

**Nominalization of Adjectives as an Issue in Machine Analysis
of Czech Language**

We focused on units with adjectival inflection, for which the primary function is expressing an attribute and its secondary function is expressing a subject. Analysis of corpora data showed that units with adjectival inflection which perform primarily the function of a subject (hajný, pokojská) are reliably identified in the dictionary. We also found that such units with adjectival inflection which serve primarily as an attribute and only secondarily as a subject are in most cases assigned two interpretations in the dictionary, however, when disambiguating, the interpretation is selected incorrectly (strážný, výchovný). This type of expression is interpreted considerably erratically in the corpora. We propose such a solution to this problem that would enable proper disambiguation of units with adjectival inflection, but would at the same time help to avoid increasing homonymy in the morphological dictionary.

Keywords: nominalization of adjectives, corpora,
automatic morphological analysis

Mgr. Hana Žižková
Katedra českého jazyka a literatury
Pedagogická fakulta
Masarykova univerzita
Poříčí 7
603 00 Brno
zizkova@ped.muni.cz

SUBSTANTIVIZACE ADJEKTIV JAKO PROBLÉM STROJOVÉ ANALÝZY ČEŠTINY

HANA ŽIŽKOVÁ

Rozlišení slovních druhů je pro popis jazyka klíčové. Tradiční klasifikace slovních druhů vycházejí z posuzování tří kritérií: a) morfologického, které rozlišuje slovní druhy ohebné a neohebné, b) sémantického, pro které je zásadní význam a rozlišuje slova plnovýznamová a neplnovýznamová, a c) syntaktické, které se vyjadřuje k syntaktické funkci výrazů ve větě.

Nabízí se otázka, které z kritérií je pro vymezení slovního druhu zásadní a jak jsou podobné problémy řešené na rovině teoretické zohledněny v praxi automatické morfologické analýzy (AMA).

„Zásadní pro pochopení celé problematiky je to, že navzdory organizaci informací ve většině mluvnic, je slovnědruhovná charakteristika součástí lexicologie a morfologie z ní pouze vychází.“ (Cvrček – Kovářiková 2001: 113)

Slovní druhy (jako lexikologická klasifikace) rozdělují lexémy do předem stanovených skupin podle jejich nejobecnějšího významu. Toto rozdělení následně zjednodušuje popis jazyka, protože tyto skupiny reprezentují oblasti nejobecnějších významů, s jakými se můžeme setkat (Cvrček – Kovářiková 2001). Zároveň tyto skupiny reprezentují jakési prototypy hierarchicky uspořádaných syntaktických funkcí, které daný slovní druh může zastupovat.

Rozlišujeme primární a sekundární syntaktickou funkci,¹ primární syntaktickou funkci mají slova jen díky svému lexikálnímu významu (Komárek 2006: 19) a Komárek dále uvádí, že „v primární funkci je slovní druh nejvíce sám sebou a při plnění sekundární funkce je slovní druh zabarven sémantikou slovního druhu, který má tuto funkci jako primární“ (ibid.: 20).

Syntaktickou funkci podmětu plní primárně substantiva a syntaktickou funkci atributu plní primárně adjektiva. Pokud plní adjektivum syntaktickou funkci podmětu, bývá to proto, že z věty bylo vypuštěno podmětové substantivum. Tyto případy vedou k postupné substantivizaci adjektiv (srov. Nekula – Rusínová – Karlík 1995: 413). Pakliže se však ve výše uvedených případech elipsy jedná o záležitost syntaxe, není možné tento problém vyřešit na základě automatické morfologické analýzy.

Ideálně bychom tedy měli mít v případech jednotek s adjektivní flexí² ve slovníku uvedeny dvě slovnědruhovné interpretace. Ne vždy tomu tak však je. Navíc označení slovního druhu ve slovníku je jedna věc a věc druhá je,

/1/ Komárek uvádí i terciální funkci, avšak vzhledem k tomu, že se soustředíme na jednotky s adjektivní flexí (substantivizovaná adjektiva), primární a sekundární funkce jsou pro nás dostačující.

/2/ V našem článku se zaměříme pouze na ty jednotky s adjektivní flexí, pro které je primární funkcí vyjádření atributu a sekundární funkcí vyjádření podmětu.

jak jednotka s adjektivní flexí funguje v jazyce a realizuje primární a sekundární funkce.

