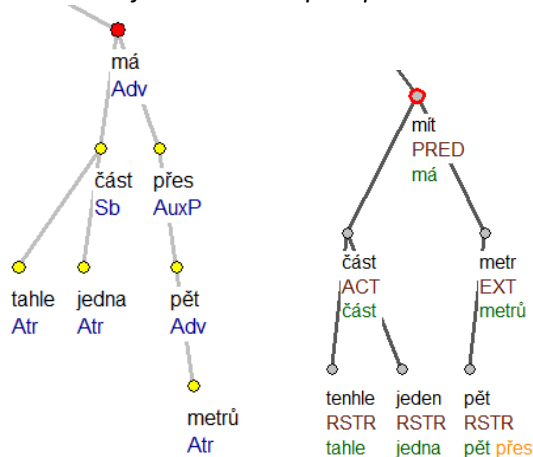


Číslovkové výrazy

Při anotaci číslovek na a-rovině usilujeme o zachycení formálních syntaktických vztahů. Syntaktická struktura číslovkových výrazů se proto může lišit od sémantického zachycení těchto výrazů na t-rovině (1).

Př. 1: *tahle jedna část má přes pět metrů*



1. Základní pravidla anotace číslovkových výrazů

U číslovkového výrazu lišíme dvě části:

- (i) **číslovku**, tj. část vyjadřující počet;
- (ii) **počítané jméno**, tj. část vyjadřující počítané věci.

Pravidlo 1. Určení řídicího členu ve spojení číslovky a počítaného jména.

Počítané jméno je řídicím členem spojení vždy, když jeho tvar odpovídá syntaktické pozici celého číslovkového výrazu. Číslovka pak funguje jako přívlastek. Srov. příklady v (2a).

Pokud tvar počítaného jména neodpovídá syntaktické pozici celého číslovkového výrazu, ale počítané jméno je v genitivu, pak počítané jméno závisí na číslovce (na kořenu číslovky; viz dále pravidlo 4); syntaktické pozici odpovídá tvar číslovky. Srov. příklady v (2b).

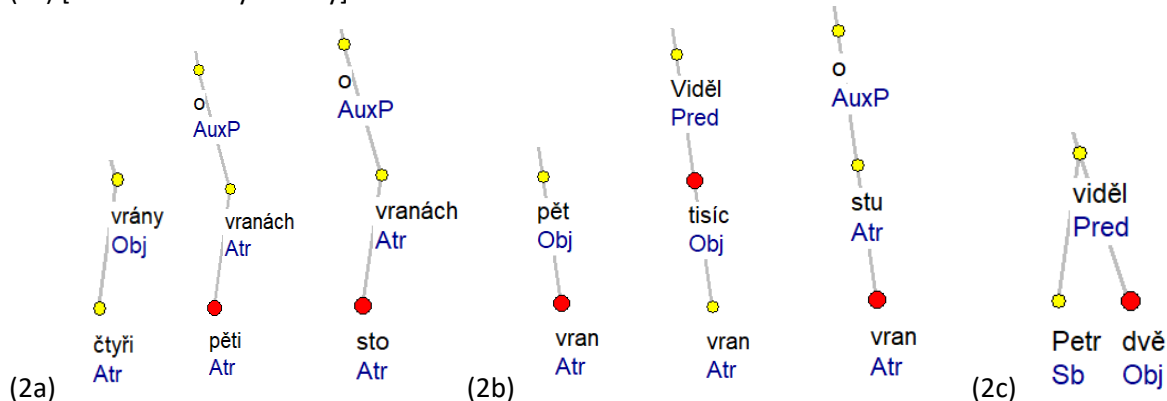
Poznámka: Pozor na případy, kdy syntaktická pozice odpovídá genitivu.

Pravidlo 2. Číslovka bez počítaného předmětu. Pokud je ve větě číslovka bez počítaného předmětu, chápeme ji vždy substantivně (tj. neanalyzujeme v takových případech elipsu počítaného jména). Srov. poslední příklad v (2c).

Př. 2: (2a) *čtyři vrány; o pěti vranách, o sto vranách;*

(2b) *pět vran; viděl tisíc vran; o stu vran;*

(2c) [Honza viděl čtyři vrány] *Petr viděl dvě*



Z formálního hlediska (podle chování ve spojení s počítaným jménem v různých pádech) rozlišujeme tři skupiny číslovek:

(A) Číslovky *jeden, nejjeden, dva, tři, čtyři, oba*

Ve spojení s počítaným jménem počítané jméno tvarově odpovídá syntaktické funkci a číslovka funguje jako **přívlastek shodný**.

Př. *viděl tři vrány, o třech vranách, bez třech vran*

Poznámka: Ve víceslovných číslovkách mohou být tyto číslovky užity sklonně i **nesklonně** (viz dále pravidlo 4)

Srov.: *viděl dvacet dvě vrány, dvacet jednu vránu; o dvaceti jedné vráně*

vs. *viděl dvacet dva vran, dvacet jedna vran; o dvaceti jedna vranách*

(B) Číslovky typu *pět*; tj. číslovky se dvěma tvary (*pět, dvacet, sedm, pětatřicet, několik, mnoho* ad.) a dále zřejmě některé číslovky nesklonně (např. *pár*).

V syntaktické pozici nominativu nebo akuzativu (tvar *pět*) je po těchto číslovkách počítané jméno je v genitivu plurálu. Řídící je číslovka, počítané jméno je chápáno jako závislé.

Srov.: *viděl pět vran, několik vran, pár vran*

V syntaktické pozici odpovídající ostatním pádům (tvar *pěti*) počítané jméno tvarově odpovídá syntaktické funkci; číslovka je závislá na počítaném jménu a funguje jako přívlastek shodný.

Srov.: *o pěti vranách, o několika vranách, o pár vranách; pěti vranám, několika vranám, bez pár vran*

(C) Substantivní číslovky *sto, tisíc, milión, miliarda, setina*

Číslovky chápeme jako substantiva. Počítané jméno je ve všech syntaktických pozicích v gen. pl.; řídící je číslovka, počítaný předmět je chápán jako závislý. Výjimkou je nesklonný tvar číslovky *sto* – viz pozn. níže.

Srov.: *viděl sto vran, tisíc vran, milión vran, setinu vran; o stu vran; o tisíci vran; o miliónu vran, o setině vran*

Ve víceslovných číslovkách se chovají jako počítaná jména, jejich sekvence vytváří genitivní řady (viz 5.2).

Srov.: *viděl pět miliónů, jeden milión*

Poznámka: Tvar *tisíc* se v genitivních pozicích chápe jako (nesklonný) genitiv (3).

Srov. *viděl pět tisíc*. Gen [= viděl pět tisíců] vs. *o pěti tisících*

Poznámka: Nesklonné *sto* (*o sto vranách*) – zde je číslovka závislá na počítaném jménu, které tvarově odpovídá syntaktické funkci. Číslovka je chápána jako atribut.

