

11 Developmental corpus Šolar

The Šolar annotation system, developed alongside the Slovene Šolar developmental corpus (Arhar Holdt et al. 2022), is designed for categorizing language corrections in texts written by pupils in Slovene primary schools and students in Slovene secondary schools. The system's initial framework for annotating corrections was established in the corpus's first edition (Kosem et al. 2012), and significantly enhanced in its 2.0 version (Kosem et al. 2016), which also saw the initial development of its annotation guidelines (Arhar Holdt et al. 2018). The system is structured hierarchically into three levels: it starts by identifying corrections at the linguistic level, then classifies the general type of correction, and finally pinpoints the specific linguistic issue. This three-tiered tagging approach ensures both robust and nuanced application across various contexts.

- [Introduction to Tags](#)
- [Annotation Guidelines](#)
- [References and Links](#)

Introduction to Tags

This chapter summarises the Šolar tags. A more detailed presentation can be found in the guidelines in the Annotation Guidelines chapter.

Tag	Linguistic level	Category of correction	Specific language problem
Č/VOK/odveč	Spelling	Vowels	Superfluous vowel
Č/VOK/izpust	Spelling	Vowels	Omitted vowel
Č/VOK/menjava-ao	Spelling	Vowels	AO substitution
Č/VOK/menjava-ei	Spelling	Vowels	EI substitution
Č/VOK/menjava-uo	Spelling	Vowels	UO substitution
Č/VOK/menjava-drugo	Spelling	Vowels	Other vowel substitutions
Č/KONZ/odveč	Spelling	Consonants	Superfluous consonant
Č/KONZ/izpust	Spelling	Consonants	Omitted consonant
Č/KONZ/menjava-sz	Spelling	Consonants	Substitution SZ
Č/KONZ/menjava-td	Spelling	Consonants	Substitution TD
Č/KONZ/menjava-kgh	Spelling	Consonants	Substitution KGH
Č/KONZ/menjava-mn	Spelling	Consonants	Substitution MN
Č/KONZ/menjava-šž	Spelling	Consonants	Substitution ŠŽ
Č/KONZ/menjava-strešice	Spelling	Consonants	Substitution of DIACRITIC
Č/KONZ/menjava-drugo	Spelling	Consonants	Other consonant substitution
Č/W/začetek	Spelling	Labio-velar approximant w	Word-initially
Č/W/sredina	Spelling	Labio-velar approximant w	Word-medial
Č/W/konec	Spelling	Labio-velar approximant w	Word-final
Č/W/v	Spelling	Labio-velar approximant w	Prepositional V
Č/SKLOP/zlog	Spelling	Letter clusters	A syllable is missing or is superfluous
Č/SKLOP/lj	Spelling	Letter clusters	Cluster LJ
Č/SKLOP/nj	Spelling	Letter clusters	Cluster NJ
Č/SKLOP/ij	Spelling	Letter clusters	Cluster IJ
Č/SKLOP/podvojene	Spelling	Letter clusters	Doubled letters

