

Presentation/ representation of entries in dictionaries — Requirements, recommendations and information

ICS 01.020; 01.080.99

National foreword

This British Standard was published by BSI. It is the UK implementation of ISO 1951:2007.

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.

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 31 May 2007

© BSI 2007

ISBN 978 0 580 50745 8

Amendments issued since publication

Amd. No.	Date	Comments

INTERNATIONAL STANDARD

**ISO
1951**

Third edition
2007-02-15

Presentation/representation of entries in dictionaries — Requirements, recommendations and information

*Présentation/représentation des entrées dans les dictionnaires —
Exigences, recommandations et information*



Reference number
ISO 1951:2007(E)

AVA Documents
Document Sharing Hub

Contents

Page

Foreword	iv
Introduction.....	v
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 Formal description of dictionary entries.....	3
4.1 An overview of data elements and compositional elements	3
4.1.1 Data elements	4
4.1.2 Hierarchical structures: dictionary and entries	8
4.1.3 Compositional elements	9
4.2 Formal structure of a dictionary entry	13
4.2.1 Formal grammar for high-level structures.....	13
4.2.2 Formal grammar for other lexical unit containers	16
4.2.3 Formal grammar for blocks	17
4.2.4 Formal grammar for other comment containers	17
4.2.5 Formal grammar for lexical units.....	18
4.2.6 Formal grammar for comments	19
4.3 Content models	20
4.3.1 Content elements	20
4.3.2 Embedded elements.....	21
4.3.3 Basic elements	22
4.3.4 Pointer	23
4.3.5 Formal grammar for embedded containers.....	23
4.3.6 Formal grammar for data category content.....	24
4.4 General qualifiers	25
5 Means of presentation	25
5.1 Layout aids.....	25
5.1.1 General	25
5.1.2 Layout aids for dictionaries	25
5.2 Compacting mechanisms	26
5.2.1 General abbreviations.....	26
5.2.2 Abbreviated headword repetitions	26
5.2.3 Repeat symbols (Tilde or Dash).....	26
5.2.4 Nesting.....	27
Annex A (informative) Arabic, Roman and Hellenic numbering system	28
Annex B (informative) Tables of functions of lexicographical symbols	30
Annex C (informative) Examples of XML encoding	45
Annex D (informative) Assigning layout aids to dictionary components and compacting dictionary entries.....	68
Bibliography.....	76

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 1951 was prepared by Technical Committee ISO/TC 37, *Terminology and other language and content resources*, Subcommittee SC 2, *Terminographical and lexicographical working methods*.

This third edition cancels and replaces the second edition (ISO 1951:1997), the scope of which has been extended in order to address publishers' and users' needs by taking into account various types of electronic dictionaries and the constraints of single sourcing for producing dictionaries, as well as disseminating and reusing data in lexicographical practice.

Real dictionary entries used as examples in this International Standard only illustrate the principles of XML representation of lexicographical data and their associated presentations. They do not engage the publishers' responsibility.

Introduction

General aim of this International Standard

During the past decade, dictionary-making processes have undergone important changes due to the spread of electronic dictionaries. Consequently, lexicographers are faced with a growing diversification of methods during dictionary preparation and publishing.

This revised International Standard aims to support the creation and management of various types of dictionaries. It takes into account different ways of using dictionaries, especially such new functionalities of electronic documents as hyperlinks.

To allow dictionary content to be reused in different printed and electronic formats, lexicographers increasingly tend to create a single well-structured lexicographical source or data repository. In addition to reproducing all the typographical conventions described in the former edition of ISO 1951, this revised International Standard provides a specific model based on current best professional practices, in order to allow necessary production, exchange and management procedures.

In the text of this International Standard, the use of the auxiliary verb “shall” indicates a requirement or specification that is to be met precisely as stated; the use of the auxiliary verb “should” indicates a recommendation of a good way to do something that is to be followed unless a better way can be demonstrated to have been adopted; and the use of the auxiliary verb “can” indicates information that the user may find useful.

blank

Presentation/representation of entries in dictionaries — Requirements, recommendations and information

1 Scope

This International Standard deals with monolingual and multilingual, general and specialized dictionaries. It specifies a formal generic structure independent of the publishing media and it proposes means of presenting entries in print and electronic dictionaries. The relationship between the formal structure and the presentation of entries used by publishers and read by users is explained in examples provided in the informative annexes.