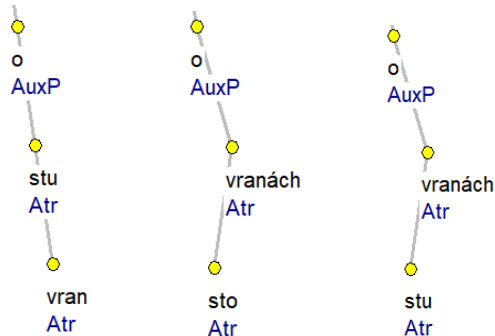
Srov. *o stu vran* vs. *o sto vranách* (4)

Poznámka: V nestandardních (chybných) spojeních jako *o tisíci vranách, o stu vranách* (na místo *o tisíci vran, o stu vran*) postupujeme podle pravidla 1: řídící je počítané jméno, které tvarově odpovídá syntaktické funkci; číslovka je zachycena jako atribut (4).

Př. 3: vydělal pět tisíc



Př. 4: o stu vran vs. o sto vranách vs. o stu vranách



2. Víceslovné/víceřádové číslovky

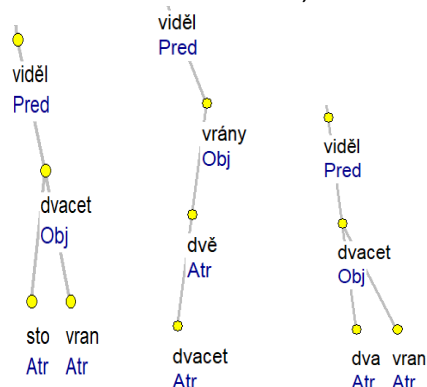
Při určování, zda řídícím členem číslovkového výrazu bude číslovka nebo počítané jméno, se řídíme pravidlem 1. Následující pravidla se týkají struktury víceslovné/víceřádové číslovky.

Pravidlo 3. Číslovka tvoří jeden podstrom.

Pravidlo 4. Určení kořenu podstromu číslovky. Kořenem podstromu číslovky je **nejnižší vyjádřený řád**. Srov.: *viděl sto dvacet vran*; *viděl dvacet dvě vrány* (5).

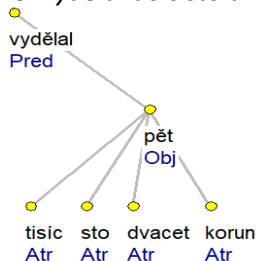
Výjimkou jsou nesklonně užitá číslovky *jedna, dva, tři, čtyři*, které nezachycujeme jako kořen číslovky, ale jako závislé na kořenu číslovky (skupiny/řádu), kterým je v tomto případě až druhý nejnižší vyjádřený řád. Srov.: *viděl dvacet dva vran* (5).

Př. 5: *viděl sto dvacet vran*; *viděl dvacet dvě vrány*; *viděl dvacet dva vran*



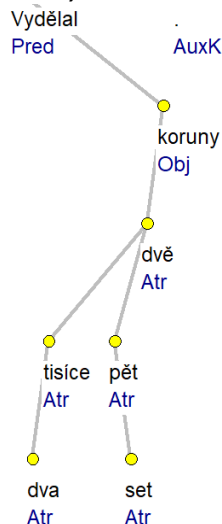
Pravidlo 5. Závislost ostatních částí číslovky. Ostatní části číslovky **závisí na kořenu číslovky** po skupinách/po řádech **jako sestry** a mají afun *Atr* (6).

Př. 6: *vydělal tisíc sto dvacet pět korun*



Pravidlo 6. Struktura skupin/řádů. Struktura jednotlivých řádů (stovek, tisíců, miliónů) se posuzuje pro každou skupinu zvlášť podle pravidel 1-5 (7).

Př. 7: *Vydělal dva tisíce pět set dvě koruny*



Příklady: Desítky a jednotky

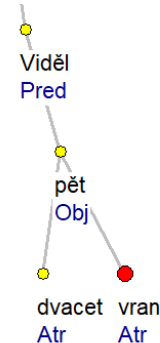
Př. 8: Viděl dvacet dvě vrány



Př. 9: Viděl dvacet dva vran



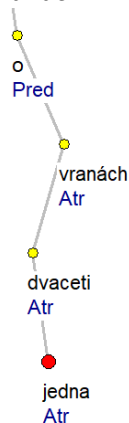
Př. 10: Viděl dvacet pět vran



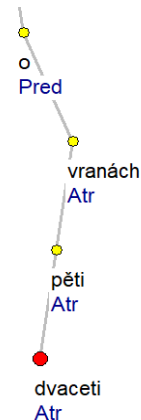
Př. 11: o dvaceti jedné vráně



Př. 12: o dvaceti jedna vranách

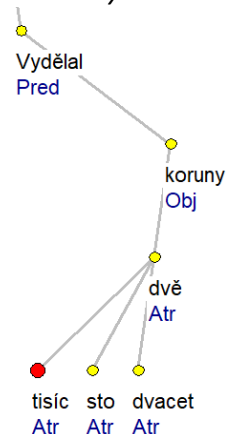


Př. 13: o dvaceti pěti vranách

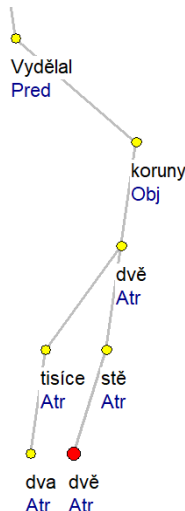


V příslušném pádu syntaktické pozice je počítané jméno

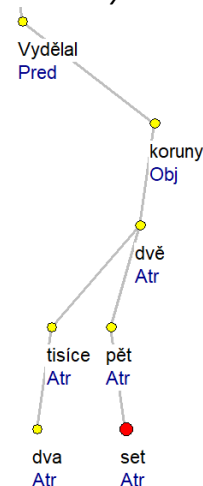
Př. 14: Vydělal tisíc sto dvacet dvě koruny.



Př. 15: Vydělal dva tisíce dvě stě dvě koruny.

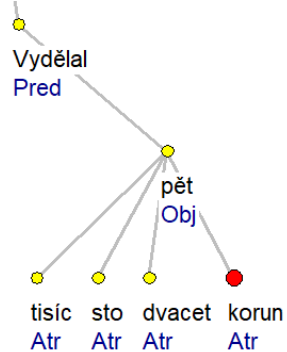


Př. 16: Vydělal dva tisíce pět set dvě koruny.

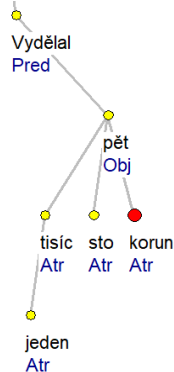


Počítané jméno není v příslušném pádu syntaktické pozice (ale je v genitivu)

Př. 17: Vydělal tisíc sto dvacet pět korun.

