





Recenzenti

prof. PaedDr. Jana Kesselová, CSc.

Mgr. Michaela Mošat'ová, PhD.





**Jazykovedný ústav Ľ. Štúra Slovenskej akadémie vied
Slovenský národný korpus**

**Mária Šimková – Katarína Gajdošová –
Beáta Kmeťová – Marek Debnár**

**Slovenský národný korpus
Texty, anotácie, vyhľadávania**

**Bratislava
2017**





Hlavná redaktorka

PhDr. Mária Šimková, Ph.D.

Autori

PhDr. Mária Šimková, Ph.D.

Mgr. Katarína Gajdošová, Ph.D.

Mgr. Beáta Kmeťová

Mgr. Marek Debnár, PhD.

Publikácia vznikla v rámci projektu *Budovanie Slovenského národného korpusu a elektronizácia jazykovedného výskumu na Slovensku* podporovaného Ministerstvom kultúry SR, Ministerstvom školstva, vedy, výskumu a športu SR a Slovenskou akadémiou vied na základe zmluvy č. 0372/2012. Vydanie bolo financované zo zdrojov poskytnutých rovnakými partnermi na základe Zmluvy o združení prostriedkov na tvorbu a rozvoj Slovenského národného korpusu č. 0323/2017.

Tlačové podklady

Mgr. Vladimír Radik

Návrh obálky

Vanda Stowasserová z výtvarných podkladov Viliama Slaminku vytvorených pre Slovenský národný korpus JÚĽŠ SAV.



© Mária Šimková, Katarína Gajdošová, Beáta Kmeťová, Marek Debnár
Vydal: Jazykovedný ústav Ľ. Štúra Slovenskej akadémie vied, 2017

Mikula, s. r. o.

Vydanie: prvé

Všetky práva vyhradené. Dielo ani žiadna jeho časť sa nesmie reprodukovať, ukladať do informačných systémov ani inak rozširovať bez súhlasu majiteľov autorských práv.

ISBN 978-80-88814-98-6





Obsah

1. Úvod.....	7
2. Spôsob tvorby, štruktúra a sprístupňovanie textových a jazykových zdrojov Slovenského národného korpusu	13
Čo je korpus.....	15
Jazykové a textové zdroje SNK a ich sprístupňovanie.....	15
Tvorba a štruktúra korpusov SNK	19
3. Bibliografická a štýlovo-žánrová anotácia dokumentov a textov v Slovenskom národnom korpuse	33
Úvod	35
Východiská bibliografickej a štýlovo-žánrovej anotácie SNK.....	35
Anotácia dokumentov v archíve SNK.....	38
Konverzia textov v SNK.....	39
Anotácia textov v banke SNK.....	41
4. Lematizácia a morfológická anotácia slov jednotiek v Slovenskom národnom korpuse	61
Úvod	63
Token, lema, tag.....	65
Systém morfológických kategórií a značiek v SNK	73
Zastúpenie slovných druhov a tried v textoch SNK	82
5. Možnosti vyhľadávania v korpusových dátach.....	91
Úvod	93
Registrácia.....	94
Vstup do NoSketch Engine a výber korpusu	94
Zadanie hľadaného výrazu, konkordancia	98
CQL vyhľadávanie	103
Filtrovanie vyhľadaných konkordancií	107
Hľadanie podľa morfológických kategórií a typu textov.....	111
Hľadanie ustálených spojení.....	122
Hľadanie výslovnostných realizácií v Slovenskom hovorenom korpuse	125
Hľadanie v paralelnom korpuse.....	128
6. Výberový slovník termínov a pojmov z korpusovej lingvistiky ...	135
7. Výberový zoznam metaznakov, operátorov a podmienok na vyhľadávanie	151
8. Literatúra a zdroje	159