The objective of this International Standard is to facilitate the production, merging, comparison, extraction, exchange, dissemination and retrieval of lexicographical data in dictionaries. Following a lexicographical lemma-oriented approach, it does not deal with concept-oriented works as defined in ISO 704.

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 704:2000, *Terminology work — Principles and methods*

ISO 1087-1:2000, *Terminology work — Vocabulary — Part 1: Theory and application*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 704 and ISO 1087-1 and the following apply.

These definitions concern basic and unambiguous terms of dictionary structure and presentation, common in most types of dictionaries, which are within the scope of this International Standard. Terms considered specific to certain dictionaries have not been included here.

3.1

comment

metalinguistic information describing a **lexical unit** (3.8) by means of lexicographical **data elements** (3.3) or **compositional elements** (3.2)

3.2

compositional element

composite information unit made of elements

NOTE There are three families of compositional elements: **blocks** (3.2.1), **containers** (3.2.2) and **groups** (3.2.3).

3.2.1

block

factorizing structure

compositional element (3.2) used to factorize **elements** (3.5) that are shared as refiners by many instances of a specific element

NOTE Examples of blocks are provided in Tables 6 to 14.

3.2.2

container

refining structure

compositional element (3.2) used to supply additional information about one single specific **data element** (3.3) by the mean of other **elements** (3.5)

EXAMPLE A headword container is used for giving the pronunciation or the part of speech which refines a **headword** (3.6) which is itself the refined data element.

NOTE 1 Adapted from ISO 16642:2003, C.4.5.

NOTE 2 An example of a container is provided in Tables 4 and 5.

3.2.3

group

compositional element (3.2) used to aggregate several independent **elements** (3.5)

EXAMPLE A sense is described by a **group** of **elements** such as definition, subject field, etc.

NOTE An example of a group is provided in Tables 15 and 16.

3.3

data element

data category

unit of data for which the definition, identification, representation, and permissible values are specified by means of a set of attributes

[ISO/IEC 11179-1:2004, definition 3.3.8]

NOTE Lists of data elements are provided in Tables 1 and 2.

3.4

dictionary entry

lexicographical entry

entry

part of a dictionary which contains information related to one **lemma** (3.7) and its variants

3.5

element

any **data element** (3.3) or **compositional element** (3.2)

3.6

headword

entry word

lemma (3.7) that serves as the heading for an entry in a dictionary

3.7

lemma

base word

lexical unit (3.8) chosen according to lexicographical conventions to represent the different forms of an inflectional paradigm

EXAMPLE "Sell" is the lemma of the paradigm "sells, sold, selling, etc."

3.8**lexical unit**

unit of language, belonging to the lexicon of a given language and which is described or mentioned in a dictionary

3.9**lexicographical symbol**

letter, punctuation mark, other typographical or graphic symbol or group of symbols or any combination thereof used to represent certain lexicographical or terminological data as displayed or output either singularly or in conjunction with another item of lexicographical data

3.10**nested entry**

grouping structure for related dictionary entries that share a common headword

4 Formal description of dictionary entries

For the sake of clarity, the following formal model, thereafter called XmLex, is illustrated by short examples encoded according to an XML Definition Type Document called XmLex_V00 (for more information, see informative Annex C).

4.1 An overview of data elements and compositional elements

Dictionary entries can be seen as comments about topics, that are lexical units. An entry has a main topic (the headword); other topics (e.g. variants, translations) are said to be “related topics”. Topics and comments are data elements. Each data element has a content model. Data elements are grouped into compositional elements in order to produce an unambiguous and fully computable entry. Open lists of data elements and compositional elements are provided herein, and are extendable by the user for specific purposes.

Printed dictionaries generally use typographical conventions (normal/bold/italic), spatial disposition (before/after) and punctuation (comma or semicolon) to indicate relations between topics and comments. In the XmLex model,

- the position of an element is never used for expressing relations between two elements,
- there are no markers equivalent to typographical signs (comma or semicolon).

