In this document we give the guidelines that were used to annotate the morphological makeup of verbs and instructions for the use of the database.

Root

The cell 'root' can be used as an indicator which verbs belong to the same word family. Crucially, we do not claim that the annotated form is the exact phonological form of the root of the verb. Rather, this should be taken as a cell that can help the users group together verbs that are related by some common core (the chosen form could in principle be replaced by a random number/index).

Still, we tried to capture what we believe is the common core of all the related verbs. The material in the cell 'root' therefore refers to the verb without prefixes, thematic vowels or verbal suffixes.

Note that verbs that display suppletion between the perfective and imperfective form are annotated as having different roots. In cases where the suppletive form is related to another existing verb (*oditi* 'to go away.pfv', *odhajati* 'to go away.ipfv' where haj has the same root as *hoditi* 'to walk') the suppletive form and the related verb are marked as having the same root (in this case h-d, see below for hyphen).

Examples (for both BCMS and Slovenian):

Slo/BCMS	pisati	pis	'to write' the verb has no prefix, removed are the TV *a* and the infinitival ending
Slo/BCMS	napisati	pis	'to write.pfv' removed are the prefix *na*-, the TV *a* and the infinitival ending
Slo	prepisovati	pis	'to write.ipfv' removed are the prefix -*na* and the verbal suffix *ov*, the TV *a* and the infinitival ending
BCMS	prepisivati	pis	'to write.ipfv' removed are the prefix -*na* and the verbal suffix *iv*, the TV *a* and the infinitival ending

As evident from the table above, the meaning of verbs was taken as a motivating factor to determine that verbs with homophonous material belong in the same word family. In addition, another motivation for the choice of the root were alternation patterns. If multiple verbs have the same alternation pattern in their paradigm and in the aspectual pair, we treated them as having the same root.

Examples from Slovenian

Slo	razložiti	l-g	'to explain.pfv', 'to put down.pfv'
Slo	razlagati	l-g	'to explain.ipfv', 'to put down.ipfv'
Slo	položiti	l-g	'to lay down.pfv'
Slo	polagati	l-g	'to lay down.ipfv'
BCMS	odabrati	b-r	'to chosse.pfv'
BCMS	odabirati	b-r	'to chose.ipfv'
BCMS	sabrati	b-r	'to sum.pfv'
BCMS	sabirati	b-r	'to sum.ipfv'

If the verb displays root allomorphy, we choose one of the forms of the root. While we gravitated towards what plausibly is the underlying form, this is not necessarily always so. Crucially, the same form of the root is then used with all the verbs within the same word family.

Slo	počiti	pok	'to burst.pfv'
Slo	pokati	pok	'to burst.ipfv'
Slo	napočiti	pok	'to arrive.pfv'
BCMS	izbeći	beg	'to avoid.pfv'
BCMS	izbegnuti	beg	'to avoid.pfv'
BCMS	izbegavati	beg	'to avoid.ipfv'
BCMS	bežati	beg	'to run away.ipfv'

If there are two potential homophonous roots and only one of them displays an alternation, we use this to separate it from the other root. We return to the hyphen.

Slo	dotakniti	t-k	'to touch.pfv'
Slo	dotikati	t-k	'to touch.ipfv'
Slo	točiti	t-č	'to pour.ipfv'
Slo	natakati	t-č	'to pour.ipfv'
BCMS	primaknuti	m-k	'to get closer.pfv'
BCMS	primicati	m-k	'to get closer.ipfv'

BCMS	umočiti	m-č	'to dip.pfv'
BCMS	umakati	m-č	'to dip.ipfv'

Notes about annotation

Numbers

Where two groups of verbs seem to have homophonous roots that cannot be connected with a similar meaning, we differentiate roots with numbers.

