ISO 24611-1:2025-11 (E)

Language resource management - Morphosyntactic annotation framework (MAF) - Part 1: Core model

Contents

Page

Foreword ... **iv**

Introduction ... **v**

1 Scope ... **1**

2 Normative references ... **1**

3 Terms and definitions ... **1**

4 MAF metamodel ... **6**
4.1 Levels of description in the MAF metamodel ... 6
4.2 MAF in the standards landscape ... 6
4.3 Metadata ... 7
4.4 Structural ambiguities ... 7
4.5 MAF metamodel in detail ... 8

5 Token-level segmentation ... **10**
5.1 General remarks ... 10
5.2 Formal description: <seg> ... 10
5.3 Embedding notation ... 11
5.4 Stand-off notation ... 12
5.5 Normalization and script conversion ... 14
5.6 Inline token annotation strategies for token separation ... 15
5.6.1 General remarks ... 15
5.6.2 Adjacent tokens in embedded mode ... 15
5.6.3 Overlapping tokens ... 15

6 Word-forms as linguistic units ... **16**
6.1 General remarks ... 16
6.2 Formal description: ... 17
6.3 Token attachment ... 17
6.3.1 One token: one word-form ... 17
6.3.2 Several contiguous tokens: one word-form ... 18
6.3.3 Several discontinuous tokens: one word-form ... 18
6.3.4 Zero token: one word-form ... 19
6.3.5 One token: several word-forms ... 19
6.4 Referencing lexical entries ... 19
6.5 Compound word-forms ... 20
6.6 Identification of word-forms ... 21

7 Morphosyntactic content ... **21**
7.1 General remarks ... 21
7.2 Using feature structures ... 21
7.3 Compact morphosyntactic tags ... 22
7.4 FSR libraries ... 22
7.5 Designing morphosyntactic tagsets ... 23

8 Handling ambiguities ... **24**
8.1 General ... 24
8.2 Word-form content ambiguities ... 24
8.3 Lexical and structural ambiguities ... 25

9 Conformance ... **25**

Annex A (informative) **Examples** **27**
Annex B (informative) **Referencing externally defined data categories** **31**
Bibliography **34**