

Automatické rozpoznávání substantivního dativu při morfologické disambiguaci českých textů¹

Vladimír Petkevič

Ústav teoretické a počítačové lingvistiky FFUK

Vladimir.Petkevic@ff.cuni.cz

ABSTRACT: The paper deals with the identification of the dative/non-dative interpretation of a non-prepositional noun during a rule-based automatic morphological disambiguation of Czech sentences. Based on the valency considerations including syntactic functions of the complements of verbs (these functions being an object or an adverbial) and of the resulting lists of verbs having / not having object dative valency, negative and positive non-heuristic (safe) as well as heuristic rules are presented. The negative rules assign a noun a non-dative interpretation using a list of verbs that cannot tolerate a dative noun in their close vicinity, whereas the positive ones specify a given noun as being in the dative case on the basis of the verbs having object dative valency, the object being either the only object of the verb, or its second indirect object. The paper concerns itself only with the non-prepositional nouns which are much more difficult to disambiguate than the prepositional ones.

KEYWORDS: non-prepositional nominal dative, Czech texts, corpus, morphological disambiguation

KLÍČOVÁ SLOVA: nepředložkový substantivní dativ, české texty, korpus, morfologická disambiguace

0. Úvod

Při automatické morfologické disambiguaci českých textů je největším problémem určit správný pád **jmen** (substantiv, adjektiv, zájmen, číslovka). Tvar náležící nějakému jmennému lexému může být:

(a) homonymní se členem paradigmatu jiného lexému (*kárám* = 1. os. sg. přítomnosti lexému *kárat*, nebo Dpl substantivního lexému *kára*),

¹ Tento příspěvek byl podpořen z grantu GAČR P406/10/0434.

(b) homonymní uvnitř paradigmatu téhož lexému (*souvislosti* = Gsg, Dsg, Lsg, Npl, Apl, Vpl lexému *souvislost*).

Složitější je disambiguace typu (b) a my zde konkrétně předvedeme negativní i pozitivní pravidla disambiguace nepředložkového dativu substantiv.

1. Motivace. Ve snaze automaticky rozpoznat dativ substantiva ve větě/klauzi na základě kontextu jsme motivováni třemi cíli:

- přispět k teoretickému popisu povrchové syntaxe češtiny z hlediska morfologické disambiguace
- zlepšit disambiguaci českých textů
- díky kvalitnější morfologické disambiguaci zlepšit i syntaktickou analýzu českých textů.

2. Frekvence dativu substantiv

Podle údajů v publikaci Statistiky češtiny (srov. Bartoň et al., 2009: 132) zaujímá substantivní dativ frekvenčně předposlední místo mezi pády:

Tabulka 1. Frekvence pádů substantiv v korpusu SYN2005

nominativ:	29,1 %
genitiv:	27,5 %
DATIV:	3,6 % (75 % sg, 25 % pl)
akuzativ:	19,2 %
vokativ:	0,35 %
lokál:	11,8 %
instrumentál:	8,5 %
Celkem:	100,0 %

V této tabulce není rozlišen dativ předložkový (prep) a nepředložkový (neprep); v dalším budeme tyto typy dativu jasně odlišovat. Zajímavý je frekvenční poměr obou těchto typů dativu (srov. Bartoň et al., 2009: 136; Jelínek, 2012: 108)

Tabulka 2. Poměr předložkového a nepředložkového dativu substantiv v korpusu SYN2005

	prep	neprep
substantivní dativ:	54 %	46 %

3. Hlavní typy homonymie dativu v substantivních paradigmatech

V této části se zaměříme jen na homonymii paradigmatickou, způsobenou pádovým synkretismem. Poměrně velkou homonymii vykazují singulárové dativní tvary, mnohem menší pak tvary plurálové. Například tvar dativu singuláru (Dsg) životných maskulin je homonymní s jinými pády u těchto substantivních vzorů životných maskulin:

typ: *muži*

Dsg: Vsg, Lsg, Npl, Vpl, Ipl

typ: *pánu, předsedovi, chlapíkovi*

Dsg: Lsg

typ: *soudci*

Dsg: Lsg, Npl, Vpl, Ipl

typ: *synu, Georgi*

Dsg: Vsg, Lsg

Velkou homonymii vykazují i substantiva ostatních jmenných rodů, zvláště feminina singuláru vzoru *kost*.

