

# CZECH POSITIONAL MORPHOLOGICAL TAGS

## 1. PART OF SPEECH

- A** Adjectives
- C** Numerals
- D** Adverbs
- I** Interjection
- J** Conjunction
- N** Noun
- P** Pronoun
- V** Verb
- R** Preposition
- T** Particle
- X** Unknown, Not Determined, Unclassifiable
- Z** Punctuation (also used for the Sentence Boundary Token)

## 2. SUB PART OF SPEECH

- #** Sentence boundary
- %** Author's signature, e.g. haš-99\_:B\_;S
- \*** Word krát (lit.: "times")
- ,** Conjunction subordinate (incl. "aby", "kdyby" in all forms)
- }** Numeral, written using Roman numerals (XIV)
- :** Punctuation (except for the virtual sentence boundary word ###, which uses "Sub part of speech" #)
- =** Number written using digits
- ?** Numeral "kolik" (lit. "how many"/"how much")
- @** Unrecognized word form
- ^** Conjunction (connecting main clauses, not subordinate)
- 4** Relative/interrogative pronoun with adjectival declension of both types (soft and hard) ("jaký", "který", "čí", ..., lit. "what", "which", "whose", ...)
- 5** The pronoun he in forms requested after any preposition (with prefix n-: "něj", "něho", ..., lit. "him" in various cases)
- 6** Reflexive pronoun "se" in long forms ("sebe", "sobě", "sebou", lit.

"myself" / "yourself" / "herself" / "himself" in various cases; "se" is personless)

- 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) This should be done somehow more consistently, virtually any word can have this contracted -s ("cos", "polívkus", ...)
- 8** Possessive reflexive pronoun "svůj" (lit. "my"/"your"/"her"/"his" when the possessor is the subject of the sentence)
- 9** Relative pronoun "jenž", "již", ... after a preposition (n-: "něhož", "niž", ..., lit. "who")
- A** Adjective, general
- B** Verb, present or future form
- C** Adjective, nominal (short, participial) form "řád", "schopen", ...
- D** Pronoun, demonstrative ("ten", "onen", ..., lit. "this", "that", "that", ... "over there", ...)
- E** Relative pronoun "což" (corresponding to English which in subordinate clauses referring to a part of the preceding text)
- F** Preposition, part of; never appears isolated, always in a phrase ("nehledě (na)", "vzhledem (k)", ..., lit. "regardless", "because of")
- G** Adjective derived from present transgressive form of a verb
- 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
- I** Interjections
- J** Relative pronoun "jenž", "již", ... not after a preposition (lit. "who", "whom")
- K** Relative/interrogative pronoun "kdo" (lit. "who"), incl. forms with affixes -ž and -s (affixes are distinguished by the category "Variant" (for -ž) and "Person" (for -s))
- L** Pronoun, indefinite "všechn", "sám" (lit. "all", "alone")
- M** Adjective derived from verbal past transgressive form
- N** Noun (general)
- O** Pronoun "svůj", "nesvůj", "tentam" alone (lit. "own self", "not-in-mood", "gone")
- 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
- Q** Pronoun relative/interrogative "co", "copak", "cožpak" (lit. "what",

"isn't-it-true-that")

