

Téma 2 – Lingvistické a sémantické problémy obsahové analýzy

1. Co je to obsahová analýza (OA)?
2. Ontologie OA
Co je to obsah?
Kde je obsah?
3. Gnozeologie OA
Jak zjistit obsah?
Jak vyjádřit obsah?
4. Praxe OA
Jak chápat/modelovat obsah?

ÚISK – OBA T02

1

1

Gnozeologické problémy obsahové analýzy

- **Kognitivní** – jak to poznat?
 - Jak zjistit potřebu informace (= obsah informační potřeby)
 - Jak zjistit obsah informačního zdroje
- **Sémanticko-lingvistické** – jak to vyjádřit?
 - Jak vyjádřit potřebu informace
 - Jak vyjádřit obsah zdroje

ÚISK – OBA T02

2

2



3

3

Charles Sanders Peirce

1839 – 1914



prvost
druhost
třetost

- Triadické pojetí znaku
1. znak
 2. označované
 3. interpretant
- 1 zastupuje 2 vzhledem k 3

- Triadická typologie znaků
- ikón
 - index
 - symbol

ÚISK – OBA T02

4

4

USA



Charles William Morris 1901 – 1979

Základy teorie znaku. 1938

- **syntaxe**
- **sémantika**
- **pragmatika**


↓

„**warrant**“ = zdůvodněný,
ospravedlněný, oprávněný


ÚISK – OBA T02 5

5

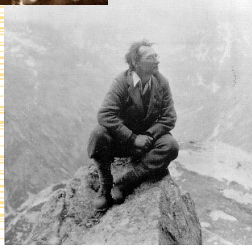
Evropa



Ivor Armstrong Richards 1893–1979 Charles Kay Ogden 1889–1957



Richards, I. A. a Ogden, Ch. K.
The meaning of meaning. 1923



- referent
- reference
- symbol

ÚISK – OBA 102 6

6

Evropa



Ferdinand de Saussure 1857 – 1913

Kurs obecné lingvistiky. 1916

- **langage** (řeč ≈ schopnost jazyka)
= **langue** (jazyk – systém) + **parole** (pro/mluva – text)
- diachronie – synchronie (viz korpusy [Diakorp](#) – [SYN](#))
- paradigma – syntagma

Duální pojetí znaku

- znak = **signifiant** (označující) + **signifié** (označované)

ÚISK – OBA T02 7

7

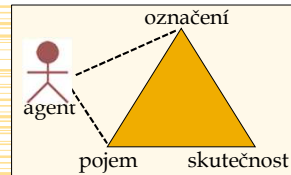
Systémové pojetí jazyka

prvky	■ lexikální subsystém (slovník)	
	■ pravidla (gramatika, syntaxe)	
vztahy	■ paradigmatické	
	■ syntagmatické	
funkce	■ poznávací	referenční
	■ komunikační	
	■ jazyk = skutečnost myšlenky	
	■ emotivní, poetická, konativní ad.	

ÚISK – OBA T02 8

8

Typologie jazykových systémů



- čeština
- TDKIV
- AGROVOC
- HTML

- **Přirozený jazyk**
„měkký“, fuzzy systém
pravidla se formulují během používání
- **Speciální jazyk (terminologie)**
- **Jazyk s řízeným / strukturovaným slovníkem**
jednoznačné prvky
- **Umělý jazyk**
„tvrdý“ systém
platí pravidla formální logiky
jednoznačné prvky i vztahy
notace

ÚISK – OBA T02

9

Přirozený jazyk

Systém

- **univerzální** (neomezený obsahem sdělení)
- **složitý**, jeho používání vyžaduje **inteligenci**
- **otevřený, neustále se měnící**

Vývoj přirozeného jazyka

- **smskovat, vygooglit**
- **čanit, lepá**
- **email, žába, kočka, cookie**

ÚISK – OBA T02

10

Lexikální subsystém jazyka – způsoby zaznamenání

Korpus
(*parole*)
syntagmatické

vztahy

Tezaurus
(*langue*)
paradigmatické

„přirozený jazyk
v počítači“
– Český národní korpus
<http://www.korpus.cz/>



ÚISK – OBA T02

11

11

Peter Mark Roget 1779 – 1869



Anglický polyhistor:

lékař, lexikograf, autor četných publikací z oboru medicíny a optiky, přispěvatel do encyklopedií, vynálezce logaritmického pravítka...