4. Hlavní syntaktické funkce nepředložkového substantivního dativu

Určit předložkový dativ je úkol poměrně snadný, mnohem obtížnější je disambiguace dativu nepředložkového, a v dalším se proto budeme zabývat pouze tímto dativem. Substantivum v dativu může mít tyto tři povrchově-syntaktické funkce:

- *předmět, adverbiale* – substantivum v dativu tu rozvíjí sloveso, adjektivum nebo adverbium
- *přívlastek* – substantivum v dativu rozvíjí syntaktické substantivum

a z toho budeme při disambiguaci vycházet. Pro další úvahy předpokládáme, že ve větě, kterou se snažíme disambiguovat, je u každého substantiva vyznačeno, že jeho pád je/není určen předložkou (substantivum tedy je/není řídicím členem příslušné předložkové skupiny). Toto rozlišení jsme obecně schopni stanovit s 95% úspěšností, v případě dativu pak s úspěšností dokonce překračující 99,5 %.

4.1 Předmět

Dativ může vyjadřovat (srov. Šmilauer, 1966: 213nn.):

- hloubkově obligatorní jediný valenční přímý **předmět** slovesa/adjektiva/adverbia (*podobný hlasu-3 dítěte*)
- hloubkově obligatorní valenční **nepřímý předmět**: dativ vyjadřuje druhý předmět u slovesa/adjektiva, přičemž první předmět je v akuzativu, výjimečně v genitivu (*přikázal to-4 podřízenému-3, jednomu kavárníkovi-3 se zželelo Adriany-2*)
- předmět jako nevalenční, volný dativ (*člověku-3 se chce zvracet*).

4.2 Adverbiale

Volný adverbialní dativ může být (srov. Šmilauer, 1966: 221nn.; Jelínek, 2012: 153):

- prospěchový: *Můžete ho natrhat mamince-3 do vázy.*
- zřetelový: *Majiteli-3 způsobil škodu 50 tisíc.*
- přívlastňovací: *Anně-3 se zavíraly oči.*

4.3 Přívlastek

Substantivum v dativu rozvíjející syntaktické substantivum je z definice jeho přívlaskem (*poplatnost nové konvenci-3*).

5. Pozitivní a negativní morfologická disambiguace substantiva v potenciálním dativu

Pozitivní disambiguací dativu míníme určení dativu u příslušného substantiva, negativní disambiguací naopak vyloučení jeho dativní pádové interpretace. Lze přitom využít mj. těchto faktů:

- existence morfologicky jednoznačných jmen (= subst / adj / pron / num) v dativu v okolí určovaného substantiva (např. *dívkám, novému, jim, jednomu*)
- existence morfologicky sice nejednoznačných jmen (*touze, nejlepším, té, pěti*), ale zjednoznačňných v průběhu dosavadní disambiguace dané věty
- určovaný substantivní tvar je v dativu sice jen potenciálně, ale preponované adjektivum je v jednoznačném dativu: *starému-3 hradu-346*
- dativ je pád velmi nepravděpodobný (jen cca 3,6 % substantiv), čehož lze využít alespoň heuristicky.

5.1. Hlavní pravidla pro disambiguaci nepředložkového dativu

Při návrhu pravidel vycházíme z těchto předpokladů:

1. Nepředložkový substantivní dativ má tyto funkce:
 - přímý/nepřímý předmět
 - adverbiale zřetele/prospěchu/posesivity
 - přívlask.
2. Každé slovo ve větě může být rozvito nejvýše jedním nepředložkovým dativním doplněním (na rozdíl třeba od doplnění v akuzativu, instrumentálu či genitivu).
3. Ve vstupní větě je již vyznačeno, že každé substantivum je/není „předložkové“.

Pro co nejspolehlivější disambiguaci je nutno mít k dispozici seznamy lexémů s těmito valenčními vlastnostmi:

Seznamy sloves:

- a) slovesa s jediným hloubkově obligatorním předmětem, a to v dativu: *čelit-3, podobat-3 se, odolat-3, holdovat-3...*
- b) slovesa se dvěma předměty ve svém valenčním rámci: hloubkově obligatorním přímým akuzativním předmětem a hloubkově obligatorním nepřímým dativním předmětem: *darovat-34, umožnit-34, ...*
- c) slovesa s fakultativním volným dativem ve funkci předmětu/adverbiále: *nalít (nalil jí-3 skleničku), obrátit (obrátil zajatci-3 kapsy)*
- d) slovesa nepřipouštějící dativ ve svém okolí: *bolet-3, nenávidět-3, použít-3, pochopit-3, snažit-3 se, týkat-3 se* a další.

