



SLOVENSKI STANDARD

SIST ISO 24617-4:2018

01-september-2018

**Upravljanje z jezikovnimi viri - Ogradje za semantično označevanje (SemAF) - 4.
del: Semantične vloge (SemAF-SR)**

Language resource management -- Semantic annotation framework (SemAF) -- Part 4:
Semantic roles (SemAF-SR)

Gestion de ressources linguistiques -- Cadre d'annotation sémantique (SemAF) -- Partie
4: Rôles sémantiques (SemAF-SR)

Ta slovenski standard je istoveten z: ISO 24617-4:2014

ICS:

01.020	Terminologija (načela in koordinacija)	Terminology (principles and coordination)
35.060	Jeziki, ki se uporabljajo v informacijski tehniki in tehnologiji	Languages used in information technology

SIST ISO 24617-4:2018

en,fr,de

INTERNATIONAL
STANDARD

ISO
24617-4

First edition
2014-08-01

**Language resource management —
Semantic annotation framework
(SemAF) —**

**Part 4:
Semantic roles (SemAF-SR)**

*Gestion de ressources linguistiques — Cadre d'annotation sémantique
(SemAF) —*

Partie 4: Rôles sémantiques (SemAF-SR)



Reference number
ISO 24617-4:2014(E)

© ISO 2014



COPYRIGHT PROTECTED DOCUMENT

© ISO 2014

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword	iv
Introduction	v
1 Scope	1
2 Terms and definitions	1
2.1 Formal semantic units.....	1
2.2 Conceptual semantic units and relations.....	2
2.3 General linguistic units.....	3
3 Abbreviated terms	4
4 Purpose and justification	4
5 The nature of semantic roles	4
5.1 General.....	4
5.2 Typical examples of semantic roles.....	6
6 Metamodel	7
6.1 Key concepts.....	7
6.2 Introduction to a metamodel for semantic role annotation.....	11
6.3 Criteria for distinguishing semantic roles.....	12
6.4 Defining eventuality frames.....	13
7 Interactions	14
7.1 Semantic roles and semantic types.....	14
7.2 Complexities.....	14
7.3 The role of context in semantic role assignment.....	14
7.4 Fuzzy boundaries between roles.....	15
7.5 Multiple classification.....	15
7.6 Inheritance relations between semantic roles.....	15
8 Guidelines for developing new semantic role frameworks for languages and/or domains	16
8.1 General.....	16
8.2 Mapping from VerbNet to LIRICS.....	17
Annex A (informative) Specification of ISO semantic roles	19
Annex B (informative) Review of existing frameworks	28
Annex C (informative) Specification of the annotation language	37
Bibliography	41

ISO 24617-4:2014(E)

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information

The committee responsible for this document is ISO/TC 37, *Terminology and other language and content resources*, Subcommittee SC 4, *Language resource management*.

ISO 24617 consists of the following parts, under the general title *Language resource management — Semantic annotation framework (SemAF)*:

- *Part 1: Time and events (SemAF-Time, ISO-TimeML)*
- *Part 2: Dialogue acts*
- *Part 4: Semantic roles (SemAF-SR)*
- *Part 5: Discourse structure (SemAF-DS)*
- *Part 7: Spatial information (ISO-Space)*

The following parts are under preparation:

- *Part 8: Semantic relations in discourse (SemAF-DRel)*

Principles of semantic annotation (SemAF-Basics) will form the subject of future Part 6.

Introduction

This part of ISO 24617 aims to specify criteria for defining semantic roles (SRs), and is the outcome of an agreement that the various semantic role frameworks being used to support data annotation (e.g. FrameNet, VerbNet, PropBank, EngVallex, and LIRICS, to name only a few examples for English) have strong underlying compatibilities. The goal is to provide both an explanation of these compatibilities and a loose mapping between definitions of individual semantic roles, as listed in the different frameworks, that will benefit the community as a whole.

The current specification has been developed under the aegis of the ISO Semantic Annotation Framework (SemAF), where it is known as SemAF-SR.

The main parts of ISO 24617-4 consist of the following:

- Scope;
- Normative references;
- Terms and definitions;
- motivation and requirements;
- basic concepts and metamodel specifications;
- examples of mapping existing frameworks to the metamodel.

