BS ISO 24617-4:2014

BSI Standards Publication

Language resource management — Semantic annotation framework (SemAF)

Part 4: Semantic roles (SemAF-SR)

bsi.

...making excellence a habit.™

Document Sharing Hub

Full document details: BS ISO 24617-4:2014

National foreword

This British Standard is the UK implementation of ISO 24617-4:2014.

The UK participation in its preparation was entrusted to Technical Committee TS/1, Terminology.

A list of organizations represented on this committee can be obtained on request to its secretary.

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

© The British Standards Institution 2014. Published by BSI Standards Limited 2014

ISBN 978 0 580 73557 8

ICS 01.020

Compliance with a British Standard cannot confer immunity from legal obligations.

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 30 September 2014.

Amendments issued since publication

Date	Text affected

Documents
Document Sharing Hub

Full document details: BS ISO 24617-4:2014

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)

Documents
Document Sharing Hub
© ISO 2014

Full document details: BS ISO 24617-4:2014

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

© ISO 2014 – All rights reserved

Documents
Document Sharing Hub

Full document details: BS ISO 24617-4:2014

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 2014 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 24617-4:2014

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.

© ISO 2014 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 24617-4:2014

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.

© ISO 2014 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 24617-4:2014

Documents
Document Sharing Hub

INTERNATIONAL STANDARD ISO 24617-4:2014(E)

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 2014 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 24617-4:2014

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.

© ISO 2014 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 24617-4:2014

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 2014 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 24617-4:2014

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)*.

© ISO 2014 – All rights reserved
AVA Documents Document Sharing Hub
Full document details: BS ISO 24617-4:2014

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;

© ISO 2014 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 24617-4:2014

— machine learnable;

— able to provide an appropriate foundation for inferencing.

The purpose of the specifications for semantic role definitions in this part of ISO 24617 is to provide these attributes.

5.2 Typical examples of semantic roles

A list of the best-known roles and the properties usually associated with them is given below. They are taken from the EAGLES discussion on Standardizing Subcategorization (see Reference [61]). Comments in parentheses have been added for clarification purposes and include comparisons with LIRICS (Linguistic Infrastructure for Interoperable Resources and Systems), and occasionally with VerbNet.

NOTE These role definitions, and the ones for LIRICS, are quite general and can cover a wide range of participant types. The specific preposition associated with an individual semantic role quite often adds nuances of meaning in addition to that conveyed by the semantic role itself. For instance, in "*Eat the fish with caution*," the phrase "with caution" would typically be labelled as MANNER. Another example of a MANNER role could be the phrase "in three bites" from "*Eat the fish in three bites*." Clearly the interpretation of these two phrases will be quite different, as will be their impact on the representation of the *eating* eventuality. Semantic role labels cannot be expected to clarify all such types of subtle differences in meaning, and additional research on the definitions of prepositions and their interactions with semantic roles, such as Srikumar and Roth, 2013[48], is needed.

Agent

A participant that the meaning of the verb specifies as doing or causing something, possibly intentionally; for example, as the subject of *kill*, *eat*, *hit*, *smash*, *kick,* and *watch*. (LIRICS has a similar Agent, which acts intentionally or consciously.)

Patient

A participant that the verb characterizes as having something happen to it, and as being affected by what happens to it; for example, as the object of *kill*, *eat,* and *smash*, but not of *watch*, *hear,* and *love*. (If someone watches television, the television is not affected by the watching, so it would be a Theme rather than a Patient. LIRICS has a similar Patient.)

Experiencer

A participant that is characterized as aware of something; for example, as the subject of *love* or as the object of *annoy*. (LIRICS uses an Agent for these verbs and has no Experiencer; VerbNet has the same role, but only when the Experiencer is affected by the event.)

Theme

A participant that is characterized as changing its position or condition, or as being in a state or position; for example, as the object of *give* and *hand* or as the subjects of *walk* (in line with the policy of labelling the object in motion as a Theme) and of *die*. (According to the EAGLES definition, people who die of old age would not be considered to be Patients because an Agent has not acted upon them. EngVallex stays more syntactically oriented and marks such participants as Actors or Patients, depending on their syntactic position.)

Location

The thematic role associated with the Noun Phrase expressing the location in a sentence with a verb of location [perhaps in a Prepositional Phrase (PP)]; for example, the subjects of *keep*, *own*, *retain*, and *know*, and locative PPs. (For EAGLES, the location of the thing *kept* or *owned* is considered to be with the *keeper* or the *owner*, and so on. LIRICS has a Location role, but uses it in a more restrictive way, and uses Agent or Pivot for several of these verbs instead of Location, as does VerbNet.)

© ISO 2014 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 24617-4:2014

Source

Object from which motion proceeds; for example, as the subject of *sell* and *promise* or as the objects of *deprive*, *free,* and *cure*. Note that the motion can be abstract. (LIRICS uses Source for these abstract examples, but uses Initial-Location instead for physical motion verbs.)

