

# CJBB84

4.11.2010

## Značkování slovesných tvarů a možnosti vyhledávání složených slovesných tvarů v ČNK

- Vf.\* sloveso infinitiv
- VB.\* sloveso přítomný/futurní (indik.)
- Vt.\* sloveso přítomný/futurní arch. tv. (indik.)
- Vi.\* sloveso imperativ
- Vp.\* sloveso l-ové příděje (vč. tvarů s –s)
- Vq.\* sloveso l-ové příděje (vč. tvarů na –t)
- Vs.\* sloveso pasivní příděje (vč. tvarů s –s)
- Ve.\* sloveso přechodník přítomný
- Vm.\* sloveso přechodník minulý
- Vc.\* sloveso kondicionál sl. být (*bych, ...*)
- J,.\* spojka spojky podřadicí vč. *aby, ... kdyby, ...*

# Korpusy a slovesné tvary v češtině

- Syntetické tvary
- indikativ prézenta aktiva
- imperativ
- indikativ futura aktiva dokonavých sloves
- syntetické futurum
- 3. os. indikativu minulého času aktiva

# Korpusy a slovesné tvary v češtině

- Analytické tvary
- Opisné/analytické futurum nedokonavých sloves aktiva
- 1. a 2. osoba minulého času indikativu aktiva
- opisné pasivum (všechny časy)
- kondicionál (přít. a min.)
- zvratné pasivum

# Analytické tvary

- tvary s infinitivem (Vf.\*)
- tvary s l-ovým příčestím (V[pq].\*)
- tvary s pasivním příčestím (Vs.\*)
- tvary s kondicionálem slovesa *být* (Vc.\*)

# Jak najdeme tvary opisného futura ?

- Menu: Korpus > Implicitní atribut > lemma
- Dotazovací řádek: být
- Menu: Korpus > Implicitní atribut > word
- P filtr : KWIC >0 >0
- Dotazovací řádek: .\*bud.\*
- P filtr : : KWIC >0 >5
- Menu: Korpus > Implicitní atribut > tag
- Dotazovací řádek: : Vf.\*

# KORPUS

The screenshot shows the Bonito software interface. The title bar reads "Bonito". The menu bar includes "Manažer", "Korpus", "Dotaz", "Konkordance", "Zobrazení", "Výběr", and "Nápověda". The search bar shows "P filtr" and "jméno: syn2000". The search parameters are set to "První", "Poslední", "Od 1", "Pozic", "od", "KWIC >", "Do 5", "Pozic", "od", "KWIC >".

The main text area displays a concordance search for the word "bude". The search results are shown in a table format with three columns: the original text, the word "bude" (highlighted in red), and the surrounding context. The text is from a story about a man named Vítězslav and his relationship with a woman named Zelenat.

hvíli , kdy bude hajný Vítězslav dýchat a trubička se ně uctívý ? To nebude samo sebou . Bud'to po mně de něco chtít , snad srnce na svatbu dcery , nebo mi ral , má - li v brašně detekční trubičky . " Dneska se hal domů . Těšil se , jak celý tento podařený kousek Vítězslave ! " " Ubytuju ho na lovecké chatě u mě a lá lovecký listek , že při známostech , které mají , se kazovi Vítězslava se slovy : " To je lesník , který vás " Bajhirš a odstřelový . Připrav se , jak přijde blíž a ' Večer ho namažu jak gardového dlaždiče , že ráno otu pěti bodů . Na terč srnce si může střelec zvolit , manželka tušila , o čem bude řeč , čeho se to radění ONSKÉ KUKAČKY " a s ním všichni jak jsme tady sí dnes něco zařídit v okresním městě a že večer jej	<b>bude</b> <b>bude</b> <b>bude</b> <b>budete</b> <b>bude</b> <b>budeme</b> <b>nebude</b> <b>bude</b> <b>budeš</b> <b>nebude</b> <b>bude</b> <b>bude</b> <b>budeme</b> <b>bude</b>	zelenat . Šmejdil za ním jako barvář po pobar něco <b>chtít</b> , snad srnce na svatbu dcery , nebo <b>chtít</b> něco nekalého provést . " Navenek se t holky <b>zelenat</b> , " a téměř zamilovaně je pohlad za chvíli <b>líčit</b> své manželce a dcerce . Ráno to <b>chodit</b> nad ni do patnáctek . Troubí to tam do <b>muset</b> ani moc učit a tak se stalo , že z obch <b>doprovázet</b> . Zná revír jako své vlastní boty a se <b>cítit</b> , střílej ! " O Fridolína Dřevokaze se po <b>vědět</b> o světě , " kul plány hajný - lovecký do - li <b>střílet</b> z volné ruky , nebo použije - li jako o asi <b>týkat</b> . Nemýlila se . " Děvčata , " začal V <b>pochodovat</b> k fořtovi na polesí . " Snad ne již <b>doprovázet</b> jeho kolega hajný Miloslav . Prof
--	---	--