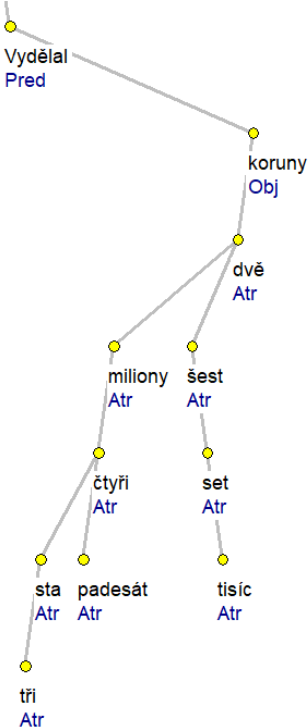


Př. 18: Vydělal jeden tisíc sto pět korun.

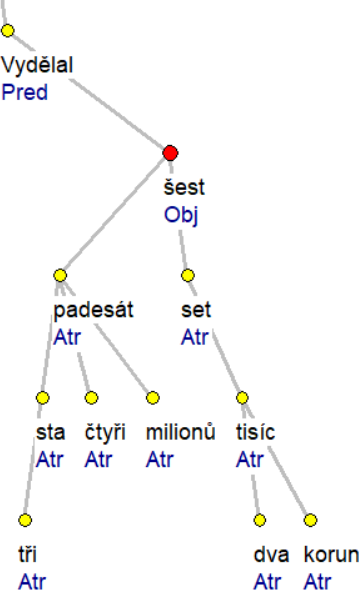


Víceslovné/víceřádové číslovky

Př. 19: Vydělal tři sta padesát čtyři miliony šest set tisíc dvě koruny. (354 600 002)



Př. 20: Vydělal tři sta padesát čtyři milionů šest set tisíc dva koruny. (354 600 002)



3. Číslovky psané číslicemi

Číslovky psané číslicemi visí ve stromě na tom místě, kde by visel **kořen číslovky zapsané slovy**.

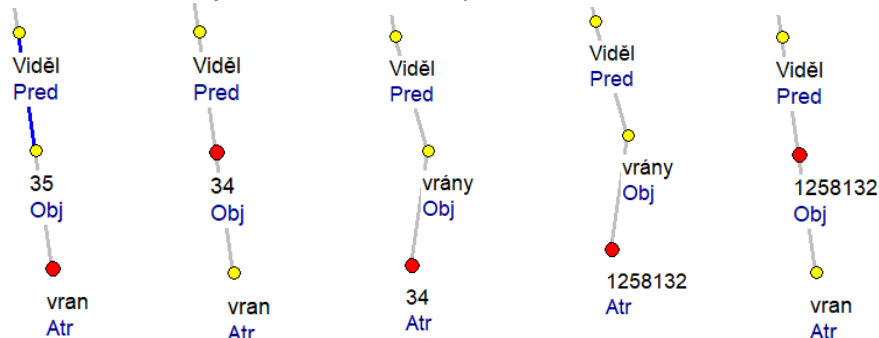
Pro správné „čtení“ číslovky zpravidla pomáhá následující počítané jméno. Srov. (21).

Př. 21: *viděl 35 [třicet pět] vran*

viděl 34 [třicet čtyři] vran vs. *viděl 34 [třicet čtyři] vrány*

viděl 1 258 132 [jeden milion dvě stě padesát osm tisíc sto třicet dvě] vrány

vs. *viděl 1 258 132 [jeden milion dvě stě padesát osm tisíc sto třicet dva] vran*



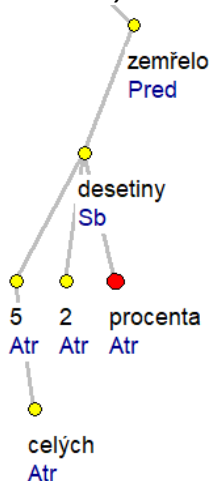
4. Desetinná čísla a zlomky

(a) vypsaná slovy

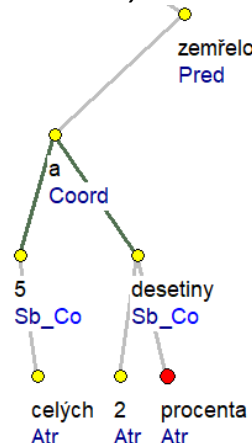
Je celá řada možností, jak desetinné číslo nebo zlomek číst a zapsat slovy, avšak desetinná čísla a zlomky se slovy běžně nepiší. Při anotaci pro celý zápis desetinného čísla vybereme jedno slovo jako kořen (pravidlo 3) a v souladu s pravidly 4-6 na něj zavěsíme po skupinách další části jako sestry. Kořen je zpravidla dílová číslovka (*pětina*, *setina*), která je substantivní (na morfologické rovině má substantivní tag), následující počítané jméno v genitivu visí ve stromě na této číslovce.

Srov. (22) a (23).

Př. 22: *5 celých 2 desetiny procenta zemřelo*



Př. 23: *5 celých a 2 desetiny procenta zemřelo*



(b) psané číslicemi

Desetinné číslo psané číslicemi je reprezentováno jako jeden uzel. U zlomku psaného s lomítkem je každá část samostatným uzlem. Zlomky a desetinná čísla psaná číslicemi a znaky zachycujeme analogicky k odpovídajícím strukturám psaným slovy. Srov. (24) – (27).

Poznámka: Matematické operace jako sčítání, odčítání, dělení, tedy i „lomeno“ jsou na t-rovině zachyceny jako koordinace se speciální funkcí OPER. Srov. (27).

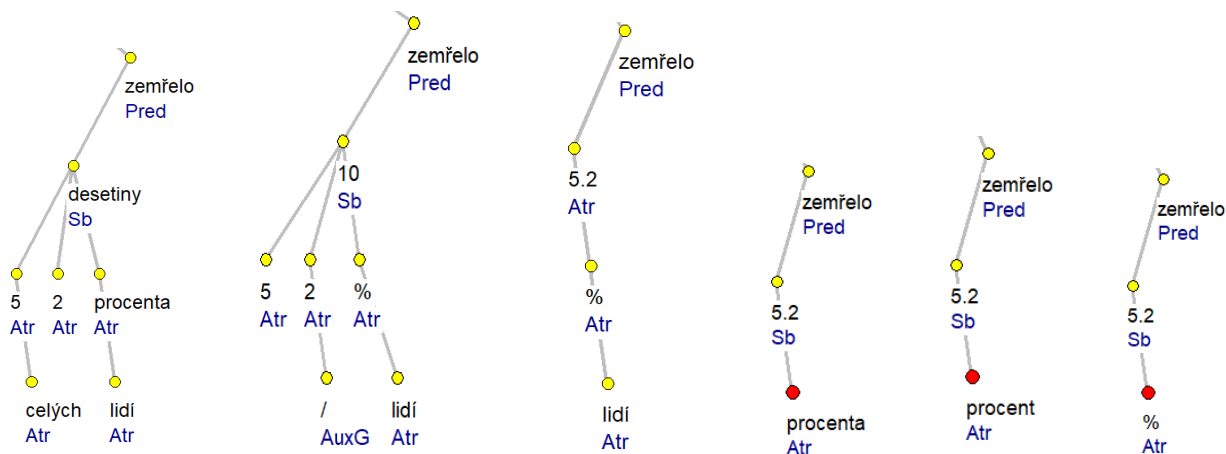
Př. 24: 5 celých 2 desetiny procenta lidí zemřelo