Goal

Object to which motion proceeds (e.g. the Path-prepositional phrase for caused-motion verbs like *throw*, and for metaphorical motion events), the subject of *receive* and *buy* and the dative object of *tell* and *give*. (Adapted from Dowty, 1989.[71] LIRICS has a for abstract objects and uses Final-Location for physical motion.)

Although these semantic roles were initially defined in relation to the arguments and syntactic adjuncts of verbs, many theories recognize semantic roles associated with nouns, particularly event nominals related to verbs. *The Romans*, *Jerusalem,* and *70 C.E*. therefore play the same roles in both examples (9) and (10):

(9) *The Romans destroyed Jerusalem in 70 C.E.*

(10) *The destruction of Jerusalem by the Romans in 70 C.E. is famous.*

Nombank, which extends PropBank to nouns, handles this by using the same roleset for both the verb and the noun; FrameNet handles it by including *destroy* and *destruction* in the same semantic frame, and annotating each with the same frame elements.

6 Metamodel

6.1 Key concepts

The discussion of a metamodel begins by defining the following concepts:

— eventuality frames;

— arguments/adjuncts;

— granularity of word senses and semantic roles;

— semantic classes;

— entailment.

6.1.1 Eventuality frames

This part of ISO 24617 adopts the classic type-token distinction that is accepted by philosophical and knowledge representation communities, which is to say that a concept is distinct from the objects that are instances of it. With respect to eventualities, an individual utterance can correspond to an instance, or a token, of a particular eventuality frame (2.2.3). (Note that an utterance does not necessarily describe an eventuality, e.g. *Oh, no*! or *Ouch!*) All utterances that describe eventualities involving objects being *thrown* can be considered to be tokens of the eventuality frame '*throwing of objects*'. Semantic role labelling frameworks require the creation of lexical entries for each eventuality frame to be annotated; these lexical entries are described at the type level and illustrated with several naturally occurring sentences that describe tokens. For instance, for each lexeme PropBank defines a Frame File with one or more framesets (also called rolesets) corresponding to individual eventuality frames (FrameNet associates it with a Frame, VerbNet with a class, and EngVallex with a valency entry). These are used to guide the annotation process.

© ISO 2014 – All rights reserved

Documents
Document Sharing Hub
Full document details: BS ISO 24617-4:2014

6.1.2 Arguments and adjuncts

The argument/adjunct distinction is central to a definition of semantic roles because of the different relations that arguments and adjuncts (defined in Clause 2 as eventuality modifiers) have with regard to the predicate. This distinction is not always easy to grasp, but in practice, nearly all successful semantic role annotation projects assume some variation of this distinction. Arguments are characterized by the following three properties. The first property is obligatoriness. A predicate is central to an eventuality description, but the arguments are the necessary elements without which the eventuality description is incomplete. This makes arguments obligatory, at least prototypically. Obligatoriness cannot be defined on purely syntactic grounds, since arguments that are essential to a predicate are routinely dropped in surface realizations; for example, in example (7), repeated as example (11), the thing that is *eaten by John* is not realized syntactically, but there is still a strong sense that it is an important element in the eventuality description, and a core role has to be assigned to this argument. By contrast, although the *eating* event must have occurred at a certain time and in a certain place, the time and place are not necessary for the eventuality description to be complete, and are considered adjuncts, or eventuality modifiers.

(11) *John ate.*

Being obligatory also seems to be linked to a second property of arguments, which is that the types of arguments associated with a predicate can be specific to that predicate since they are based on its semantics. In other words, different predicates tend to take different sets of arguments. A Recipient is a receiver of an item or a message, as in *He gave a book to Mary*. This role is natural in these sentences but would be very odd in the description of an *eating* eventuality. This contrasts with typical adjuncts like time and location, which can occur with a wide variety of quite different predicates. To put it another way, arguments tend to co-vary with the predicate while adjuncts do not. Statistically speaking, arguments are also more likely than adjuncts to co-occur with the predicate.

The third property of arguments is that they are often assumed to be unique, and multiple arguments are not expected to fill the same role and have the same semantic relationship with the predicate. A notable exception to this can occur with plurals, which for many verbs can also be split into separate argument slots. For example, *John and Mary met*, can also be readily phrased as *John met Mary*, and with such subtle differences, if any, in focus and agentivity that they are sometimes labelled Agent and Co-Agent. A significant number of 'reciprocal' predicates such as *fight*, *match*, *marry*, *play* (in the sense of playing a two-person game like tennis), and *tie* (to score equal points) share this property. Reciprocality is not confined to Agents and could also characterize Themes (*match),* and perhaps other roles as well.

Different semantic role annotation schemes, which are described in Annex B, often choose to formalize all, or a subset, of these properties, as discussed in that annex. In theoretical linguistics, there is much controversy over where the boundary should be drawn between arguments and adjuncts. For a fairly radical view setting out a fluid characterization of the boundary between arguments and adjuncts, see Reference [15]. Generally speaking, arguments are considered to influence the interpretation of the verb while adjuncts are not.