Tag	Linguistic level	Category of correction	Specific language problem
Č/SKLOP/premet	Spelling	Letter clusters	Metathesis
Č/PRED/sz	Spelling	Variable (allophonic) prepositions	Preposition s/z
Č/PRED/kh	Spelling	Variable (allophonic) prepositions	Preposition k/h
O/KAT/sklon-rt	Morphology	Categorical corrections	Case: genitive-accusative
O/KAT/sklon-dm	Morphology	Categorical corrections	Case: dative-locative
O/KAT/sklon-mo	Morphology	Categorical corrections	Case: locative-instrumental
O/KAT/sklon-drugo	Morphology	Categorical corrections	Other case substitutions
O/KAT/število-em	Morphology	Categorical corrections	Number: singular-plural
O/KAT/število-dm	Morphology	Categorical corrections	Number: dual-plural
O/KAT/število-ed	Morphology	Categorical corrections	Number: single-dual
O/KAT/spol	Morphology	Categorical corrections	Gender
O/KAT/vid	Morphology	Categorical corrections	Aspect
O/KAT/čas	Morphology	Categorical corrections	Tense
O/KAT/oseba	Morphology	Categorical corrections	Person
O/KAT/nedoločnik-kratki	Morphology	Categorical corrections	Reduced infinitive
O/KAT/nedoločnik-namenilnik	Morphology	Categorical corrections	Infinitive and supine
O/KAT/nedoločnik-osebna	Morphology	Categorical corrections	Infinitive and finite verb
O/KAT/povratnost	Morphology	Categorical corrections	Reflexivity
O/KAT/naklon	Morphology	Categorical corrections	Mood
O/KAT/način	Morphology	Categorical corrections	Voice
O/KAT/oblika-zaimka	Morphology	Categorical corrections	Pronomial form
O/KAT/določnost	Morphology	Categorical corrections	Definiteness
O/KAT/stopnjevanje	Morphology	Categorical corrections	Comparison
O/PAR/glagolska-osnova	Morphology	Paradigmatic Corrections	Verbal root
O/PAR/glagolska-končnica	Morphology	Paradigmatic Corrections	Verbal ending
O/PAR/neglagolska-osnova	Morphology	Paradigmatic Corrections	Non-verbal root
O/PAR/neglagolska-končnica	Morphology	Paradigmatic Corrections	Non-verbal ending
O/PAR/neobstojni-vokal	Morphology	Paradigmatic Corrections	Epenthetic vowel
O/PAR/preglas-in-cč	Morphology	Paradigmatic Corrections	Umlaut and cč

Tag	Linguistic level	Category of correction	Specific language problem
O/DOD/variante	Morphology	Additional Annotation	Morphological variants
O/DOD/besede-mati-hči	Morphology	Additional Annotation	Mati-hči
O/DOD/besede-otrok	Morphology	Additional Annotation	Otrok
B/SAM/napačno-lastno	Vocabulary	Nouns	Erroneous proper noun
B/SAM/lastno-občno	Vocabulary	Nouns	Proper and common name
B/SAM/občno-besedišče	Vocabulary	Nouns	Common vocabulary
B/GLAG/predpona	Vocabulary	Verbs	Verbal prefixes
B/GLAG/moči-morati	Vocabulary	Verbs	Substitution moči-morati
B/GLAG/naklonski	Vocabulary	Verbs	Other substitutions of modal verbs
B/GLAG/drugo	Vocabulary	Verbs	Other substitutions of verbs
B/ZAİM/povratna-svojilnost	Vocabulary	Pronoun	Reflexive possessive
B/ZAİM/ki-kateri	Vocabulary	Pronoun	Substitution ki -- kateri
B/ZAİM/oziralni	Vocabulary	Pronoun	Other problems with relative pronouns
B/ZAİM/noben	Vocabulary	Pronoun	Substitution of negative pronouns
B/ZAİM/drugo	Vocabulary	Pronoun	Other pronomial substitutions
B/PRED/glagolske-zveze	Vocabulary	Preposition	Prepositions in verbal phrases
B/PRED/neglagolske-zveze	Vocabulary	Preposition	Prepositions into non-verbal phrases
B/PRED/lokacijske-dvojnice	Vocabulary	Preposition	Locative doublets
B/PRED/drugo	Vocabulary	Preposition	Other substitutions of prepositions
B/VEZ/in-pa-ter	Vocabulary	Conjunction	Substitutions of in-pa-ter
B/VEZ/protivni	Vocabulary	Conjunction	Coordinating adversative conjunction
B/VEZ/sprememba-odnosa	Vocabulary	Conjunction	Change to subordination
B/VEZ/drugo	Vocabulary	Conjunction	Other substitutions of conjunctions
B/PRID/drugo	Vocabulary	Adjective	All problems related to adjectives
B/PRISL/drugo	Vocabulary	Adverb	All problems related to adverbs