Pro **adjektiva** je nutno mít obdobné typy seznamů jako u sloves, pro **adverbia** nepřipadá v úvahu typ b) výše.

Seznamy substantiv:

- a) substantiva s fakultativním dativním přívlastkem: *přednost-3, pomoc-3, pocta-3, služba-3, ...*
- b) substantiva nepřipouštějící dativ ve svém okolí: *naděje-3, chápání-3, aplikace-3, obchod-3, ...*

Máme-li k dispozici uvedené seznamy, dá se jich – jak ukážeme – využít v pravidlech, omezíme se však jen na slovesa.

I. Negativní pravidla

Uveďme nejprve jistá (= neheuristická) negativní disambiguační pravidla JN1–JN3:

Pravidlo JN1. Nechť určované substantivum v možném dativu (dále N?3) v klauzi není předložkové. Nechť navíc v klauzi jsou ze sloves, adjektiv, substantiv a adverbii jen tvary lexémů nepřipouštějících dativ jako své doplnění.

Pak N?3 není v dativu.

Příklad 1

- (1) Je možné použít **termínu-23** transdukce.
(2) Svěřil se mi, že pohrdá **každým-73**, kdo...

Pravidlo JN2. Mějme v klauzi (její začátek, resp. konec je vyznačen jako <K>, resp. </K>) konfiguraci:

<K> ... Vne-3 N?3 VA-ne3* </K>

přičemž platí:

- a) necht' Vne-3 je sloveso, které nepřipouští dativ
b) v posloupnosti VA-ne3* se nikde za N?3 se nenachází tvar slovesného/adjektivního lexému připouštějícího dativ jako své doplnění ani předložka z množiny POSTP = {*napospas, naproti, navzdory, vstříc*} (jsou to předložky s dativní valencí, jež mohou stát za svým jménem).

Pak N?3 není v dativu.

Příklad 2

- (3) Naznačil, že je ovládána **velrybáři-73**.

Pravidlo JN3. V češtině neexistuje slovo, které by mělo ve svém rámci dvě dativní doplnění. Vycházejme tedy z těchto negativních trigramů:

W3 VAND N?3
N?3 VAND W3

kde:

- W3 je slovo v jednoznačném dativu
- N?3 je slovo, které může být v nepředložkovém dativu
- VAND je sloveso/adjektivum/substantivum/adverbium

Proto platí pravidla **JN3a** a **JN3b**:

Pravidlo JN3a. Necht' klauze K má jednu z těchto struktur:

<K> ne3* W3 VAND N?3 ne3* </K>

<K> ne3* N?3 VAND W3 ne3* </K>

a navíc:

- a) N?3 není v předložkovém dativu
- b) W3 je slovo v jistém nepředložkovém dativu, ale nikoli tvar *si*
- c) v posloupnosti V-ne3* není ani před W3, ani za N?3 tvar lexému připouštějícího dativ jako své doplnění ani předložka z množiny POSTP.

Pak N?3 není v dativu.

Příklad 3

(4) *Hůl mu3 ulehčí **chůzi-34** a dalekohled není tak těžký.*

(5) *Právě její neumělosti a **neznalosti-13** jsou mi-3 drahé.*

Pokud ovšem nejsou splněny uvedené podmínky, N?3 může být v dativu:

(6) *Podářilo se mu-3 ubránit **touze-3** přivinout ji těsně k sobě...*

(7) *Musel odolávat **pokušení-3** vysvětlit jim-3, jak...*

Uvedené podmínky je však ještě zpřísnit, neboť je nutno vyloučit komparativní konstrukce s lexémy *jako, než, srov.:*

(8) *Věděla jsem, že jako Židovi-3 nebude **tatínkovi-3** dovoleno podnikat...*

(9) *Navíc mému **synovi-3** nebylo samotnému-3 ještě šestnáct, když...*

II. Pozitivní pravidla

Na rozdíl od pravidel negativních přisoudí následující pozitivní pravidla substantivu dativní interpretaci. Vycházíme tu hlavně z valenčních seznamů sloves, která ve svém rámci obsahují

- buď jediný přímý předmět v obligatorním dativu, např. *čelit, holdovat...*
- nebo dva předměty, přímý a nepřímý, z nichž nepřímý předmět je v dativu, např. slovesa *nabídnout, vnutit*.