5 2/10 % lidí zemřelo

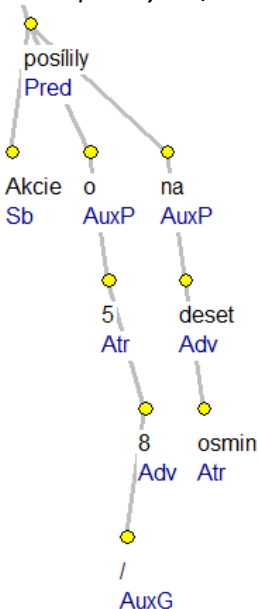
5.2 % lidí zemřelo

5.2 procenta zemřelo

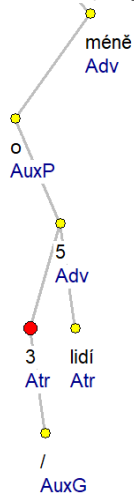
5.2 procent zemřelo



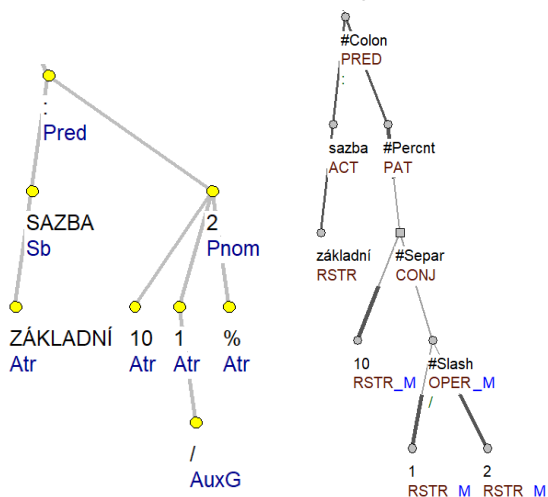
Př. 25: Akcie posílily o 5/8 na deset osmin.



Př. 26: o 3/5 [tři pětiny] lidí méně



Př. 27: ZÁKLADNÍ SAZBA: 10 1/2 %



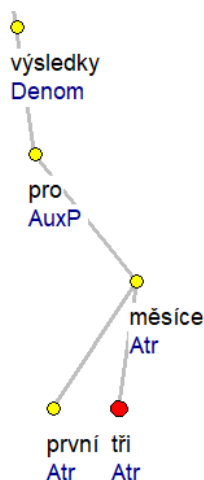
5. Další pravidla

5.1 Rozvíjení číslovkového výrazu přívlastkovým adjektivem

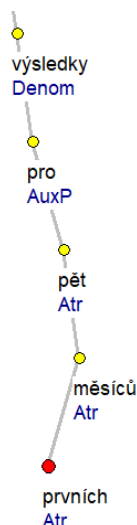
Adjektivní přívlastek rozvíjející celé spojení číslovky a počítaného jména se v některých případech neshoduje se slovem, které se podle pravidla 1 a 3 vybere jako řídící člen celého spojení. Stává se tak zejména v případech, kdy řídícím členem celého spojení jsou číslovky typu *pět* apod. (i *tisíc* a *sto*) v nominativu a akuzativu. Tyto číslovky „neumí přidělit shodu“ adjektivu, adjektivní přívlastek se shoduje až s počítaným jménem. Adjektivní přívlastek ve stromě vždy závisí na jménu, se kterým se shoduje v rodě, čísle a pádu. Srov. (30) – (32).

Srov.: *výsledky pro první tři měsíce* vs. *výsledky pro prvních pět měsíců*

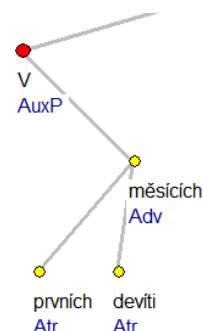
Př. 30: *výsledky pro první tři měsíce*



Př. 31: *výsledky pro prvních pět měsíců*



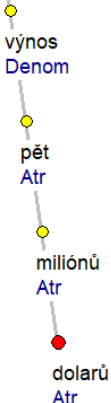
Př. 32: *V prvních devíti měsících*



5.2 Genitivní řady

V genitivních řadách (substantivních číslovkách a počítaných jmen) je pak obecně těžké určit závislosti (k čemu který genitiv patří). V nejasných případech je řešením postupná závislost u číslovkových substantiv (*milión*, *tisíc*, *sto* ad. a v nominativu a akuzativu i číslovky typu *pět*) a počítaného jména. Srov. (33) – (35).

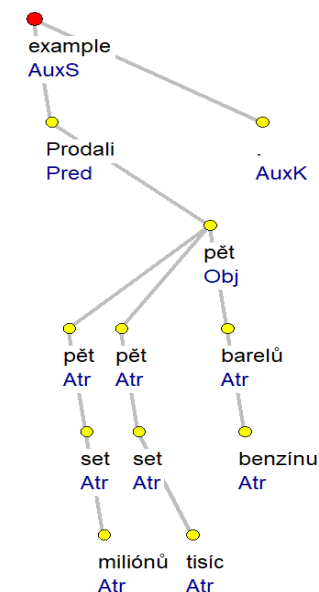
Př. 33: *výnos pět milionů dolarů*



Př. 34: *bez pěti milionů dolarů*



Př. 35: *Prodali pět set milionů pět set tisíc pět barelů benzínu.*

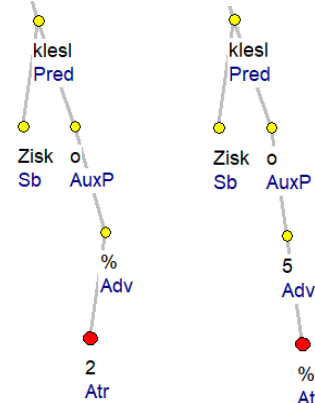


5.3 Znak nebo zkratka zastupující počítané jméno (% , €, \$, Kč, USD)

Znak nebo zkratka zastupující počítané jméno (% , €, \$, Kč, USD) visí ve stromě v místě, kde by viselo slovo, které znak nebo zkratka zastupuje. Srov. (36), (37) a (38).

