DRAFT INTERNATIONAL STANDARD
ISO/DIS 24617-2

ISO/TC **37**/SC **4**

Secretariat: **KATS**

Voting begins on:
2019-11-13

Voting terminates on:
2020-02-05

Language resources management — Semantic annotation framework (SemAF) —

Part 2: Dialogue acts

Gestion des ressources langagières — Cadre d'annotation sémantique (SemAF) —

Partie 2: Actes de dialogue

ICS: 01.020

This document is circulated as received from the committee secretariat.

Reference number
ISO/DIS 24617-2:2019(E)

COPYRIGHT PROTECTED DOCUMENT

Contents

Page

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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

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.

ISO 24617-2 was prepared by Technical Committee 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*

— *Part 1: Time and events*

— *Part 2: Dialogue acts*

— *Part 4: Semantic roles*

— *Part 6: Principles of semantic annotation*

— *Part 7: Spatial information*

— *Part 8: Semantic relations in discourse*

— *Part 9: Reference annotation framework*

— *Part 11: Measurable quantitative information*

— *Part 12: Quantification*

Introduction

Since its publication in 2012, ISO 24617-2 has been used in a number of annotation efforts as well as in the development of language-based interactive systems. These experiences have brought to light (1) that the standard allowed dialogue act annotations that are slightly inaccurate in some respects, (2) that some applications would benefit from the availability of mechanisms for customizing the set of concepts defined in the standard, and (3) that certain use cases require the representation of functional dialogue act information to be extended with semantic content information. This second edition seeks to remedy the noted inaccuracies, and to provide mechanisms (a) for customizing the set of defined concepts, and (b) for extending the information types in dialogue act annotations.

The improved accuracy of this second edition concerns the annotation of semantic dependence relations of dialogue acts and their scopes, and of rhetorical relations between dialogue acts. The mechanisms for extending and customizing the standard for a specific application concern most notably the annotation of information about the (domain-specific) semantic content of dialogue acts, the introduction of application-specific dialogue act types, the addition of communicative functions for fine-grained specification of feedback, and the annotation of speaker emotions.

This second edition is downward compatible with the original ISO 24617-2:2012 in the sense that every annotation made with the original version is a valid annotation according to the second edition. Existing annotations do not need to be revised in order to be compliant with the second edition.

Language resources management —Semantic annotation framework (SemAF) — Part 2: Dialogue acts

1   Scope

This international standard provides a set of empirically and theoretically well-motivated concepts for dialogue annotation; a formal language for expressing dialogue annotations, the Dialogue Act Markup Language (DiAML); and a method for segmenting a dialogue into semantic units. This allows the manual or automatic annotation of dialogue segments with information about the communicative actions which the participants perform by their contributions to the dialogue. The standard supports multidimensional annotation, in which units in dialogue are viewed as having multiple communicative functions. The DiAML language has an XML-based representation format, and a formal semantics which makes it possible to apply inference to DiAML representations. The standard specifies data categories for reference sets of communicative functions and dimensions of dialogue analysis, and provides mechanisms for customizing these sets of concepts, extending them with application- or domain-specific concepts, or selecting relevant coherent subsets of them. Additionally, the standard provides guidelines for annotators and annotated examples. This standard is applicable to spoken, written, and multimodal dialogues involving two or more participants.

2   Normative references

The following referenced documents are indispensable for the application of this document.

ISO 12620:2009 *Terminology and other language resources – Specification of data categories and management of a Data Category Registry for language resources*.

ISO 24612:2011 *Language resource management – Linguistic annotation framework*.

ISO 24610-1:2006 *Language resource management -- Feature structures, Part 1: Feature structure representation*.

ISO 24617-6:2015 *Language resources management —Semantic annotation framework (SemAF) — Part 6: Principles of semantic annotation*.

3   Terms and definitions

For the purposes of this document, the following terms and definitions apply.[1]

3.1
addressee
dialogue (3.6) **participant** (3.14) oriented to by the **sender** (3.20) in a manner to suggest that his **utterance**s (3.23) are particularly intended for this participant, and that some response is therefore anticipated from this participant, more so than from the other participants.