Pravidlo JP1. Necht' klauze K má jednu z těchto struktur:

<K> ne-np3* V3 ne-val3* N?3 ... </K>
<K> ne-np3* N?3 ne-val3* V3 ... </K>

a navíc necht' jsou splněny tyto podmínky:

- a) V3 je sloveso mající ve svém rámci obligatorní přímý předmět v dativu
- b) N?3 je substantivum v potenciálním nepředložkovém dativu
- c) posloupnost ne-np3* neobsahuje nepředložkový dativ
- d) posloupnost ne-val3* neobsahuje nepředložkový dativ ani tvar lexému s obligatorní/fakultativní dativní valencí.

Pak N?3 je (s velmi vysokou pravděpodobností!) v dativu.

Příklad 4

(10) *Nejtěžší je odolat **chuti-32** na cigaretu.*

Po bezpečných pravidlech uvedeme nyní pravidla heuristická, přičemž vycházíme z následující teze:

Obecná heuristická teze. Je-li v blízkosti slovesa vyžadujícího přímý předmět v dativu substantivum typu N?3 v morfologicky možném dativu, pak N?3 je skutečně v dativu.

Pravidlo JPHeu1. Necht' klauze K má tuto strukturu:

<K> V-ne3* N4 V34 N?3 V-ne3* </K>

a necht' navíc:

(a) N4 je substantivum v jednoznačném nepředložkovém akuzativu, které nevyjadřuje čas, míru ani způsob

(b) V34 je sloveso s akuzativní i dativní valencí

(c) v posloupnosti V-ne3* není ani před N4, ani za N?3 tvar lexému připouštějícího dativ jako své doplnění, ani předložka z množiny POSTP.

Pak N?3 je (s velmi vysokou pravděpodobností!) v dativu.

Příklad 5

(11) *Svůj veškerý životní zájem-4 věnovala **ochraně-36** práv zvířat.*

III. Heuristická negativní pravidla

Negativní heuristická disambiguační pravidla jsou motivována především následující úvahou: v klauzi s velkou pravděpodobností nejsou v těsném sousedství dvě substantiva v dativu, která netvoří syntagma. Jinými slovy: jsou-li vedle sebe dvě substantiva v dativu, tvoří velmi pravděpodobně (adordinační) syntagma.

Pravidlo HN. Mějme v klauzi negativní bigram čili konfiguraci:

N3 N?3 nebo N?3 N3

kde:

a) N3 je substantivum v jistém dativu

b) N?3 je substantivum v možném dativu

c) substantiva N3 a N?3 se neshodují:

- buď ve jmenném rodě,
- a/nebo v čísle.

Pak substantiva N3 a N?3 velmi pravděpodobně netvoří syntagma, a tedy N?3 ve skutečnosti není v dativu.

Příklad 6

(12) *Chce zachránit synovi-3 **kůži-34**.*

Pravidlo **HN** však má své výjimky, kdy sousední substantiva tvoří adordinační syntagma, ale neshodují se buď v rodě a/nebo v čísle:

(13) v [řece-3fem **Jordánu-3mascinan**]

(14) *svému [spojenci-3mascinan **Francii-4fem**]*

6. Závěr

V tomto příspěvku byla představena hlavní disambiguační pravidla, jež na základě valenčních vlastností slov, zejména sloves, negativně i pozitivně disambiguují substantiva v morfologicky potenciálním dativu.

Literatura

Bartoň, T. et al. (2009): *Statistiky češtiny*. Praha: Nakladatelství Lidové noviny.

Jelínek, T. (2012): *Forma a funkce substantiv v češtině: vztah pádu a syntaktické funkce. Na materiálu korpusu současné psané češtiny (SYN2005)*. Disertační práce. Praha: FF UK.

Šmilauer, V. (1966): *Novočeská skladba*. Praha: SPN.