Tag	Linguistic level	Category of correction	Specific language problem
B/OST/drugo	Vocabulary	Other parts of speech	All problems related to other parts of speech
B/MEN/polnopomenska-v-zaimek	Vocabulary	Substitutions beyond the confines of part of speech	Lexical word or phrase changed into pronoun
B/MEN/zaimek-v-polnopomensko	Vocabulary	Substitutions beyond the confines of part of speech	Pronoun to a lexical word or phrase
B/MEN/veznik-zaimek	Vocabulary	Substitutions beyond the confines of part of speech	Substitution of conjunction and pronoun
B/MEN/besedna-družina	Vocabulary	Substitutions beyond the confines of part of speech	Word family
B/MEN/samostalnik-bz	Vocabulary	Substitutions beyond the confines of part of speech	Noun and phrase
B/MEN/glagol-bz	Vocabulary	Substitutions beyond the confines of part of speech	Verb and phrase
B/MEN/prislov-pridevnik-bz	Vocabulary	Substitutions beyond the confines of part of speech	Adverb/adjective and phrase
B/MEN/drugo	Vocabulary	Substitutions beyond the confines of part of speech	Other types of substitutions
B/DOD/zaznamovano	Vocabulary	Substitutions beyond the confines of part of speech	Stylistically marked vocabulary
S/BR/povedek-osebek	Syntax	Word order	Constituent order: predicate-subject
S/BR/povedek-predmet	Syntax	Word order	Constituent order: predicate-object
S/BR/povedek-prislovno-določilo	Syntax	Word order	Order: sentence-adverbial determiner
S/BR/členek	Syntax	Word order	Order: particle
S/BR/znotraj-stavčnega-člena	Syntax	Word order	Order within clausal constituents
S/BR/naslonski-niz-znotraj	Syntax	Word order	Clitic string: order of clitics
S/BR/naslonski-niz-prirednost-podrednost	Syntax	Word order	Clitic string: independent-subordinate
S/BR/drugo	Syntax	Word order	Other changes to word order
S/IZPUST/samostalnik-občno-ime	Syntax	Omitted constituents	Noun: common noun
S/IZPUST/samostalnik-lastno-ime	Syntax	Omitted constituents	Noun: proper noun
S/IZPUST/glagol-bit	Syntax	Omitted constituents	The verb bit

Tag	Linguistic level	Category of correction	Specific language problem
S/IZPUST/glagol-drugo	Syntax	Omitted constituents	Other omitted verbs
S/IZPUST/veznik-pa	Syntax	Omitted constituents	The word pa
S/IZPUST/veznik-drugo	Syntax	Omitted constituents	Other omitted conjunctions
S/IZPUST/predlog-ponovljen	Syntax	Omitted constituents	Repeated prepositions
S/IZPUST/predlog-drugo	Syntax	Omitted constituents	Other omitted prepositions
S/IZPUST/zaimek-osebni	Syntax	Omitted constituents	Personal pronoun
S/IZPUST/zaimek-drugo	Syntax	Omitted constituents	Other omitted pronouns
S/IZPUST/pridevnik	Syntax	Omitted constituents	Adjective
S/IZPUST/prislov	Syntax	Omitted constituents	Adverb
S/IZPUST/členek	Syntax	Omitted constituents	Particle
S/IZPUST/stavek	Syntax	Omitted constituents	Sentence
S/ODVEČ/ponavljanje	Syntax	Superfluous constituents	Literal repetition
S/ODVEČ/samostalnik-občno-ime	Syntax	Superfluous constituents	Noun: common noun
S/ODVEČ/samostalnik-lastno-ime	Syntax	Superfluous constituents	Noun: proper noun
S/ODVEČ/glagol-bit	Syntax	Superfluous constituents	The verb biti
S/ODVEČ/glagol-drugo	Syntax	Superfluous constituents	Other superfluous verb
S/ODVEČ/veznik-pa-vezniki	Syntax	Superfluous constituents	The word pa with another conjunction
S/ODVEČ/veznik-pa-drugo	Syntax	Superfluous constituents	Other examples including the word pa
S/ODVEČ/veznik-začetek	Syntax	Superfluous constituents	Conjunction at the beginning of a sentence
S/ODVEČ/veznik-dvojni	Syntax	Superfluous constituents	Doubled conjunction
S/ODVEČ/veznik-drugo	Syntax	Superfluous constituents	Other superfluous conjunction
S/ODVEČ/predlog	Syntax	Superfluous constituents	Preposition
S/ODVEČ/zaimek-osebni	Syntax	Superfluous constituents	Personal pronoun
S/ODVEČ/zaimek-kazalni	Syntax	Superfluous constituents	Demonstrative pronoun
S/ODVEČ/zaimek-svojilni	Syntax	Superfluous constituents	Possessive pronoun
S/ODVEČ/zaimek-drugo	Syntax	Superfluous constituents	Other superfluous pronouns
S/ODVEČ/pridevnik	Syntax	Superfluous constituents	Adjective