NOTE Source: Goffman (1981). This definition is a *de facto* standard in the linguistics literature. It has been slightly modified here, in replacing 'speaker' by 'sender' and avoiding the use of ambiguous pronouns. Goffman's original definition says: "dialogue participant oriented to by the speaker in a manner to suggest that his utterances are particularly intended for him, and that some response is therefore anticipated from him/her, more so than from the other participants".

---

[1] In this document, 'he', 'him' and 'his' are used in a generic sense, without implying any gender-related distintinctions.

3.2
allo-feedback act
feedback act (3.8) where the **sender** (3.20) elicits information about the **addressee**'s (3.1) processing of an **utterance** (3.23) that the sender contributed to the **dialogue** (3.6), or where the sender provides information about his perceived processing by the addressee of an utterance that the sender contributed to the dialogue before

EXAMPLE   1. A: Now move up.
                   2. B: Slightly northeast you mean?
                   3. A: Slightly yeah
With utterance 3, A performs an allo-feedback act signalling that he thinks B understood utterance 1 correctly.

3.3
auto-feedback act
feedback act (3.8) where the **sender** (3.20) provides information about his own processing of an utterance contributed to the **dialogue** (3.6) by another **participant** (3.14)

EXAMPLE   B's utterance in the example dialogue fragment in (3.2) signals that he is uncertain whether he understood the previous utterance correctly.

3.4
communicative function
property of certain stretches of communicative behaviour, describing how the behaviour changes the **information state** (3.12) of an understander of the behaviour

NOTE A communicative function may be 'qualified', i.e. one or more **qualifiers** (3.15) may be associated with it. For example, an answer may be qualified as 'uncertain', and the acceptance of a request may be 'conditional'. See Clause 10.3 for explanation and examples.

3.5
context
synonym for **information state** (3.13)

3.6
dialogue
exchange of **utterances** (3.23) between two or more persons or artificial conversational systems

3.7
dialogue act
communicative activity of a **dialogue** (3.6) **participant** (3.14), interpreted as having a certain **communicative function** (3.4) and **semantic content** (3.18)

NOTE A dialogue act may additionally also have certain **functional dependence relations** (3.10), **rhetorical relations** (3.17) and **feedback dependence relation**s (3.9) with other units in a **dialogue** (3.6).

3.8
dimension
class of **dialogue acts** (3.6) that are concerned with a particular aspect of communication, corresponding to a particular category of semantic content

NOTE Examples are (1) dialogue acts advancing the task or activity that motivates the dialogue (the 'Task' dimension) ; (2) dialogue acts providing and eliciting feedback (the Auto- and Allo-Feedback dimensions); (3) dialogue acts for allocating the speaker role (the Turn Management dimension).

See clauses 5, 7, and 9 for discussion and more examples.

3.9
feedback act
dialogue act (3.6) which provides or elicits information about the **sender**'s (3.20) or the **addressee**'s (3.1) processing of something that was uttered in the dialogue

NOTE Two classes of feedback are distinguished in this standard: **auto-feedback acts** (3.2) and **auto-feedback acts** (3.3).

3.10
feedback dependence relation
relation between a **feedback act** (3.8) and the stretch of communicative behaviour whose processing the act provides or elicits information about

EXAMPLE   In the example that accompanies definition (3.2), both the allo-feedback act expressed by utterance 3 and the auto-feedback act expressed by utterance 2 have a feedback dependence relation to utterance 1.

3.11
functional dependence relation
relation between a given **dialogue act** (3.6) and a preceding dialogue act on which the semantic content of the given dialogue act depends due to its **communicative function** (3.4)

EXAMPLE   The relation between an answer and the corresponding question, such as between utterance 3 and utterance 2 in the example accompanying definition (3.2), or the relation between the acceptance of an offer and the corresponding offer.

NOTE A dialogue act A2 may also depend on another dialogue act A1 occurring earlier in a dialogue because of relations between their semantic contents, e.g. because A2 contains a reference to an element occurring in A1. This is not a functional dependence relation, since it is not due to A2's communicative function.

3.12
functional segment
minimal stretch of communicative behaviour that has one or more **communicative functions** (3.4)

NOTE   The condition of being 'minimal' ensures that functional segments do not include material that does not contribute to the expression of a communicative function that identifies the segment.

EXAMPLE   The functional segment corresponding to the answer given by S in the following dialogue fragment does not include the parts *"Just a moment please"* and   *".... let me see..."* but only the parts "*the first train to the airport on Sunday morning is"* and *"at 5:45*.