4. číslo	s	jednotné	slovo, ryba, ústav, muž
	p	množné	slová, ryby, ústavy, muži/mužovia
5. pád	1	nominatív	muž, vedúci, matka, Tichá, more, mláďa
	2	genitív	muža, vedúceho, matky, Tichej, mora, mláďaťa
	3	datív	mužovi, vedúcemu, matke, Tichej, moru, mláďaťu
	4	akuzatív	muža, vedúceho, matku, Tichú, more, mláďa
	5	vokatív	mami, Soni, oci
	6	lokál	mužovi, matke, Tichej, mori, mláďati
	7	inštrumentál	mužom, vedúcim, matkou, Tichou, morom, mláďaťom

Ukážka:

Aby dostala autoritatívny charakter /SSiS4 , nie je uvedený jej autor / SSms1 , ale jej Predmluva /SSfs1 je podpísaná šiestimi vedúcimi osobnosťami /SSfp7 slovenského národného hnutia /SSns2 ...

Adjektívum

Pozícia	Znak	Hodnota	Príklad
1. slovný druh	A	adjektívum	milý, svieži, priateľkin, psí
2. paradigma	A	adjektívna	pravý, cudzí, páví
	F	zmiešaná	otcov, matkin
	U	neúplná	super, nanič, hoden, rád, rada
3. kongruencia v rode	m	mužský životný	láskavý (otec), svieži (muž), matkin (známy)
	i	mužský neživotný	láskavý (pohl'ad), svieži (vietor), matkin (plášť)
	f	ženský	pravá (láska), zimná (vetrovka), otcova (kniha)
	n	stredný	nové (srdce), super (predstavenie), čierne (psíča)
4. kongruencia v čísle	s	jednotné	láskavý (otec), láskavý (pohl'ad), otcova (kniha), super (predstavenie), čierne (psíča)
	p	množné	láskaví (otcovia), láskavé (pohl'ady), otcove (knihy), super (predstavenia), čierne (psíčatá)
5. kongruencia v páde	1	nominatív	milý (priateľ), drahá (mama)
	2	genitív	milého (priateľa), drahej (mamy)
	3	datív	milému (priateľovi), drahej (mame)
	4	akuzatív	milého (priateľa), drahú (mamu)
	5	vokatív	milý (priateľ/priateľa), drahá (mami)
	6	lokál	milom (priateľovi), drahej (mame)
	7	inštrumentál	milým (priateľom), drahou (mamou)





6. stupeň	x	pozítív	vzácný, drahá, otcov, psí, horúci
	y	komparatív	vzácnejší, drahší, drevenejší (tanečník)
	z	superlatív	najvzácnejší, najdrahší, najdrevenejší (tanečník)

Ukážka:

Aby dostala autoritatívny /AAis4x charakter , nie je uvedený jej autor , ale jej Predmluva je podpísaná šiestimi vedúcimi osobnosťami slovenského /AAns2x národného /AAns2x hnutia ...

Pronomen

Pozícia	Znak	Hodnota	Príklad
1. slovný druh	P	pronominum	akýkoľvek, onen, jeho, kadiaľ
2. paradigma	S	substantívna	koľkátko, vsetučko
	A	adjektívna	aký, ktorá, inakšie, samý
	P	zámenná	ja, ty, my, vy, seba, sebe
	F	zmiešaná	on, ona, ono, kto, ten, môj, všetok, čo, žiaden
	U	neúplná	koľko, jeho, jej, ich
	D	príslovková	ako, kam, kde, kade, vtedy, začo
3. rod, resp. kongruencia	m	mužský životný	oni, moji, tí, nejakí (priatelia)
	i	mužský neživotný	ony, moje, tie, nejaké (knihy)
	f	ženský	ona, moja, tá, nejaká (priateľka)
	n	stredný	ono, moje, to, nejaké (vyjadrenie)
	h	všeobecný	ja, ty, my, vy, seba
4. číslo, resp. kongruencia	s	jednotné	ja, ktorý, ten, nikto, nejaký
	p	množné	my, ktorí, tí, všetci, nijakí
5. pád, resp. kongruencia	1	nominatív	ja, svoj, ktorý, nič, on
	2	genitív	mňa, svojho, ktorého, ničoho, neho
	3	datív	mne, svojmu, ktorému, ničomu, nemu
	4	akuzatív	mňa, svojho, ktorého, nič, oňho
	5	vokatív	môj (pane, priateľu), naša (mami)
	6	lokál	mne, svojom, ktorom, ničom, ňom
	7	inštrumentál	mnou, mojím, ktorým, ničím, ním
6. aglutinovanosť	g	aglutinované	preňho, naňho, oň, zaň, doň