The bottom status bar shows: "Zobrazeno: 1+500/168845 (0%) Řádek: 1".

# Další možnosti

- Je rozsah 1. – 5. pozice napravo od pomocného slovesa být jediná možnost ?
- Co může přinést pokus rozšířit interval ?
- Je možné prohledávat i levé okolí ?



# KORPUS

**Bonito**  
Manažer Korpus Dotaz Konkordance Zobrazení Výběr Nápověda

P filtr [ ] jméno: [ ] syn2000 [ ]

První Poslední Od -1 Pozic [ ] od < KWIC [ ] Do -1 Pozic [ ] od < KWIC [ ]

hospodský od pultu . " Neslavil jsem nic , ale teprve **slavit** a natoč mi ještě jedno a pošli velkého r  
z prominentní lovecký host musí být protiva . " Ale více **pít** , musíme do lesa , je nejvyšší čas . Něč  
t , že to není ten jeho jelen a že jiného než toho svého **lovit** . Že když ho neuloví , tak že se nic nest  
i doběhne pro hrnec a podá mi ho . Já vám tady **aportovat** . Až mi bude hrnec podán , tak vám řel  
rázdnou hospodu ? A večer k pivu nikdo chleba **přikusovat** . " Tak tedy hospodský přinesl bochník  
nesnese odpor , tak byl raději zticha . Stejně v tom **chodit** , vždyť by ho nepoznal ani vlastní pes .  
řihla a prohlásila , že voda smrdí jak sračky a že více to **pít** . Po požití léčivé vody ale dostala nárar  
ete fořte slavný , " prohlásil hajný Miloš . " Nic se **roznášet** . Kdo to vykecá , je bez přemíí ! " odpo  
řitězslav mu řekl , že takové šufany již vícekrát **doprovázet** , a byl rozčilený , což se u něj často nes  
ravé ruky , posunul brýle na nose a řekl : " Na nic se **čekat** . Jde se na to hned ! Ten trouba Dusba  
idla a komu že na něm vlastně záleží , a že už si s ním **hrát** , když nehraje pořádně , a kdo ho tu vla  
o mě nácvik leteckého poplachu . O podrobnostech se **šřít** . Ale neustále se mi vrací jedna předsta  
dvanáctivoltová a ještě jinak voltová baterie buď **fungovat** nebo nebude . Já myslel , že on ví , co c  
ekl a mávl rukou , " tak tebe , Bille , se dnešní hodina **týkat** . Ale pro ty , kteří se ještě nerozhodli a

> Query : "být"  
> P-filter: ".\*bud.\*"  
> P-filter: "Vf.\*"

Zobrazeno: 1+500/4911 (10%) Řádek: 1

# složitější formulace dotazů

- [tag="VB.....F.\*"] []\* [tag="Vf.\*"] within <s>
- [tag="Vf.\*"] []\* [tag="VB.....F.\*"] within <s>
- Co se jimi chce říct?
- Hledáme v rámci jedné věty všechny kombinace slovesa v budoucím čase a infinitivu

# KORPUS

The screenshot shows the Bonito software interface. The title bar reads "Bonito". The menu bar includes "Manažer", "Korpus", "Dotaz", "Konkordance", "Zobrazení", "Výběr", and "Nápověda". The main window has a search bar with "Nový dotaz" and a dropdown menu. The search criteria are "jméno:" and "syn2000". The search results are displayed in a list of phrases, all in red text, centered in the window. The phrases are:

- bude se vracet
- budou mít
- nebude trvat
- nebudou pomlouvat
- budeme moci
- budeme moci v neděli řádně prospat
- bude hajný Vítězslav dýchat
- bude hajný Vítězslav dýchat a trubička se bude zelenat
- bude zelenat
- bude něco chtít
- bude něco chtít , snad srnce na svatbu dcery , nebo mi bude chtít
- bude něco chtít , snad srnce na svatbu dcery , nebo mi bude chtít něco nekalého provést
- budete holky zelenat
- bude za chvíli ličit

At the bottom of the window, the status bar shows "Počet výskytů: 254163" and the query: "> Query : [tag="VB.....F.\*"] []\* [tag="Vf.\*"] within <s>". The bottom-most status bar shows "Zobrazeno: 1+500/254163 (0%) Řádek: 1".

# Pokus se podobným způsobem najít další analytické formy

- minulý čas
- kandidáty na pasivní konstrukce
- kondicionál
- tvary s volným morfémem „se“ / „si“

# Jak najdeme tvary minulého času ?

- Menu: Korpus > Implicitní atribut > lemma
- Dotazovací řádek: být
- Menu: Korpus > Implicitní atribut > tag
- P filtr : KWIC <0 >0
- Dotazovací řádek: VB.....[12].\*
- P filtr : : KWIC >0 >5
- Menu: Korpus > Implicitní atribut > tag
- Dotazovací řádek: : V[pq].\*

# KORPUS

The screenshot shows the Bonito software interface. The title bar reads "Bonito". The menu bar includes "Manažer", "Korpus", "Dotaz", "Konkordance", "Zobrazení", "Výběr", and "Nápověda". The search bar contains "P filtr" and "jméno: syn2000". The search parameters are set to "První", "Poslední", "Od 0", "Pozic", "od", "KWIC >", "Do 5", "Pozic", "od", "KWIC >". The main text area displays a concordance for the word "být", with the word highlighted in red in the original image. The concordance entries are:

e pokusím přiblížit čtenáři některé z jejich kousků , které	<b>jsem</b>	bud' <b>zažil</b> přímo s nimi , nebo slyšel z jejich
fořte , služba je služba a povinnost je povinnost . Jednou	<b>jsem</b>	se <b>dohodli</b> , že mám přijít ve dvě hodiny a .
ia to . A slezte dolů , poněvadž tady jsou tři zajáci , které	<b>jsem</b>	vám <b>přinesl</b> a ruce mě bolí , jak jsem se s :
sou tři zajáci , které jsem vám přinesl a ruce mě bolí , jak	<b>jsem</b>	se s nimi <b>musel</b> produčirovat až sem . " Fc
uzávěrka je dávno hotová , musel jsem ji udělat sám . Ty	<b>jsi</b>	tady <b>měl</b> být ve dvě hodiny odpoledne a ne
í hodiny v noci ! " " No jó , fořte , " pravil Vítězslav , " to	<b>jste</b>	<b>měl</b> říct včera , že mám přijít ve čtrnáct h
1 , a byl bych tady s přesností švýcarských hodin . Vy	<b>jste</b>	ale <b>řekl</b> ve dvě a kolik je ? " Podíval se na
inky . " No prosím , dvě hodiny , dvacet tři minuty . A to	<b>jsem</b>	vás nejméně půl hodiny <b>budil</b> . Podívejte si
ů . " Venku pak dodal , smutně zíraje na tu spoušť : " Ty	<b>jsi</b>	tu uzávěrku pěkně <b>uzavřel</b> , pojed' dovnitř !
svému podřízenému , hajnému Vítězslavovi . " Společně	<b>jsem</b>	do hospody <b>přišli</b> a společně se odchází . F
a v posteli . " Jeleni mají na troubení ještě čas , ale včera	<b>jsem</b>	<b>troubil</b> já a natroubený jsem ještě teď , " e
í blesko hlavou : " Musel mi vypadnout v té škarpe , jak	<b>jsem</b>	<b>havaroval</b> . " Usedl tedy na motocykl a uj
oval si pro sebe a zajel před nádražní restauraci . " Čím	<b>jsem</b>	se <b>pokazil</b> , tím se musím spravit . " A jelik
ého druha morálně podpořit . " Přiznáváte se k tomu , že	<b>jste</b>	<b>řídil</b> dopravní prostředek po požití alkohoh

The bottom status bar shows: "Zobrazeno: 1+500/313365 (0%) Řádek: 1".