1. U: What time is the first train to the airport on Sunday morning please?
2. S: Just a moment please... the first train to the airport on Sunday morning is .... let me see... at 5:45.

A consequence of this definition is that functional segments may be discontinuous, may overlap or be embedded, and may contain parts from more than one turn.

3.13
information state
context
the totality of a **dialogue** (3.5) **participant**'s (3.14) beliefs, assumptions, expectations, goals, preferences, hopes, and other attitudes that may influence the participant's interpretation and generation of communicative behaviour

3.14
participant
person or artificial agent involved in the exchange of **utterance**s (3.23)

3.15
qualifier
predicate that can be associated with a **communicative function** (3.4)

EXAMPLE A: Would you like to have some coffee?
        B: Only if you have it ready.

B's utterance accepts A's offer under a certain condition; this can be described by qualifying the communicative function Accept Offer with the predicate 'conditional'. See Clause 10.3 for more examples.

3.16
reference segment
stretch of communicative behaviour that a **feedback dependence relation** (3.11) refers to and that is not a **functional segment** (3.11)

3.17
rhetorical relation
discourse relation
semantic or pragmatic relation between two **dialogue acts** (3.6) or their **semantic content**s (3.20)

NOTE Relations such as *elaboration, explanation, justification, cause,* and *concession* have been studied extensively in the analysis of (monologue) text, where they are often called 'rhetorical relations' or 'discourse relations', and are mostly viewed either as relations between text segments or as relations between events or propositions, described in text segments. See e.g. Hovy and Maier, 1992; Lascarides & Asher, 2007; Mann & Thompson, 1988. Many of these relations also occur in dialogue, either as relations between dialogue acts or between the semantic contents of dialogue acts.

EXAMPLE In the following example, the statement in the second utterance provides a *motivation* for the question in the first utterance:

A: Can you tell me what flights there are to Sydney on Saturday? I'd like to attend my mother's 80th birthday.

EXAMPLE A rhetorical relation between the semantic contents of two dialogue act occurs in the following, where the content of B's statement mentions a *cause* for the content of A's statement:

A: I can never find these stupid remote controls
B: That's because they don't have a fixed location

3.18
semantic content
information, situation, action, event, or objects that a stretch of communicative behaviour refers to

3.19
semantic content category
semantic content type
kind of information, situation, action, event, or objects that form the **semantic content** (3.20) of a **dialogue act** (3.7)

EXAMPLES The various dimensions (see 3.8) defined in this standard correspond to categories of semantic content. In particular, the Task dimension corresponds to the category of task-specific actions and information; the Allo- and Auto-Feedback dimensions correspond to the categories of information about the processing by the current speaker or by the addressee, respectively, of something that was said before; the Turn Management dimension corresponds to the the category of information about the allocation of the speaker role, and so forth.

3.20
sender
dialogue (3.5) **participant** (3.14) who produces a **dialogue act** (3.7)

3.21
speaker

sender (3.20) of a **dialogue act** (3.6) in the form of speech, possibly combined with nonverbal communicative behaviour

NOTE A dialogue participant may say something while another participant occupies the **speaker role** (3.22), therefore the term 'speaker' is not synonymous with 'participant who occupies speaker role'.

3.22
speaker role

role occupied by a **dialogue** (3.5) **participant** (3.14) who has temporary control of the **dialogue** (3.6) and speaks for some period of time

NOTE Source: DAMSL Revised Manual.

3.23
turn unit

stretch of communicative activity produced by one **participant** (3.14) who occupies the **speaker role** (3.21), bounded by periods where another participant occupies the speaker role

3.23
utterance

anything said, written, keyed, gesticulated, or otherwise expressed

NOTE An utterance is mostly a part of what a sender contributes in a turn unit.

4 Purpose, justification, and use cases

4.1 Purpose and justification

The notion of a dialogue act plays a key role in the analysis of spoken and multimodal dialogue, as well as in the design of spoken dialogue systems and embodied conversational agents. These activities all depend on the availability of dialogue corpora, annotated with dialogue act information.