Automatická morfologická analýza probíhá ve třech krocích: v prvním se jedná o rozčlenění na slovní tvary (tokenizace), ve druhém jde o přiřazení jedné, většinou však více interpretací z morfologického slovníku, a ve třetím kroku se jedná o desambiguaci (zjednoznačnění), tedy přiřazení jedné interpretace.

Analýza dat z korpusu³ ukázala, že ve slovníku jsou spolehlivě označeny ty jednotky s adjektivní flexí, které plní primárně funkci podmětu (typ *hajný*, *pokojská*).

druhy jsou jetei prostřední (Trifolium medium), černýs	hajný /NNMP1-----A-----	(Melampyrum nemorosum) – rostlina charakteristická též pro světlé
. </p><p> Novinář čekají, že pytlák, který se stal	hajným /NNMP3-----A-----	, neodolá a vrátí se ke svému původnímu povolání,
traurig bin ...! no, nechme toho. Znáte	Hajného /NNMS4-----A-----	? " </p><p> Několik studentů věstě Richarda přitakalo. </p><p> „L
náměstíčkem patřila svého času k zámeckému panství, žili tady	hajný /NNMP1-----A-----	, kteří spravovali les i nedealeko oboru. Po dvou
Vidíme ho přespávat u vetešníka Eichlera v Kuřívodch, u	hajného /NNMS2-----A-----	Josefa Vítky v myslivně u Maníkovic, v chládně za
škole ve Vsetíně. </p><p> Otec Adolfa Kubička byl lesníkem a	hajným /NNMS7-----A-----	. Sám měl proto k lesu také velmi blízko.
jem často v terénu. Například jsem s lesníkem	hajným /NNMS7-----A-----	v lese nebo na poli. Nesedím osm hodin denně
. Mártý vyrostl doslova v lese. Jeho otec byl	hajnej /NNMS1-----A---G-----	a on do Vínáckí let žil v hájence. Pobíhal
na náklady majitele. </p><p> Milan Bureš: Můj děda byl	hajný /NNMS1-----A-----	a tchán myslivec. Měli vychovaného českého fouska. Dnešní
<p> František Černý s Jaroslavem Kozlem, hlavní aktéři potyčky s	hajným /NNMS7-----A-----	u Želešic, byli obviněni také z pokusu o vraždu
; Rozdávěj ! " Táhnou ji dvě party:	hajní /NNMP1-----A-----	, kteří jdou bahnem, zatímco pěšáci táhnou po souši
nezmiňuje.) Babinský a Hoffmannovou si našli noclch u	hajného /NNMS2-----A-----	Josefa Vítky, který bydlel v myslivně nedealeko Maníkovic. </p>
budou jednou dělat naše dcery. </p><p> NÁŠ EXPERT PHDr. Martin	Hajný /NNMS1-----A-----	psycholog a psychoterapeut. </p><p> Polévky, které nejlíp zasýtlí </p>
o původu nestěší. Byl v podezření i jeden z	hajných /NNMP2-----A-----	a některé jiné osoby, že na hraběte s úmyslem

Obr. 1: hajný, SYN2015

strávil několik nocí ve vyšetřovací vazbě kvůli obvinění ze znečištění	pokojská /NNFS2-----A-----	v luxusním newyorském hotelu, Francouzský socialista, který patřil
svého existenciálního postavení, takže při troše štěstí ušetřím vinou	pokojskou /NNFS4-----A-----	nepřijemného překvapení. Vidíš, já opravdu myslím na prosté
potajmu vafší, nesmíš uschovávat v kufru. Tam ji	pokojská /NNFS1-----A-----	objeví. Lépe je odklidit plotnyku do skříně na šaty
pomocná síla v laboratorii, a našla si práci jako	pokojská /NNFS1-----A-----	v hotelu Palace, kde ji hosté nechávali na nočních
, hraběnky umývaly v hotelových kuchyních nádobí a pracovaly jako	pokojské /NNFS2-----A-----	v Berlíně, Paříži, Římě a nakonec i
, které Mills očekával. Pak zavraždili Ralfa, zranili	pokojskou /NNFS4-----A-----	a ukradli obraz. " Zmlkla a po chvíli se
přesel k bílé pohovce, na níž ta příšerně nesmělá	pokojská /NNFS1-----A-----	jenom obrátila polštáře, aby skryla mokré skvrny, a
splněno, aféra italskou legendu politiky zničila. </p><p> VYŘÍZENO.	Pokojská /NNFS1-----A-----	Nassifatou Diallová obvinila šéfa MFF ze sexuálního útoku. Strauss-K
uklizen. Chvilu jsem přemýšlel, že bych zavinil na	pokojskou /NNFS4-----A-----	a ukázal ji vrstvu prachu na stínítku lampy, ale
brigády: servírka, barmanka, zaměstnanec v kavárně,	pokojská /NNFS1-----A-----	obvyklý plat: 10 000 – 17 640 korun čistého měsíčně ubytování
šlil. Doma si namísto služby představuje majordoma a oddanou	pokojskou /NNFS4-----A-----	: " Pokojská: Hrabě Pericoloso Sporgersi mi nabídl pět
který tu zaručeně narážte, jsou buď prachdři, nebo	pokojské /NNFS3-----A-----	. Řekneme, že můj abuelo tu nikdy nebyl,
<p> „ Ano, koule se objevuje na Dürerově Melancholli I.	Pokojská /NNFS1-----A-----	Millsovi řekla, že sir Alistair Kermeide byl vášnivým obdivovatelem
hlase profesor. „ Žádný telefon tady nevidím. " </p><p>	Pokojská /NNFS1-----A-----	se rukou chabě dotkla kapsy na zástěře. Volala