Ukážka:

Aby dostala autoritatívny charakter , nie je uvedený jej /PUfs1 autor , ale jej /PUfs1 Predmluva je podpísaná šiestimi vedúcimi osobnosťami slovenského národného hnutia ...

Numerále

Pozícia	Znak	Hodnota	Príklad
1. slovný druh	N	numerále	jeden, dva, raz, sto, prvý, dvojmo
2. paradigma	S	substantívna	nula, milión, státisíce, raz
	A	adjektívna	jediný, prvý, dvojitý, obojaký
	N	číslovková	dva, dvaja, oba, obaja, obidva, tri, štyri
	F	zmiešaná	jeden, jedna, jedno
	U	neúplná	sto, tisíc, dvoje, tisícero, byť, veľa, plno, málo
	D	príslovková	prvýkrát, sedemkrát, dvojmo, neraz, mnohorako
	3. rod, resp. kongruencia	m	mužský životný
	i	mužský neživotný	jeden (tvorca), jedny (nohavice), dva (domy)
	f	ženský	jedna (žemľa), jedny (dvere), dve (priateľky), piatim (známkam)
	n	stredný	jedno (okno), jedny (plúca), dve (puzdrá), päť (srdc), (pracovalo) päť (mužov)
4. číslo, resp. kongruencia	s	jednotné	jeden (muž), druhá (osoba), trojaké (víno)
	p	množné	jedny (osoby), druhé (miesta), trojaké (vína)
5. pád, resp. kongruencia	1	nominatív	jeden (pán), druhá (osoba), trojaké (víno)
	2	genitív	jedného (pána), druhej (osoby), trojakého (vína)
	3	datív	jednému (pánovi), druhej (osobe), trojakému (vínu)
	4	akuzatív	jedného (pána), druhú (osobu), trojaké (víno)
	5	vokatív	(človeče) jeden!
	6	lokál	jednom (pánovi), druhej (osobe), trojakom (víne)
	7	inštrumentál	jedným (pánom), druhou (osobou), trojakým (vínom)

Ukážka:

Prvým /NAms7 držiteľom bol Darryl F . Zanuck , ktorý si toto ocenenie odniesol neskôr ešte dva /NNip4 razy /NSip4. Na opravu jedného /NFis2 bytu





treba priemerne od 200 - do 400 – tisíc /NUns4 korún , a práve tu štát štátnou
prémiou iniciuje občanov k sporiteľským aktivitám.