Over the years a variety of dialogue act annotation schemes have been developed, such as those of the TRAINS human-computer dialogue project, (Allen et al., 1994); of the Map Task studies of human-human dialogue (Anderson et al., 1991); of the Verbmobil speech translation project (Alexandersson et al., 1998), and of the ICSI-MRDA and AMI/AMIDA meeting browsing technology projects (Shriberg et al., 2004; Carletta et al., 2005). These schemes were developed for specific purposes and application domains. They contain overlapping sets of concepts and make use of often mutually inconsistent terminology, sometimes employing different terms for the same concept, or the same term for different concepts.

The multidimensional DIT scheme (Bunt, 1994) was developed for information-seeking dialogues without depending on a particular domain. The DAMSL scheme (Dialogue Act Markup using Several Layers, Allen and Core,1997; Core et al., 1998) constitutes an application-independent multidimensional annotation scheme. The DIT++ scheme (Bunt, 2006; 2009) combines the DIT scheme with concepts from DAMSL and other schemes into a comprehensive general-purpose annotation scheme. The annotation scheme of the NICT Kyoto corpus (Ohtake and Mizukami, 2017) is a variant of the DIT++ scheme.

In the European project LIRICS (Linguistic Infrastructure for Interoperable Resources and Systems, Romary et al., 2007) a set of dialogue acts defined in the DIT++ taxonomy was selected and redefined in the form of data categories, following ISO standard 12620 for concept definitions. This set of concepts has been tested for its usability and coverage (a) in the manual annotation of spoken dialogues in English, Dutch and Italian, and (b) in the automatic annotation of spoken and multimodal dialogue in English, and forms a significant part of the background of this standard.

The main purpose of this standard is to define a reference set of domain-independent basic concepts for dialogue act annotation, plus a formal language for representing such annotations. This formal language, the Dialogue Act Markup Language (DiAML) has a formal semantics, which makes it possible to apply techniques for automatic reasoning to DiAML annotations.

4.2   Use cases

The set of concepts defined in ISO 24617-2:2012 is based on the DIT++ taxonomy, which was originally developed to serve a double purpose: on the one hand for the articulate functional description of communicative activity in natural human dialogue, and on the other hand to provide a basis for the design of dialogue management modules in interactive systems. As part of the ISO Semantic Annotation Framework (SemAF), , the a strong focus in ISO 24617-2:2012 came to lie on its use for annotation. Still, the concepts of ISO 4617-2, like those of DIT++, have multiple use cases, which can be grouped into four types:

- UC1: Manual annotation of spoken, written, or multimodal human-human or human-computer dialogue.
- UC2: Automatic annotation of spoken, written, or multimodal human-human or human-computer dialogue starting from transcriptions or recordings of raw spoken or multimodal communicative user behaviour.
- UC3: Recognition of dialogue acts in spoken, written, nonverbal, or multimodal communicative user behaviour in human-computer interaction.
- UC4: Generation of dialogue acts by the dialogue manager component of a dialogue system.

The different use cases bring different requirements and desiderata:

- UC1: Manual dialogue act annotation has the advantage of producing annotations of the highest quality if performed by experts, but has the drawback of being very costly and only feasible for limited amounts of data. Expert manual annotation delivers the highest quality of annotations since expert human annotators are not only skilled in recognizing the relevant features of communicative behaviour, but also have a wealth of context information, general world knowledge, and common-sense reasoning abilities to infer speaker beliefs and intentions. Expert annotators are therefore able to assign fine-grained characterizations to segments of dialogue behaviour with high accuracy. In order to support manual annotation, the annotation scheme should therefore include fine-grained concepts with the level of detail that expert annotators can use.
- UC2: Automatic annotation of human-human dialogue, or of the user's contributions in a human-computer dialogue, typically lacks the general world knowledge and the skills of expert human annotators, and typically has access to context information only as far as represented in the dialogue history. Automatic annotation therefore in general cannot reliably characterize dialogue behaviour with the same level of detail as expert human annotation. To effectively support automatic annotation, the annotation scheme should therefore contain concepts that are more coarse-grained than those needed for manual annotation.
- UC3: The automatic recognition of dialogue acts in user behaviour in an interactive system is a very similar task as automatic dialogue act annotation, except that in an interactive system the semantic contents of dialogue acts play a prominent role, often determined by structural properties of the application domain. For a given application, it may be beneficial to have a tight coupling between communicative functions and semantic content, and to define application-specific functions for specific types of content. For effectively supporting this use case, it may be beneficial to extend the (application-independent) annotation scheme with application-specific concepts.
- UC4: The generation of dialogue acts in an interactive system concerns the decision how to continue a dialogue when it is the system's turn, and this is the main task of the system's dialogue manager component. This is typically a two-stage process, where the first stage is to decide on the communicative functions and semantic contents of one or more possible dialogue acts, and the second is to decide on an appropriate realization in linguistic, nonverbal, or multimodal form. In contrast with human dialogue participants, who may be somewhat vague or unspecific about their beliefs and intentions, a system's dialogue manager typically works with precise beliefs and goals, and generates, in the first of these two stages, dialogue acts with fine-grained communicative functions, possibly even more fine-grained than human participants commonly use. This happens for instance for feedback acts, since the system may report a processing problem with great accuracy. This calls for the annotation scheme to include very fine-grained functions, however not more fine-grained than humans can understand.

