

---

---

**Language resource management —  
Feature structures —**

Part 1:  
**Feature structure representation**

*Gestion des ressources linguistiques — Structures de traits —  
Partie 1: Représentation de structures de traits*



**PDF disclaimer**

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

This document is a preview generated by EVS

© ISO 2006

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing 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](mailto:copyright@iso.org)  
Web [www.iso.org](http://www.iso.org)

Published in Switzerland

# Contents

Page

Foreword.....	v
Introduction.....	vi
1 Scope.....	1
2 Normative references.....	1
3 Terms and definitions.....	1
4 General characteristics of feature structure.....	4
4.1 Overview.....	4
4.2 Use of feature structures.....	4
4.3 Basic concepts.....	5
4.4 Notations.....	5
4.4.1 Overview.....	5
4.4.2 Graph notation.....	6
4.4.3 Matrix notation.....	7
4.4.4 XML-based notation.....	8
4.5 Structure sharing.....	10
4.6 Collections as complex feature values.....	12
4.6.1 Overview.....	12
4.6.2 Lists as feature values.....	12
4.6.3 Sets as feature values.....	14
4.6.4 Multisets as feature values.....	15
4.7 Typed feature structure.....	16
4.7.1 Overview.....	16
4.7.2 Types.....	16
4.7.3 Notations.....	16
4.8 Subsumption: relation on feature structures.....	18
4.8.1 Overview.....	18
4.8.2 Definition.....	18
4.8.3 Condition A on path values.....	19
4.8.4 Condition B on structure sharing.....	19
4.8.5 Condition C on type ordering.....	20
4.9 Operations on feature structures and feature values.....	21
4.9.1 Overview.....	21
4.9.2 Compatibility.....	21
4.9.3 Unification.....	22
4.9.4 Unification of shared structures.....	22
4.10 Operations on feature values and types.....	23
4.10.1 Concatenation and union operations.....	23
4.10.2 Alternation.....	24
4.10.3 Negation.....	25
4.11 Informal semantics of feature structures.....	27
5 XML Representation of feature structures.....	29
5.1 Overview.....	29
5.2 Organization.....	29
5.3 Elementary feature structures and the binary feature value.....	30
5.4 Other atomic feature values.....	32
5.5 Feature and feature-value libraries.....	35
5.6 Feature structures as complex feature values.....	37
5.7 Re-entrant feature structures.....	40
5.8 Collections as complex feature values.....	41

5.9	Feature value expressions .....	44
5.9.1	Overview .....	44
5.9.2	Alternation .....	44
5.9.3	Negation .....	47
5.9.4	Collection of values .....	48
5.10	Default values .....	48
5.11	Linking text and analysis .....	50
<b>Annex A</b>	<b>(informative) Formal definitions and implementation of the XML representation of feature structures .....</b>	<b>54</b>
A.1	Overview .....	54
A.2	RELAX NG specification for the module .....	54
<b>Annex B</b>	<b>(informative) Examples for illustration .....</b>	<b>60</b>
<b>Annex C</b>	<b>(informative) Type inheritance hierarchies .....</b>	<b>62</b>
C.1	Overview .....	62
C.2	Definition .....	62
C.3	Multiple inheritance .....	64
C.4	Type constraints .....	64
<b>Annex D</b>	<b>(informative) Denotational semantics of feature structure .....</b>	<b>66</b>
D.1	Feature structure signatures .....	66
D.2	Feature structure algebra .....	66
D.3	FS domains .....	67
D.4	Feature structure interpretations .....	68
D.5	Satisfiability .....	68
D.6	Subsumption .....	68
D.7	Unification .....	69
<b>Annex E</b>	<b>(informative) Use of feature structures in applications .....</b>	<b>70</b>
E.1	Overview .....	70
E.2	Phonological representation .....	70
E.3	Grammar formalisms or theories .....	70
E.4	Computational implementations .....	71
<b>Bibliography</b>	.....	<b>75</b>

This document is a preview generated by EVS

## 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 24610-1 was prepared by Technical Committee ISO/TC 37, *Terminology and other language and content resources*, Subcommittee SC 4, *Language resource management*.

ISO 24610 consists of the following parts, under the general title *Language resource management — Feature structures*:

— *Part 1: Feature structure representation*

The following part is under preparation:

— *Part 2: Feature system declaration*

## Introduction

This part of ISO 24610 results from the agreement between the Text Encoding Initiative Consortium (TEI) and the ISO TC 37/SC 4 that a joint activity should take place to revise the two existing chapters on feature structures and feature system declaration in *The TEI Guidelines* called *P4*.

It is foreseen that ISO 24610 will have the following two parts.

- Part 1, *Feature structure representation*, describes feature structures and their representation. It provides an informal but explicit overview of their basic characteristics and formal semantics. In addition, part 1 defines a standard XML (eXtended Markup Language) vocabulary for the representation of untyped feature structures, feature values, and feature libraries. It thus provides a reference format for the exchange of feature structure representations between different application systems.
- Part 2, *Feature system declaration*, discusses ways of validating typed feature structures which are conformant to part 1, and of enforcing application-specific constraints. It proposes an XML vocabulary for the representation of such constraints with reference to a set of features and the range of values appropriate for them, and thus facilitates representation and validation of a type hierarchy as well as other well-formedness conditions for particular applications, in particular those related to the goal of language resource management.

This document is a preview generated by EVS

# Language resource management — Feature structures —

## Part 1: Feature structure representation

### 1 Scope

Feature structures are an essential part of many linguistic formalisms as well as an underlying mechanism for representing the information consumed or produced by and for language engineering applications. This part of ISO 24610 provides a format for the representation, storage and exchange of feature structures in natural language applications concerned with the annotation, production or analysis of linguistic data. It also defines a computer format for the description of constraints that bear on a set of features, feature values, feature specifications and operations on feature structures, thus offering a means of checking the conformance of each feature structure with regards to a reference specification.

### 2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 8879, *Information processing — Text and office systems — Standard Generalized Markup Language (SGML)*, as extended by TC 2 (ISO/IEC JTC 1/SC 34 N029:1998-12-06).

ISO 19757-2, *Information technology — Document Schema Definition Language (DSDL) — Part 2: Regular-grammar-based validation — RELAX NG*

NOTE The first reference permits the use of XML and the second, RELAX NG, provides a specification for XML modules. RELAX NG is a schema language for XML, standing for REGular Language for XML for Next Generation, and simplifies and extends the features of DTDs, Document Type Definitions.

### 3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 8879 and ISO 19757-2 and the following apply. This list is provided to clarify the terminology relating to feature structures used throughout this part of ISO 24610. Terminology derived from XLM and other formal languages is not defined here.

#### 3.1 alternation

operation on feature **values** (3.23) that returns one and only one of the values supplied as its argument

NOTE Given a feature specification  $F : a|b$ , where  $a|b$  denotes the alternation of  $a$  and  $b$ ,  $F$  has either the value  $a$  or the value  $b$ , but not both.

#### 3.2 atomic value

**value** (3.23) without internal structure, i.e. value other than **feature structure** (3.10) and **collection** (3.4)