2.3 The Positional Tag System for Czech

The Positional Tag System is a newer and preferred method of recording the morphological information associated with the usage of word forms. Therefore, it should be used in the future for all new annotation and other projects involving Czech morphology described herein. We also suppose that in the near future, the positional tag system will be extended to cover more categories important to natural language processing tasks, such as aspect (which in fact does have some morphological implications, currently hidden in the paradigm naming scheme), name entity type (such as "geographical name"), currently only part of the Term field (Sect. 4.8), style features, now part of the Style field (Sect. 4.7) while simultaneously unmarked or mixed together in the tag, word-forming derivation information etc. Such an extension will leave the current tag positons intact (even though possibly obsolete) for upward compatibility.

2.3.1 Format and Columns

Every tag is represented as a string of 15 symbols. Each position (also called "column") in the string corresponds to one morphological category according to a more or less traditional system of formal morphology. The part of the tag corresponding to a single position is also referred to as a **subtag**[3]. Every value in each category is thus represented as a single symbol, mostly an uppercase letter of the English alphabet (for example, `P` for plural), sometimes also another symbol (`f` for an infinitive, `^` for conjunctions). Non-applicable values are denoted by a single hyphen (`-`).

The traditional system of morphological categories (and their values) has never been consistently and completely described. We have taken a different approach than the traditional one also in cases where consistency (or suitability) of description would be unnecessarily complicated. For example, all nouns carry the category "negativeness", even though only some are commonly used in negative (with the prefix `ne-`); the reason is that there are two paradigms of nouns which regularly form negation: those derived from adjectives (and from verbal adjectives) with the "property" meaning (e.g. `ospalost` (lit. "sleepiness"), or `dělitelnost` (lit. "dividability"); traditional paradigm name: `kost`), and simple verbal nouns derived from passive participle (e.g. `vaření` (lit. "cooking" (N)); traditional paradigm name: `stavení`). It could be also argued that any noun can be negated in Czech (in sentences like `Strach nestrach,`

[3]As usual, we will sometimes use the term subtag also for the value found at the given position in some tag, since in practice this does not lead to confusion.

`musel odjet.`, (lit. "Fear (or) no-fear, (he) had to-leave")), even though this is treated only on a case-by-case basis in the Czech morphological dictionary.

A principle of "minimal" description is followed wherever possible. For example, when a category (in our case, the "detailed POS" category, which is used for this purpose) determines which other categories are relevant for a given word form, it is not repeated elsewhere. The same holds for the appropriate lemma, even though some compromises have been introduced here: namely, part of speech (`POS`) and the detailed part of speech (`SUBPOS`) subtags are *always* present, even though `POS` and in many cases[4] also `SUBPOS` could be deterministically derived from the lemma. Sometimes the same is true for other categories (e.g., gender (`GENDER`) of nouns, person (`PERSON`) of personal pronouns (which are represented by three distinct lemmas (`já, ty, on` (lit. "I, you, he") instead of a single one), etc. For most verb forms however, the `SUBPOS` category is independent of the lemma - rather, it determines which other categories are applicable for the corresponding word form. For example, the `SUBPOS` value `f` is used for the infinitive, for which no category except negativeness is relevant (and therefore the remaining 12 subtags are marked by a hyphen ("`-`") as non-applicable).

On the other hand, in some cases no distinction among "traditional" values is being made where the possibility of correctly distinguishing them is low based on local context. This could certainly be considered a kind of "cheating" from the tagging point of view, but one has to draw some boundary anyway (in the English tagging efforts there is also no distinction between e.g. persons in plural verbal forms!). For example, possessive pronouns in third person plural are not distinguished in gender and agreement number, nor in case; passive participles (both active and passive) in masculine are not distinguished in animateness, etc. Typically, a letter `X` is used where all possible values might be considered in a more detailed tagset, or a special letter is used with a more restricted choice (e.g. `Y` is used for masculine animate/masculine inanimate "non-distinction").[5] In particular, the letter `X` in the part-of-speech position is used to denote that no categories have been found for a particular word form (either because the form has not been identified as a correct Czech word form, or it has been identified but the morphological categories could not be assigned, for technical or other reasons).

[4]See the considerations in the `SUBPOS` category description below, Sect. 2.3.5.

[5]Recently, some difficulties have been discovered when applying this tagset in machine translation applications between closely related languages (Hajič, Hric and Kuboň, 2000). Therefore it might be necessary to develop a more detailed tagset with on-the-fly $(1 : n)$-type of mapping between the basic tagset described here and the more detailed one. Then it will be possible to apply always only the tagset suitable for a given application.

2.3.2 Description of the Values of the Categories

The following 15 categories are used as subtags in the positional tag system[6]:

Table 2.1: Categories (Subtag Names)

#	Category (Subtag Name)	Description in English	Description in Czech
1	POS	Part of Speech	Slovní druh
2	SUBPOS	Detailed Part of Speech	Slovní poddruh
3	GENDER	Agreement Gender	Rod
4	NUMBER	Agreement Number	Číslo
5	CASE	Case	Pád
6	POSSGENDER	Possessor's Gender	Rod vlastníka
7	POSSNUMBER	Possessor's Number	Číslo vlastníka
8	PERSON	Person	Osoba
9	TENSE	Tense	Čas
10	GRADE	Degree of Comparison	Stupeň
11	NEGATION	Negation (by prefix)	Negace
12	VOICE	Voice	Slovesný rod
13	RESERVE1	Reserved for future use	Rezerva
14	RESERVE2	Reserved for future use	Rezerva
15	VAR	Variant, Style, Register	Varianta, styl

The subtag names are the same names used in the processing software, including the tagger, as well as for online processing (see Sect. 7 for references).

[6]For the description of the compact tag system, see Sect. 2.4. For a complete 1:1 table with positional-to-compact tag mapping, refer to the electronic resources (Sect. 7).

2.3.3 Co-occurrence Tables

In the following section, three types of co-occurrence tables are used for a better understanding of the interaction of the individual categories in Czech:

- Category `POS`

  For each of the major part-of-speech `POS` values, a list of detailed part-of-speech category (i.e. `SUBPOS`) is given which are associated with it.

- Category `SUBPOS`

  For each value of the `SUBPOS` category, a table consisting of at most two parts is presented. First, the major part-of-speech value (`POS`) is given to which the current subcategory value uniquely belongs[7]. Then, each subsequent row of the table shows one category which contains at least once a *non*-N.A. ("not applicable", i.e. hyphen) value. All the values of such a category are listed in the right-hand column, including the N.A. value (if it has ever been used for the current `SUBPOS`). There might be no such additional rows in case there are no non-N.A. values used with that `SUBPOS` value in a tag (for example, for particles, punctuation, etc.).

- Other categories

  The tables presented for each of the non-N.A. values of all the remaining 13 categories are "reversed" `SUBPOS` tables, i.e. they contain a list of all `POS`/`SUBPOS` pairs (such as `AA`) for which the particular category (such as `GENDER`) *and* its value (such as `F`) has been used at least once. Although a bit redundant, the `POS` value common to several `SUBPOS` values is being shown in the left-hand column for easy reference, whereas the list of the possible `SUBPOS` values is in the right-hand column.

[7]Therefore, the `POS` value is in fact redundant, as it can be unambiguously determined from the `SUBPOS` value; however, for reasons of tradition and, more importantly, for reasons related to reliable estimates of parameters of statistically-based disambiguation (see Chapter 6), the major part-of-speech category is being kept as a category of its own.