Př. 36: *zisk klesl o 2 %* [dvě procenta]

vs. *zisk klesl o 5 %* [pět procent]

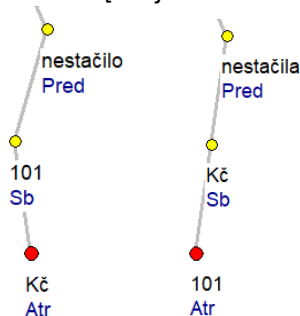


Př. 37: *vystačili s 90 \$* [s devadesáti dolary]



Př. 38: *101 Kč* [sto jedna korun] *nestačilo*

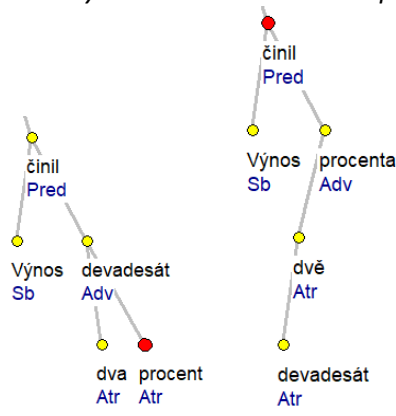
vs. *101 Kč* [sto jedna koruna] *nestačila*

**5.4 Více možných „čtení“ číslovkového výrazu**

V některých případech může být sporné, jak číslovkový výraz (zapsaný číslicemi a znaky) "číst", a tedy jaká má být struktura. Může si konkurovat:

(i) **sklonné** a **nesklonné čtení** (se stejným významem). Rozdíl ve „čtení“ může, ale nemusí mít vliv na strukturu stromu. Pokud má „čtení“ číslovkového výrazu dopad na strukturu (srov.: *výnos činil 92 %* = [devadesát dvě procenta] vs. [devadesát dva procent]), pak je při anotaci upřednostněna **struktura pro „sklonné čtení“** (39).

Př. 39: *výnos činil devadesát dva procent; výnos činil devadesát dvě procenta; výnos činil 92 %*



(ii) „genitivní“ a „akuzativní“ čtení, tj. význam „čeho?“ a „kolik“. Anotaci volíme podle smyslu (výsledná struktura opět může, ale nemusí být rozdílná). Při nejasném významu volíme strukturu, ve které počítané jméno zachytíme jako řídicí člen spojení.

Srov. *výnos 400 [čtyř set] fondů* (výnos čeho? Gen) vs. *výnos 400 [čtyři sta] fondů* (výnos kolik? Akuz) - stejná struktura (spíše význam „výnos čeho“); (40).

Srov.: *vklad 5 USD* [pěti dolarů] (*vklad čeho?* Gen) vs. [pět dolarů] (*vklad kolik?* Akuz) - rozdílná struktura (možné obojí „čtení“); (41).

Srov.: *Spadl z výšky 5 m* [pět metrů] vs. [pěti metrů]

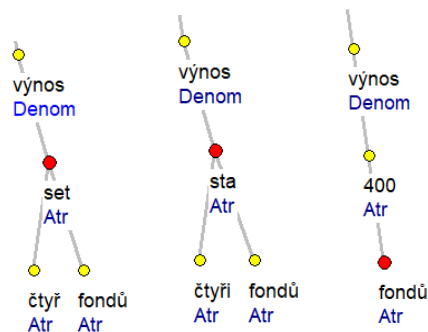
- rozdílná struktura (možné obojí „čtení“); (42).

Srov. též příklady (5a) a (5b) v příloze.

Př. 40: *výnos čtyř set fondů*

vs. *výnos čtyři sta fondů*

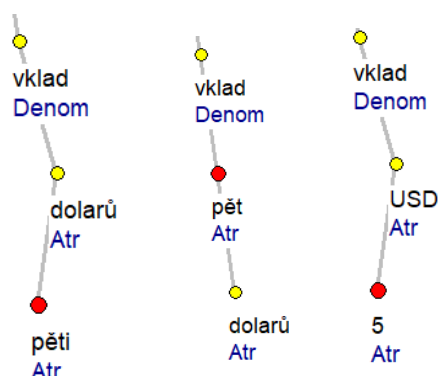
vs. *výnos 400 fondů* [čtyř set]



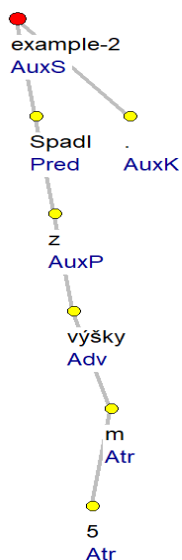
Př. 41: *vklad pěti dolarů*

vs. *vklad pět dolarů*

vs. *vklad 5 USD*

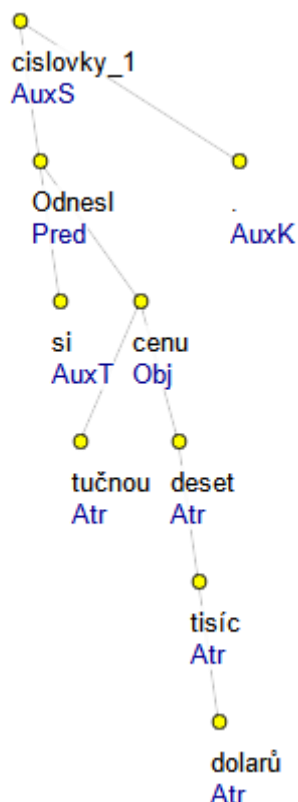


Př. 42: *Spadl z výšky 5 m.*

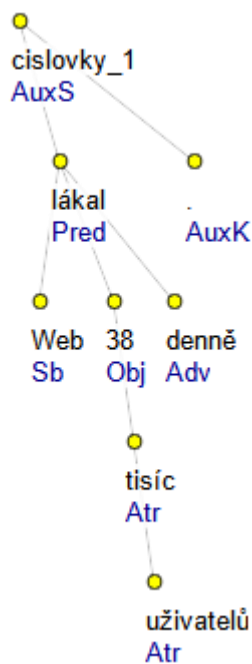


Příloha A. Příklady z korpusu

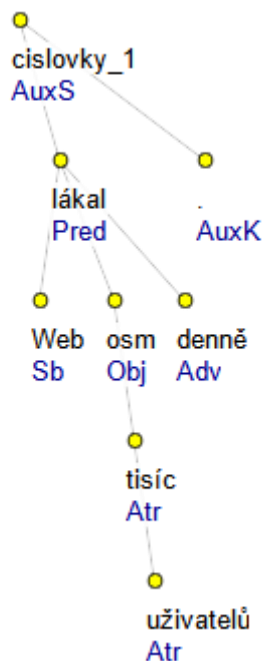
1. Odnosl si tučnou cenu deset tisíc dolarů.



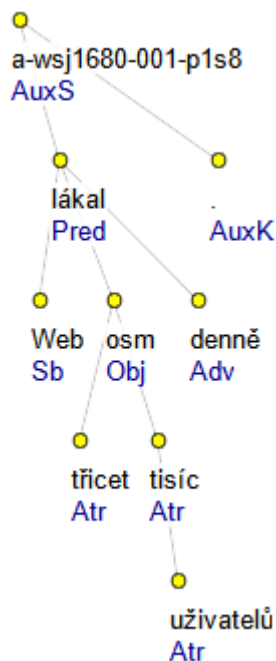
2a. Web lákal 38 tisíc uživatelů denně.



