální slovesa *vyvést z rovnováhy, lézt na nervy, dohánět k zuřivosti*, nalezená při vyhledávání slovního spojení „TO mě/ho“.

(45) *To mě vyvedlo z rovnováhy.*

*To mně lezlo na nervy*

*Jen ho to dohání k zuřivosti.*

## ZÁVĚR

Při mapování výskytu tvaru TO v českých textech jsme zjistili, že automatická desambiguace je opravdu složitá a obecná pravidla pro automatickou desambiguaci jsou nespolehlivá. Při procházení některých výskytů jsme narazili na neobvyklá vyjádření, nová slova, která nejsou automatickou morfologií rozpoznána. Podařilo se nám určit pouze pár základních pravidel, která pomohou při určení (desambiguaci) tvaru TO v konkrétních případech. Soustředili jsme se na dvě oblasti: jednak na frazémy a ustálená sousloví, jednak na valenci sloves, která se kombinují se slovem TO. Pokusili jsme se ukázat sporné případy různého použití slova TO, poukázali jsme na některá specifika sloves, která se kombinují s nadbytečně užitým slovem TO. Studie desambiguace slova TO nás dovedla k určení případů, kdy současná automatická desambiguace špatně interpretuje slovo TO, a přivedla k dalším hlediskům, která mohou sloužit k určení valenčního rámce některých sloves.

## LITERATURA:

*Český národní korpus – SYN2010*. Ústav Českého národního korpusu FF UK, Praha 2010.  
<<http://www.korpus.cz>>.

Hnátková, M. – Kopřivová, M. (2012): From a Dictionary to a Corpus. *Konference EUROPHRAS 2012 – Phraseology and Culture*. Maribor, v tisku.

Hnátková, M. (2011): Výsledky automatického vyhledávání frazémů v autorských korpusech. In: *Korpusová lingvistika* Praha 2011, sv. 3: Gramatika a značkování korpusů. Nakladatelství Lidové noviny / Ústav českého národního korpusu, 171–185.

Čermák, F. et al. (2009): *Slovník české frazeologie a idiomatiky 4. Výrazy větné*. Praha: Leda.

## GRANTOVÁ PODPORA:

Tento příspěvek byl podpořen z grantu GAČR P406/10/0434.