Compositional elements (containers, blocks and groups) are used to encode logical relations between comments and topics so that it is always possible, on one hand, to generate automatically any printed presentation and, on the other hand, to compute automatically all the relations between elements when transforming data (for inverting a bilingual dictionary for instance) or when reusing data in other contexts like translation memory systems or lexical databases.

This part of this International Standard

- describes data elements and their grouping in compositional elements needed for representing most common dictionary entries;¹⁾
- specifies a formal dictionary model expressed in Extended Backus-Naur form which is often used as a formal notation to describe the syntax of a given language;
- gives in annexes examples of implementation and means of validation using XML, Xpointer, XSL and XHTML specifications.

1) For each data element, a conventional name and description is provided, based as far as possible on ISO 12620:1999. Free data elements, the type of which is definable by the user, allow extensions of the model for “negotiated interchange”.

4.1.1 Data elements

4.1.1.1 Lexical units

The following table gives the list of lexical units and comments that should be used in a standardized dictionary entry.

The first column contains a designation of the data element. The second gives its generic identifier as used in the formal model. The third column gives a short explanation, and the fourth refers to the first example (if available) of the data element in the annexes (the first number points to the example, the second indicates the line).

Table 1 — List of lexical units

Name	Generic identifier	Explanation	See: Annex Example Line
abbreviated form	AbbreviatedForm	Lexical unit formed by omitting words or letters from a longer form [...]. [Adapted from ISO 1087-1:2000, definition 3.4.9]	C 7 3
analogy	Analogy	Lexical unit having some similarity of meaning with the current lexical unit. [Adapted from ISO 1087-1:2000, definition 3.4.9]	C 10 4
antonym	Antonym	Lexical unit for which the concept constitutes the opposite of the concept represented by the current lexical unit. [Adapted from ISO 12620:1999, A.10.18.6]	C 5 45
compositional phrase	CompositionalPhrase	Any recurrent and conventional juxtaposition of words such as collocation, proverb, saying, etc.	C 4 15
derivation	Derivation	A change in the form of a lexical unit, usually modification in the base/root or affixation which signals a change in part-of-speech information.	C 5 9
example	Example	An instance that is typical of a lexical unit's usage in a specific sense.	C 4 10
false friend	FalseFriend	A lexical unit in one language that only appears to have formal or semantic similarity with a lexical unit in another language, but that does not represent the same concept. [Adapted from ISO 12620:1999, A.3.2]	C 15 6
free topic	FreeTopic	Lexical unit whose type is not defined in this International Standard.	C 18 5
full form	FullForm	The complete representation of a lexical unit for which there is an abbreviated form. [Adapted from ISO 12620:1992, A.2.1.7]	C 20 3
headword	Headword	Lemma heading a dictionary entry.	C 1 4

Table 1 — List of lexical units *(continued)*

Name	Generic identifier	Explanation	See: Annex Example Line
inflection	Inflection	The modification of the form of a word to express the different grammatical relationships into which it may enter.	C 22 5
international scientific term	InternationalScientificTerm	A term that is part of an international scientific nomenclature as adopted by an appropriate scientific body. [Adapted from ISO 12620:1999, A.2.1.4]	C 39 12
multiword unit	MultiWordUnit	A lexical unit made of more than one word and conveying only one sense.	C 1 42
symbol	Symbol	A designation of a concept by letters, numerals, pictograms or any combination thereof. [Adapted from ISO 12620:1999, A.2.1.13]	C 23 7
synonym	Synonym	A lexical unit that represents the same or a very similar concept as the headword in a dictionary entry. [Adapted from ISO 12620:1999, A.2.1.2]	C 5 27
translation	Translation	An equivalent lexical unit belonging to a target language.	C 1 15
variant	Variant	One of the alternative forms of a lexical unit. [Adapted from ISO 12620:1999, A.2.1.9]	C 32 5

Table 2 — List of comments

Name	Generic identifier	Explanation	See: Annex Example Line
attestation	Attestation	Date or period when a lexical unit has been observed.	C 11 8
case	Case	The form of a lexical unit (noun, pronoun, or modifier) that indicates its grammatical relationship to other words in a clause or sentence.	C 24 7
citation	Citation	Quotation from a book, article or document.	C 4 25
complement	Complement	Ancillary part of a lexical unit (the "to" preposition for an English verb for instance). [Adapted from ISO 16642:2003, C.4.10.2]	