Verbum

Pozícia	Znak	Hodnota	Príklad
1. slovný druh	V	verbum	klásť, čítať, vidieť, činiť
2. slovesná forma	I	infinitív	byť, hriať, volať, viesť, hovoriť
	K	prézent (indikatív)	je, hreje, volá, vedie, hovorí
	M	imperatív	bud'!, hrej!, volajte!, veďte!, hovor!
	H	prechodník	súc, hrejúc, volajúc, vedúc, hovoriac
	L	l-ové prídavné	bol, hrialo, volali, viedla, hovorili
	B	futúrum byť	budem, budeš, bude, budeme, budete, budú, poletím, povedú
3. vid	d	dokonavý	zohrejem, zavolám, povieme
	e	nedokonavý	budem hriať, volala som, bola by hovorila
	j	obojvidové sloveso	aplikovať, počuť
4. číslo	s	jednotné	je, hrialo, bude, poviem
	p	množné	sú, volali, budeme, hovorili
5. osoba	a	prvá	som, sme, hrejme!, volali sme, budem hovoriť
	b	druhá	si, ste, hriali ste, volajte!, budeš viesť, hovoril by si
	c	tretia	je, sú, hrejú, volalo, povedie, hovoria
6. kongruencia v rode	m	mužský životný	(otec) prišiel, (priatelia) volali
	i	mužský neživotný	(strom) rozkvitol, zafúkal (vietor)
	f	ženský	(známka) sa prilepila, (dievky) tancovali
	n	stredný	(dielo) vyšlo, (šetrenie) pripísalo (dôležitosť)
	h	všeobecný	(vy ste) prišli
	o	neurčený/ neurčiteľný	(chlapi, ženy i deti) sa tešili
7. negácia	-	negácia	nebyť, neprichádzať, nepriatelieť sa, nebať sa
	+	afirmácia	prichádzať, priatelieť sa, rásť

Ukážka:

Pri každom jedle sa ho doprosovala /VLescf+ , aby nebol /VLescm- sebec-
ký a aby si uvedomil /VLdscm+ , že malá sestrička je /VKesc+ chorá a takisto
potrebuje /VKesc+ jesť /Vle+ , ale márne .





Formálne particípium

Pozícia	Znak	Hodnota	Príklad
1. slovná trieda	G	participium	robiaci, sediaci, naložený, zohriaty
2. druh	k	aktívne	pracujúci, visiaci, píšuci, platiaci
	t	pasívne	robený, kosený, obratý, zožatý
3. kongruencia v rode	m	mužský životný	sediaci (otec), pracujúci (muž), vyhodенý (známy)
	i	mužský neživotný	kričiaci (pohľad), pracujúci (stroj), pokrčený (plášť)
	f	ženský	čistená (ulica), strhujúca (kampaň), ušitá (košeľa)
	n	stredný	opustené (mesto), písané (potvrdenie), skáčuce (mača)
4. kongruencia v čísle	s	jednotné	píšuci (otec), uplakaný (pohľad), vypratá (košeľa)
	p	množné	píšuci (otcovia), vypraté (košele), skáčuce (mačatá)
5. kongruencia v páde	1	nominatív	kreslený (film), bežiaci (reklama), ušité (sako)
	2	genitív	kresleného (filmu), bežiacej (reklamy), ušitého (saka)
	3	datív	kreslenému (filmu), bežiacej (reklame), ušitému (saku)
	4	akuzatív	kreslený (film), bežiacu (reklamu), ušité (sako)
	5	vokatív	(synu môj) neveriaci!
	6	lokál	kreslenom (filme), bežiacej (reklame), ušitom (saku)
	7	inštrumentál	kresleným (filmom), bežiacou (reklamou), ušitým (sakom)
6. stupeň	x	pozitív	kričiaci, hodená, skáčuce
	y	komparatív	uplakanejší, strhujúcejší
	z	superlatív	najuplakanejší, najstrhujúcejší

Ukážka:

Je to oblačnosť súvisiaca /Gkfs1x so spomínaným /Gtis7x studeným frontom . Vie sa len , že sochy budú umiestnené /Gtfp1x na terase v Ulici spievajúceho /Gkns2x mora .

Adverbium

Pozícia	Znak	Hodnota	Príklad
1. slovný druh	D	adverbium	prísne, milo, pravidelne, prázdno
2. stupeň	x	pozitív	draho, vzácne
	y	komparatív	drahšie, vzácnejšie
	z	superlatív	najdrahšie, najvzácnejšie





Ukážka:

A robilo sa tak správne /Dx , lebo sa vyžadovali , najmä na politické účely , krátke , úsečné slová s jednoznačným významom , rýchlo /Dx vysloviteľné a rezonujúce v myslí rečníka čo najkratšie /Dz .