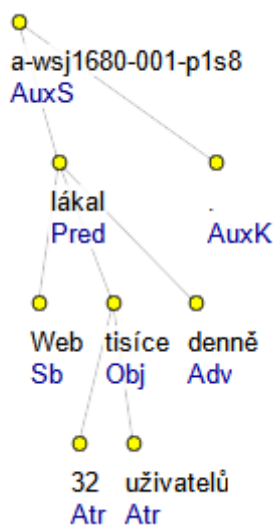
2b. Web lákal osm tisíc uživatelů denně.



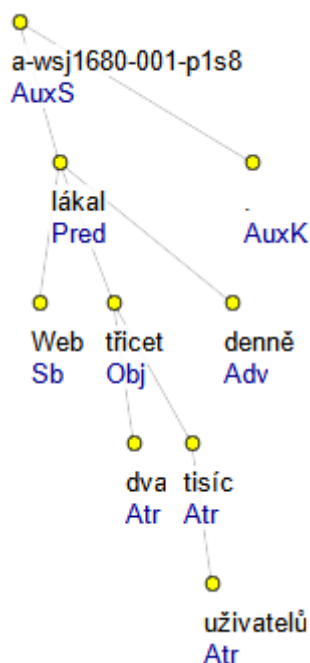
2c. Web lákal třicet osm tisíc uživatelů denně.



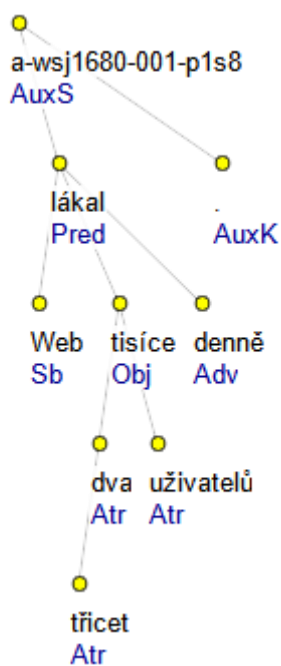
3a. Web lákal 32 tisíce uživatelů denně.



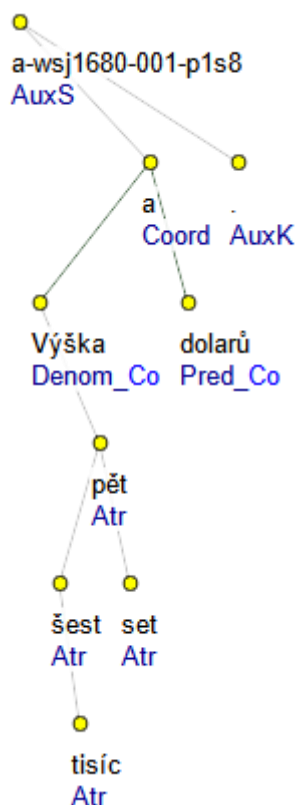
3b. Web lákal třicet dva tisíc uživatelů denně.



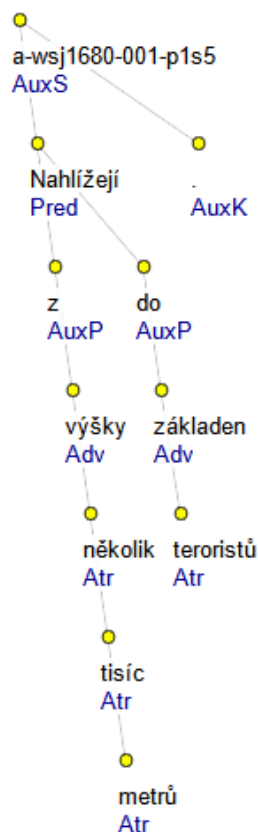
3c. Web lákal třicet dva tisíce uživatelů denně.



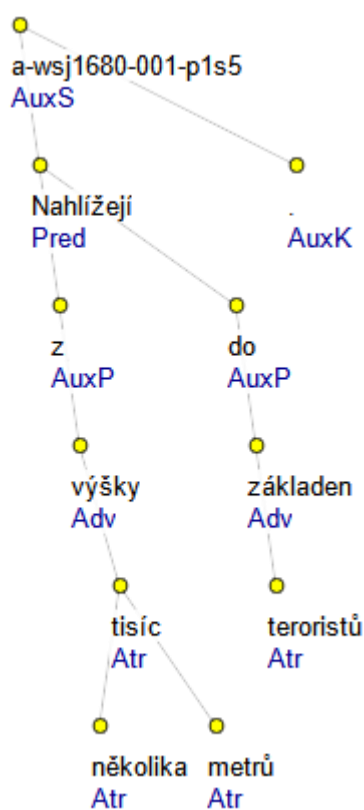
4. (Tlak paliva klesá. Tlak oleje klesá.) Výška šest tisíc pět set a klesá.



5a. Nahlížejí z výšky několik tisíc metrů do základen teroristů



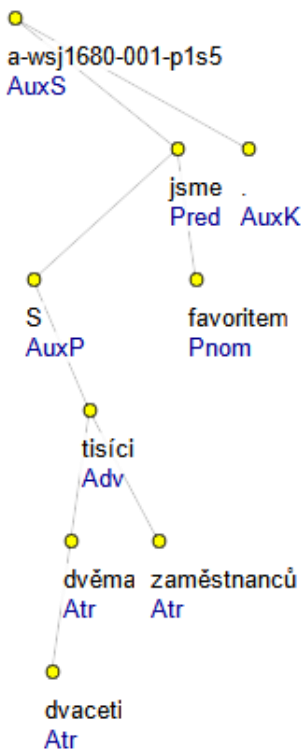
5b. Nahlížejí z výšky několika tisíc metrů do základen teroristů.



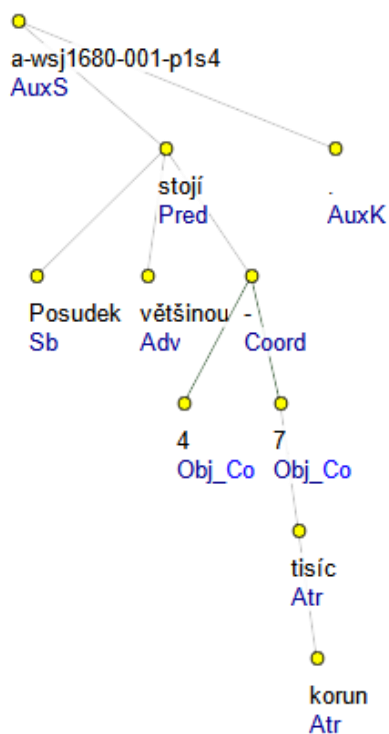
6. Země letos přijme 37,5 tisíce žadatelů o azyl