Thesaurus of English Words and Phrases, Classified and Arranged so as to Facilitate the Expression of Ideas and Assist in Literary Composition.

1. vydání 1852

- slovník synonym a antonym
- kategorizace slov

ÚISK – OBA T02

12

12

Roget's Thesaurus

Main Entry: cat
Part of Speech: noun
Definition: feline animal, sometimes a pet
Synonyms: bobcat, cheetah, cougar, grimalkin, jaguar, kitten, kitty, leopard, lion, lynx, malkin, mouser, ocelot, panther, puma, puss, pussy, tabby, tiger, tomcat

Zdroj: Roget's New Millennium™ Thesaurus, First Edition (v 1.3.1)
 Copyright © 2006 by Lexico Publishing Group, LLC. All rights reserved.

ÚISK – OBA T02 13

13

Roget's Thesaurus

Existence

End, entity, being, exist. ce | Nonentity, nullity, nihility
 Essence, quintess. ce, quiddess. ce | nonexistent. ce, nothing, nought
 Nature, thing, substance | void, zero, cypher, blank
 course, World, fraction | constitution
 empty
 Reality, (v. truth) actual | unsubstantial
 exist. ce - fact, | Unreal, ideal, imaginary
 course of things, under y sure | visionary, fabulous
 extant, present | fictitious, supposititious
 Positive, affirmation, absolute | absent, shadow, dream
 substantive | phantom, phantasm
 intrinsic, inherent | Negative, virtual, extrinsic
 To be, exist, obtain, stand | potential, adjective
 pass, subsist, prevail, lie
 - on foot, on y tapis
 to constitute, form, compose, to consist of
 scope, habitude, temperament
 Slate, mode of exist. ce condition, nature, constitution, habit, place
 Affection, predicament, situation, posture, contingency
 Circumstances, case, plight, term, time, - point, degree
 Juncture, conjuncture, pass, emergency, exigency

ÚISK – OBA T02 14

14

Výkladový slovník vs. jazykový tezaurus

slovník	tezaurus
způsob použití	
slovo (znak)	pojem (význam)
↓ Co znamená?	↓ Jak vyjádřit?
pojem (význam)	slovo (znak)
základní (primární) uspořádání	
abecední	systematické
základní (primární) proces	
sémaziologický	onomaziologický

ÚISK – OBA T02 15

15

Co je to pojem?

Jednota intenze (obsah) a extenze (rozsah)

inverzní poměr

Zjednodušené chápání obsahu: množina charakteristik (vlastností, atributů)

ÚISK – OBA T02 16

16

Co je to lexikální jednotka? též **lexém**, *lexical unit*, *token*

obecně

- sekvence písmen, číslic a dalších specifických znaků používaná pro označení určitého **pojmu**
- jednotka jazyka, která je nositelem významu nebo funkce
- **z hlediska významového tvoří jeden nedělitelný celek**

v organizaci znalostí

- slovo nebo slovní spojení ve formě buď
 - výrazu přirozeného jazyka, nejčastěji substantiva nebo substantivního spojení
 - notačního znaku

ÚISK – OBA T02

17

Typologie vztahů v jazyce

Paradigmatické sémantické trvalé, <i>a priori</i> bezkontextové	Syntagmatické syntaktické ad hoc, <i>a posteriori</i> kontextové
organizace označením vztahy významů slov např. ve slovníku	organizace umístěním vztahy umístění slov v promluvě / v textu
<i>langue</i> (systém jazyka)	<i>parole</i> (úzus jazyka, promluva, mluva)
<ul style="list-style-type: none"> ▪ ekvivalence ▪ hierarchie (obecná / generická, partitivní, instanční) ▪ asociace 	<ul style="list-style-type: none"> ▪ kolokace ▪ koordinace

ÚISK – OBA T02

18

Paradigmatické a syntagmatické vztahy

skákal	pes	přes oves	osa substituce (paradigmatická) význam nezávislý na kontextu
poskakoval	Alík	ovesným polem	
běžela	šelma psovítá	skrz pole s obilím	
osa kombinace (syntagmatická)			

ÚISK – OBA T02

19

19

Paradigmatické a syntagmatické vztahy

skákal	pes	přes oves	osa substituce (paradigmatická) význam nezávislý na kontextu
poskakoval	Alík	ovesným polem	
běžela	šelma psovítá	skrz pole s obilím	
život je	pes	počasí	
osa kombinace (syntagmatická) význam určený kontextem			

