CJBB75
7. 3. Různé korpusy a rozdíly v anotačních schématech
Tokenizace, automatická anotace, d[ei]sambiguace
Anotace velkých synchronních korpusů ČNK
Anotace mluvených korpusů
Anotace KSK
Pražský a brněnský systém anotací
Specifika anotací SYN2005
Co se skrývá za označením slovní druh XX.* a X@.*

Doporučná četba pro zájemce o probíranou problematiku:

Jelínek, T.: Nové značkování v Českém národním korpusu. *Naše řeč 91*, 2008, s. 13–20.

Jelínek, T., Petkevič, V.: Systém jazykového značkování korpusů současné psané češtiny. In Petkevič, V. – Rosen, A. (eds.) *3. Gramatika a značkování korpusů*, Praha : Nakladatelství Lidové noviny/Ústav Českého národního korpusu, 2011, s. 154–170.

Osolsobě, K.: Popis gramatických významů (hodnot) jednoduchých slovesných tvarů v anotacích českých (slovenských) korpusů. *SPFFBU* A 55, Brno : FF MU, 2007, s. 201–218.

Petkevič, V.: Reliable Morphological Desambiguation of Czech: Rule-Based Approach is Necessary. In: Šimková, M. (ed.), *Insight into the Slovak and Czech Corpus Linguistics*, Bratislava : Veda, 2006, s. 26–44.

Petkevič, V.: Využití vidu ke zkvalitnění automatického značkování češtiny. In Bičan, A. – Klaška, J. – Macurová, P. – Zmrzlíková, J. (eds.), *Karlík a továrna na lingvistiku. Prof. Petru Karlíkovi k životnímu jubileu*, Host : Brno, 2010,s. 368–387.

14. 3. ÚKOL 2
V Internetové jazykové příručce (http://prirucka.ujc.cas.cz) v kapitole

Psaní spřežek a spřahování

http://prirucka.ujc.cas.cz/?id=130&dotaz=sp%C5%99e%C5%BEka

1.1.1 Typ *například – na příklad, zpočátku – z počátku*

Tyto příslovečné spřežky se v úzu vyskytují jak v podobě dvojslovné, tak jednoslovné bez významového rozlišení. Spřežky tedy mají stejný význam a plní ve větě tytéž funkce jako jejich dvojslovné protějšky – *na příklad* (předložka + podstatné jméno) i *například* (příslovce). Jediný rozdíl bývá v četnosti užívání, některé výrazy se v úzu objevují psané spíše jako dvě slova, u jiných převažuje podoba jednoslovná.

Další příklady: *do široka – doširoka, na boso – naboso, na čisto – načisto* (viz Psaní spřežek typu naměkko – na měkko, domodra – do modra, po anglicku), *z počátku – zpočátku, k večeru – kvečeru; na podiv – napodiv*, *nade všecko – nadevšecko, na čase – načase* atp.

Psaní spřežek typu naměkko – na měkko, domodra – do modra, po anglicku

Do roku 1993 (tedy do posledního vydání Pravidel českého pravopisu, která přinesla změny v některých oblastech pravopisu) se doporučovalo psát spojení tohoto typu zvlášť při vyjadřování děje (*vymalovat strop do modra*) a dohromady při vyjádření stavu (*šaty domodra*). Toto rozlišování bylo PČP v roce 1993 opuštěno.

Nyní jsou jako správné chápány obě varianty psaní – zvlášť (jako spojení předložky a jména: *na černo, na kyselo, do křupava*) i dohromady (jako tzv. spřežka Psaní spřežek a spřahování: *načerno, nakyselo, dokřupava*). Mezi příslovečným předložkovým výrazem a spřežkou není žádný významový rozdíl.

Možnost tvořit nové spřežky je zde téměř neomezená, v tomto směru nám nejširší pole působnosti poskytují výrazy s předložkami *do* a *na*. Pokud si budeme zapisovat recepty, máme situaci poměrně jednoduchou. Nemůžeme udělat chybu, protože tyto spřežky můžeme psát dvojím způsobem. Píšeme tedy: *nasekáme na hrubo* i *nahrubo, na jemno* i *najemno; smažíme do zlatova* i *dozlatova, vaříme do měkka* i *doměkka, zpracujeme za tepla* i *zatepla, za sucha* i *zasucha; vejce na tvrdo* i *natvrdo; kapr na černo* i *načerno* atp.

V názvech, které jsou svým původem ustrnulá spojení předložky s jmenným tvarem přídavného jména (např. *kapr po mlynářsku, pstruh po námořnicku*), píšeme předložku zvlášť (na rozdíl od ustálených spojení *poslepu* i *po slepu, postaru* i *po staru, podomácku* i *po domácku,* která můžeme psát dvojím způsobem); zeměpisná jména v nich píšeme s malým počátečním písmenem: *vepřové po německu, nákyp po francouzsku;* stejně tak lze *zmizet po anglicku.*

1. Najděte v korpusech řady SYN případy, které jsou kodifikovanými dubletami dokládajícími proces adverbializace.
2. Vytvořte tabulku a uveďte čísla (procenta) dublet.
3. Nastudujte v manuálu manažeru Bonito kapitolu o Konkordace/Statistiky/Rozložení a podívejte se na užití dublet (vyberte si jednu dvojici s vyššími frekvencemi) s ohledem na relevantní ukazatele (např. žánr).

Např. takto ....

	do křupava	***tagy***	*dokřupava*	***tag***
SYN2000	6	RR.* NN.*	4	Db.*
SYN2005				
SYN2010				
SYN2006PUB				
SYN2009PUB				