Prepozícia

Pozícia	Znak	Hodnota	Príklad
1. slovný druh	E	prepozícia	po, pre, na, do, cez, medzi
2. forma	v	vokalizovaná	so, zo, odo, podo
	u	nevokalizovaná	s, z, od, pod, prostredníctvom
3. spojenie s pádom	2	genitív	od, do, z
	3	datív	k, ku
	4	akuzatív	za, na
	6	lokál	o, na
	7	inštrumentál	s, so, medzi

Ukážka:

Zo /Ev2 sna ťa prebudí bozk mládenca , ktorý ťa bude nado /Ev4 všetko ľúbiť. A to najmä v /Eu6 súvislosti s /Eu7 cenou vstupného.

Konjunkcia

Pozícia	Znak	Hodnota	Príklad
1. slovný druh	O	konjunkcia	a, ale, alebo, či, pretože, že
2. kondicionálnosť	Y	kondicionálnosť	aby, keby, čoby, žeby

Ukážka:

Odišiel som , ako /O som prišiel , sťaby /OY dajaký ostrieľaný sprisahanec . Verím ti , počula som jasne mamu , akoby /OY stála nad nami .

Partikula

Pozícia	Znak	Hodnota	Príklad
1. slovný druh	T	partikula	azda, nuž, bodaj, sotva, áno, nie
2. kondicionálnosť	Y	kondicionálnosť	kiežby, žeby

Ukážka:

Ozdravenie bánk je nepochybne /T prvým a zrejme /T rozhodujúcim krokom pri vytváraní priaznivého podnikateľského prostredia. STV však /T nie /T je v štandardnej situácii a musí bojovať o každého diváka.

Interjekcia

Pozícia	Znak	Hodnota	Príklad
1. slovný druh	J	interjekcia	fíha, bác, bums, dokelu, ahoj, cveng, plesk





Ukážka:

Ach /J , prosím vás , nekričte tak! Deti sa chytili za ruky , urobili okolo Maji s Vilkom kruh a spievali : dzum /J , dzum /J , dzum /J ...

Reflexívum

Pozícia	Znak	Hodnota	Príklad
1. slovná trieda	R	reflexívum	sa, si

Ukážka:

Ako ľahko by sa /R dalo uverit' , keby sa /R človek nemusel obzerať okolo seba , pomyslel si /R Winston ...

Kondicionálová morféma

Pozícia	Znak	Hodnota	Príklad
1. slovná trieda	Y	kondicionálová morféma	by

Ukážka:

To by /Y sa mali naučiť' aj kluby .

Abreviácie, značky

Pozícia	Znak	Hodnota	Príklad
1. slovná trieda	W	abreviácie, značky	km, kg, atď., H2O, SND

Ukážka:

Na prelome júla a augusta , v období tzv /W . psích dní , u nás leto vrcholí .

Interpunkcia

Pozícia	Znak	Hodnota	Príklad
1. slovná trieda	Z	interpunkcia	., !, (, +

Ukážka:

Na prelome júla a augusta , /Z v období tzv . /Z psích dní , /Z u nás leto vrcholí . /Z Deti sa chytili za ruky , /Z urobili okolo Maji s Vilkom kruh a spievali : /Z dzum , /Z dzum , /Z dzum . /Z . /Z . /Z

Neurčiteľný slovný druh

Pozícia	Znak	Hodnota	Príklad
1. slovná trieda	Q	neurčiteľný slovný druh	bielo(-čierny), New (York)

Ukážka:

Otvorene vnímam hodnotenie ťažkej sociálno /Q - ekonomickej situácie väčšiny občanov na Slovensku.





