06 JOS-SYN Syntax

The JOS-SYN system, which was crafted during the Linguistic Annotation of Slovene: Methods and Resources project (Erjavec et al. 2010) and later applied in the Communication in Slovene initiative (Krek et al. 2020), is designed to mark syntactic relations in Slovene sentences. Drawing on the scholarly groundwork of Slovene linguistics, particularly Toporišič's 2004 grammar, it also integrates core principles from established dependency tagging frameworks. Its hallmark is the incorporation of insights from the JOS morphosyntactic annotation or its updated version, MULTEXT-East v6 (Erjavec 2012). The system enriches the data with only the essential information not captured by morphosyntax, thus creating a robust and clear annotation framework.

- Introduction to Tags
- Annotation Guidelines
- References and Links

Introduction to Tags

This chapter summarises the JOS-SYN syntax tags. A more detailed presentation can be found in the guidelines in the Annotation Guidelines chapter.

Tag	Description
Atr (Attribute)	Atr is used to link heads and their dependents in word phrases. The source is the head of the phrase, the target is its dependent. Typically it is used in noun phrases, adjectival and adverbial phrases or to connect parts of complex verb phrases with modal verbs and non-finite verb forms, as well as to link subject or object complements to the verb.
PPart (Predicate part)	PPart forms a link between elements without a dependency relation in the usual head-dependent sense which are consequently defined merely as parts of a word phrase. Typically it is used to link parts of verb phrases with the finite verb form or a participle ending in -l, as the source, and morphemes »ne«, »se«, »si«, »bi«, or the forms of the auxiliary verb be used to form future and past tenses, i. e. »bo«, »je«, etc., as the target.
Coord (Coordination)	Coord is used to link parts of coordinate structures on phrase level. It forms a link between the head of the first part of the coordinate structure and the head of the second part of the structure. The source is always the head in the left part of the structure and the target is the head in the right part of the structure.
Conj (Conjunction)	Conj is used in combination with the Coord relation to link three elements – connected with Coord and Conj – in a triangle. Conj is used to link the head of the second part of the coordinate structure on the phrase level, as the source, and the coordinating conjunction or punctuation mark (if it functions as the coordinating conjunction), as the target.
MWU (Multi-word unit)	MWU is used to link words which have a very strong tendency to appear together as a group forming a multiword unit and do not show characteristics of a head-dependant phrase structure. Typically, this relation is used to link words with a variant spelling with or without a space, some multi-word conjunctions and similar elements.

Tag	Description
Sb (Subject)	Sb is used to link parts of clauses or sentences that can be defined as traditional subjects. However, the nodes linked with this relation do not comply entirely with the definition of a subject in traditional grammars. On the clause level, it forms a link between the predicate node and the subject node, with the head of the verb phrase in the predicate, as the source, and the head of the noun phrase or other kinds of phrases in the subject, as the target. On the sentence level, it forms a link between the main clause and the subject clause with the head of verb phrase in the main clause, as the source, and the head of the verb phrase in the subject clause, as the target.
Obj (Object)	Obj is used to link parts of clauses or sentences that can be defined as traditional objects. However, the nodes linked with this relation do not comply entirely with the definition of an object in traditional grammars. On the clause level, it forms a link between the predicate node and the object node, with the head of the verb phrase in the predicate, as the source, and the head of the noun phrase or other kinds of phrases in the object, as the target. On the sentence level, it forms a link between the main clause and the object clause with the head of verb phrase in the main clause, as the source, and the head of the verb phrase in the object clause, as the target.
AdvM (Adverbial of manner)	AdvM is used to link parts of clauses or sentences that can be defined as traditional adverbials of manner. However, the nodes linked with this relation do not comply entirely with the definition of such adverbials in traditional grammars. On the clause level, it forms a link between the predicate node and the adverbial node, with the head of the verb phrase in the predicate, as the source, and the head of the noun phrase or other kinds of phrases in the adverbial, as the target. On the sentence level, it forms a link between the main clause and the adverbial clause with the head of verb phrase in the main clause, as the source, and the head of the verb phrase in the adverbial clause, as the target.
AdvO (Adverbial, other)	AdvO is used to link parts of clauses or sentences that can be defined as traditional adverbials, with the exception of adverbials of manner. However, the nodes linked with this relation do not comply entirely with the definition of such adverbials in traditional grammars. On the clause level, it forms a link between the predicate node and the adverbial node, with the head of the verb phrase in the predicate, as the source, and the head of the noun phrase or other kinds of phrases in the adverbial, as the target. On the sentence level, it forms a link between the main clause and the adverbial clause with the head of verb phrase in the main clause, as the source, and the head of the verb phrase in the adverbial clause, as the target.