ÚISK – OBA T02

20

20

„Pěkný“ v Tezauru jazyka českého paradigmatické vztahy (*a priori*)



krásný, překrásný, čarokrásný, kouzelný, **pěkný**, šumný, onaký, hezký, přehezky, hezounky, hezoučký, báječný, fantastický, znamenitý; **půvabný**, okouzlující, milý, roztomilý, slíčný, lepý, ladný, spanilý, líbezný, rozkošný, čínaný, hezký jako obrázek, malovaný, k pomilování, šarmantní, graciézní; **pohledný**, vzhledný, úhledný, švarný, fešný, dobře vypadající, vznosný, urostlý, ztepilý, souměrný, útvarný, pěkně tvarovaný, dobře rostlý, dobře stavěný, udělaný, formózní; **přitažlivý**, zaoblený, vlnadný, bujný, bujných tvarů, lepých tvarů, sošný, svůdný, rajcovný, ženský, sexy, atraktivní, beroucí dech; **elegantní**, vytříbený, vybraný, dokonalý, vkusný, módní, šik, apartní, mondénní, nenucený, uhlazený, dobře udržovaný, pěstěný, upravený, úhledný, jako ze škatulky, vystrojený, vyšňořený, švihácký; **nádherný**, velkolepý, znamenitý, skvělý, báječný, kultivovaný, oslnivý, zářivý, úchvatný, senzační, úžasný, božský, kouzelný, skvostný, splendidní, magnificentní;

ÚISK – OBA T02

21

21

Připomenutí z předmětu KSA

Paradigmatické vztahy

ekvivalence	obsahu
	formy
hierarchie	obecná (širší – užší extenze / intenze)
	závislost (když..., tak...), historická (předchozí – následující)
	„speciální“ (s inferencí a dědičností) <ul style="list-style-type: none"> ■ generická (rod – druh) ■ partitivní (celek – část) ■ instanční (obecné – individuální)
asociace	sekvenční (prostor, čas)
	příčinná
	...

ÚISK – OBA T02

22

22

„Pěkný“ v Českém národním korpusu syntagmatické vztahy (*a posteriori*)

Shuffle: ✓

hole "omezeni/syn2015" | ARF: 1,076.41 | Result is shuffled 1 / 55

, ale spíš sháním něco jiného . Když uvidím nějaký **pěkný** kousek za rozumnou cenu a mám peníze , rá
 . Jestli o nás jenom cekne , tak poletí z **pěkný** vejšky . "</p><p>„ Jo , ale ...
 (o něm více v následující kapitole) , upoutá **pěkný** starý dům U Splavnů čp. 36/III – s podlouhý
 hodinky přímo od dealerů . Ti , když mají nějaký **pěkný** model , se na mě už rovnou obracají . "</p><p>
 jaký lidi potkává na chodbě . " Aspoň jsou tam **pěkný** holky , " řekl jsem jí , když odcházela .
 jakousi bedničku a pronesl improvizované uvítání . Byl to velice **pěkný** , spontánní úvod k samotnému zahájení . Dř
 Mell stála u otcovy ledové rakve .</p><p><p>Byl na ni **pěkný** pohled , ale Žestokost se jen s pohledem ne
 si neřekla : „ Tak a teď si udělám svůj **pěkný** portrét a dám ho na odív . " Spíše na
 to rohový pozemek a má kouzelnou zahrádku . Květiny a **pěkný** trávnik . Chtějí za něj tři sta tisíc . Ale
 která právě unikla smrti . K nejbližšímu záhybu chodby doběhla **pěkný** kus před Valerí , a když k němu dorazila vzle
 , který skončil v síti . Druhý už byl ale **pěkný** . Doufám , že tak budu pokračovat i dál .

ÚISK – OBA T02

23

23

Jak vyjádřit obsah / potřebu informace?