Slovenský národný korpus

Server: bonito.korpus.sk

Konkordancia

- [Nové hľadanie](#)
- [Zoznam slov](#)

Dokumentácia

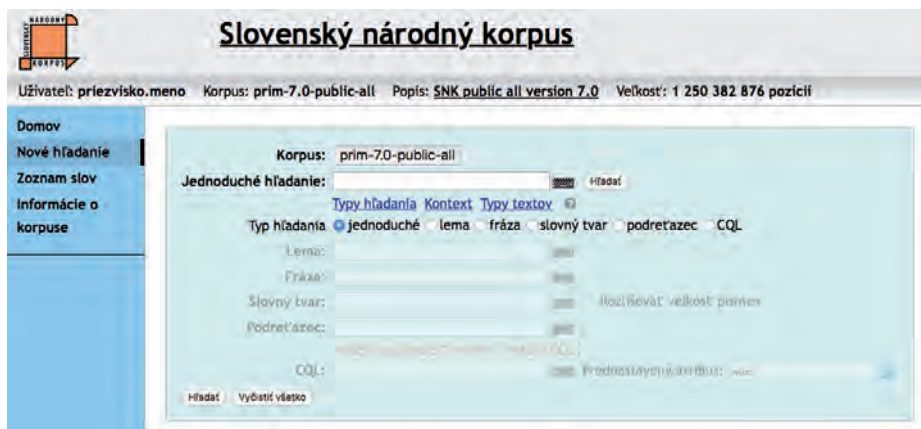
- [Morfologická anotácia SNK \(sk\)](#)
- [Bibliografická a štýlovo-žánrová anotácia SNK \(sk\)](#)
- [Sketch Engine Wiki \(en\)](#)
- [Sketch Engine Documentation Index \(en\)](#)

Nástroje

- [Zmena hesla](#)

Obrázok č. 1

Po kliknutí na položku **Nové hľadanie** sa otvorí výzva na zadanie prihlasovacích údajov. Úspešnosť prihlásenia sa prejaví zobrazením stránky s ponukou na vyhľadávanie:

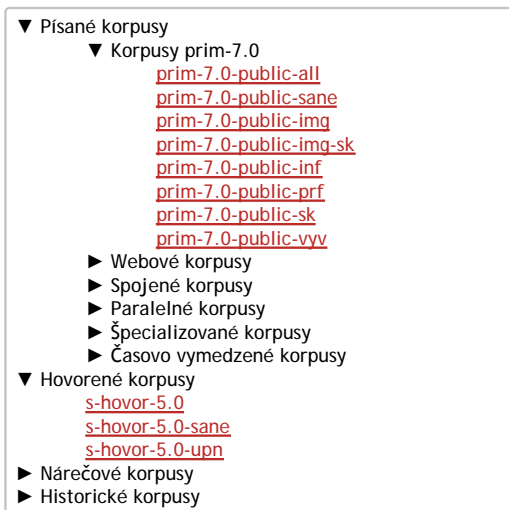


Obrázok č. 2

Na lište pod logom SNK sa nachádza prihlasovacie meno používateľa a názov korpusu, s ktorým sa bude pracovať. Ďalej je stručný popis daného korpusu, ktorý zároveň slúži ako odkaz na príslušnú podstránku s podrobnejšími informáciami o konkrétnom korpuse na stránke SNK. Poslednou informáciou je veľkosť korpusu uvedená ako počet pozícií (počet pozícií = počet tokenov, čiže všetkých slov a iných jednotiek nachádzajúcich sa v korpuse).



Po kliknutí na obdĺžnik za názvom Korpus: sa zobrazí ponuka verejne dostupných korpusov (porov. aj tabuľku v kap. 2). Jednotlivé korpusy a podkorpusy možno zobrazit' rozbalením skupiny po kliknutí na trojuholníček pred názvom množiny korpusov.