Obr. 2: pokojská, SYN2015

Dále jsme zjistili, že jednotky s adjektivní flexí, které primárně plní funkci atributu a sekundárně plní funkci podmětu, mají ve většině případů ve slovníku přiřazeny dvě interpretace, ale na úrovni desambiguace je interpretace zvolena chybně (typ: *strážný*, *výchovný*). Tento typ výrazů je v korpusu interpretován značně chaoticky.

/3/ Pracovali jsme s korpusem SYN2015 z ÚČNK.

tmy . Po obou stranách měla věžičky s okénky pro	strážné /NNMP4-----A-----	, kteří však svá místa už dávno opustili pod pa
vlastně se promění v živá želízka , takže v sobě	strážné /NNMP4-----A-----	uvězní . </p><p> Nadháněčka drží v ruce prop
na zed' . Když všechno dobře půjde , nemusí si	Strážný /AAIS4-----1A-----	nějaký čas vůbec všimnout , že jsme pryč . </p>
pustil láhev . </p><p> " Stojko a Umanský si povídají se	strážnými /NNMP7-----A-----	, " Uhodila pěsti do otevřené dlaně . " Griškin
v dálce . </p><p> Na půl cesty se potkali s postupujícími	strážnými /NNMP7-----A-----	a po šeptem vedeném pohovoru společně vešli
, že vězni pracují ve správě si žijí líp než	strážní /NNMP1-----A-----	, </p><p> Za jízdy naskočíš na nákladní vago
. Opírela se o stěnu za sebou a jeden ze	strážných /NNMP2-----A-----	se natáhl , aby ji přidržel . „ To jsem
že to byl on , kdo jim dal echo . </p><p>	Strážní /NNMP1-----A-----	ho poslali zpátky do cely . </p><p> Připadalo
. " Přikývl jsem . Denisa byla má láska a	strážný /AAMS1-----1A-----	anděl . Bez ní bych nikdy nedostal tohle místo
ž dávno poučily a do dveří si nechaly nainstalovat hrbaté	strážné /NNMP4-----A-----	, zařízení proti zlodějům . Takže byla tady , v
, jsem syna utišíš nedokázala , zatímco ona ano .	Strážná /AAFS1-----1A-----	ani nic nefekla , jen mě udeřila pažbou do zad
půlnoci za mrazivé noci nechal nastoupit asi 800 vězňů i	strážné /NNMP4-----A-----	. Ke všem nastoupeným měl velitel tábora Jar
pěželi ve tmě do prázdného baráku , ještě neobsazeného .	Strážný /AAMS1-----1A-----	neslyšel nic a psi tam našťěstí neměli . A v

Obr. 3: strážný, SYN2015

NÁVRHY ŘEŠENÍ

V souladu s Komárkovým tvrzením, že „lingvistická klasifikace prvků jazyka musí být obrazem skutečných jazykových kategorií projevujících se při fungování jazyka“ (Komárek 2006: 15), je naší snahou nalézt takové řešení uvedeného problému, aby bylo možné správně desambigovat jednotky s adjektivní flexí, ale zároveň se vyhnout nárůstu slovnědruhovému homonymie znamenáním těchto tvarů do morfologického slovníku. Vycházíme z toho, že navržené řešení musí být konzistentní a systémové, a cílem je, aby koncový uživatel dostal jednotnou a správnou informaci.