Slo	prati	p-r	'to wash.ipfv'
Slo	sprati	p-r	'to wash away.pfv'
Slo	spirati	p-r	'to wash away.ipfv'
Slo	odpreti	p-r1	'to open.pfv'
Slo	odpirati	p-r2	'to open.ipfv'
BCMS	biti	bi	'to be.ipfv'
BCMS	zbivati	bi	'to happen.ipfv'
BCMS	ubiti	bi2	'to kill.pfv'
BCMS	ubijati	bi2	'to kill.pfv'

Hyphen

We use a hyphen to indicate roots that display apophony (i.e. don't have stable vowels), for example between the imperfective and perfective form of the verb.

Slo	brati	b-r	'to read.ipfv'
Slo	prebrati	b-r	'to read.pfv'
Slo	prebirati	b-r	'to read.pfv'
BCMS	ubosti	b-d	'to stab.ipfv'
BCMS	ubadati	b-d	'to stab.ipfv'
BCMS	bosti	b-d	'to stab.ipfv'

Plus

While we remove any strictly verbal material (prefixes, TV, suffixes) from the verb when annotating the 'root',multifunctional affixes are included in this column and are marked with a plus (+). Plus is also used in compounds (these are also annotated with 1 in the column Compounds).

Slo	beračiti	ber+ač	'to beg.ipfv' berač 'beggar', where -ač is a common normalizer
Slo	krvaveti	krv+av	'to bleed.ipfv' krvav 'bloody' where -av is a common adjectiviser
Slo	telovaditi	tel+o+vad	'to exercise.ipfv' tel+o+vadba 'exercise' from tel-o 'body' + vad-b-a 'training'
BCMS	krvariti	krv+ar	'to bleed.ipfv'
BCMS	besniti	bes+n	'to rage.ipfv'
BCMS	obelodaniti	bel+o+dan	'to disclose.pfv'

Colon

In the BCMS Database, the colon mark is used in cases where the annotator has the intuition there is a prefix, but the root reconstructed on this analysis does not appear in other words. e.g. *u:zeti* 'take', where *zeti* does not appear in any other verb. The colon is used in both the root column (as *u:ze*) and the columns listing prefixes (as *u:*).

In Slovenian, the colon is used when the base potentially includes a prefix, but the verb itself is not derived by prefixation (these instances are also marked as potentially prefixed). E.g., *odmevati* 'to echo' derived from the noun *odmev* 'echo' (whereby **mevati* is not a verb and *od-* is a well-attested preposition/prefix) is marked as having the root *od:mev*.

Prefixes

In what follows we describe how prefixation is annotated in the database.

The general guideline that we followed in determining if a verb is prefixed or not is 'if it can be taken to be a prefix, then it is a prefix', where the starting point were existing lists of prefixes in BCMS and Slovenian (Babić 1986, Klajn 2002 for BCMS, Toporišič 2000 for Slovenian).

To this end, the first step was to determine if a verb that seems to be prefixed has an unprefixed pair, e.g. Slovenian *prebrati* 'to finish reading' has the pair *brati* 'to read', the BCMS *pročitati* 'to finish reading' has the pair *čitati* 'to read'. If the seemingly prefixed verb does not have an unprefixed pair, we have checked if a verb with the same root, but a different prefix exists. If it does, the verb was taken to be prefixed and annotated as such. Examples of this kind are, for example, verbs with the root *četi* in Slovenian - *za-četi* 'start', *po-četi* 'to do', *na-četi za-četi* 'start', pri-četi, *za-četi* 'start' (even though *četi* does not exist). Finally, if there are no corresponding unprefixed verbs or verbs with the same root but other prefixes, verbs were still marked as prefixed if their root is independently attested in other categories (e.g. nouns, adjectives, compounds). Examples of this kind include e.g. *uz-nemiriti* 'upset' in BCMS, where *nemiriti* is not attested on its own or with other prefixes, but *nemir* 'unrest' exists as a noun.

In general, the annotation shows if the verb is prefixed and, if it is, which prefixes are present on the verb (see below). In addition to the columns described below, two other columns are relevant. The

column Prefixed_verb shows if a verb has a prefix or prefixes (1) or not (0). This information is also relevant for the column Simplex_verb too - in this column either a prefix or any type of an affix, it is marked with a 0.