This part of ISO 24617 contains three informative annexes. In [Annex A](#), the ISO semantic roles are specified. In [Annex B](#), information is provided both on past and current activities in semantic role annotation and on tools and frame files. [Annex C](#) contains the abstract and concrete syntax for the metamodel.

Language resource management — Semantic annotation framework (SemAF) —

Part 4: Semantic roles (SemAF-SR)

1 Scope

The aim of this part of ISO 24617 is to propose a consensual annotation scheme for semantic roles; that is to say, a scheme that indicates the role that a participant plays in an event or state, as described mostly by a verb, and typically providing answers to questions such as “‘who’ did ‘what’ to ‘whom’”, and ‘when’, ‘where’, ‘why’, and ‘how’. This includes not only the semantic relations between a verb and its arguments but also those relations that are relevant for other predicative elements such as nominalizations, nouns, adjectives, and predicate modifiers; the predicating role of adverbs and the use of coercion fall outside the scope of this part of ISO 24617.

NOTE In linguistics, **coercion** occurs when the grammatical context causes the language-user to reinterpret all or parts of the semantic and/or formal features of a lexeme that appear in that context.^[60]

2 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

2.1 Formal semantic units

2.1.1

argument

formal semantic unit that is an essential element of a *predicate argument structure* (2.1.3) and can have variable instantiations depending on the utterance

Note 1 to entry: An argument corresponds to a *participant* (2.2.5) of an *eventuality* (2.2.2) described by the *predicate argument structure* (2.1.3).

Note 2 to entry: Arguments typically satisfy certain argument positions and can be described as being syntactico-semantic notions, whereas *participants* (2.2.5) are semantico-conceptual. The standard view is that subsets of the *participants* associated with an *eventuality* (2.2.2) are selected as arguments by the verb (or nominal or adjective) expressing the *eventuality* (2.2.2). Other *participants* (2.2.5) are either incorporated or realized as *eventuality modifiers* (2.2.4).

Note 3 to entry: Natural language predicates typically have one, two, or three arguments, although they can have more.

2.1.2

predicate

formal semantic unit that represents a semantic relation between one or more *arguments* (2.1.1) in a *predicate argument structure* (2.1.3)

Note 1 to entry: Predicates are indicated by predicative linguistic elements such as verbs, nouns, and adjectives.

ISO 24617-4:2014(E)

2.1.3

predicate argument structure

formal representation of the core semantic content of an utterance, consisting of a *predicate* (2.1.2) constant, and its *arguments* (2.1.1)

Note 1 to entry: In classical logic-based semantics, this corresponds to predicate argument structures in first-order predicate logic.

Note 2 to entry: One of the *arguments* (2.1.1) can be a variable uniquely identifying the instance of the predicate argument structure to allow references to it in other predicate argument structures.

Note 3 to entry: The representation of event semantics is subject to many variations; some of them, such as in Reference [41], can have separate *predicates* (2.1.2) for each *semantic role* (2.2.6) relation. In this case, the predicate argument structure of an utterance is the sum of the individual predicate *semantic role* (2.2.6) assertions representing the semantic content of the utterance.

2.2 Conceptual semantic units and relations

2.2.1

entity

conceptual semantic unit that typically functions as a *participant* (2.2.5)

Note 1 to entry: An entity is an individual such as a person, organization, physical object, or logical entity, as well as, on occasion, a number, quantity, dimension, or a reification of an eventuality, a property, or a quality, e.g. emotion (anger, love), the value of a colour, etc.

2.2.2

eventuality

event, state, process, or action which can have *participants* (2.2.5) and which is being referred to by a verbal, adjectival, or nominal description in an utterance

Note 1 to entry: The formal representation of an eventuality is a *predicate argument structure* (2.1.3).

Note 2 to entry: See ISO 24617-1. An eventuality can also be described as ‘something that can be said to obtain or hold true, to happen or to occur’, as in ISO 24617-1. As such, they can be actual, hypothetical, or generic, covering situations such as “You should go home,” or “He might be John’s brother.”