2.3.4 Part of speech (`POS`)

The `POS` category denotes the major[8] part of speech, according to the traditional Czech scheme known from both comprehensive as well as high-school grammars. However, the assignment of the `POS` values is driven mainly by the requirements of consistency in further processing, therefore it is not always in line with such grammars. This problem concerns, primarily, particles (vs. adverbs, pronouns, nouns, etc.) and pronouns (vs. adjectives). Also, in cases where such traditional grammars do not agree or use different word sense[9] distinctions, or in cases where they escape unambiguous identification using part of speech classification such as “pronominal adverb”, the part of speech which can be assigned most consistently in local context is used.

Even though the `POS` category can take on a handful of values only, there are some difficulties with the “traditional” `POS` category values (apart from those vague designations found in many dictionaries). They stem from the fact that although Part of Speech is a (surface-) syntax category, it is often related to morphological properties of the lemma. Therefore, confusion exists as to what to assign to thoseadjectives which can systematically be used as nouns in Czech (and there are only a few which cannot)[10] in order to make them at least loosely fit what grammarians consider a reasonable (surface) syntax grammar of Czech. To be more consistent, we should assign `N` (noun) to almost all adjectives (as an alternative to `A`, for adjective), keeping other morphological information intact. However, since – again – part of speech does not influence the “inner working” of the formal morphological system at all, we would like to keep the invariant “one lemma corresponds to one part of speech” intact.

Obviously, this invariant in `POS` assignment is closely related to derivations. Since we have decided not to deal with derivations in the current work[11], we will simply define that if a root change, or adding a suffix, prefix etc. changes a `POS` category, then it is considered a derivation (there are also other derivations which do not change `POS`, but that has no influence on the `POS`/lemma invariant discussed here). This ensures the invariant holds throughout the dictionary.

[8] Or “main” in the linguistic terminology.

[9] Sometimes referred to also as lexical meaning.

[10] One could argue that this is a case of (syntactic) ellipsis, where the noun is missing. It is, but there is a high number of systematic cases like this in the language, and therefore we talk – at the (surface) syntax level – about *systemic ellipsis*. Systemic ellipsis is not considered an ellipsis any more, it is rather a special grammatical construct. Yes, there is also a “real”, so-called *actual ellipsis*, which however has no ramifications for morphology, of course.

[11] Some derivational information is present in the dictionary, though; namely, for regular derivations. See Sect. 4.5.1.

For each of the POS values[12], a list of detailed part of speech (SUBPOS) categories is given which are associated with it; a detailed description of the individual SUBPOS values is given below in Sect. 2.3.5.

Value count: 12

A Adjective

POS	Detailed part-of-speech used (SUBPOS)
A	. 2 A C G M O U

Examples:

POS & SUBPOS	possible form(s)	lit. translation (description)
A2	technicko-	technic(al/o)-
AA	technický	technical
AC	rád	(be) glad
AG	kouřící	smoking (adj.)
AM	zvítězivší	having-won (adj.)
AO	svůj	himself (to be h.)
AU	Martinin	Martina's
AU	Havlův	Havel's
A.		obsolete

[12]For printing space-related and aesthetic reasons, the co-occurrence table for verbs has been moved before prepositions, even though V should alphabetically go between T and X.

`C` Numeral

POS	Detailed part-of-speech used (SUBPOS)
C	3 = ? a d h j k l n o r u v w y z }

Examples:

POS & SUBPOS	possible form(s)	lit. translation (description)
C=	1.23	numbers (using digits)
Ca	mnoho	many
Cd	čtverý	four-fold (adj. declension)
Ch	jedny	one-kind-of
Cj	čtvero	four-kinds-of (noun usage)
Ck	čtvery	four-kinds-of (adj. usage)
Cl	tři	three (cardinal numeral 1-4)
Cn	pět	five (cardinal numeral > 4)
Co	tolikrát	that-many-times (indefinite)
Cr	druhý	second (any ordinal numeral)
Cu	kolikrát	how-many-times (interrog./rel.)
Cv	sedmkrát	seven-times (definite count)
Cw	nejeden	not-only-one (adj. declension, nom.)
Cy	desetina	one-tenth (fractions)
Cz	kolikátý	at-what-position (interrog./rel.)
C}	XIV	14 (Roman numerals)
C3		obsolete

`D` Adverb

POS	Detailed part-of-speech used (SUBPOS)
D	! b g

Examples:

POS & SUBPOS	possible form(s)	lit. translation (description)
Db	nahoru	up (no degrees of comparison)
Dg	rychle	quickly (degrees of c. possible)
D!		obsolete

I Interjection

POS	Detailed part-of-speech used (SUBPOS)
I	I

Example:

POS & SUBPOS	possible form(s)	lit. translation (description)
II	ach	oh!

J Conjunction

POS	Detailed part-of-speech used (SUBPOS)
J	* , ^

Examples:

POS & SUBPOS	possible form(s)	lit. translation (description)
J*	krát	times (binary math. operations)
J,	že	that (subordinate)
J^	a	and (coordinating)

N Noun

POS	Detailed part-of-speech used (SUBPOS)
N	; N

Example:

POS & SUBPOS	possible form(s)	lit. translation (description)
NN	robot	robot (any noun incl. proper)
N;		obsolete

P Pronoun

POS	Detailed part-of-speech used (SUBPOS)
P	0 1 4 5 6 7 8 9 D E H J K L P Q S W Y Z

Examples:

POS & SUBPOS	possible form(s)	lit. translation (description)
P0	naň	on-him (compound with -ň)
P1	jehož	whose (in relative clause)
P4	jaký	what
P4	který	which
P5	něj	him (he, after prep. only)
P6	sebe	himself (long form)
P7	se, si	refl. pronouns
P8	svůj	his (poss. refl. pronoun)
P9	něhož	who, in rel. clause, after prep.
PD	tento	this (demonstrative)
PE	což	which (in rel. clause)
PH	mě	me (pers. pron. clitic)
PJ	jenž	who, in rel. clause
PK	kdo	who (rel./interrogative)
PL	všechen	all
PP	ty	you (personal)
PQ	co	what (rel./interrogative)
PS	můj	my (possessive)
PW	nic	nothing (negative)
PY	oč	about-what (compound with -č)
PZ	nějaký	some
PZ	něco	something

`V` Verb

POS	Detailed part-of-speech used (SUBPOS)
V	B c e f i m p q s t ~

Examples:

POS & SUBPOS	possible form(s)	lit. translation (description)
VB	dělám	(I) do (present/future forms)
Vc	bychom	(we) would (conditional)
Ve	dělajíce	(they-)doing (transgressive pres.)
Vf	dělat	(to) do (infinitive)
Vi	dělejme	(let's) do (imperative)
Vm	udělav	(he-)having-done (transgr. past)
Vp	dělali	(they) did (past part.)
Vq	dělalť	(he) did (archaic with -ť)
Vs	děláno	(it) was-being-done (passive part.)
Vt	dělámť	(I) do (archaic with -ť)

`R` Preposition

POS	Detailed part-of-speech used (SUBPOS)
R	F R V

Examples:

POS & SUBPOS	possible form(s)	lit. translation (description)
RF	nehledě	regardless (part of a compound)
RR	v	in
RV	ve	in (with vocalization)

`T` Particle

POS	Detailed part-of-speech used (SUBPOS)
T	T

Examples:

POS & SUBPOS	possible form(s)	lit. translation (description)
TT	jen	only

`X` Unknown, Not Determined, Unclassifiable

POS	Detailed part-of-speech used (SUBPOS)
X	@ X x

Examples:

POS & SUBPOS	possible form(s)	lit. translation (description)
X@	qqq	form not found/recognized
X@	x	also: one-letter “graphics”
XX	Srebrenica	found, tag missing in dict.
Xx	red.	found as abbrev., no tag

`Z` Punctuation (also used for the “Sentence Boundary” token)

POS	Detailed part-of-speech used (SUBPOS)
Z	# :

Examples:

POS & SUBPOS	possible form(s)	lit. translation (description)
Z#	<s>, ###	sentence boundary
Z:	,	comma (any punctuation)

2.3.5 Detailed part of speech (SUBPOS)

This category is the most detailed one; it contains values for fine-grained distinction of the major part of speech category. Its primary technical purpose, however, is to serve as an indicator of applicability/non-applicability of other categories (i.e. the categories GENDER, NUMBER, CASE, etc. up to the last category, VAR). We are using unique values so that the value of the major speech category can be determined unambiguously from the value of the SUBPOS category.