Prefixes are marked in 4 columns (Prefixes (list), 1, 2 and 3). In the column Prefixes (list), we show all the prefixes that the verb contains, separated by a "+". The next three columns show specific prefixes, as well as the position of each individual prefix with respect to the root, where column 1 refers to the position closest to the root, column 2 the second one, and column 3 the prefix that is most distant from the root. This means that the order of prefixes in the columns is reversed compared to their linear order in the verb itself (and in column Prefixes (list)). If there is no prefix in the respective position, we annotate this with 0. More examples are given below.
Note that this annotation only marks the linear position of the prefix and that we are making no claims about the syntactic position of each prefix (whether it is lexical, intermediate, super-lexical etc.).

BCMS	prefixes			Slovenian	prefixes		
	1	2	3		1	2	3
brati	0	0	0	brati	0	0	0
odbrati	od	0	0	nabrati	na	0	0
otpočeti	po	od	0	razočarati	o	raz	0
sporazumeti	raz	po	s	prerazporediti	po	raz	pre

Additional information

Potentially prefixed

If the verb cannot be annotated as prefixed using the criteria described above, but there is a strong intuition that the verb contains a prefix, this is marked in column Potentially_prefixed_verbs with 1. For instance, in BCMS, the verb *za-taškati* 'to cover up.pfv' seems to include the prefix *za-* (which retains the same meaning as in other verbs with *za-*, e.g. *za-kopati* 'burry'), but the base *taškati* is not attested independently or with other prefixes).

Language	Example	translation	Potentially prefixed	Explanation
BCMS	zataškati	'to cover up'	1	no *taškati* on its own or with other prefixes
Slo	odmevati	'to echo'	1	*za-mevati, *na-mevati, *mevati

Finally, verbs that are taken to be potentially prefixed are not taken to be simplex (i.e. they are marked with a 0 in column Simplex_verbs).

Issues

In some cases it is unclear whether the item in question is a single prefix or if it could be further separated. One such example is Slovenian *izpod-* where prefixes *iz-* and *pod-* also exist (as do prepositions *iz*, *pod* and izpod). In the database we decompose these prefixes.

Loan prefixes were not annotated as prefixes (e.g., *re-* in *reciklirati* 'recycle' (both languages) or *de-* in the BCMS *destabilizovati* 'destabilize').

Compounds

The column marks if the base is a compound (1) or not (0). The relevant criteria here is if the base consists of more than one lexical root, e.g., (Slo/BCMS) *blag-o-slov-i-ti* 'to bless'.

Suffixes

In these columns we annotate individual suffixes that appear in the verbs. Specifically, we mark the presence of a suffix with a 1 and we in general separate three types of phenomena:

1. Verbal suffixes, i.e. 1) secondary imperfectivisers which consist of more than just theme vowels, 2) verbalisers used (also) in loanword integration which consist of more than just theme vowels, and 3) other suffixes with a clear semantic contribution in the verbal domain (most typically aspect).
2. Multifunctional suffixes (these are also annotated in the column Root)
3. Suffix-like and other

To annotate these items we first created a list of candidates for each type. In what follows we give the lists of all three types of suffixes for BCMS and Slovenian separately, as the two have different suffixes.

Column 'Suffixed' reflects a summary of all this annotation - verbs that have no suffix are marked with a 0, verbs that have either a verbal suffix, a potential non-verbal suffix or an suffix-like element get a 1.

Column 'Simplex_verbs' marks with 1 verbs that do not have any of the abovementioned suffixes or prefixes and are not compounds, otherwise, it is marked with a 0.