Tag	Linguistic level	Category of correction	Specific language problem
S/ODVEČ/prislov-mera	Syntax	Superfluous constituents	Adverb of degree
S/ODVEČ/prislov-drugo	Syntax	Superfluous constituents	Other superfluous adverbs
S/ODVEČ/členek	Syntax	Superfluous constituents	Particle
S/ODVEČ/stavek	Syntax	Superfluous constituents	Clause
S/ODVEČ/poved	Syntax	Superfluous constituents	Sentence
S/STR/svojina-od	Syntax	Structure	Possessives with od
S/STR/svojina-rodilnik	Syntax	Structure	Possessives with the genitive
S/STR/ločilo-veznik	Syntax	Structure	Substitution punctuation-conjunction
S/STR/združevanje-stavkov	Syntax	Structure	Merged clauses
S/STR/deljenje-stavkov	Syntax	Structure	Separation of clauses/sentences
S/STR/besedna-zveza-stavek	Syntax	Structure	Word/phrase instead of clause and vice versa
S/STR/preoblikovanje-stavka	Syntax	Structure	Reworked clause
S/DOD/pleonazem	Syntax	Additional Annotation	Pleonasm
S/DOD/vsebina-drugo	Syntax	Additional Annotation	Superfluous content
S/DOD/vsebina-napake	Syntax	Additional Annotation	Erroneous content
S/DOD/pomensko-prazni	Syntax	Additional Annotation	Semantically null
Z/MV/pridevnik-ski	Orthography	Capital/lowercase letters	Adjectives ending in -ski
Z/MV/pridevnik-drugo	Orthography	Capital/lowercase letters	Other adjectives
Z/MV/občna-imena	Orthography	Capital/lowercase letters	Common noun
Z/MV/osebna-imena	Orthography	Capital/lowercase letters	Personal name with lowercase letter
Z/MV/narodnost	Orthography	Capital/lowercase letters	Nationality with lowercase letter
Z/MV/zemljepisna-imena	Orthography	Capital/lowercase letters	Geographical name with lowercase letter
Z/MV/stvarna-imena	Orthography	Capital/lowercase letters	Proper nouns with lowercase letter
Z/MV/premi-govor	Orthography	Capital/lowercase letters	Direct speech
Z/MV/začetek-povedi	Orthography	Capital/lowercase letters	Sentence initial