For each value of the SUBPOS category, a table consisting of at most two parts is presented. Since the SUBPOS subtag in the positional tag system is the functional counterpart to the "prefixes" of the tags in the compact tag system, the corresponding compact tag prefix is shown in the heading of each table for easy reference. First, the major part-of-speech value is given to which the current subcategory value uniquely belongs. Then (if needed), each subsequent row of the table shows one category which contains at least once a *non*-N.A. ("not applicable", i.e. hyphen) value. All the values of such category are listed in the right-hand column, including the N.A. value (if it has ever been used for the current SUBPOS).

It should be noted that in most cases, this category can be deterministically recovered from the lemma alone. Generally, the exceptions are only verbs and some numerals, where the primary purpose of this category (i.e., to determine which other relevant categories are applicable) prevails and therefore every verbal and some numeral lemmas map to a *set* of SUBPOS values, whereas most other lemmas map to exactly one SUBPOS value. In both cases, however, the POS is determined unambiguously (see also Sect. 2.3.4 for more on POS).

Value count: 75

! Abbreviation used as an adverb (now obsolete)[13]

Compact prefix	DABBR
Category	Values used
POS	D

Category co-occurrence for SUBPOS = !

[13]Abbreviations are now being handled using standard tags for noun, adjectives, adverbs etc., typically with several Xs as category values, and VAR value of 8 to denote the fact that a given form is an abbreviation.

Sentence boundary (for the "virtual" word ###)

Compact prefix	ZSB
Category	**Values used**
POS	Z

Category co-occurrence for SUBPOS = #

* Word krát (lit.: "times") (POS: C, numeral)

Compact prefix	JC
Category	**Values used**
POS	J

Category co-occurrence for SUBPOS = *

, Conjunction subordinate (incl. aby, kdyby in all forms)

Compact prefix	J*[**]
Category	**Values used**
POS	J
NUMBER	- P S X
PERSON	- 1 2 3
VAR	- 1

Category co-occurrence for SUBPOS = ,

. Abbreviation used as an adjective (now obsolete)

Compact prefix	AABBR
Category	**Values used**
POS	A

Category co-occurrence for SUBPOS = .

0 Preposition with attached -ň (pronoun něj, lit. "him"); proň, naň, ... (POS: P, pronoun)

Compact prefix	PPD
Category	**Values used**
POS	P

Category co-occurrence for SUBPOS = 0

1 Relative possessive pronoun jehož, jejíž, ... (lit. “whose” in subordinate relative clause)

Compact prefix	PSE*****
Category	Values used
POS	P
GENDER	F I M N X Z
NUMBER	D P S X
CASE	1 2 3 4 6 7 X
POSSGENDER	F X Z
POSSNUMBER	P S
PERSON	3
VAR	- 2

Category co-occurrence for SUBPOS = 1

2 Hyphen (always as a separate token)

Compact prefix	HYPH[*]
Category	Values used
POS	A
NEGATION	A N
VAR	- 1

Category co-occurrence for SUBPOS = 2

3 Abbreviation used as a numeral (now obsolete)

Compact prefix	CABBR
Category	Values used
POS	C

Category co-occurrence for SUBPOS = 3

4 Relative/interrogative pronoun with adjectival declension of both types ("soft" and "hard") (jaký, který, čí, ..., lit. "what", "which", "whose", ...)

Compact prefix	PQF***
Category	**Values used**
POS	P
GENDER	F I M N X Y Z
NUMBER	D P S X
CASE	1 2 3 4 6 7 X
VAR	- 3 6 7

Category co-occurrence for SUBPOS = 4

5 The pronoun "he" in forms requested after any preposition (with prefix n-: něj, něho, ..., lit. "him" in various cases)

Compact prefix	PP3R***
Category	**Values used**
POS	P
GENDER	F N X Z
NUMBER	P S
CASE	2 3 4 6 7
PERSON	3
VAR	- 1

Category co-occurrence for SUBPOS = 5

6 Reflexive pronoun se in long forms (sebe, sobě, sebou, lit. "myself" / "yourself" / "herself" / "himself" in various cases; se is personless)

Compact prefix	PRX*
Category	**Values used**
POS	P
NUMBER	X
CASE	2 3 4 6 7

Category co-occurrence for SUBPOS = 6

7 Reflexive pronouns `se` (`CASE` = 4), `si` (`CASE` = 3), plus the same two forms with contracted `-s`: `ses, sis` (distinguished by `PERSON` = 2; also number is singular only)

Compact prefix	PR***[**]
Category	Values used
POS	P
NUMBER	S X
CASE	3 4
PERSON	- 2

Category co-occurrence for `SUBPOS` = 7

8 Possessive reflexive pronoun `svůj` (lit. "my"/"your"/"her"/"his" when the possessor is the subject of the sentence)

Compact prefix	PRS***
Category	Values used
POS	P
GENDER	F H I M N X Y Z
NUMBER	D P S
CASE	1 2 3 4 5 6 7
VAR	- 1 6 7

Category co-occurrence for `SUBPOS` = 8

9 Relative pronoun `jenž, již, ...` after a preposition (`n-`: `něhož, niž, ...`, lit. "who")

Compact prefix	PAER***
Category	Values used
POS	P
GENDER	F N X Z
NUMBER	P S
CASE	2 3 4 6 7
VAR	- 1 2 3

Category co-occurrence for `SUBPOS` = 9

: Punctuation (except for the virtual sentence boundary word `###`, which uses the `SUBPOS #`)

Compact prefix	`ZIP`
Category	Values used
`POS`	`Z`

Category co-occurrence for `SUBPOS` = `:`

; Abbreviation used as a noun (now obsolete)

Compact prefix	`NABBR`
Category	Values used
`POS`	`N`

Category co-occurrence for `SUBPOS` = `;`

= Number written using digits (`POS`: `C`, numeral)

Compact prefix	`ZNUM`
Category	Values used
`POS`	`C`

Category co-occurrence for `SUBPOS` = `=`

? Numeral `kolik` (lit. "how many"/"how much")

Compact prefix	`CQ*`
Category	Values used
`POS`	`C`
`CASE`	`1 2 3 4 6 7`

Category co-occurrence for `SUBPOS` = `?`

@ Unrecognized word form (`POS`: `X`, unknown)

Compact prefix	`NOMORPH[*]`
Category	Values used
`POS`	`X`
`VAR`	`- 0 1`

Category co-occurrence for `SUBPOS` = `@`

A Adjective, general

Compact prefix	A*****
Category	Values used
POS	A
GENDER	F I M N X
NUMBER	D P S X
CASE	1 2 3 4 5 6 7 X
GRADE	1 2 3
NEGATION	A N
VAR	- 1 6 7 8 9

Category co-occurrence for SUBPOS = A

B Verb, present or future form

Compact prefix	V****
Category	Values used
POS	V
NUMBER	P S
PERSON	1 2 3
TENSE	F P
NEGATION	A N
VOICE	A
VAR	- 1 2 3 4 5 6 7 8

Category co-occurrence for SUBPOS = B

C Adjective, nominal (short, participial) form rád, schopen, ...