<u>Verbal suffixes</u>

We consider verbal suffixes to have verbalizing, aspectual and/or modifying function (e.g., diminutive). Verbal suffixes are annotated in two ways. In column <u>Verbal suffix (y/n)</u> it is simply shown if a verb has a verbal suffix (1) or does not have a verbal suffix (0). It is then specified which verbal suffix the verb has in one of the following columns (with 1 if the verb has that suffix), where each of the columns refers to s specific suffix. We give the lists of verbal suffixes below. Note that the list for Slovenian and BHS are slightly different.

Verbal suffixes in BCMS

Verbal suffix	Example	Translation
av+a	iz-beg-av-a-ti	'to avoid.ipfv'
ek+a	prd-ek-a-ti	'to fart.impf'
et+a	lup-et-a-ti	'to fool around.ipfv'
ir+a	aktiv-ir-a-ti	'to activate'
is+a	afirm-is-a-ti	'to affirm'
iv+a	do-ček-iv-a-ti	'to welcome.ipfv'

k+a	čač-k-a-ti	'to poke around.ipfv'
	pre-ćut-k-iv-a-ti	'to keep quiet repeatedly.ipfv'
n+0	trep-nu-ti, trep-n-e-m	'to blink, I blink (pfv)'
	sta-∅-ti, sta-n-e-m	'to stop, I stop (pfv)'
	na-dah-nj-iv-a-ti, na-dah-nj-uj-e-m	'to inspire, I inspire (ipfv)'
ov+a	ver-ov-a-ti	'to believe.ipfv'
t+a/a	za-huk-t-a-ti	'to heat up.pfv'
t+a/i	drht-a-ti	'to tremble.ipfv'
t+a/je	tre-p-t-a-ti	'to blink.ipfv'
t+e/i	drh-t-je-ti	'to tremble.ipfv'
uc+a	pij-uc-k-a-ti	'to sip.ipfv '

Verbal suffixes in Slovenian

Verbal suffix	Example	Translation	Notes
ni/ne	kih-ni-ti	'to sneeze.pfv'	
av+a	o-dobr-av-a-ti	'to approve.ipfv'	
ov+a/u+je	vz-dih-ov-a-ti dež-ev-a-ti	'to sigh.ipfv' 'to rain.ipfv'	In some context, ev+a/u+je
ev+a/ev+a	pre-hit-ev-a-ti	'to overhaul.ipfv'	
ir+a	identific-ir-a-ti	'to identify.ipfv'	
ik+a	svet-l-ik-a-ti	'to gleam.ipfv'	
t+a	tep-t-a-ti	'to trample.ipfv'	
t+e	iz-puh-t-e-ti	'to evaporate.ipfv'	
k+a	tip-k-a-ti	'to type.ipfv'	
et+a	lesk-et-a-ti	'to glitter.ipfv'	
lj+a	pri-cur-lj-a-ti	'to trickle out'	
v+a	pri-dob-i-v-a-ti	'to extract.ipfv'	

Multifunctional suffixes

Suffixes that do not have a verbalizing, aspectual or modifying function, are annotated as multifunctional as they (i) potentially originate from a different category and (ii) have a vague meaning. These suffixes are annotated in a single column called 'Multifunctional suffixes'. We mark if the verb has one one of the following suffixes (see the list for BCMS and Slovenian below) with 1 (if not, 0). To see which multifunctional suffix the verb has, see the column Root. As above, the list of suffixes are different for BCMS and Slovenian.