Jedním z návrhů řešení problematického značení substantivizovaných adjektiv bylo využití hybridní kategorie. Termín hybridní kategorie je „český ekvivalent anglického termínu *mixed category* a používá se pro označení různých gramatických fenoménů, jejichž společným rysem je, že jeden jazykový výraz má vlastnosti dvou kategorií, nejčastěji se termín *mixed category* používá v užším smyslu pro výrazy kombinující rysy dvou lexikálních kategorií“ (Karlík – Nekula – Pleskalová, v rukopisu).

Ukázalo se však, že možnost zavedení hybridní kategorie do morfologického značkování není ideální, a to proto, že do hybridní kategorie patří substantiva typu *hajný*: mají distribuci substantiv, syntaktické chování substantiv, ale mají adjektivní skloňování. Většina slov této kategorie (*porybný*, *pokojská* ad.) není homofonní s adjektivy, a proto určení slovního druhu z hlediska AMA je konzistentní a jednoznačné. Není tedy třeba, abychom v těchto případech zaváděli nové značení, pokud stávající plně vyhovuje. Naopak problematické, z hlediska určení slovního druhu pomocí AMA, jsou výrazy spadající do hybridní kategorie, které jsou homofonní s adjektivy (*výchovné*, *trvalá* ad.). Jedná se o otevřenou skupinu slov, u nichž by správná interpretace slovního druhu uživatelům jistě pomohla.

Další z návrhů řešení bylo stanovení formálních pravidel. Zjistili jsme však, že ani tento návrh není vyhovující, protože ne vždy zobecnění funguje.

Zde pro ilustraci uvádíme příklady návrhů formálních pravidel a jejich ověřování v korpusu SYN2015:

- jednotka s adjektivní flexí následovaná slovesem = substantivum

svým složením sboru si učitelé spojují příznivější **výchovné** /AANS4----1A---- působení na žáky. („ Ti chlapi jsou
 </p><p> Vztah cíle jazykového, vzdělávacího a **výchovného** /NNNS2----A---- je možné vyjádřit i touto rovnicí :
 statečně, mnohá muzea dosud nepochopila své **výchovné** /AAFP4----1A---- možnosti a nezapojují se ani do e

Obr. 4: výchovné, SYN2015

Ukázalo se, že i pro poučeného lingvistu je bez kontextu v mnohých případech obtížné slovní druh určit, například ve větě: *To je výchovné.* (Bud' se jedná o částku placenou za výchovu, tedy substantivum, nebo se jedná o adjektivum *výchovný*.)

- jednotka s adjektivní flexí následovaná interpunkcí = substantivum

použít jako stolní gril . foto : weber 2 . **skladné** /AANS4----1A---- řešení Jednoduchá elektrická plotýnka
 nizela dlouhá rukojeť , moderní hrnce jsou oblíbené a **skladné** /AAIP1----1A---- . A dá se nastavit intenzita páry podle :
 </p><p> „ Výhodavakujevtom , že jsou relativně **skladné** /AAIP1----1A---- . Vše potřebné se vejde na jednu paletu

Obr. 5: skladné, SYN2015

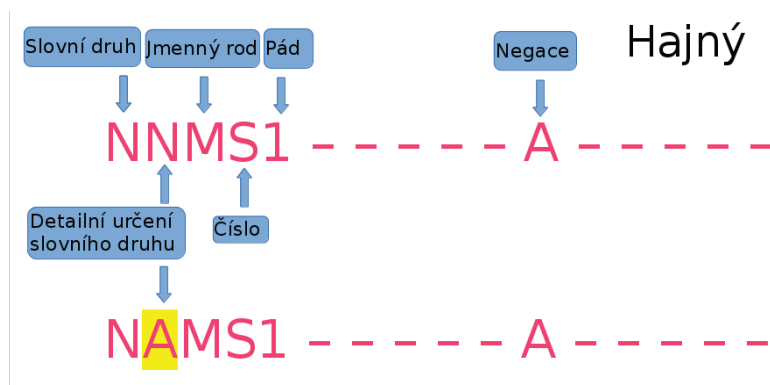
- jednotka s adjektivní flexí následovaná vlastním jménem/zkratkou = substantivum