Compact prefix	AC***[*]
Category	Values used
POS	A
GENDER	F M N Q T Y
NUMBER	P S W
CASE	- 4
NEGATION	A N

Category co-occurrence for SUBPOS = C

D Pronoun, demonstrative (`ten, onen, ...`, lit. “this”, “that”, “that ... over there”, ...)

Compact prefix	PD***
Category	**Values used**
POS	P
GENDER	F I M N X Y Z
NUMBER	D P S
CASE	1 2 3 4 6 7
VAR	- 1 2 5 6 8

Category co-occurrence for SUBPOS = D

E Relative pronoun `což` (corresponding to English “which” in subordinate clauses referring to a part of the preceding text)

Compact prefix	PE*
Category	**Values used**
POS	P
CASE	1 2 3 4 6 7

Category co-occurrence for SUBPOS = E

F Preposition, part of; never appears isolated, always in a phrase (`nehledě (na), vzhledem (k), ...`, lit. “regardless”, “because of”)

Compact prefix	RF
Category	**Values used**
POS	R

Category co-occurrence for SUBPOS = F

G Adjective derived from present transgressive form of a verb

Compact prefix	AVG****
Category	**Values used**
POS	A
GENDER	F I M N
NUMBER	D P S
CASE	1 2 3 4 5 6 7
NEGATION	A N
VAR	- 6 7

Category co-occurrence for SUBPOS = G

H Personal pronoun, clitical (short) form (mě, mi, ti, mu, ...); these forms are used in the second position in a clause (lit. "me", "you", "her", "him"), even though some of them (mě) might be regularly used anywhere as well

Compact prefix	PP****[*]
Category	Values used
POS	P
GENDER	- Z
NUMBER	S
CASE	2 3 4
PERSON	1 2 3

Category co-occurrence for SUBPOS = H

I Interjections (POS: I)

Compact prefix	I
Category	Values used
POS	I

Category co-occurrence for SUBPOS = I

J Relative pronoun jenž, již, ... *not* after a preposition (lit. "who", "whom")

Compact prefix	PAE***
Category	Values used
POS	P
GENDER	F I M N X Y Z
NUMBER	D P S
CASE	1 2 3 4 6 7
VAR	- 1 2 3

Category co-occurrence for SUBPOS = J

K Relative/interrogative pronoun kdo (lit. "who"), incl. forms with affixes -ž and -s (affixes are distinguished by the category VAR (for -ž) and PERSON (for -s))

Compact prefix	PQ***[**]
Category	Values used
POS	P
GENDER	M
CASE	1 2 3 4 6 7
PERSON	- 2
VAR	- 2

Category co-occurrence for SUBPOS = K

L Pronoun, indefinite všechnen, sám (lit. "all", "alone")

Compact prefix	PL***
Category	Values used
POS	P
GENDER	F I M N X Y Z
NUMBER	D P S
CASE	1 2 3 4 5 6 7
VAR	- 1 6

Category co-occurrence for SUBPOS = L

M Adjective derived from verbal past transgressive form

Compact prefix	AVV****
Category	Values used
POS	A
GENDER	F I M N
NUMBER	D P S
CASE	1 2 3 4 5 6 7
NEGATION	A N
VAR	- 6 7

Category co-occurrence for SUBPOS = M

N Noun (general)

Compact prefix	N****
Category	**Values used**
POS	N
GENDER	F I M N X
NUMBER	D P S X
CASE	1 2 3 4 5 6 7 X
NEGATION	A N
VAR	- 1 2 3 4 5 6 7 8 9

Category co-occurrence for SUBPOS = N

O Pronoun svůj, nesvůj, tentam alone (lit. “own self”, “not-in-mood”, “gone”)

Compact prefix	A1**
Category	**Values used**
POS	A
GENDER	F I M N Y
NUMBER	P S
VAR	- 1 6

Category co-occurrence for SUBPOS = O

P Personal pronoun já, ty, on (lit. "I", "you", "he") (incl. forms with the enclitic -s, e.g. tys, lit. "you're"); gender position is used for third person to distinguish on/ona/ono (lit. "he/she/it"), and number for all three persons[14]

Compact prefix	PP***[*]
Category	Values used
POS	P
GENDER	- F I M N X Y Z
NUMBER	P S
CASE	1 2 3 4 5 6 7
PERSON	1 2 3
TENSE	- P
NEGATION	- A
VOICE	- A
VAR	- 1 2 6

Category co-occurrence for SUBPOS = P

Q Pronoun relative/interrogative co, copak, cožpak (lit. "what", "isn't-it-true-that")

Compact prefix	PQC*
Category	Values used
POS	P
CASE	1 2 3 4 6 7
VAR	- 9

Category co-occurrence for SUBPOS = Q

R Preposition (general, without vocalization)

Compact prefix	R*
Category	Values used
POS	R
CASE	1 2 3 4 6 7 X
VAR	- 1 3 8

Category co-occurrence for SUBPOS = R

[14] It has been decided to allow for the enclitic -s only for verbs (in the regular paradigms), even though it is in principle possible to add the -s to *any* Czech word form.

S Pronoun possessive `můj`, `tvůj`, `jeho` (lit. “my”, “your”, “his”); gender position used for third person to distinguish `jeho`, `její`, `jeho` (lit. “his, her, its”), and number for all three pronouns

Compact prefix	PS*****[*]
Category	Values used
POS	P
GENDER	F H I M N X Y Z
NUMBER	D P S X
CASE	1 2 3 4 5 6 7 X
POSSGENDER	- F X Z
POSSNUMBER	P S
PERSON	1 2 3
VAR	- 1 6 7 8

Category co-occurrence for SUBPOS = S

T Particle (POS: T, particle)

Compact prefix	T
Category	Values used
POS	T
VAR	- 1

Category co-occurrence for SUBPOS = T

U Adjective possessive (with the masculine ending `-ův` as well as feminine `-in`)

Compact prefix	AS****
Category	Values used
POS	A
GENDER	F I M N X
NUMBER	D P S X
CASE	1 2 3 4 5 6 7 X
POSSGENDER	F M
VAR	- 1 6 7 8

Category co-occurrence for SUBPOS = U

V Preposition (with vocalization `-e` or `-u`): (`ve, pode, ku`, ..., lit. “in”, “under”, “to”)

Compact prefix	RV*
Category	Values used
POS	R
CASE	1 2 3 4 6 7
VAR	- 1

Category co-occurrence for SUBPOS = V

W Pronoun negative (`nic`, `nikdo`, `nijaký`, `žádný`, `...`, lit. “nothing”, “nobody”, “not-worth-mentioning”, “no/none”)

Compact prefix	PN**[**]
Category	Values used
POS	P
GENDER	- F I M N X Y Z
NUMBER	- D P S
CASE	1 2 3 4 5 6 7
VAR	- 2 6 7