Tag	Linguistic level	Category of correction	Specific language problem
Z/MV/hiperkorekcija-ločila	Orthography	Capital/lowercase letters	Hypercorrection following a period
Z/MV/drugo	Orthography	Capital/lowercase letters	Other problems with initial letters
Z/SN/skupaj-glagol	Orthography	Together/separate	Verb together
Z/SN/skupaj-predlog	Orthography	Together/separate	Preposition together
Z/SN/narazen-predlog	Orthography	Together/separate	Preposition separate
Z/SN/skupaj-prislov	Orthography	Together/separate	Adverb together
Z/SN/narazen-prislov	Orthography	Together/separate	Adverb separate
Z/SN/narazen-pridevnik	Orthography	Together/separate	Adjective separate
Z/SN/narazen-drugo	Orthography	Together/separate	Other separate
Z/SN/skupaj-drugo	Orthography	Together/separate	Other together
Z/KR/drugo	Orthography	Abbreviations	All problems related to abbreviations
Z/ŠTEV/drugo	Orthography	Numbers	All problems related to numbers
Z/LOČ/nerazvrščeno	Orthography	Punctuation	Unclassified punctuation corrections
Z/LOČ/vzorec-vejica-stavki	Orthography	Punctuation	Comma before subordinate clauses
Z/LOČ/vzorec-vejica-stavčni-členi	Orthography	Punctuation	Comma between parts-of-speech
Z/LOČ/vzorec-vejica-vezniki	Orthography	Punctuation	Comma and multi-word conjunctions
Z/LOČ/vzorec-vejica-kot	Orthography	Punctuation	Comma and comparative structures
Z/LOČ/vzorec-vejica-pristavki	Orthography	Punctuation	Comma and appositions etc.
Z/LOČ/vzorec-vejica-vrinjen-odvisnik	Orthography	Punctuation	Comma and inserted subordinate clauses
Z/LOČ/vzorec-vejica-priedja-zvez	Orthography	Punctuation	Comma and coordinate phrases
Z/LOČ/vzorec-vejica-priedja-odvisnikov	Orthography	Punctuation	Comma and coordinate clauses
Z/LOČ/vzorec-vejica-pridevniški-niz	Orthography	Punctuation	Comma in adjective strings

Tag	Linguistic level	Category of correction	Specific language problem
Z/LOČ/vzorec-vejica-elipsa-povedka	Orthography	Punctuation	Comma and predicate ellipsis
Z/LOČ/vzorec-vejica-kopičenje-ločil	Orthography	Punctuation	Comma and punctuation accumulation
Z/LOČ/vzorec-vejica-kopičenje-veznikov	Orthography	Punctuation	Comma and conjunction accumulation
Z/LOČ/vzorec-vejica-navajanje	Orthography	Punctuation	Comma and quotation
P/OBL/drugo	Related corrections	Related morphology corrections	All corrections related to morphology
P/SKLA/osebek	Related corrections	Related syntax corrections	Corrections of subject
P/SKLA/drugo	Related corrections	Related syntax corrections	Other corrections related to syntax
P/ZAP/mala-velika	Related corrections	Related orthography corrections	Corrections of initial letter
N//nečitljivo	Illegible and dubious examples		Illegible examples
N//preveri	Illegible and dubious examples		Dubious examples

Annotation Guidelines

This chapter summarizes the annotation guidelines for semantic-role labelling as applied to Slovene texts. The guidelines are arranged from the latest, up-to-date version to the oldest version.

Version 1.2 (22/11/2023)

Project [Empirical foundations for digitally-supported development of writing skills](#)

ARHAR HOLDT, Špela, LAVRIČ, Polona, ROBLEK, Rebeka, GOLI, Teja, BON, Mija, 2023: *Categorizing Teachers' Corrections: Guidelines for Annotating the Šolar Corpus*. Version 1.2. Prepared in the project Empirical foundations for digitally-supported development of writing skills. [\[DOCX\]](#) [\[PDF\]](#)

Version 1.1 (12/8/2022)

Project [Development of Slovene in a Digital Environment](#)

ARHAR HOLDT, Špela, LAVRIČ, Polona, ROBLEK, Rebeka, GOLI, Teja, 2022: *Categorizing Teachers' Corrections: Guidelines for Annotating the Šolar Corpus*. Version 1.1. Prepared in the project Development of Slovene in a Digital Environment. [\[DOCX\]](#) [\[PDF\]](#)

Version 1.0 (16/12/2018)

Project [Upgrade of Šolar Corpus](#)

ARHAR HOLDT, Špela, LAVRIČ, Polona, ROBLEK, Rebeka, GOLI, Teja, 2018: *Kategorizacija učiteljskih popravkov: Smernice za označevanje korpusa Šolar 2.0*. Različica 1.0. Rezultat projekta Nagradnja korpusa Šolar. [\[PDF\]](#) - only in Slovene

References and Links

This chapter compiles relevant references and provides links to projects where the Šolar system has been developed and applied to Slovene texts.