Table 2 — List of comments (continued)

Name	Generic identifier	Explanation	See: Annex Example Line
definition	Definition	A statement that describes a concept and permits its differentiation from other concepts within a system of concepts. [Adapted from ISO 12620:1999, A.5.1]	C 4 61
display	Display	Synthetic text that can be shown instead of separate topics or comments.	C 2 33
etymology	Etymology	Information on the origin of a word and the development of its meaning. [Adapted from ISO 12620:1999, A.2.4.2]	C 4 6
formula	Formula	Figures, symbols or the like used to express a concept briefly, such as a mathematical or chemical formula. [Adapted from ISO 12620:1999, A.2.1.14]	C 16 3
frequency	Frequency	The relative commonness with which a lexical unit occurs. [Adapted from ISO 12620:1999, A.2.3.4]	C 19 3
free comment	FreeComment	A free metalinguistic used for describing a lexical unit.	C 17 1
geographical usage	GeographicalUsage	Lexical unit usage reflecting regional differences. [Adapted from ISO 12620:1999, A.2.3.2]	C 21 10
grammatical gender	GrammaticalGender	A set of two or more grammatical categories into which the nouns of certain languages are divided.	C 1 15
grammatical number	GrammaticalNumber	In many languages, the grammatical distinction that indicates the number of objects referred to by the lexical unit. [Adapted from ISO 12620:1999, A.2.2.3]	C 2 48
grammatical pattern	GrammaticalPattern	Grammatical structure in which the linguistic unit frequently occurs.	C 34 5
guide phrase	GuidePhrase	A phrase used to indicate a specific application of a word or sense.	C 2 50
insert	Insert	A text, table or picture describing some grammatical, encyclopaedic, scientific or cultural knowledge related to the dictionary entry or to several dictionary entries. This insert can be totally independent from the dictionary text.	
mood	Mood	A property of verbs that indicates the attitude of the speaker about the factuality or likelihood of what is expressed.	C 25 6

Table 2 — List of comments (*continued*)

Name	Generic identifier	Explanation	See: Annex Example Line
normative status	NormativeStatus	A term status qualifier assigned by an authoritative body, such as a standards body or a governmental entity with a regulatory function. [Adapted from ISO 12620:1999, A.2.9.1]	C 26 4
note	Note	Supplemental information pertaining to any other element in the data collection. [Adapted from ISO 12620:1999, A.8]	C 3 24
part of speech	PartOfSpeech	A category assigned to a lexical unit based on its grammatical and semantic properties. [Adapted from ISO 12620:1999, A.2.2.1]	C 1 11
person	Person	An indication of the grammatical person (1st, 2nd, 3rd, etc.) associated with a given inflected lexical unit.	C 25 7
pronunciation	Pronunciation	The representation of the manner by which a lexical unit is articulated. [Adapted from ISO 12620:1999, A.2.5] It can be represented either phonetically or phonologically.	C 11 3
range of application	RangeOfApplication	Scope within which a sense is true.	C 1 14
register	Register	Classification indicating the relative level of language individually assigned to a lexical unit. [Adapted from ISO 12620:1999, A.2.3.3]	C 24 9
search form	SearchForm	A lexical unit entered in a lexicographical entry for purposes of retrieval. [Adapted from ISO 12620:1999, A.10.6.3]	C 1 16
see	See	A cross-reference to a headword which is a synonym of the current headword.	C 3 5
see also	SeeAlso	A cross-reference to a related headword.	C 6 19
sense qualifier	SenseQualifier	Any indication about a sense (figurative, literary, old...).	C 1 45
sort key	SortKey	A lexical unit entered in a lexicographical entry for purposes of sorting when the order of entries does not follow the character-set collating sequence.	
source language	SourceLanguage	The language of a lexical unit that is to be translated into another language.	C 1 1

Table 2 — List of comments (continued)