Category co-occurrence for SUBPOS = W

X (temporary) Word form recognized, but tag is missing in dictionary due to delays in (asynchronous) dictionary creation

Compact prefix	X
Category	Values used
POS	X
VAR	- 8

Category co-occurrence for SUBPOS = X

Y Pronoun relative/interrogative `co` as an enclitic (after a preposition) (`oč, nač, zač`, lit. “about what”, “on/onto what”, “after/for what”)

Compact prefix	PQD
Category	Values used
POS	P

Category co-occurrence for SUBPOS = Y

Z Pronoun indefinite (`nějaký`, `některý`, `číkoli`, `cosi`, `...`, lit. "some", "some", "anybody's", "something")

Compact prefix	PI**[**]
Category	Values used
POS	P
GENDER NUMBER CASE VAR	- F I M N X Y Z - D P S 1 2 3 4 5 6 7 - 1 2 4 6 7

Category co-occurrence for SUBPOS = Z

^ Conjunction (connecting main clauses, not subordinate)

Compact prefix	JE
Category	Values used
POS	J
VAR	- 1 2 8

Category co-occurrence for SUBPOS = ^

a Numeral, indefinite (`mnoho`, `málo`, `tolik`, `několik`, `kdovíkolik`, `...`, lit. "much/many", "little/few", "that much/many", "some (number of)", "who-knows-how-much/many")

Compact prefix	CI*
Category	Values used
POS	C
CASE VAR	1 2 3 4 5 6 7 X - 1

Category co-occurrence for SUBPOS = a

b Adverb (without a possibility to form negation and degrees of comparison, e.g. `pozadu`, `naplocho`, ..., lit. "behind", "flatly"); i.e. both the NEGATION as well as the GRADE attributes in the same tag are marked by - (Not applicable)

Compact prefix	DB
Category	Values used
POS	D
VAR	- 1 2 4 6 7 8 9

Category co-occurrence for SUBPOS = b

c Conditional (of the verb být (lit. "to be") only) (by, bych, bys, bychom, byste, lit. "would")

Compact prefix	VC**
Category	Values used
POS	V
NUMBER	P S X
PERSON	1 2 3
VAR	- 6

Category co-occurrence for SUBPOS = c

d Numeral, generic with adjectival declension (dvojí, desaterý, ..., lit. "two-kinds/...", "ten-...")

Compact prefix	CD***
Category	Values used
POS	C
GENDER	F I M N X Y
NUMBER	D P S
CASE	1 2 3 4 5 6 7
VAR	- 1 2 6 7

Category co-occurrence for SUBPOS = d

e Verb, transgressive present (endings -e/-ě, -íc, -íce)

Compact prefix	VG***
Category	Values used
POS	V
GENDER	H X Y
NUMBER	P S
NEGATION	A N
VAR	- 2

Category co-occurrence for SUBPOS = e

`f` Verb, infinitive

Compact prefix	VF*
Category	Values used
POS	V
NEGATION	A N
VAR	- 1 2 3 4 6

Category co-occurrence for SUBPOS = f

`g` Adverb (forming negation (NEGATION set to A/N) and degrees of comparison GRADE set to 1/2/3 (comparative/superlative), e.g. `velký`, `zajímavý`, ..., lit. "big", "interesting"

Compact prefix	DG**
Category	Values used
POS	D
GRADE	1 2 3
NEGATION	A N
VAR	- 1 2 3 6 8

Category co-occurrence for SUBPOS = g

`h` Numeral, generic; only `jedny` and `nejedny` (lit. "one-kind/sort-of", "not-only-one-kind/sort-of")

Compact prefix	CD1***
Category	Values used
POS	C
GENDER	F I M N X Y
NUMBER	D P
CASE	1 2 3 4 5 6 7

Category co-occurrence for SUBPOS = h

`i` Verb, imperative form

Compact prefix	VM***
Category	Values used
POS	V
NUMBER	P S
PERSON	1 2 3
NEGATION	A N
VAR	- 1 2 3 4 6 7 8 9

Category co-occurrence for SUBPOS = i

`j` Numeral, generic greater than or equal to 4 used as a syntactic noun (`čtvero, desatero, ...`, lit. "four-kinds/sorts-of", "ten-...")

Compact prefix	CDJS*
Category	Values used
POS	C
NUMBER	S
CASE	1 2 3 4 5 6 7
VAR	- 1

Category co-occurrence for SUBPOS = j

`k` Numeral, generic greater than or equal to 4 used as a syntactic adjective, short form (`čtvery`, ..., lit. "four-kinds/sorts-of")

Compact prefix	CD2P*
Category	Values used
POS	C
NUMBER	P
CASE	1 2 3 4 5 6 7

Category co-occurrence for SUBPOS = k

l Numeral, cardinal jeden, dva, tři, čtyři, půl, ... (lit. "one", "two", "three", "four"); also sto and tisíc (lit. "hundred", "thousand") if noun declension is not used

Compact prefix	CG***
Category	Values used
POS	C
GENDER	F H I M N X Y Z
NUMBER	D P S
CASE	1 2 3 4 5 6 7
VAR	- 1 2 6 9

Category co-occurrence for SUBPOS = l

m Verb, past transgressive; also archaic present transgressive of perfective verbs (ex.: udělav, lit. "(he-)having-done"; arch. also udělaje (VAR = 4), lit. "(he-)having-done")

Compact prefix	VV***
Category	Values used
POS	V
GENDER	H X Y
NUMBER	P S
NEGATION	A N
VAR	- 4

Category co-occurrence for SUBPOS = m

n Numeral, cardinal greater than or equal to 5

Compact prefix	CB**
Category	Values used
POS	C
NUMBER	P S
CASE	1 2 3 4 5 6 7 X
VAR	- 1

Category co-occurrence for SUBPOS = n

o Numeral, multiplicative indefinite (-krát, lit. (“times”): mnohokrát, tolikrát, . . . , lit. “many times”, “that many times”)

Compact prefix	CIM
Category	Values used
POS	C
VAR	- 1

Category co-occurrence for SUBPOS = o

p Verb, past participle, active (including forms with the enclitic -s, lit. “’re” (“are”))

Compact prefix	VR***[**]
Category	Values used
POS	V
GENDER	F M N Q T Y
NUMBER	P S W
PERSON	2 X
TENSE	R
NEGATION	A N
VOICE	A
VAR	- 1 3 6 8

Category co-occurrence for SUBPOS = p

q Verb, past participle, active, with the enclitic -t̆, lit. (perhaps) “-could-you-imagine-that?” or “but-because-” (both archaic)

Compact prefix	VRE***
Category	Values used
POS	V
GENDER	M N Q T Y
NUMBER	P S W
PERSON	X
TENSE	R
NEGATION	A N
VOICE	A
VAR	2 3

Category co-occurrence for SUBPOS = q

`r` Numeral, ordinal (adjective declension without degrees of comparison)

Compact prefix	`CR***`
Category	Values used
`POS`	`C`
`GENDER`	`F I M N`
`NUMBER`	`D P S`
`CASE`	`1 2 3 4 5 6 7`
`VAR`	`- 6 7`

Category co-occurrence for `SUBPOS` = `r`

`s` Verb, past participle, passive (including forms with the enclitic `-s`, lit. "'re" ("are"))

Compact prefix	`VS***[**]`
Category	Values used
`POS`	`V`
`GENDER`	`F M N Q T Y`
`NUMBER`	`P S W`
`CASE`	`- 4`
`PERSON`	`2 X`
`TENSE`	`H X`
`NEGATION`	`A N`
`VOICE`	`P`
`VAR`	`- 2`

Category co-occurrence for `SUBPOS` = `s`

`t` Verb, present or future tense, with the enclitic `-t̆`, lit. (perhaps) "-could-you-imagine-that?" or "but-because-" (both archaic)

Compact prefix	`V*****`
Category	Values used
`POS`	`V`
`NUMBER`	`P S`
`PERSON`	`1 2 3`
`TENSE`	`F P`
`NEGATION`	`A N`
`VOICE`	`A`
`VAR`	`2 3 9`

Category co-occurrence for `SUBPOS` = `t` u Numeral, interrogative kolikrát, lit. "how many times?"