Tag	Description
Root (Root dependency)	Root forms a link between the abstract node of the clause or sentence, as the source, with elements which form further connections in a dependency tree. The targets are typically clause predicates, predicateless elliptical parts of sentences or independent particles within a sentence. Furthermore, it forms a link with all other tokens (word or punctuation) without an explicit syntactic role in a sentence.

Annotation Guidelines

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

Version 2.0 (02-2023)

Project Development of Slovene in a Digital Environment

ARHAR HOLDT, Špela, TERČON, Luka, KREK, Simon, LEDINEK, Nina, MOŽE, Sara, SAKSIDA, Amanda, HOLZ, Nanika, 2023: *Navodila za skladenjsko označevanje slovenščine po sistemu JOS-SYN.*

Različica 2.0. Rezultat projekta Razvoj slovenščine v digitalnem okolju. [DOCX] [PDF] - only in Slovene

Version 1.0 for non-standard Slovene (21-12-2016)

Project Janes - Resources, Tools and Methods for the Research of Nonstandard Internet Slovene

HOLOZAN, Peter, KREK, Simon, PIVEC, Matej, RIGAČ, Simon, ROZMAN, Simon, VELUŠČEK, Aleš, ARHAR HOLDT, Špela, 2016: *Smernice za označevanje z odvisnostnim sistemom JOS: nestandardna slovenščina.* Različica 1.0. Rezultat projekta Viri, orodja in metode za raziskovanje nestandardne spletne slovenščine. [DOCX] [PDF] - only in Slovene

Version 1.0

Project Communication in Slovene

HOLOZAN, Peter, KREK, Simon, PIVEC, Matej, RIGAČ, Simon, ROZMAN, Simon, VELUŠČEK, Aleš, 2008: *Specifikacije za učni korpus.* Različica 1.0. Kazalnik K2 projekta Sporazumevanje v slovenskem jeziku. [PDF] - only in Slovene

References and Links

This chapter compiles relevant references and provides links to projects where the JOS-SYN syntax has been developed and applied to Slovene texts.

Projects, in which the system has been developed or applied

JOS - Linguistic Annotation of Slovene: Methods and Resources

Communication in Slovene

Janes - Resources, Tools and Methods for the Research of Nonstandard Internet Slovene

Development of Slovene in a Digital Environment

Training corpora containing manually revised JOS-SYN tags

Arhar Holdt, Špela; Krek, Simon; Dobrovoljc, Kaja; Erjavec, Tomaž; Gantar, Polona; Čibej, Jaka; Pori, Eva; Terčon, Luka; Munda, Tina; Žitnik, Slavko; Robida, Nejc; Blagus, Neli; Može, Sara; Ledinek, Nina; Holz, Nanika; Zupan, Katja; Kuzman, Taja; Kavčič, Teja; Škrjanec, Iza; Marko, Dafne; Jezeršek, Lucija; Zajc, Anja, 2022, Training corpus SUK 1.0, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, http://hdl.handle.net/11356/1747.

The Q-CAT tool for manual annotation following the JOS-SYN system

Brank, Janez, 2022, Q-CAT Corpus Annotation Tool 1.4, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, http://hdl.handle.net/11356/1684.

References

Arhar Holdt, Špela; Fišer, Darja; Erjavec, Tomaž in Krek, Simon. Syntactic annotation of Slovene CMC: first steps. Proceedings of the 4th Conference on CMC and Social Media Corpora for the Humanities, 27.–28. september 2016, Ljubljana, Slovenia, 2016, str. 3–6. http://nl.ijs.si/janes/cmc-corpora2016/proceedings/.

Erjavec, Tomaž; Fišer, Darja; Krek, Simon in Ledinek, Nina. 2010. The JOS Linguistically Tagged Corpus of Slovene. V: Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Valeta, Malta, Maj. European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2010/pdf/139_Paper.pdf

Erjavec, Tomaž. 2012. MULTEXT-East: morphosyntactic resources for Central and Eastern European languages. Language Resources and Evaluation, 46(1): 131–142.

Krek, Simon; Erjavec, Tomaž; Dobrovoljc, Kaja; Gantar, Polona; Arhar Holdt, Špela; Čibej, Jaka in Brank, Janez. The ssj500k training corpus for Slovene language processing. V: Fišer, D. in Erjavec, T. Jezikovne tehnologije in digitalna humanistika: zbornik konference: 24.-25. september 2020, Ljubljana, Slovenija. Ljubljana: Inštitut za novejšo zgodovino, 2020. Str. 24–33.

http://nl.ijs.si/jtdh20/pdf/JT-DH_2020_Krek-et-al_The-ssj500k-Training-Corpus-for-Slovene-Language-Processing.pdf.

Toporišič, Jože. (2004): Slovenska slovnica. Maribor: Obzorja.