2.2.3

eventuality frame

generalized abstract specification of the *word sense* (2.3.6) associated with an *eventuality* (2.2.2) in an utterance

Note 1 to entry: The frame consists of the specification of (a) a *predicate* (2.1.2) that can participate in a class hierarchy if such a hierarchy is specified, and (b) the *arguments* (2.1.1) that this *predicate* (2.1.2) expects along with their *semantic roles* (2.2.6).

2.2.4

eventuality modifier

particular type of *participant* (2.2.5) that completes the description of an *eventuality* (2.2.2) but is optional and not essential

Note 1 to entry: Eventuality modifiers are distinct from other types of participants in that they are used in supplying information that is typically more peripheral and more general, for example, situating the eventuality in time or space.

Note 2 to entry: In FrameNet, these would be peripheral frame elements and in PropBank, ArgM’s.

Note 3 to entry: Eventuality modifiers typically correspond to syntactic adjuncts.

2.2.5**participant**

conceptual semantic unit referred to by one or more lexical items in an utterance, which is or can be involved in an *eventuality* (2.2.2)

Note 1 to entry: Both *entities* (2.2.1) and *eventualities* (2.2.2) can function as participants.

2.2.6**semantic role**

mode of involvement of a *participant* (2.2.5) in an *eventuality* (2.2.2)

Note 1 to entry: Semantic roles for specific eventualities are often associated with prototypical semantic relations, e.g. if *John* causes a *breaking* event, he is the *Agent*; if he uses a *hammer*, it is the *Instrument*; and someone who *receives* something is a *Recipient* (see [Clause 5](#) for descriptions).

2.3 General linguistic units**2.3.1****lemma****lemmatized form**

conventional form chosen to represent a *lexeme* (2.3.2)

Note 1 to entry: See ISO 24611.

2.3.2**lexeme**

fundamental unit, generally associated to a set of word forms sharing a common meaning

Note 1 to entry: See ISO 24611.

2.3.3**lexical entry**

container for managing a set of word forms and possibly one or several meanings [*word senses* (2.3.6)] to describe a *lexeme* (2.3.2)

Note 1 to entry: See ISO 24611.

2.3.4**lexicon**

resource comprising a collection of *lexical entries* (2.3.3) for a language

Note 1 to entry: See ISO 24611.

2.3.5**utterance**

stretch of speech about which no assumptions have been made in terms of linguistic theory

Note 1 to entry: See Reference [12].

2.3.6**word sense**

meaning associated with a *lexeme* (2.3.2) in a context

Note 1 to entry: The 'river bank' sense of *bank* and the 'financial institution' sense of *bank* are considered to be two different word senses, or *lexical units*, with the same word form, or *lexeme* (2.3.2). *I called him on the radio* and *Call me a taxi* are associated to different word senses of the *lexeme* (2.3.2) *call*. Unrelated senses, as in *bank*, are called *homonyms*. Senses of the same word form or *lexeme* which are clearly related (and can be difficult to distinguish) are called *polysemes*, e.g. *Coins with an image of the king, preoccupied with body image, evokes a strong mental image*.

ISO 24617-4:2014(E)

3 Abbreviated terms

EngVallex	English Valency Lexicon
LIRICS	Linguistic Infrastructure for Interoperable Resources and Systems
PropBank	Proposition Bank
SR	semantic roles
SRL	semantic role labelling
WSD	word sense disambiguation

4 Purpose and justification

Semantic roles are arousing increasing interest in the information-processing community because they make explicit the key conceptual relations of participation between a verb and its arguments; that is to say, they specify ‘who’ did ‘what’ to ‘whom’, and ‘when’, ‘where’, ‘why’, and ‘how’. For English alone, there are already several different semantic role frameworks, including FrameNet, VerbNet, LIRICS, EngVallex, and PropBank. Although these have been developed independently, there are strong underlying compatibilities between them, and they share a central definition of what a semantic role is, and what its span is, within an individual sentence. In addition to defining key concepts, this part of ISO 24617 aims to clarify and specify these underlying compatibilities and provide, where possible, a mapping between similar semantic roles across different frameworks. This mapping is intended to serve as an illustration of how different semantic role definitions can be linked to each other across frameworks, and presupposes a specification of clearly defined criteria for distinguishing semantic roles.