Compact prefix	CQM
Category	**Values used**
POS	C

Category co-occurrence for SUBPOS = u

v Numeral, multiplicative, definite (-krát, lit. "times": pětkrát, ..., lit. "five times")

Compact prefix	CM
Category	**Values used**
POS	C
VAR	- 1

Category co-occurrence for SUBPOS = v

w Numeral, indefinite, adjectival declension (nejeden, tolikátý, ..., lit. "not-only-one", "so-many-times-repeated")

Compact prefix	CIF***
Category	**Values used**
POS	C
GENDER NUMBER CASE	F I M N X Y Z D P S 1 2 3 4 5 6 7

Category co-occurrence for SUBPOS = w

x Abbreviation, part of speech unknown/indeterminable (now obsolete)

Compact prefix	XABBR
Category	**Values used**
POS	X

Category co-occurrence for SUBPOS = x

y Numeral, fraction ending at `-ina` (`POS`: `C`, numeral); used as a noun (`pětina`, lit. "one-fifth")

Compact prefix	`CFF**`
Category	Values used
`POS`	`C`
`GENDER`	`F`
`NUMBER`	`P S`
`CASE`	`1 2 3 4 5 6 7`
`VAR`	`- 6`

Category co-occurrence for `SUBPOS` = `y`

z Numeral, interrogative `kolikátý`, lit. "what" ("at-what-position-place-in-a-sequence")

Compact prefix	`CQF***`
Category	Values used
`POS`	`C`
`GENDER`	`F I M N X Y Z`
`NUMBER`	`D P S`
`CASE`	`1 2 3 4 6 7`

Category co-occurrence for `SUBPOS` = `z`

} Numeral, written using Roman numerals (`XIV`)

Compact prefix	`CX`
Category	Values used
`POS`	`C`
`VAR`	`- 1 2`

Category co-occurrence for `SUBPOS` = }

~ Abbreviation used as a verb (now obsolete)

Compact prefix	`VABBR`
Category	Values used
`POS`	`V`

Category co-occurrence for `SUBPOS` = ~

2.3.6 Gender (GENDER)

Grammatical gender is being described here (both the *lexical* gender of nouns, as well as the *agreement* gender of verbs, adjectives, pronouns and numerals).

Czech grammatical gender is considered to have four different values: masculine animate, masculine inanimate, feminine, and neuter (M, I, F and N, respectively, which are called the basic gender tags).

This category is heavily using tags which are not fully disambiguated when looking from the traditional (and sometimes an application) point of view. For example, with certain classes of verbs forms (such as past participle, both active and passive) the gender is denoted as Y (masculine), whereas in most other cases the animateness of masculine forms is determined, too, using I for inanimate masculine gender and M for masculine animate gender.

Value count: 11

- Not applicable

F Feminine

Used only with:

POS	SUBPOS
A	A C G M O U
C	d h l r w y z
N	N
P	1 4 5 8 9 D J L P S W Z
V	p s

H Feminine or Neuter

Used only with:

POS	SUBPOS
C	l
P	8 S
V	e m

`I` Masculine inanimate

Used only with:

POS	SUBPOS
A	A G M O U
C	d h l r w z
N	N
P	1 4 8 D J L P S W Z

`M` Masculine animate

Used only with:

POS	SUBPOS
A	A C G M O U
C	d h l r w z
N	N
P	1 4 8 D J K L P S W Z
V	p q s

`N` Neuter

Used only with:

POS	SUBPOS
A	A C G M O U
C	d h l r w z
N	N
P	1 4 5 8 9 D J L P S W Z
V	p q s

`Q` Feminine (with singular only) or Neuter (with plural only); used only with participles and nominal forms of adjectives

Used only with:

POS	SUBPOS
A	C
V	p q s

T Masculine inanimate or Feminine (plural only); used only with participles and nominal forms of adjectives

Used only with:

POS	SUBPOS
A	C
V	p q s

X Any of the basic four genders

Used only with:

POS	SUBPOS
A	A U
C	d h l w z
N	N
P	1 4 5 8 9 D J L P S W Z
V	e m

Y Masculine (either animate or inanimate)

Used only with:

POS	SUBPOS
A	C O
C	d h l w z
P	4 8 D J L P S W Z
V	e m p q s

Z Not fenimine (i.e., Masculine animate/inanimate or Neuter); only for (some) pronoun forms and certain numerals

Used only with:

POS	SUBPOS
C	l w z
P	1 4 5 8 9 D H J L P S W Z

2.3.7 Number (NUMBER)

The number category mostly takes on only one of the two standard values: S for singular or P for plural. Nevertheless, Czech still uses so-called *dual* number for several nouns used for describing symmetrical body parts: eyes, arms, legs, ears (but not, for example, knees, at least not in standard Czech). Due to strong agreement rules, this distinction applies then also to adjectives and pronouns. However, as it appears only in the instrumental case (see below, CASE value 7), it is not necessary to make this distinction for verb number (as verb/nominal agreement concerns only the nominative case).

Also, as in all the other cases (nominative through locative, see below the description of the CASE category) this distinction does not appear on the surface, we also do not make this distinction there. Moreover, all the nouns displaying the dual forms are in feminine, thus the D value is actually used only in tags having also the feminine value of the gender category (this is true regardless of the POS category).

The special value W is used in connection with the GENDER value Q, in order to distinguish the combination of (feminine, singular) and (neuter, plural) from the simple cases.

The value X is used only with undeclinable nouns and adjectives, and with pronouns with systematically ambiguous number. It is also used for abbreviations, foreign words, etc., where the grammatical number distinction is systematically hidden.

Value count: 6

- Not applicable

D Dual

Used only with:

POS	SUBPOS
A	A G M U
C	d h l r w z
N	N
P	1 4 8 D J L S W Z

`P` Plural

Used only with:

POS	SUBPOS
A	A C G M O U
C	d h k l n r w y z
J	,
N	N
P	1 4 5 8 9 D J L P S W Z
V	B c e i m p q s t

`S` Singular

Used only with:

POS	SUBPOS
A	A C G M O U
C	d j l n r w y z
J	,
N	N
P	1 4 5 7 8 9 D H J L P S W Z
V	B c e i m p q s t

`W` Singular for feminine gender, plural with neuter; can only appear in participle or nominal adjective form with gender value `Q`

Used only with:

POS	SUBPOS
A	C
V	p q s

`X` Any

Used only with:

POS	SUBPOS
A	A U
J	,
N	N
P	1 4 6 7 S
V	c

2.3.8 Case (`CASE`)

Case is one the most interesting (and difficult to learn) categories in Czech (as well as in other highly inflective languages). Czech traditionally distinguishes among seven cases[15]: nominative, genitive, dative, accusative, vocative, locative and instrumental. Also traditionally, cases are numbered (in the order presented; see also below).