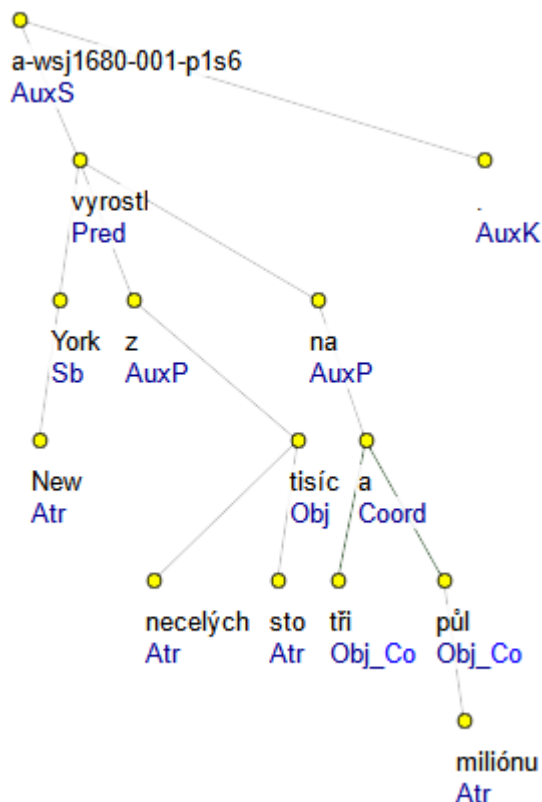
7. S dvaceti dvěma tisíci zaměstnanců jsme favoritem



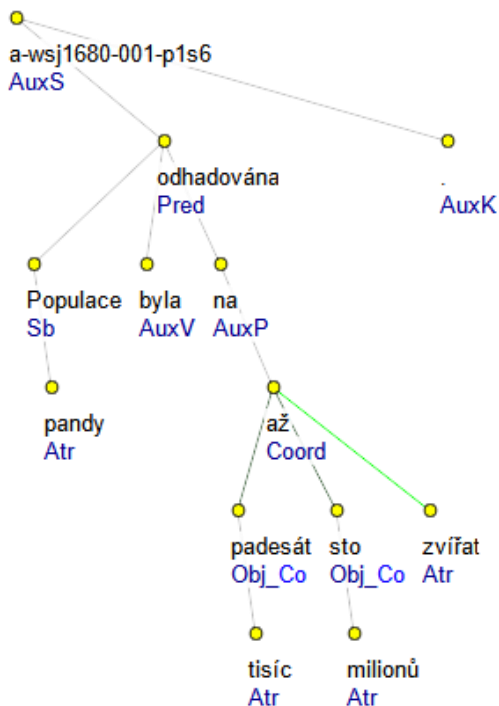
8. Posudek většinou stojí 4 - 7 tisíc korun.



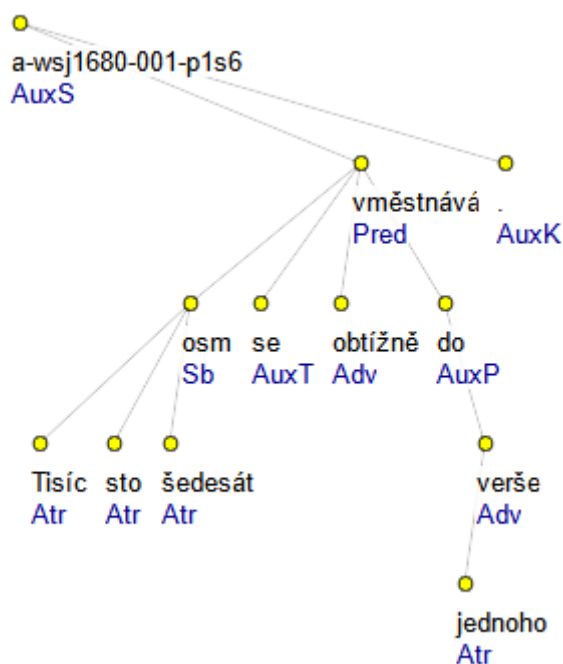
9. New York vyrostl z necelých sto tisíc na tři a půl miliónu.



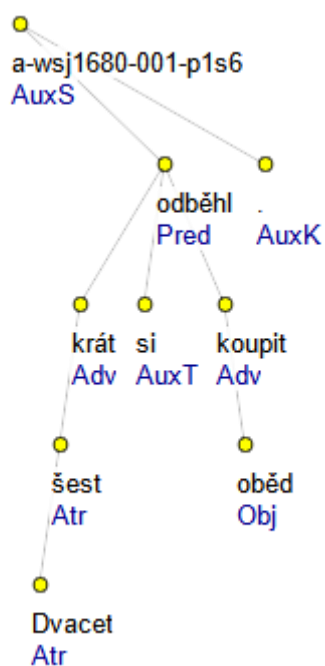
10. Populace pandy byla odhadována na padesát tisíc až sto milionů zvířat.



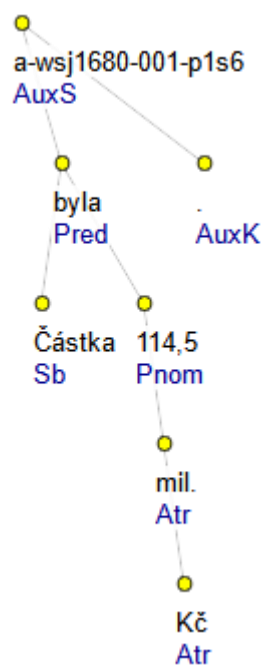
11. Tisíc sto šedesát osm se obtížně vměstnává do jednoho verše.



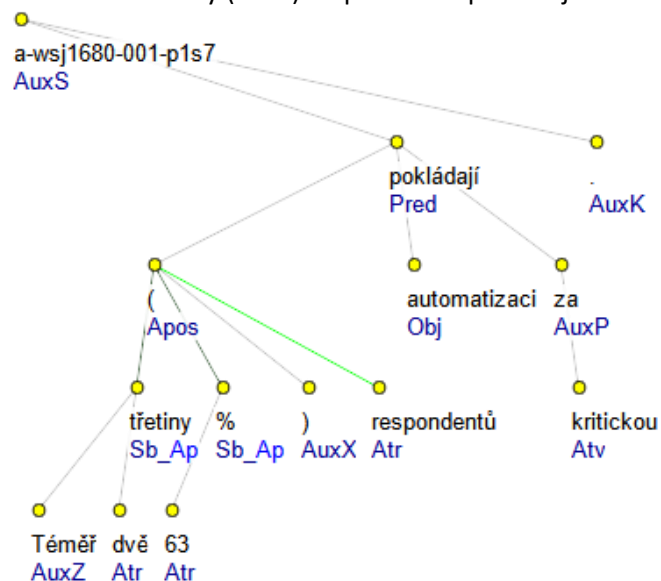
12. Dvacet šestkrát si odběhl koupit oběd.