3

Lingvistické a sémantické problémy organizace znalostí – asymetrické vztahy 1 : N

- 1) **synonymie** (1 význam – N lexémů)
- 2) **homonymie** (1 lexém – N významů)
XOR, buď / nebo, vylučující
- 3) **koordinace – složené pojmy / složená témata** (N významů)
AND, a zároveň, partitivní
- 4) **slovní spojení** (1 význam – N lexémů)

ÚISK – OBA T02

24

24

4

Rekapitulace

Model znaku - trojúhelník reference (sémiotický trojúhelník)

J znak
(jazykový výraz) - lexikální jednotka

vyjádření ↗

označení (denotace) ↘

smysl S (koncept, pojem) ← odraz

D denotát
(skutečnost, objekt, věc)

ÚISK - OBA T02 25

25

Ideál - vzájemně jednoznačná korespondence znak - smysl - denotát

kyselina sírová pondělí

Problém:
víceznačnost,
synonymie,
neurčitost
významu

ÚISK - OBA T02 26

26

Problém - nejednoznačnost, neurčitost významu

Typy ekvivalence v jazyce

synonymie

jazykový výraz → denotát

homonymie

jazykový výraz → denotát

ÚISK - OBA T02 27

27

Synonymie

fotbal - kopaná doktor - lékař

ÚISK - OBA T02 28

28

Typy synonymie

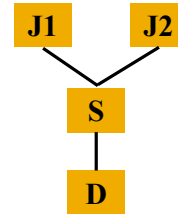
- lexikální (věcná, sémantická)**
 - „pravá“ synonymie (ekvivalence)
cesta – silnice – dálnice – vozovka – stezka – pěšina
 - hierarchie**
domácí zvíře – kočka – Macek
 - asociace**
nevolnost – nemoc, plný – prázdný
- gramatická (morfologická)**
1 slovo ve více tvarech – *cesta, cesty, cestě, cestu...*
- mezijazyková ekvivalence**
kočka – die Katze – кошка – cat
- formální**
budova, budou, ubudou, budoucnost, vybudovat...

ÚISK – OBA T02

29

29

Ekvivalence v synonymii



2406 **ZNAK** (logický) čeho (pojmu, druhu, jedince, pravdy ap.; z. základní, podstatný, nezbytný, nutný, odlišující, roz-, význačný, charakteristický, jistý, bezpečný, nemylný, neo-), podstatná, příznačná, charakteristická **známka**, znamení čeho (vzdělanosti, poctivosti), význačná, rozlišující **vlastnost**, **podstatný**, charakteristický **rys** čeho, *kniž.* význak, **přívlastek** (p-y lidské; p-y hluboké povahy), **atribut** (pojmu, realismu), **charakteristikon** (národnosti, génia), **signum** (neomylné s. pravého umění), **pečeť** (individuality, pravdy), *zast.* **oznak** (hlavní o. pravdy), **přímět(ek)**, *zř.* **poznatek** (egoismu); *log. z.*, určující prvek při stanovení pojmu
viz též typ, typický

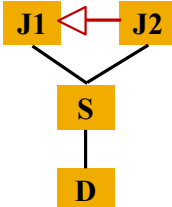
ÚISK – OBA T02

30

30

Hierarchie v synonymii

archilexém



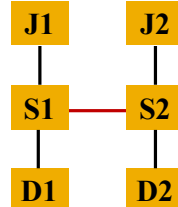
2406 **ZNAK** (logický) čeho (pojmu, druhu, jedince, pravdy ap.; z. základní, podstatný, nezbytný, nutný, odlišující, roz-, význačný, charakteristický, jistý, bezpečný, nemylný, neo-), podstatná, příznačná, charakteristická **známka**, znamení čeho (vzdělanosti, poctivosti), význačná, rozlišující **vlastnost**, **podstatný**, charakteristický **rys** čeho, *kniž.* význak, **přívlastek** (p-y lidské; p-y hluboké povahy), **atribut** (pojmu, realismu), **charakteristikon** (národnosti, génia), **signum** (neomylné s. pravého umění), **pečeť** (individuality, pravdy), *zast.* **oznak** (hlavní o. pravdy), **přímět(ek)**, *zř.* **poznatek** (egoismu); *log. z.*, určující prvek při stanovení pojmu
viz též typ, typický

ÚISK – OBA T02

31

31

Asociace v synonymii



2406 **ZNAK** (logický) čeho (pojmu, druhu, jedince, pravdy ap.; z. základní, podstatný, nezbytný, nutný, odlišující, roz-, význačný, charakteristický, jistý, bezpečný, nemylný, neo-), podstatná, příznačná, charakteristická **známka**, znamení čeho (vzdělanosti, poctivosti), význačná, rozlišující **vlastnost**, **podstatný**, charakteristický **rys** čeho, *kniž.* význak, **přívlastek** (p-y lidské; p-y hluboké povahy), **atribut** (pojmu, realismu), **charakteristikon** (národnosti, génia), **signum** (neomylné s. pravého umění), **pečeť** (individuality, pravdy), *zast.* **oznak** (hlavní o. pravdy), **přímět(ek)**, *zř.* **poznatek** (egoismu); *log. z.*, určující prvek při stanovení pojmu
viz též typ, typický