Once a case is relevant for some detailed part of speech category (more precisely, for some `SUBPOS`), it can take any of the seven values (with negligible – and questionable – exceptions for vocative, `5`, for personal pronouns). Virtually all nouns, adjectives, pronouns, and numerals express case, and for agreement purposes, we use the case category also for all prepositions. Having decided that verb valency (subcategorization) is not part of morphological processing (it is bound to be part of lexical semantics, and certainly is not related in any way to the analysis of word forms), there is no case agreement information linked to verbs. However, there is one case where `CASE` is present at a verb form: in passive participle feminine forms, an accusative ending might be attached if the participle is used as an nominal adjective form (typically in a *verbal attribute* syntactic function).

Value count: 9

`-` Not applicable

`1` Nominative

Used only with:

POS	SUBPOS
A	A G M U
C	? a d h j k l n r w y z
N	N
P	1 4 8 D E J K L P Q S W Z
R	R V

[15] Slovak has also seven, although there is no distinction between vocative and nominative; from the same reason, Russian drops vocative altogether, having only six cases; Polish has six, too. For comparison, English and French have only one common case ("nominative" for nominal parts of speech, even though for some pronouns, accusative case can be identified ("him"), German has four (nominative, genitive, dative, accusative)

2 Genitive

Used only with:

POS	SUBPOS
A	A G M U
C	? a d h j k l n r w y z
N	N
P	1 4 5 6 8 9 D E H J K L P Q S W Z
R	R V

3 Dative

Used only with:

POS	SUBPOS
A	A G M U
C	? a d h j k l n r w y z
N	N
P	1 4 5 6 7 8 9 D E H J K L P Q S W Z
R	R V

4 Accusative

Used only with:

POS	SUBPOS
A	A C G M U
C	? a d h j k l n r w y z
N	N
P	1 4 5 6 7 8 9 D E H J K L P Q S W Z
R	R V
V	s

5 Vocative

Used only with:

POS	SUBPOS
A	A G M U
C	a d h j k l n r w y
N	N
P	8 L P S W Z

6 Locative

Used only with:

POS	SUBPOS
A	A G M U
C	? a d h j k l n r w y z
N	N
P	1 4 5 6 8 9 D E J K L P Q S W Z
R	R V

7 Instrumental

Used only with:

POS	SUBPOS
A	A G M U
C	? a d h j k l n r w y z
N	N
P	1 4 5 6 8 9 D E J K L P Q S W Z
R	R V

X Any

Used only with:

POS	SUBPOS
A	A U
C	a n
N	N
P	1 4 S
R	R

2.3.9 Possessor's gender (`POSSGENDER`)

Possessive adjectives and possessive pronouns refer to a possessor, and an possessive agreement rule applies for number and gender of the possessor. For example, in the phrase `Petr opravil Robertovi jeho auto` (lit. "Peter repaired (for) Robert his (=Robert's) car") the possessive pronoun `jeho` (lit. "his") refers to Robert and therefore there is a (possessive) agreement in number and gender between Robert (singular, masculine animate) and the possessive pronoun `jeho`. Of course, there is also the regular (morphological) agreement between the form `jeho` and `auto` (lit. "car"), namely, in gender (neuter), number (singular) and case (accusative). Had Peter repaired the car for Susan instead of Robert, the form of the possessive pronoun would have to be `její` (lit. "her"), with the possessive agreement gender changed to feminine.

Even though it could be argued (similarly to the categories `POS` and `SUBPOS`) that this category is more lexically than inflectionally based, it has been decided to treat it as all the other morphological categories.

For possessive adjectives, given the nature of possible possessors, it can separately take on only the values of masculine animate and feminine.

For possessive pronouns, both masculine genders and the neuter gender are always homonymous. Thus (cf. also the method used for the `GENDER` category) we do not use any of the three basic values (`M`, `I`, `N`) – instead, the value of `Z` is used. If a possessive pronoun usage relates to the subject of the sentence (forms of `svůj` (lit. also "his") rather than `jeho`), then there is no gender distinction at all, and therefore the value of `X` is used.

Value count: 5

`-` Not applicable

`F` Feminine possessor

Used only with:

POS	SUBPOS
A	U
P	1 S

`M` Masculine animate possessor (adjectives only)

Used only with:

POS	SUBPOS
A	U

`X` Any gender

Used only with:

POS	SUBPOS
P	1 S

`Z` Not feminine (both masculine or neuter)

Used only with:

POS	SUBPOS
P	1 S

2.3.10 Possessor's number (`POSSNUMBER`)

This category is closely related to the `POSSGENDER` category (cf. above). The same reasons and rules apply here.

This category is used for both possessive adjectives as well as possessive pronouns. For possessive adjectives, however, there is no plural possessive - the possessor, lexically present in the form, is always considered to be in singular. For example, in the phrases `Janiny zahrady` (lit. "Jane's gardens") or `kocourovo oko` (lit. "(he-)cat's eye"), the possessive number is always singular, as there is alwasy one Jane and one cat – in order to express the possessive relation between a group of possessors and some object, one has to use a different syntactic construction as Czech morphology has no means to do so directly.

Value count: 3

`-` Not applicable

`P` Plural (possessor)

Used only with:

POS	SUBPOS
P	1 S

`S` Singular (possessor)

Used only with:

POS	SUBPOS
P	1 S

2.3.11 Person (PERSON)

This category bears no surprises. It expresses the person of verb forms (if applicable), and person of personal pronouns. Given the basic principles of the morphological analyzer (see Section 2.1.2), there is no attempt to assign person to participles, transgressives, etc., although traditionally one would do so[16]. Thus person is assigned only to present (and future) verb forms, and to personal pronouns[17].

Value count: 5

- Not applicable

1 1^{st} person

Used only with:

POS	SUBPOS
J	,
P	H P S
V	B c i t

2 2^{nd} person

Used only with:

POS	SUBPOS
J	,
P	7 H K P S
V	B c i p s t

3 3^{rd} person

Used only with:

POS	SUBPOS
J	,
P	1 5 H P S
V	B c i t

[16]It has been done, for example, in the manually tagged corpus of contemporary Czech developed in the 70's in the Institute of the Czech Language, see (Těšitelová, 1985; Těšitelová, 1992; Těšitelová, Petr and Králík, 1985; Těšitelová, Petr and Králík, 1986)

[17]Personal pronouns, however, have separate lemmas for the three persons: `já`, `ty`, `on`. This redundancy is however seen as not-so-disturbing as we get clear agreement relations between pronouns used as clause subjects and their syntactically related verb forms

`X` Any person

Used only with:

POS	SUBPOS
V	p q s

2.3.12 Tense (TENSE)

Tense belongs to verb forms only. Contrary to traditional tense category assignment (and for the same reasons as discussed earlier in the section on Person), tense is meant here in the purely morphological sense, without regard to the actual tense of an analytical verb form used within a particular sentence. Thus e.g. the verb form `pracoval` (lit. "(he) worked") is assigned only the past tense value, even though it can appear as a part of present conditional (`pracoval by` (lit. "(he) would work") as well as real past tense (or "perfective" tense, as Czech does not really distinguish these two): `pracoval jsem` (lit. "(I) (have) worked").

