

## Introduction

- ▶ Russian has rich inflection; therefore its general morphological tagsets are necessarily large.
- ▶ We have developed a positional tagset for Russian with about 2000 tags
- ▶ Inspired by the Czech Positional Tagset (Hajič 2004)
- ▶ The tagset has been used in a resource-light tagger (Feldman & Hana 2010)

## Positional Tagset

- ▶ A tag is a composition of tags each coming from a much smaller and simpler atomic tagset associated with a particular morpho-syntactic property (e.g., gender or tense).
- ▶ All tags have the same length, encoding all the features distinguished by the tagset,
- ▶ Features not applicable for a particular word have a N/A value.
- ▶ Which features are applicable is generally determined by fine-grained part of speech (SubPOS)
- ▶ E.g., AAFAS4-----2A-- encodes adjective (A), feminine gender (F), animate (A), singular (S), accusative (4), comparative (2), not-negated (A)

## Positions of the Russian tagset

Position	Abbr	Name	Nr. of values
1	p	Part of Speech	12
2	s	SubPOS (Detailed Part of Speech)	42
3	g	Gender	4
4	y	Animacy	3
5	n	Number	3
6	c	Case	7
7	f	Possessor's Gender	4
8	m	Possessor's Number	2
9	e	Person	4
10	r	Reflexivity	2
11	t	Tense	4
12	b	Verbal aspect	3
13	d	Degree of comparison	3
14	a	Negation	2
15	v	Voice	2
16	i	Variant, Abbreviation	7

## Restrictions

- ▶ Gender in plural is distinguished by nouns only
- ▶ The X wild-card values are used in the following cases only:
  - ▷ Gender: agreement gender in plural (adjectives, participles, etc.), plurale-tantum nouns, non-declinable adjectives, personal pronouns in 3rd person plural.
  - ▷ Animacy: Except for nouns, in all forms except accusative masculine singular and accusative plural of all genders. Non-declinable words.
  - ▷ Number: non-declinable nouns, adjectives and verbs, 3rd person possessive pronouns.
  - ▷ Case: non-declinable nouns and adjectives, 3rd person poss. pronouns.
  - ▷ Possessor's Gender: for the 3rd person plural possessive pronoun.
  - ▷ Person: for non-declinable verbs
  - ▷ Tense: for passive long participles (AG).
  - ▷ Aspect: bi-aspectual verbs, e.g., *ispol'zovat'* 'to use'.

## Participles

Participles are classified as adjectives:

<i>čitajuščij</i>	AGMXS1---IPI-AA-	active (A) present (P) participle
<i>čitavšij</i>	AGMXS1---IRI-AA-	active (A) past (R) participle
<i>pročitavšij</i>	AGMXS1---IRP-AA-	active (A) past (R) participle
<i>čitaemyj</i>	AGMXS1---IXI-AP-	passive (P) long (imperf/perf) participle
<i>pročitan</i>	AcM-S-----I-P-AP-	passive (P) perf. short participle

All *-nyj* participles/adjectives (*ostavlennyj* 'deserted', *varenyj* 'cooked') are considered to be general adjectives, because it is very hard to draw the line between their purely adjectival and participial use.