ÚISK – OBA T02

32

32

Kvazisynonymie, antonymie

fotbalista - Jan Kovář

tvrdý - měkký

ÚISK - OBA T02 33

33

Formální a kontextová synonymie - hypotézy

- **ekvivalence označením**
slova, která se podobně píšou, mají podobný význam
- **ekvivalence umístěním**
slova, která jsou v textech blízko sebe, mají významovou souvislost

škola (W) hrou

Až tento školní rok se pro Petra stala vzpomene si i na třídního Ritu a jeho stně jsem školu neměl rád! Nebyla to vi Komenskému a jeho schole ludus , kými jsou trojivá škola , činná škola , ni řemesly a pracovišti v duchu hesla blesky u památníku Prokopa Diviše ,

škola hrou
školu hrou
škola hrou
škole hrou
škola hrou
škola hrou
škola hrou

a poprvé přinese domů samé jedničky. V | . Druhý den je na vratech gymnázia seznan , já tu hru na školách moc často nevidím . , jako jí není ani teď . Už samotný výhled . Učení je prostě efektivnější , pokud se ječ , otevřená , volná , svobodná , netradiční ál . Žáci takto získají konkrétní představu o u * , propagoval učitel národů Jan Amos Kor

Hypotéz*

hypotéz

Hypotéza

hypotézách

Hypotézám

hypotézami

hypotéze

hypotézo

hypotézou

hypotézu

hypotézy

34

34

Typy synonymie v bázi národních věcných autorit

Temat. autority - Úplné zobrazení záznamu

Zvolte formát: [Standardní](#) -- [MARC](#)

Záznam 1 z 1

Ident. číslo:

Záhlaví: **zvukové dokumenty**

Odkaz. forma: **dokumenty zvukové**

Nadřazený termín: **dokumenty**

Podřazený termín: **audioknihy**

Asoc. termín: **audiovizuální dokumenty**
katalogizace zvukových dokumentů

MDT: **025.177 (knihovní fondy)**
371.683 (učební pomůcky)
002.1-028.17 (obecně)

Angl. ekvivalent: **sound recordings**

Skupina Konspektu: **025 - Organizace knihoven a knihovnických fondů [12]**
371 - Školství (organizace) [22]

35

35

Homonymie / mnohovýznamovost (polysémie)

jazyk - jazyk

polysémie

los - los

homonymie

ÚISK - OBA T02 36

36

Typy homonymie

a) lexikální

stát – stát – stát – stát, bez – bez, Let let, léta letí, a – a

- zvuková (*mít – mýt, ve při – vepři, sním – s ním*)
- grafická (*email – email, . . .*)

b) syntaktická

vhazování domácích

církev chce vrátit ukradený majetek

moc slov – moc slov

c) sémantická – ironie, metafora, metonymie

to se ti povedlo

šli na skleničku

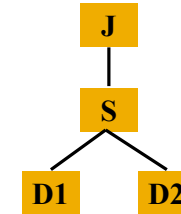
ÚISK – OBA T02

37

37

Homonymie – mnohoznačnost (dvojsmysl)

to je pěkné – to je pěkné



ÚISK – OBA T02

38

38

Homonymie – metafora



Nedělala bych si marné naděje. Vitr v peněžence
ještě nikdy žádnou vrtuli neroztočil...

39

Homonymie – metonymie

Čínská filosofie
je v C244
(1. patro vpravo a zadním
schodištěm do 2.p.)