In Czech, future tense is traditionally assigned also to present perfective verb forms. However, this has nothing to do with morphology; morphologically, these forms are simply present tense. On the contrary, there are real future tense forms (but only for a handful of verbs), such as `pojedu` (lit. "(I) will go"), which is a future tense form of the verb `jet` (lit. "to go") (vs. `jedu` (lit. "(I) go")), and of course `budu`, `budeš`, ... (lit. "(I) will", "(you) will",...) – future forms of the verb `být` (lit. "to be"), used both in the existential sense as well as an auxiliary verb form for analytically expressed future tense of imperfective verbs.

Value count: 6

`-` Not applicable

`F` Future

Used only with:

POS	SUBPOS
V	B t

`H` Past or Present

Used only with:

POS	SUBPOS
V	s

`P` Present

Used only with:

POS	SUBPOS
P	P
V	B t

`R` Past

Used only with:

POS	SUBPOS
V	p q

`X` Any (Past, Present, or Future)

Used only with:

POS	SUBPOS
V	s

2.3.13 Degrees of Comparison (GRADE)

Degrees-of-Comparison subtag is used for adjectives and adverbs and it is considered to be an inflectional category (as opposed to a derivational category). For certain types of adjectives, such as possessive adjectives, this category is considered irrelevant (the value of *Not applicable* (-) is used). On the other hand, the value *positive* (1) is used for adjectives which do not form comparatives or superlatives only on semantic grounds (such as the often discussed adjective optimální (lit. "optimal")). In most cases, however, the morphological analysis would not reject even the comparative or superlative forms of such adjectives, marking their degree of comparison correctly and consistently according to their prefix and suffix.

Czech can express degrees of comparison by using analytical forms (similar to English "more + positive adjective form"). However, these are not treated by the morphological analyzer (based on the principles stated in Sect. 2.1.2, and similarly to any other analytical form (non-)treatment).

Traditionally, numbers are being used for the degrees of comparison in Czech morphology.

Value count: 4

- Not applicable

1 Positive

Used only with:

POS	SUBPOS
A	A
D	g

2 Comparative

Used only with:

POS	SUBPOS
A	A
D	g

3 Superlative

Used only with:

POS	SUBPOS
A	A
D	g

2.3.14 Negation (NEGATION)

Similarly to the degree of comparison category, this category is treated as fully inflectional category, as negation is expressed in Czech by a prefix, as opposed to a special particle. Negation can be attached to verbs, adjectives, some pronouns and numerals, and in principle, also to nouns (although regularly, only verbal nouns are typically used with negation).

Value count: 3

- `-` Not applicable
- `A` Affirmative (not negated)

  Used only with:

POS	SUBPOS
A	2 A C G M
D	g
N	N
P	P
V	B e f i m p q s t

- `N` Negated

  Used only with:

POS	SUBPOS
A	2 A C G M
D	g
N	N
V	B e f i m p q s t

2.3.15 Voice (Active/Passive) (`VOICE`)

This category is used for verb participles only. It is *not* used for verbal adjectives, even if they are derived from passive participle forms.

Value count: 3

`-` Not applicable

`A` Active

Used only with:

POS	SUBPOS
P	P
V	B p q t

`P` Passive

Used only with:

POS	SUBPOS
V	s

2.3.16 `RESERVE1`

This (non-)category is intended for various technical uses, and also for other possible future uses, such as the aspect category, which is being neglected in the present system (it is not considered an inflectional category, but it could be extracted from the verb lemma itself, as it is lexically based).

Value count: 1

`-` Not applicable

2.3.17 `RESERVE2`

Unused.

Currently unused; same comments apply as for the `RESERVE1` category above.

Value count: 1

`-` Not applicable

2.3.18 Variant, register, style, colloquiality, etc. (`VAR`)

This is a "waste basket" type of category – all features of inflected forms which do not belong to the categories presented so far have been assigned one or more values of this last category.

Variant information is used whenever some combination of categories can be expressed by more than one word form. For example, Czech masculine inanimate nouns can typically have one or two variants in the singular genitive, singular locative, plural nominative, and plural locative cases (the presence or absence and the number of variants differs based on their paradigm, and often is just lexically based). These variants are, however, not quite equivalent in usage; typically, one would prefer one to the other in standard writing (although the other variants are *not* colloquial – see below for those variants).

Another important distinction is a style of the form. Often, bookish and archaic endings are used and the morphology must be able to handle them properly, i.e. to distinguish them from their contemporary counterparts.

Last but not least, colloquial style, which is now almost standard in spoken Czech, must be handled as well. Many standard noun, adjectival, pronominal, numeral as well as some verbal endings have their colloquial counterpart[18].

Lexically based colloquial style is not handled by the morphological processor itself. The information on the lexical colloquiality (e.g. `bejt`, lit. "to be" (coll.), or `vokno`, lit. "window" (coll.)) is typically present in the dictionary and passed to the output in the form of "lexical style information" (cf. 4.7).

Value count: 10

- `-` Not applicable (basic variant, standard contemporary style; also used for standard forms allowed for use in writing by the Czech Standard

[18] Colloquial Czech is used thorough the geographical area of the Czech republic in an extent not found in other languages, so that in fact almost every Czech knows two languages: the *standard* Czech, which is supposed to be used in writing and official oral use, and *colloquial* (called also *general*) Czech, which is being spoken at home, with friends, and in any not-so-strictly formal occasion. This phenomenon, called diglosia, would be a theme for another research. See also (Sgall Hronek)

Orthography Rules despite being marked there as colloquial)

1 Variant, second most used ("less frequent"), still standard

Currently used with:

POS	SUBPOS
A	2 A O U
C	a d j l n o v }
D	b g
J	, ^
N	N
P	5 8 9 D J L P S Z
R	R V
T	T
V	B f i p
X	@

2 Variant, rarely used, bookish, or archaic

Currently used with:

POS	SUBPOS
C	d l }
D	b g
J	^
N	N
P	1 9 D J K P W Z
V	B e f i q s t

3 Very archaic, also archaic + colloquial

Currently used with:

POS	SUBPOS
D	g
N	N
P	4 9 J
R	R
V	B f i p q t

4 Very archaic or bookish, but standard at the time

Currently used with:

POS	SUBPOS
D	b
N	N
P	Z
V	B f i m

5 Colloquial, but (almost) tolerated even in public

Currently used with:

POS	SUBPOS
N	N
P	D
V	B

6 Colloquial (standard in spoken Czech)

Currently used with:

POS	SUBPOS
A	A G M O U
C	d l r y
D	b g
N	N
P	4 8 D L P S W Z
V	B c f i p

7 Colloquial (standard in spoken Czech), less frequent variant

Currently used with:

POS	SUBPOS
A	A G M U
C	d r
D	b
N	N
P	4 8 S W Z
V	B i

8 Abbreviations

Currently used with:

POS	SUBPOS
A	A U
D	b g
J	^
N	N
P	D S
R	R
V	B i p
X	X

9 Special uses, e.g. personal pronouns after prepositions etc.

Currently used with:

POS	SUBPOS
A	A
C	l
D	b
N	N
P	Q
V	i t

2.3.19 The Tag Correspondences

The morphological dictionary is still maintained in the compact tag system, but the actual files which are used for morphological processing as described in the Chapter 5 contain *both* the (now preferred) positional *and* the compact tags. Therefore it can be determined dynamically which tag system to use for the output of the morphological processing[19].

For quick lookup, see the above `SUBPOS` co-occurrence tables, where the compact tag "prefix" is displayed in the table heading.

[19] An external tag table is being used in the original version of the software. It is available as part of the morphological analyzer software; see Sect. 7 for references.