The specification will be used in two different situations:

- in annotations where the semantic roles are recorded in annotated corpora;
- as a dynamic structure produced by automatic systems, a process typically known as semantic role labelling (SRL).

The objectives of this specification are to provide

- a reference set of data categories that define a structured collection of semantic roles with an explicit semantics,
- a pivot representation based on a framework for defining semantic roles that can facilitate mapping between different formalisms (alternative semantic role representations/syntactic theories/eventually different languages) and, in the future, between different languages, and
- guidelines for creating new resources that will be immediately interoperable with pre-existing resources.

5 The nature of semantic roles

5.1 General

For computers to make effective use of information encoded in text, they must be able to detect the eventualities that are being described and the eventuality participants. The processing of a sentence like *John threw a ball to Mary in the park* should result in the identification of a throwing event involving *John* as the Agent of the event, *Mary* as the Recipient, and the *ball* as the item being *thrown*; the location of the throwing event, or where it took place, is *the park*. This description of the event specifies the conceptual relations of participation that the referents of the noun phrases play with respect to the event. The semantic notions being specified are the *roles of the participants in an eventuality (i.e. semantic roles)*.

This part of ISO 24617 establishes LIRICS (see [Annex A](#)) as a reference set of semantic roles with precise definitions. Researchers are free to define their own sets of semantic roles, but explicit information on how they can be mapped to the reference set will make resources more interoperable. Many resources currently map to PropBank, VerbNet, or FrameNet. Since this part of ISO 24617 includes mappings of these resources to LIRICS, such mappings already qualify as meeting the requirement of interoperability.

Our *throw* example seems fairly straightforward, but complexities quickly arise. English, for instance, allows not only several different syntactic constituents to present the same semantic role, but also several different semantic roles to be presented by the same syntactic constituent. For decades, a central concern of linguists has been the elucidation of the process of mapping back and forth between the syntactic analysis of the sentence and the conceptual structure and relations in the event described. For example, in the following two sentences,

(1) *The flame melted the wax.*

(2) *The wax melted.*

a standard syntactic parser represents the *wax* as the verb's direct object in the first sentence and its subject in the second. There is nothing overt to indicate that it has the same conceptual relation in both cases despite the fact that it is expressed syntactically in a different way. We can capture this by annotating the *wax* as having the same semantic role (or conceptual relation) in both sentences. It would typically be labelled the Patient, the participant undergoing a change of state. Note that both sentences are in the active voice, not the passive voice. In *The wax was melted by the flame*, the passive provides syntactic evidence that the *wax* is playing the same role (Patient) that it plays in example (1). Since the particular pair of syntactic variations illustrated by *melt* does not occur with every transitive verb [see example (5)], it is not easily predictable. Other transitive verbs can also occur in an intransitive form while maintaining the same semantic role for the subject as the transitive, as in the following example, where *soprano* is the Agent of *sing* in both sentences (the *aria* is the Theme):

(3) *The soprano sang an aria.*

(4) *The soprano sang.*

The verb *slice* can also move the Patient (*the bread*) to subject position, as in

(5) *John sliced the bread easily./ This bread slices easily.*

although other transitive verbs, such as *eat*, cannot:

(6) *John ate the apple.*

(7) *John ate.*

(8) **The apple ate crunchily in the background.*

The last sentence is starred (*) to indicate its ungrammaticality.

Accurate interpretation of the semantic roles of the verb arguments (i.e. 'Who did what to whom?') is a crucial goal for natural language processing systems. Our ability to do this automatically has improved enormously in recent years and has been largely based on the availability of annotated corpora. In fact, there are corpora, such as FrameNet and PropBank, available with quite different semantic role annotations, and this prompts questions about the nature and number of semantic roles. This part of ISO 24617 attempts to provide definitions and examples clarifying their definition.

For semantic roles to maximize the benefit to the information processing community, it is desirable that the definitions of the semantic roles should, as far as possible, have the following properties:

- consistently recognizable;
- able to clarify sense distinctions;
- generalizable;