Projects, in which the system has been developed:

[Communication in Slovene](#)

[Upgrade of Šolar corpus](#)

[Development of Slovene in a Digital Environment](#)

[Empirical foundations for digitally-supported development of writing skills](#)

Corpora containing manually revised Šolar tags:

ARHAR HOLDT, Špela, ROZMAN, Tadeja, STRITAR KUČUK, Mojca, KREK, Simon, KRAPŠ VODOPIVEC, Irena, STABEJ, Marko, PORI, Eva, GOLI, Teja, LAVRIČ, Polona, LASKOWSKI, Cyprian Adam, KOCJANČIČ, Polonca, KLEMENC, Bojan, KRSNIK, Luka, KOSEM, Iztok, 2022, *Developmental corpus Šolar 3.0*, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042,

<http://hdl.handle.net/11356/1589>.

The CJVT Svala tool for manual annotation following the Šolar system:

ARHAR HOLDT, Špela, KOSEM, Iztok, STRITAR KUČUK, Mojca, KRSNIK, Luka, JOVAN, Leon Noe, 2022: *CJVT Svala* (Kazalnik projekta Razvoj slovenščine v digitalnem okolju), v1.0,

<https://orodja.cjvt.si/svala/>, Accessed on 2 March 2023.

References:

ARHAR HOLDT, Špela and KOSEM, Iztok. *Šolar, the developmental corpus of Slovene*. 24 August 2023, PREPRINT (Version 1) available at Research Square. <https://doi.org/10.21203/rs.3.rs-3274669/v1>

ARHAR HOLDT, Špela, KOSEM, Iztok, STRITAR KUČUK, Mojca. *Metode in orodja za lažjo pripravo korpusov usvajanja jezika*. PIRIH SVETINA, Nataša (ur.), FERBEŽAR, Ina (ur.). Na stičišču svetov: slovenščina kot drugi in tuji jezik. Ljubljana: Založba Univerze, 2022. Str. 23-30, Zbirka Obdobja, 41. https://centerslo.si/wp-content/uploads/2022/11/Arhar-Holdt-et-al_Obdobja-41.pdf.

ARHAR HOLDT, Špela, KOSEM, Iztok, GANTAR, Polona, 2017: *Corpus-based resources for L1 teaching: the case of Slovene*. Ann Marcus-Quinn, Triona Hourigan (ur.): Handbook on digital learning for K-12 schools. Cham: Springer. 91-113.

KOSEM, Iztok, ROZMAN, Tadeja, ARHAR HOLDT, Špela, KOCJANČIČ, Polonca, LASKOWSKI, Cyprian Adam, 2016: *Šolar 2.0: nadgradnja korpusa šolskih pisnih izdelkov*. Tomaž Erjavec, Darja Fišer (ur.): Zbornik konference Jezikovne tehnologije in digitalna humanistika. Ljubljana: Znanstvena založba

Filozofske fakultete. 95–100. https://www.sdjt.si/wp/wp-content/uploads/2016/09/JTDH-2016_Kosem-et-al_Solar-2-0-nadgradnja-korpusa-solskih-pisnih-izdelkov.pdf

KOSEM, Iztok, STRITAR KUČUK, Mojca, MOŽE, Sara, ZWITTER VITEZ, Ana, ARHAR HOLDT, Špela, ROZMAN, Tadeja, 2012: *Analiza jezikovnih težav učencev: korpusni pristop*. Ljubljana: Znanstvena založba Filozofske fakultete. <https://e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/view/229/329/5311-1>