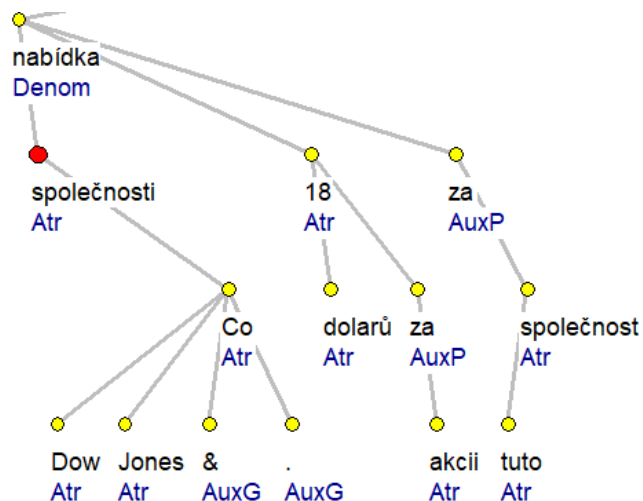
13. Částka byla 114,5 mil. Kč.



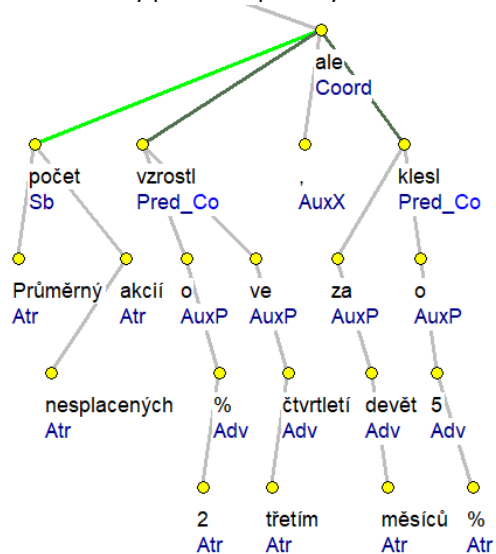
14. Téměř dvě třetiny (63 %) respondentů pokládají automatizaci za kritickou. [= 63 procenta]



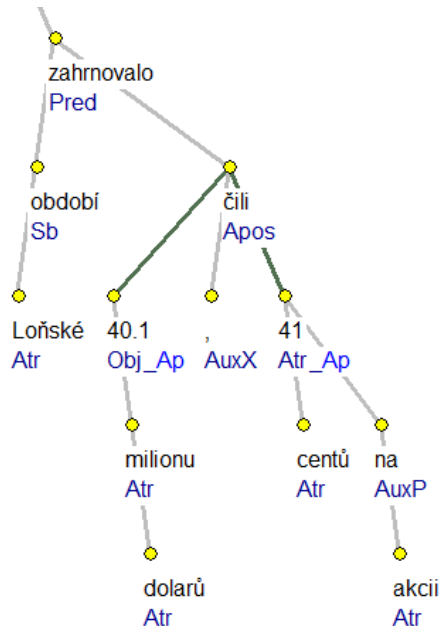
15. nabídka společnosti Dow Jones & Co. 18 dolarů [=18 dolarů] za akcii za tuto společnost



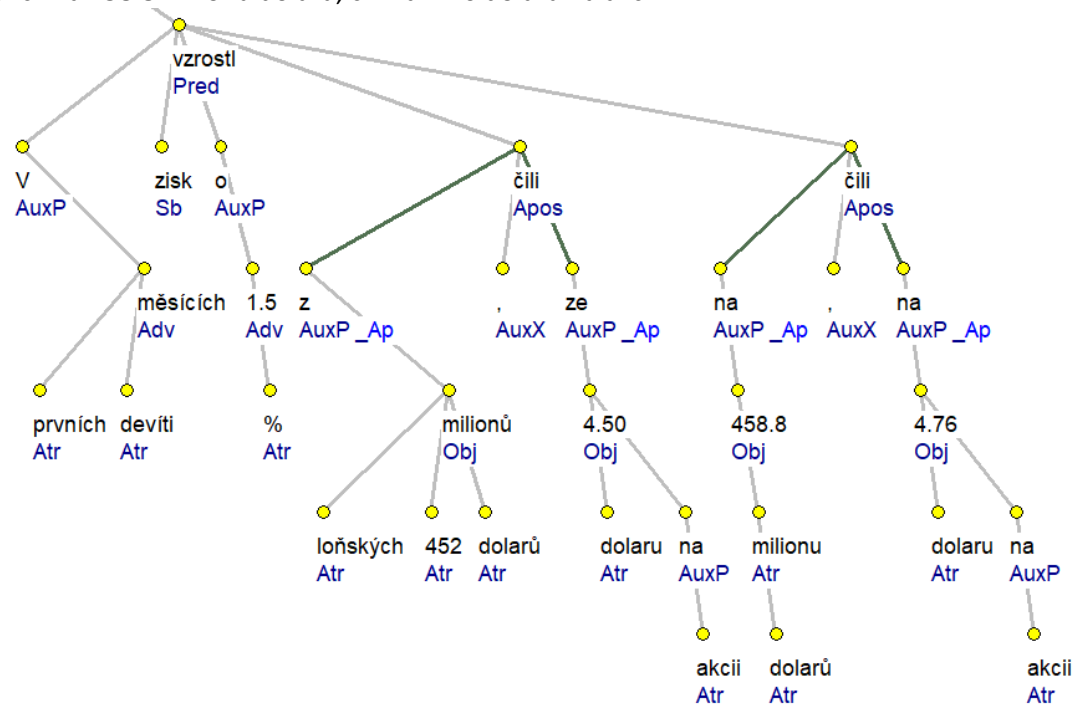
16. Průměrný počet nesplacených akcií vzrostl o 2 % ve třetím čtvrtletí, ale za devět měsíců klesl o 5 %



17. Loňské období zahrnovalo 40.1 milionu dolarů, čili 41 centů na akcii



18. V prvních devíti měsících zisk vzrostl o 1.5 % z loňských 452 milionů dolarů, čili ze 4.50 dolaru na akcii na 458.8 milionu dolarů, čili na 4.76 dolaru na akcii



Příloha B. Morfologie číslovek

Semantics & Morphological behavior		Adjectival		Nominal		Non-gendered		Adverbial	
Subtype		Tag	Example	Tag	Example	Tag	Example	Tag	Example
Cardinal	definite	Cn	<i>jeden</i>	Cz	<i>sto</i>	C1	<i>tři</i>	-	-
	indefinite	Cy	<i>nejeden</i>	-	-	Ca	<i>několik</i>	-	-
Ordinal	definite	Cr	<i>první</i>	-	-	-	-	-	-
	indefinite	Cw	<i>kolikátý</i>	-	-	-	-	-	-
Collective	definite	Cd	<i>dvoje</i>	Cj	<i>patero</i>	-	-	-	-
	indefinite	Ch	<i>kolikery</i>	Ck	<i>kolikero</i>	-	-	-	-
Multiplicative	definite	-	-	-	-	-		Cv	<i>tříkrát</i>
	indefinite	-	-	-	-	-		Co	<i>tolikrát</i>
Tag categories		GENDER NUMBER CASE (VAR)		GENDER NUMBER CASE (VAR)		CASE (NUMBER) (VAR)		(VAR)	