Multifunctional suffixes in BCMS

Multifunctional suffix	Example	Translation
(a)c	u-nov-c-i-ti (> unovčiti)	'to cash out.pfv'
ać	o-dom-ać-i-ti	'to domesticate.pfv'
ač	ras-tum-ač-i-ti	'to clarify.pfv'
aj	pre-za-log-aj-i-ti	'to have a snack.pfv'
(a)k	peš-ak-i-ti (> pešačiti)	'to hike.ipfv'
(a)lj	češ-lj-a-ti	'to comb.ipfv'
(a)n	ob-zn-an-i-ti	'to announce.pfv'
ar	krv-ar-iti	'to bleed.ipfv'
at	svoj-at-a-ti	'to appropriate.ipfv'
av	s-mrš-av-i-ti	'to lose weight.pfv'
c	jur-c-a-ti	'to run around.ipfv'
en	o-kam-en-i-ti	'to petrify.pfv'
ev	o-duš-ev-i-ti	'to enthuse.pfv'
er	trep-er-i-ti	'to flicker.ipfv'
ic	gran-ič-i-ti	'to border.ipfv'
ič	sumnj-ič-i-ti	'to suspect.ipfv'
ij	rob-ij-a-ti	'serve a sentence.ipfv'
ik	jad-ik-ova-ti	'to moan.ipfv'
in	bašt-in-i-ti	'to inherit.ipfv'
iš	jur-iš-a-ti	'to rush.ipfv'
išt	s-klad-išt-i-ti	'to store.ipfv'

iv	u-milost-iv-i-ti	'to propitiate.pfv'
iz	simbol-iz-ova-ti	'to symbolize.pfv'
j	jak-j-a-ti (> jačati)	'to strenghten.ipfv'
k	piš-k-i-ti	'to pee.ipfv'
lj	kap-lj-a-ti	'to drop.ipfv'
n	bes-n-i-ti	'to rage.ipfv'
ok	s-ved-oč-i-ti	'to testify,ipfv'
olj	mig-olj-i-ti	'to squirm.ipfv'
os	žig-os-a-ti	'to stigmatize.ipfv'
oš	o-pust-oš-i-ti	'to devastate.pfv'
ot	sram-ot-i-ti	'to disgrace.ipfv'
ov	u-tjel-ov-i-ti	'to embody.pfv'
š	mek-š-a-ti	'to soften.ipfv'
telj	s-prija-telj-i-ti	'make a friend.pfv'
uć	o-mog-uć-i-ti	'to enable.pfv'
ulj	smej-ulj-i-ti	'to smile.ipfv'
uš	pev-uš-i-ti	'to croon.ipfv'
ut	skak-ut-a-ti	'to jump around.ipfv'

Slovenian: Multifunctional suffixes

Multifunctional suffix	Example	Translation
oš	pust-oš-i-ti	'to ravage.ipfv'
ic	kop-ič-i-ti	'to accumulate.ipfv'
n	bes-n-e-ti	'to rage.ipfv'
in	koren-in-i-ti	'to root.ipfv'
it	o-plemen-it-i-ti	'to refine.pfv'
išč	sklad-išč-i-ti	'to store.ipfv'
ač	u-dom-ač-i-ti	'to domesticate.pfv'

ak	en+ak-iti [enačiti]	'to equate.ipfv'
l	raz-svet-l-i-ti	'to illuminate.pfv'
ar	koles-ar-i-ti	'to bike.ipfv'
k	slad-k-a-ti	'to sweeten.ipfv'
b	u-glas-b-i-ti	'to set music to.ipfv'
š	z-manj-š-a-ti	'to reduce.ipfv'
ot	o-sram-ot-i-ti	'to disgrace.ipfv'
oč	o-mog-oč-iti	'to enable.pfv'
lj	kap-lj-a-ti	'to dribble.ipfv'
eh	bol-eh-a-ti	'to suffer.ipfv'
telj	s-po-prija-telj-i-ti	'to befriend.ipfv'
il	krm-il-i-ti	'to steer.ipfv'
av	krv-av-e-ti	'to bleed.ipfv'
ov	raz-pol-ov-i-ti	'to halve.pfv'
ol	mrg-ol-e-ti	'to teem.ipfv'
et	klep-et-a-ti	'to chat.ipfv'

<u>Suffix-like</u>

If the verb apparently has the form (pref+)root+theme, but there are reasons to assume a more complex structure, this is indicated with a one (1) in this column.