- R** Preposition (general, without vocalization)
- 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
- T** Particle
- U** Adjective possessive (with the masculine ending -ův as well as feminine -in)
- V** Preposition (with vocalization -e or -u): ("ve", "pode", "ku", ..., lit. "in", "under", "to")
- W** Pronoun negative ("nic", "nikdo", "nijaký", "žádný", ..., lit. "nothing", "nobody", "not-worth-mentioning", "no"/"none")
- X** (temporary) Word form recognized, but tag is missing in dictionary due to delays in (asynchronous) dictionary creation
- Y** Pronoun relative/interrogative co as an enclitic (after a preposition) ("oč", "nač", "zač", lit. "about what", "on"/"onto" "what", "after"/"for what")
- Z** Pronoun indefinite ("nějaký", "některý", "číkoli", "cosi", ..., lit. "some", "some", "anybody's", "something")
- a** Numeral, indefinite ("mnoho", "málo", "tolik", "několik", "kdovíkolik", ..., lit. "much"/"many", "little"/"few", "that much"/"many", "some" ("number of"), "who-knows-how-much/many")
- 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)
- c** Conditional (of the verb "být" (lit. "to be") only) ("by", "bych", "bys", "bychom", "byste", lit. "would")
- d** Numeral, generic with adjectival declension ("dvoji", "desaterý", ..., lit. "two-kinds"/..., "ten-...")
- e** Verb, transgressive present (endings -e/-ě, -íc, -íce)
- f** Verb, infinitive
- g** Adverb (forming negation ("Negation" set to A/N) and degrees of comparison "Grade" set to 1/2/3 (comparative/superlative), e.g. "velký", "za-jí-l-mal-vý", ..., lit. "big", "interesting")
- h** Numeral, generic: only "jedny" and "nejedny" (lit. "one-kind"/"sort-of", "not-only-one-kind"/"sort-of")
- i** Verb, imperative form
- j** Numeral, generic greater than or equal to 4 used as a syntactic noun ("čtyvero", "desatero", ..., lit. "four-kinds"/"sorts-of", "ten-...")

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

**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

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

**n** Numeral, cardinal greater than or equal to 5

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

**p** Verb, past participle, active (including forms with the enclitic -s, lit. 're ("are"))

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

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

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

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

**u** Numeral, interrogative "kolikrát", lit. "how many times?"

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

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

**y** Numeral, fraction ending at -ina; used as a noun ("pětina", lit. "one-fifth")

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

### 3. GENDER

**F** Feminine

**H** {F, N} – Feminine or Neuter

**I** Masculine inanimate

**M** Masculine animate

**N** Neuter

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

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

**X** Any

**Y** {M, I} – Masculine (either animate or inanimate)

**Z** {M, I, N} – Not feminine (i.e., Masculine animate/inanimate or Neuter); only for (some) pronoun forms and certain numerals

### 4. NUMBER

**D** Dual, e.g. "nohama"

**P** Plural, e.g. "nohami"

**S** Singular, e.g. "noha"

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

**X** Any

### 5. CASE

**1** Nominative, e.g. "žena"

**2** Genitive, e.g. "ženy"

**3** Dative, e.g. "ženě"

**4** Accusative, e.g. "ženu"

**5** Vocative, e.g. "ženo"

**6** Locative, e.g. "ženě"

**7** Instrumental, e.g. "ženou"

**X** Any

### 6. POSSESSIVE GENDER

**F** Feminine, e.g. "matčin", "její"

**M** Masculine animate (adjectives only), e.g. "otců"

**X** Any

**Z** {M, I, N} – Not feminine, e.g. "jeho"

### 7. POSSESSIVE NUMBER

**P** Plural, e.g. "náš"

**S** Singular, e.g. "můj"

**X** Any, e.g. "your"

### 8. PERSON

**1** 1st person, e.g. "píšu", "píšeme"

**2** 2nd person, e.g. "píšeš", "píšete"

**3** 3rd person, e.g. "píše", "píšou"

**X** Any person

### 9. TENSE

**F** Future

**H** {R, P} – Past or Present

**P** Present

**R** Past

**X** Any

### 10. GRADE

**1** Positive, e.g. "velký"

**2** Comparative, e.g. "větší"

**3** Superlative, e.g. "největší"

### 11. NEGATION

**A** Affirmative (not negated), e.g. "možný"

**N** Negated, e.g. "nemožný"

### 12. VOICE

**A** Active, e.g. "píšící"

**P** Passive, e.g. "psaný"

### 13., 14. RESERVE 1, RESERVE 2

- Not applicable

### 15. VARIANT

- Basic variant, standard contemporary style; also used for standard forms allowed for use in writing by the Czech Standard Orthography Rules despite being marked there as colloquial

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

**2** Variant, rarely used, bookish, or archaic

**3** Very archaic, also archaic + colloquial

**4** Very archaic or bookish, but standard at the time

**5** Colloquial, but (almost) tolerated even in public

**6** Colloquial (standard in spoken Czech)

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

**8** Abbreviations

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