# Další možnosti

- Je rozsah 1. – 5. pozice napravo od pomocného slovesa být jediná možnost ?
- Co může přinést pokus rozšířit interval ?
- Je možné prohledávat i levé okolí ?

# minulý čas (složitější formulace dotazu)

- přišel jsem/ jsem přišel
- [tag="V(p|q).\*" ] [lemma="být" & tag="VB.....(1|2)P.\*" ] within <s>
- [lemma="být" & tag="VB.....(1|2)P.\*" ] []\* [tag="V(p|q).\*" ] within <s>



# KORPUS

The screenshot shows the Bonito software interface. At the top, there is a menu bar with options: Manažer, Korpus, Dotaz, Konkordance, Zobrazení, Výběr, and Nápověda. Below the menu bar, there is a search bar with the text "Nový dotaz" and a dropdown menu. To the right of the search bar, there is a field labeled "jméno:" with a dropdown menu and the text "syn2000".

The main content area is divided into three columns. The left column contains a paragraph of text from a story. The middle column contains a list of verb forms in red text. The right column contains a paragraph of text from a story.

**musel jsem**  
**Přivezl jsem**  
**šel jsem**  
**Neslavil jsem**  
**Trefil jsi**  
**Spiskal jsi**  
**Vyrobil jsem**  
**Věděl jsem**  
**Vzal jsem**  
**Díval jsem**  
**Troubil jsem**  
**Přijela jsem**  
**Koupila jsem**  
**Řekla jsem**

Počet výskytů: 161679  
> Query : [tag="V(p|q).\*"] [lemma="být" & tag="VB.....(1|2)P.\*"] within <s>

Zobrazeno: 1+500/161679 (0%) Řádek: 1

# KORPUS

The screenshot shows the Bonito software interface. The title bar reads "Bonito". The menu bar includes "Manažer", "Korpus", "Dotaz", "Konkordance", "Zobrazení", "Výběr", and "Nápověda". The search bar contains "Nový dotaz" and "jméno: syn2000". The main display area shows a list of search results in red text, which are lines of text from a corpus. The status bar at the bottom indicates "Počet výskytů: 598548" and "Zobrazeno: 1+500/598548 (0%) Řádek: 1".

Nový dotaz jméno: syn2000

**jsem bud' zažil  
jsem bud' zažil přímo s nimi , nebo slyšel  
jsi vážně tak praštěný jak vypadáš ! " řval  
jsme se dohodli  
jsem vám přinesl  
jsem vám přinesl a ruce mě boli , jak jsem se s nimi musel  
jsi tady měl  
jste měl  
jste měl říct včera , že mám přijít ve čtrnáct hodin , a byl  
jste ale řekl  
jsem vás nejméně půl hodiny budil  
jsi tu uzávěrku pěkně uzavřel  
jsme do hospody přišli  
jsem troubil**

Počet výskytů: 598548  
> Query : [lemma="být" & tag="VB.....(1|2)P.\*"] []\* [tag="V(p|q).\*"] within <s>

Zobrazeno: 1+500/598548 (0%) Řádek: 1

## a další ...

- pokuste se analogicky formulovat dotazy pro další typy analytických slovesných tvarů (kondicionál, opisné pasivum)

# Závěr

- V korpusu jsou označovány pouze jednotlivé pozice, takže značky nezohledňují funkce tvaru, který je součástí analytického tvaru
- Většina českých slovesných tvarů jsou vyjádřeny analyticky
- V korpusu lze vyhledávat „slovosledná schémata“, a tak „obejít“ nedostatek ve značkách