## Possible tags

N – Nouns			
NNgync-----a--	noun	golos	NNMIS4-----A--
A – Adjectives (incl. Participles)			
AAgync-----da--	long adjective	tjaželyj	AAMIS4-----1A--
ACg-n-----a--	short adjective	krasiv	ACM-S-----A--
AGgync---rtb-av-	long participle	čitajuščij	AGMXS1---IIP-AA-
	tv ∈ {PA, RA, XP}, i.e. present/past active, passive	smejuščajasja	AGFXS1---RRP-AA-
AUgyncf-----a--	possessive adjective	mužnin	AUMXS2M-----A--
Acg-n-----aP-	pass.perf.short participle	pročitan	AcM-S-----AP-
P – pronoun			
PP--nc--eI-----	personal pronoun e ∈ {1,2}	nam	PP--P3--1I-----
PPg-nc--3I-----	personal pronoun 3rd person	on	PPM-S1--3I-----
PP--c---R-----	personal reflexive sebj	sebj	PP--4---R-----
P5g-nc--3I-----	personal p. in prep. forms	nego	P5M-S2--3-----
PDgync-----	demonstrative	ètu	PDFXS4-----
PW--c-----	negative (nominal declension)	ničto	PW--1-----
Pwgync-----	negative (adj declension)	nikakoj	PwMXS1-----
PSgync--meI-----	possessive	moja	PSFXS1--S1I-----
PSXXXfm3I-----	possessive	ego	PSXXXMS3I-----
PSgync--R-----	possessive reflexive	svoj	PSMXS1--R-----
PQ--c-----	relative/interrogative (nom decl)	što, kto	PQ--1-----
Pqgync-----	relative/interrogative (adj decl)	kakoj	PqMXS1-----
PZ--c-----	indefinite (nom. decl.)	kogo-to	PZ--4-----
Pzgync-----	indefinite (adj. decl.)	kakoj-to	PzMXS1-----
C – Numeral			
C=-----	numbers (using digits)	3.14	C=-----
Cj-----	roman numeral	XVII	Cj-----
Cngync-----	cardinal numeral 1	odnomu	CnMAS3-----
Cngy-c-----	cardinal numeral 2, poltora	dvux	CnMA-2-----
Cn-y-c-----	cardinal numeral 3,4	trëx	Cn-A-4-----
Cn-y-c-----	cardinal numeral 5+	pjati	Cn-A-2-----
Crngync-----	ordinal	pervyj	CrMXS1-----
Cj-y-c-----	generic/collective numeral	dvoix	Cj-A-3-----
Cu--c-----	interrogative	skol'ko	Cu--x-----
Ca--c-----	indefinite numeral	neskol'ko	Ca--1-----
Ca-gync-----	indefinite num. (adj decl.)	mnogomu	CaMXS3-----
Cv-----	multiplicative	triždi	Cv-----
V – verb			
VB--n---ertb----	present (rarely fut.) finite form	otryvaes'	VB--P--2IPI----
VBg-n---rRb----	past tense	čital	VBM-S---IRI----
Ve-----r-b----	gerund	grozja	Ve-----I-I----
		napisav	Ve-----I-P----
Vf-----r-b----	infinitive	spat'	Vf-----I-I----
Vi--n---er-b----	imperative	rabotaj	Vi--S---2I-I----
D – Adverb			
Db-----	adv. not forming negation/degrees	tam	Db-----
Dg-----da--	adv. forming negation/degrees	sil'nee	Dg-----2A--
R – Preposition			
RR--c-----	nonvocalized prep. with c case	nad	RR--7-----
RV--c-----	vocalized prep. with c case	nado	RV--7-----
RF-----	part of a multiword prep.	nesmotrja	RF-----
J – Conjunction			
J^-----	coordinating conj.	i	J^-----
J,------	subordinating conj.	čto	J,------
T – particle			
TT-----	particle	net	TT-----
I – Interjection			
II-----	Interjection		II-----
Z – punctuation			
Z#-----	Sentence boundary		Z#-----
Z:------	Punctuation	!	Z:------
X – special			
X0-----	part of a multiword foreign phrase		X0-----
XX-----	unknown		XX-----

## Possible values

Position 3 – Gender						Distinguished for: N, A{ACGUc}, P{P5DLwSq8}, C{nra}, VB							
F	Feminine	M	Masculine	N	Neuter	X	Any gender						
Position 4 – Animacy						Distinguished for: N, A{AGU}, P{SDwqz}, C{nrja}							
A	Animate	I	Inanimate	X	Either								
Position 5 – Number						Distinguished for: N, A{ACGUc}, P{P5DwSq}, C{nra}, V{Bp}							
P	Plural	S	Singular	X	Any number								
Position 6 – Case						Distinguished for: N, A{AGU}, P, C{nrjua}							
1	Nominative	2	Genitive	3	Dative	4	Accusative	6	Locative	7	Instrumental	X	Any case
Position 7 – Possessor's Gender						Distinguished for: PS, AU							
F	Feminine possessor	M	Masculine possessor	N	Neuter possessor	X	Possessor of any gender						
Position 8 – Possessor's Number						Distinguished for: PP							
P	Plural possessor	S	Singular possessor										
Position 9 – Person						Distinguished for: P{P5S}, V{Bi}							
1	1st person	2	2nd person	3	3rd person	X	Any person						
Position 10 – Reflexivity						Distinguished for: AG, P{P5S}, V							
R	Reflexive	I	Irreflexive										
Position 11 – Tense						Distinguished for: A{G}, V{Bp}							
F	Future	P	Present	R	Past	X	Any tense						
Position 12 – Aspect						Distinguished for: AG, V							
P	perfective	I	imperfective	X	either aspect								
Position 13 – Degree of comparison						Distinguished for: AA, Dg							
1	Positive	2	Comparative	3	Superlative								
Position 14 – Negation						Distinguished for: N, A, Dg							
A	Not negated	N	Negated										
Position 15 – Voice						Distinguished for: AG, Ac							
A	Active	P	Passive										
Position 16 – Variant						Distinguished for: As needed							
-	Basic variant	1-8	Variants										

## Sets of tags via regular expressions

- ▶ NN [MF] AS [1-3] -----A-- – masculine or feminine singular animate noun, in nominative, genitive or dative
  - ▶ NN . AS [^14] -----A-- – singular animate noun of any gender, not in nominative nor in accusative
  - ▶ NN (MI | FA) S1 -----A-- – masc. inanimate or feminine animate noun
- Mostly used for searching, but a dot can be also used when a value is not predicted/annotated.