našem virtuálním call centru budete nabízet **předplatné** /AAFP4----1A---- Vlasty a dalších časopisů vydavatelství Sanoma M
 ak podobně . Jeden jim doporučoval odhlásit **předplatné** /AAFP4----1A---- Sedmikrásky a místo toho darovat peníze neziskov
 ze v těchto programech žijí . Pro ně může být **předplatné** /NNNS1----A---- výhodné i z toho důvodu , že se původní velká
 5 . Dále si můžete objednat roční elektronické **předplatné** /NNNS4----A---- za 685 Kč , s nímž získáte přístup ke starším
 ile jistě zaujme , že například Zoot . cz nabízí **předplatné** /AAFP4----1A---- HN se slevou 45 procent . </p><p> Z klasických sl
 souhlasíte s nimi , nabídnou vám akční roční **předplatné** /NNNS4----A---- za třicet eur měsíčně (necelých 800 Kč) plus

Obr. 6: předplatné, SYN2015

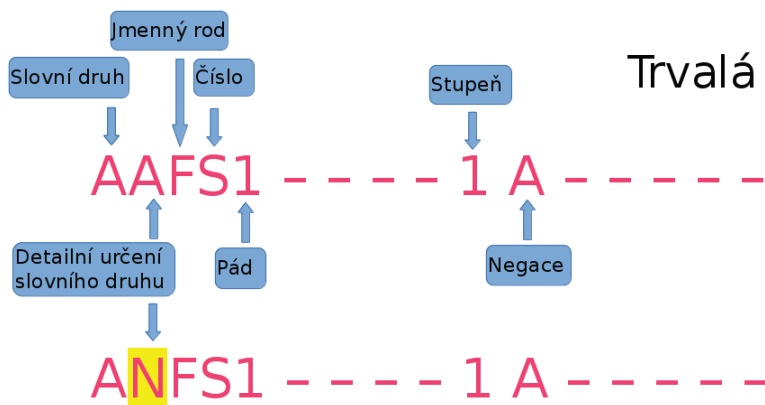
Po zvážení a ověření výše uvedených návrhů řešení se jako nejvhodnější jeví změna morfologické značky u těch výrazů, kterým je ve druhém kroku AMA přiřazena jak interpretace substantivum, tak interpretace adjektivum.

Pro substantiva s adjektivním skloňováním, která jsou výhradně substantivy, navrhujeme následující změnu morfologické značky:



Obr. 6: Návrh na změnu morfologické značky, typ hajný

V případě adjektiv, která mohou být substantivizována navrhuje následující změnu morfologické značky:



Obr. 7: Návrh na změnu morfologické značky, typ trvalá

ZÁVĚREM

Je třeba upřesnit, že navržená změna morfologické značky nebyla v době konání konference testována.

Vzhledem k tomu, že substantivizace adjektiv tvoří jen drobnou výše z problematiky slovnědruhových přechodů, je pravděpodobné, že změna morfologické značky se může a bude týkat i jiných slovních druhů. A je velmi

pravděpodobné, že druhá pozice morfologické značky těchto výrazů je již obsazena charakteristikou, kterou není vhodné libovolně nahradit jinou (týká se to například slovesných adjektiv). Naší snahou do budoucna je pečlivě popsat všechny typy slovnědruhových přechodů a zamyslet se nad konzistentním řešením ve formě změny morfologické značky. Na základě dosavadních zjištění budeme opět hledat takové řešení, které bude systémové pro všechny typy slovnědruhových přechodů, nejen pro tvrdá substantivizovaná adjektiva.

LITERATURA

Český národní korpus – SYN2015. Ústav Českého národního korpusu FF UK, Praha 2015. Dostupný z [www: http://www.korpus.cz](http://www.korpus.cz).

CVRČEK, Václav – KOVÁŘÍKOVÁ, Dominika

2011 „Možnosti a meze korpusové lingvistiky“; *Naše řeč* 94, č. 3, s. 113–133

KARLÍK, Petr – NEKULA, Marek – PLESKALOVÁ, Jana (eds.)

Encyklopedický slovník češtiny, 2. vyd., v rukopisu

KOMÁREK, Miroslav

2006 *Příspěvky k české morfologii* (Olomouc: Periplum)

NEKULA, Marek – RUSÍNOVÁ, Zdenka – KARLÍK, Petr (eds.)

1995 *Příruční mluvnice češtiny* (Praha: Nakladatelství Lidové noviny)