For Slovenian, this column is intended for secondary imperfective verbs in which we do not necessarily have phonologically overt material (which is annotated in the verbal suffixes columns). Two indicators of this additional structure are the ablaut (the root exhibits a vowel change) and a -j-. Note that while -v- has the same distribution, for Slovenian we annotate v as a verbal suffix (on par with ava, ova, eva). The Slovenian and the BCMS databases are different in this respect, because some of the arguments in Simonović et al. (2023) for non-suffixal nature of *v* in BCMS do not extend to Slovenian.

All types of examples that are annotated as suffix-like in the BCMS database are listed in the tables below (for the rationale behind the assumed segmentation, see Simonović et al. 2023).

BCMS: Suffix-like

Example	Assumed segmentation	Translation
padati	pad+Ø+a+ti	'to fall.ipfv'
skretati	skret+Ø+a+ti	'to diverge.ipfv'

davati	da+Ø+a+ti	'to give.ipfv'
duvati	du+Ø+a+ti	'to blow.ipfv'
snabdevati	snabde+Ø+a+ti	'to supply.ipfv'
primati	prim+i+μa+ti	'to receive.ipfv'
stizati	stig+Ø+[+sibilant]a+ti	'to arrive.ipfv'
zamišljati	zamisl+i+μa+ti	'imagine.ipfv'
skrivati	skri+Ø+μa+ti	'to hide.ipfv'
upijati	up+i+μa+ti	'to absorb.ipfv'
voziti	vez+Ø+o-i+ti	'to drive.ipfv'
odmicati	odm-k+i[+sibilant, -velar]-a+ti	'to move away.ipfv'
uzdisati	uz+d-h+i[+sibilant, -velar]-a+ti	'to sigh.ipfv'
ubirati	ub-r+a+i-a+ti	'to harvest.ipfv'
obasipati	obas-p+Ø+i-a+ti	'to shower.ipfv'
uplitati	uplet+Ø+i-a+ti	'to plait.ipfv'
umirati	umr+Ø+i-a+ti	'to die.ipfv'
oduzimati	uduz-m+Ø+i-a+ti	'to take away.ipfv'
sklapati	sklop+i+μa-a+ti	'to assemble.ipfv'
počinjati	[Supp]	'to begin.ipfv'
odlaziti	[Supp]	'to leave.ipfv'

Slovenian: Suffix-like

Example	Assumed segmentation	Translation
padati	pad+Ø+a+ti	'to fall.ipfv'
dosegati	doseg+Ø+a+ti	'to reach.ipfv'
ubirati	u+br+i+a+ti	'to choose.ipfv'
umirati	u+mr+i+a+ti	'to die.ipfv'
spejemati	s+pre+jem+Ø+a+ti	'to accept.ipfv'
zamišljati	zamisl+i+a+ti	'to imagine.ipfv'

poglabljati	po+glob+a+i+a+ti	'to deepen.ipfv'
prevajati	pre+ved+a+i+a+ti	'to translate.ipfv'
ogrožati	o+groz+i+a+ti	'to threaten.ipfv'
jemati	[Supp]	'to take.ipfv'
prihajati	[Supp]	'to arrive.ipfv'

Simplex verbs

This column shows which verbs are simplex (1) and which not (0) - verbs that have a prefix or a potentially prefixed, that have any of the shown suffixes or are compounds are taken *not* to be simplex.

References

Babić, Stjepan. 2002. *Tvorba riječi u hrvatskome književnome jeziku*. Zagreb: HAZU – Nakladni zavod Globus.

Klajn, Ivan. 2002. *Tvorba reči u savremenom srpskom jeziku. Prvi deo: Slaganje i prefiksacija.* Belgrade / Novi Sad: The SASA Institute for Serbian language / Matica srpska.

Simonović, Marko, Stefan Milosavljević & Boban Arsenijević. 2023. Serbo-Croatian secondary imperfectivisers consist of theme vowels. Submitted to *Journal of Slavic linguistics*.

Toporišič, Jože. 2000. *Slovenska slovnica*. Maribor: Založba Obzorja.