40

40

Metafora a metonymie = přenos významu / označení

označení A → metonymie → označení B

ÚISK – OBA T02 41

41

Metafora = přenos významu

moře moře lidí

ÚISK – OBA T02 42

42

Metonymie = přenos označení

kniha → metonymie → kniha

(část žaludku krávy)

ÚISK – OBA T02 43

43

Falešní přátelé, zrádná slova Homonyma, která se „tváří“ jako synonyma

host (česky) host (anglicky)

cesta (česky) cesta (italsky)

věcné věcné

pořádání pořádání

ÚISK – OBA T02 44

44

Složené téma / pojem

1 pojem – více jednotek obsahu

- elektrokolo
- husopaska
- teploměr
- e-learning
- samoobsluha
- narkobaron, uhlobaron
- zahraničněpolitický
- alfanumerická notace
- aktivní meziknihovní výpůjční služba
- automatická tvorba tezauru
- daň z přidané hodnoty
- matematicko-statistická teorie informace

kyber-egyptologie

ÚISK – OBA T02

45

45

Slovní spojení

těž sousloví, sdružené pojmenování, fráze, frazém, idiom (*phrase*), n-gram, víceslovný lexém, kolokace, koherentní víceslovný výraz

1 jednotka obsahu – více lexémů

Jan Neruda, černí baroni, stavební úřad, pražské jaro, láry fáry, New York City, postavení mimo hru, lze si do vlastní kapsy

- Ustálené pojmenování ze dvou nebo více slov, které **označuje jednu skutečnost** a vstupuje do věty jako hotový celek
- **Nejčastější forma:**
substantivní spojení: lexikální jednotka, kterou lze rozčlenit na oddělené složky, přičemž každá z těchto složek může být vyjádřena substantivem a nezahrnuje sloveso ve tvaru určitém

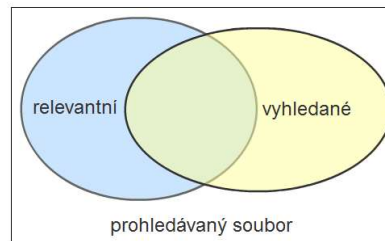
ÚISK – OBA T02

46

46

Hodnocení výsledků vyhledávání (= výběru, prozkoumání zdrojů)

- relevance
- informační šum
- **úplnost** (*recall*)
podíl vyhledaných z relevantních
- **přesnost** (*precision*)
podíl relevantních z vyhledaných



ideál:

- najít vše, co je relevantní
- vyloučit vše, co není relevantní

ÚISK – OBA T02

47

47

Problém synonymie

snižuje úplnost (*recall*):

hrozí, že nenajdeme všechny relevantní informace

řešení:

rozšiřování (expanze) požadavku/nabídky o zdroje obsahující **synonyma**

ÚISK – OBA T02

48

48

Problém homonymie

snižuje přesnost (*precision*):

hrozí, že najdeme nerelevantní informace

řešení:

zúžení požadavku/nabídky o zdroje, obsahující nerelevantní homonyma

Problém složených témat

■ **analýza**

■ **syntéza**

■ **konstrukce indexu/dotazu**

– **koordinace** Jak je vyjádřit?

} Jak zjistit jejich obsah / smysl?

Jak poznat složené téma / pojem

Zaměnitelnost

elektronická

kniha

přísluví

modlitební

cigareta

pro mládež

dětská

jízdenka

se slevou

Problém slovních spojení

Jak poznat, že jde o slovní spojení?

Kontext

stal kohoutem . Na místě , kde předtím stál , ržál **bílý kůň** . Teď už mohla začít slavnost . Trubači zatrou
lo filozof a tůň Dvakrát do téže řeky Vstupuje tiše **bílý kůň** Dvakrát do řeky co je stejná Vstupuje bělouš
světnice , honem z tebe uteku . Sníh Přišel k nám **bílý kůň** , zalehl nám celý dvůr . Sníh Přiletěl k nám bí
kůň je vlastně pokračováním souboru Jedenáctý **bílý kůň** , který vyšel ve dvou vydáních v polovině šed
středníků bůhzdarma . Jenže . Nad matičkou Prahou **bílý kůň** roztáhl svá křídla pouze jedenkrát . Od této ex
dníkatele , který v případě nejprve vystupoval jako **bílý kůň** , vinu vzal později plně na sebe , totiž nyní za
míněný Václav Franta , podle odborníků typický " **bílý kůň** " , tedy nastrčená figura , která si zřejmě ani n
muto činu bylo pravděpodobně to , že Tulej , jako **bílý kůň** používaný při různých akcích , přestal vyhov
ě jeho plochého , táhlého horizontu běžel pomalu **bílý kůň** . I dívka na něm byla bílá , i vlasy vlající ve vě
ké fejetony , shrnuté do dvou svazků (Jedenáctý **bílý kůň** , Třináctý černý kůň) , zůstávají tu však jako