6.1.3 Granularity of word senses and semantic roles

The semantic ambiguity of words is pervasive; many words, especially the most frequently used words, have multiple meanings. For example, one can *draw a gun, draw water from a well*, and *draw a diagram*. Depending on the dictionary, an entry for a word like *draw* or *run* can list dozens of separate definitions, but unless they encounter puns or other forms of word-play, people rarely have difficulty interpreting the meaning of these words. Although lexical ambiguity can present very few problems to people engaged in normal interpretation of text or speech, the same cannot be said for computers. The correct selection of the appropriate meaning of a word in context has proved to be very difficult for natural language processing (NLP) systems. Yet an accurate means of performing word sense disambiguation (WSD) would improve many NLP applications such as information extraction, information retrieval, and machine translation, and also any task that requires more complex knowledge representation and reasoning.[8] [9] [24] [47] [50] Encouraging progress has been made recently with automatic word sense disambiguation systems that approach human levels of accuracy.[68]

© ISO 2014 – All rights reserved
Documents
Document Sharing Hub
Full document details: BS ISO 24617-4:2014

This is equally true for semantic role labelling. Different senses of the same word can refer to different event types and therefore require different semantic roles. For instance, the possible association of an Agent and a Co-Agent mentioned above only applies to a particular sense of *play*, as in *John and Bill played tennis* or *John played tennis with Bill*. The acting sense, as in *Olivier played Hamlet* or in the musical performance sense, *Yo-Yo Ma played the cello*, are less likely to have Co-agents. Appropriate word sense disambiguation can improve the performance of semantic role systems by narrowing the sets of semantic labels that can apply.[53]

A fundamental problem for WSD is choosing the set of senses to be distinguished. Generally referred to as a sense inventory, the set of senses used for a WSD system is expected to be a comprehensive and fixed list of senses for every word used in the domain of the application. This conception of word senses matches people's experience with dictionaries. Dictionaries encourage to consider words as having a discrete set of senses, yet any comparison between dictionaries quickly reveals how differently a word's meaning can be divided into separate senses. Rather than having a finite list of senses, many words seem to have senses that shade from one meaning into another; where to 'draw' the line between senses often seems to be an arbitrary decision. Moreover, the determination of how many lines to draw, that is, how narrow or how general to make the senses (this is an issue of **granularity**), can vary greatly, depending on who is creating the resource. WordNet is deemed to make **fine-grained** sense distinctions, whereas the sense distinctions made by PropBank and EngVallex are particularly **coarse-grained** and are only concerned with different senses that also have different subcategorization frames.

Different choices of sense distinctions can impact on semantic role frameworks. Fine-grained sense distinctions will encourage fine-grained semantic role labels in order to capture subtle differences in the types of argument that typically occur with these senses. Examples of shifting levels of granularity can be seen in the discussion of mappings between LIRICS and PropBank, and between VerbNet, EngVallex, and FrameNet.[2] [3]. With a few exceptions, VerbNet roles are at a level of granularity that is similar to that of LIRICS, while PropBank and EngVallex are typically more coarse-grained and FrameNet is typically more fine-grained.

6.1.4 Semantic classes

One view of lexical semantics postulates that word senses can be associated with concepts and, in particular, that different senses of a word are often characterized by different semantic classes associated with the syntactic object. How these associations are created and maintained is an open question. Katz and Fodor (1963)[27] proposed one of the first linguistic theories of lexical semantics; it assumes that a linguistic grammar has a semantic component that assigns semantic representations to lexical items and, by means of recursive 'projection rules', to phrases and sentences. These representations are deemed to derive from a shared categorization of the world that includes common taxonomic elements such as PERSON, ANIMAL, PLANT, TREE, DOG, ARTEFACT, and BUILDING. This type of categorization relies heavily on subtype and supertype relations, such as *A DOG is a subtype of ANIMAL*, or *A TREE is s subtype of PLANT.*

Notwithstanding various attempts at building large-scale ontologies that capture these subtype relations in semantic class hierarchies, the field of natural language processing has never reached a consensus on a single, universal knowledge representation schema. However, the vast majority of systems rely on WordNet to fulfil this function, even though it was not developed with this in mind. The subtype/supertype relations mentioned in the previous paragraph parallel the lexical hyponym and hypernym relations that form the backbone of WordNet's associations. Other important lexical relations in WordNet are synonymy (similar, possibly substitutable, items), meronomy (part-whole), and antonymy (opposites). WordNet uses these relations extensively to create a large semantic network involving over 100 000 lexical units in the English language.[17] [38] This network has proved to be an invaluable resource for natural language processing systems seeking semantic class inheritance relations, and these systems often use the hypernym relations in an ontological form to induce supertypes. This use of WordNet blurs the distinction between lexical items and purely conceptual categories. The approach has been ported to dozens of other languages.

Another example of a broad-scale natural language processing application of assigning semantic classes to data are the widely used Named Entity types, or nominal entity types, developed under the ACE (Automatic Content Extraction) program.[51] This is basically a semantic classification task

© ISO 2014 – All rights reserved

AVA Documents
Document Sharing Hub
Full document details: BS ISO 24617-4:2014