Obrázok č. 3

Kliknutím na názov korpusu alebo podkorpusu sa príslušné dáta aktivujú na vyhľadávanie jazykových prostriedkov a informácií. Pre naše potreby si v tomto kroku vyberieme hlavný korpus prim-7.0-public-sane, ako typ hľadania vyberieme atribút lema a do vyhľadávacieho okienka zapíšeme slovo *korpus*. Ak všetko dobre funguje (na strane používateľa aj Slovenského národného korpusu), v priebehu niekoľkých sekúnd sa zobrazia všetky výskyty slova *korpus* vo všetkých tvaroch s krátkymi kontextmi a ďalšími základnými informáciami:

Hľadané korpus 1 791 (1,64 na milión) ARF: 478,00		
KBS12/10	po vykreslení jej životného príbehu a priblížení obsiahleho	korpusu /korpus/SSis2 Hildegardinho diela poukazuje na aktuálnosť jej učenia pre
KN2012/3	rokokoh Stradivari vyrábal husle s predĺženým , elegantným	korpusom /korpus/SSis7 , pevnejším kľentím a , tmavším " zvukom
SAS33	minulosti je dnes už takmer povinnou výbavou každého	korpusu /korpus/SSis2 aj nejaká miera lingvistických informácií - lematizácia ,
TY2011/0	v kontexte a dalo sa študovať . Z	korpusu /korpus/SSis2 sa potom tvoria nové slovníky a odvodzujú pravidlá
SSAV2002	súčasnú financovanie vedy na Slovensku by sa Národný	korpus /korpus/SSis1 dal len ťažko vybudovať ako vedecký projekt Jazykovedného
AŠK1	a židovských cínorinov podlieha osobitnej právnej ochrane .	Korpus /korpus/SSis1 zákonov týkajúcich sa židov je dôkazom dôkladnej znalosti
JLed1	alebo deskripcie . Spomedzi kníh nášho , obmedzeného	korpusu /korpus/SSis2 " sa nám Soboulova Francúzska revolúcia javí ako
IHos1	(vo vzťahu k frekvenciám a príznačnosti v	korpusu /korpus/SSis6 umeleckých textov) sú v týchto dvoch skupinách
SOS15	i didakticky interpretuje . Slovenského a Českého národného	korpusu /korpus/SSis2 . Výsledky porovnávania využíva aj vo výučbe voliteľných
LJan1	, opäť sme využívali zoznam lem slovenského národného	korpusu /korpus/SSis2 . Tento údaj mal podobu čísel strán ,
SOS11) . Preto v súčasnosti neprebíha len budovanie	korpusu /korpus/SSis2 textov , ale je možná už aj výskumná
SME2009/	Mandtovu šlahačku natrieme na jahody , zakryjeme druhým	korpusom /korpus/SSis7 a jemne pritlačíme . Tortu zo všetkých strán
DF1999/2	zaniknuté , monumentálne a prísne - meravé ľudské	korpusy /korpus/SSip1 z vrecoviny , sediace ako pohoniči na osi
PODV99/1	. 52 , 26 Sk - Laskonkový	korpus /korpus/SSis1 200 g
KBS12/03	. Prvou etapou bolo reštaurovanie kompozície križov s	korpusmi /korpus/SSip7 lotrov a Krista , sochy boli deinštalované v
InZ03/10	svoj účel . Mokka bomba 1 hotový tortový	korpus /korpus/SSis1 2 smotaný na šľahanie 2 čokolády na varenie
KBS07/08	stavy zastavení obnovil . Pribudol vysoký križ s	korpusom /korpus/SSis7 Ježiša Krista , postavami Panny Márie , apoštola
MYBB200	dvojča Ako správne skladovať potraviny Snažte sa pre	korpusy /korpus/SSip4 . hotové dezerty , sušienky a rôzne chutovky
SME03/05	moste performance Hojdačka , keď bude zavesená na	korpus /korpus/SSis4 Nového mosta a svoje hojdanie zopakuje v pondelok
BLOG	a tvrdosti cesta 3 - 5 mm hrubý	korpus /korpus/SSis4 . Je to rovnako ako s palacinkami .

Obrázok č. 4