ISO 24617-2 was originally designed with the main purpose of supporting interoperable dialogue act annotation, i.e., the use cases UC1 and UC2. The use cases UC3 and UC4 have been found to be potentially of equally great interest, however. Some of the modifications described in this second edition are aimed at providing effective support for these use cases.

5 Modifications compared to First Edition (ISO 24617-2:2012)

5.1 Overview

The first edition of ISO 24617-2 was published in 2012. It has been applied in annotation efforts, including the development of the DialogBank (Bunt et al., 2019), and its concepts have been used for language understanding, dialogue management, and output generation in spoken and multimodal interactive systems.[2] These applications have brought to light certain inaccuracies and limitations of the standard which are addressed in this second edition.

The development of this standard has also contributed to the specification of a framework for defining other standards for semantic annotation. This framework was established as an ISO standard in 2016 (ISO 24617-6, Principles of semantic annotation). The second edition of ISO 24617-2 follows the requirements and recommendations of this framework in improving its specification.

5.2 Downward compatibility and optional elements

An important issue in revising an annotation scheme concerns the compatibility between annotations according to the original and the revised version. Unless the revision corrects serious errors in the original scheme, it is in general desirable that 'old' annotations are still valid according to the revised version, and do not require to be re-annotated (or converted). In other words, the revised standard should preferably be *'downward compatible'* with the original version.

Designing a revised version of the ISO 24617-2 standard in a downward compatible way is greatly facilitated by the *extensibility* of the original version, which means that it allows its stock of concepts to be extended with additional concepts. ISO 24617-2 is extensible in four respects:

1. Dimensions: Due to the orthogonality of the set of dimensions, additional dimensions may be introduced as long as they are orthogonal to the already existing dimensions and to each other.
2. Communicative functions: The taxonomy of communicative functions expresses semantic relations between functions: dominance relations express different degrees of specialization; and sister relations express mutually exclusivity of functions. Communicative functions may be added to the taxonomy as long as they respect these relations.
3. Qualifiers: Like dimensions, due to the orthogonality of the qualifier attributes and their values.
4. Rhetorical relations: The ISO standard does not specify a particular set of relations, but allows any such set to be plugged in.

The extensibility of ISO 24617-2 is in turn facilitated by the optionality of some of its components. Following the ISO Principles of semantic annotation, three types of optionality can be distinguished:[3]

- Type I, semantic optionality: a component that a certain type of annotation structure may contain, but does not have to. If it does, then this provides extra information. Example: the specification of a set of `other participants' for a dialogue act.
- Type II, syntactic optionality: a component that may be but does not need to be specified in annotation representations, since it has a default value in the abstract syntax. Example: the polarity in the annotation of an event by means of an <event> element in ISO-TimeML (default "positive").
- Type III, uninterpreted optionality: a component that may be specified in annotation representations but that does not encode anything in the abstract syntax. It thus has no semantic interpretation, but it may be useful during an annotation process, or for other purposes. Example: the indication of the part of speech of an event description in ISO-TimeML.

The document specifying ISO 24617-2:2012 mentions in a number of places an element of the standard as being 'optional', without making clear in what sense the optionality should be taken. In this second edition, the mentioning of 'optional' elements is made unambiguous in this respect.

---

2 For applications in interactive systems see Malchanau (2019) and Malchanau et al. (2019).

3 See Bunt et al. (2018) for details.