Name	Generic identifier	Explanation	See: Annex Example Line
subcategorisation	Subcategorisation	The assignment of a lexical item to a subclass of its part of speech, especially with respect to the syntactic elements with which it can combine. NOTE This element only appears in grammatical containers.	C 21 4
subjectfield	SubjectField	An area of human knowledge. [Adapted from ISO 12620:1999, A.4]	C 2 11
syllabification	Syllabification	The division of a word reflecting its articulation by syllables, i.e., by uninterrupted units of pronunciation. [Adapted from ISO 12620:1999, A.2.6]	C 34 4
target language	TargetLanguage	The language into which a lexical unit is to be translated.	C 1 1
tense	Tense	A distinction of form in a verb to express distinctions of time or duration of the action or state it denotes.	C 31 5

4.1.2 Hierarchical structures: dictionary and entries

A dictionary is made of entries or nested entries.

Table 3 — List of high-level structures

Name	Generic identifier	Explanation	See: Annex Example Line
dictionary	Dictionary	A collection of dictionary entries or nested entries.	C 1 1
dictionary entry	DictionaryEntry	See definition: 3.4.	C 1 2
nested entry	NestEntry	See definition: 3.10.	D 1 8

An entry in a dictionary is made of data elements which are self-contained or combined within compositional elements.

4.1.3 Compositional elements

4.1.3.1 Containers

As far as possible, the XML encoding of an example is preceded by its printed view coming from a real dictionary. In this case, the whole entry is a boxed piece of text with a grey background on paper (green on a screen). If parts only of the entry are encoded, these parts [here “Farad n (F) DIN 1301”] appear on a white background on paper (yellow on a screen).

A container as defined in ISO 16642:2003, Annex C.4.5, is a structure used whenever a data element has to be refined by other data elements (a headword by its part of speech, a quotation by its author, a symbol by its source, etc.). Example:

Table 4 — Original data (from Annex C — Example 23)

F <phys> (unit of capacity : As/V) • Farad n (F) DIN 1301

In this English-German dictionary, although “Farad”, “n”, “F” and “DIN 1301” are printed in linear order, there are relationships (dependencies) between these elements:

“Farad” part of speech is “n”

“Farad” symbol is “F”

“F” source is “DIN 1301”

Table 5 — Encoding (from Annex C — Example 23)

```
1. < TranslationCtn>
2. <Translation>Farad</Translation>
3. <PartOfSpeech value = 'noun'/>
4. <SymbolCtn>
5.   <Symbol>F</Symbol>
6.   <Source>DIN 1301</Source>
7. </SymbolCtn>
8. < TranslationCtn>
```

A translation container (<TranslationCtn>) is used to refine the translation (Farad) by its part of speech (<PartOfSpeech>). A symbol container (<SymbolCtn>) is used to refine the symbol (F) by its source (DIN 1301). This symbol container is embedded within the translation container in order to refine the translation by its symbol.

4.1.3.2 Blocks

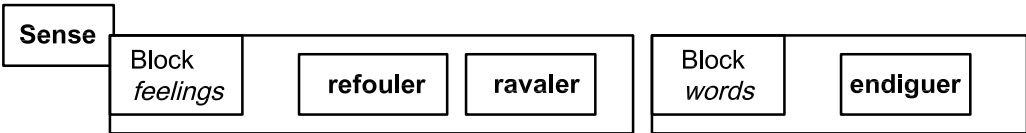
Printed dictionaries often use punctuation (comma or semicolon) to indicate relations between elements of an entry. For instance, in the following example (Table 6) “feelings”, between square brackets and before two translations which are separated by a comma, is a “range of application” that applies to these two translations. The semicolon closes the list of possible translations of “dam” when speaking of “feelings”. “Words” between square brackets opens a new “range of application”.

Blocks are used to encode this kind of logical relation.

Table 6 — Original data — Block with core components (from Annex C — Example 1)

dam ...
figurative [feelings] refouler, ravalér;
[words] endiguer

Table 7 — Schematic representation — Block with core components



There are three <Translation> sister nodes. A TranslationBlock is used to group two of them who share the same <RangeOfApplication>. Schematic representation:

Table 8 — Block encoding — Block with core components (from Annex C — Example 1)

1. <SenseGroup>
2. <SenseQualifier>figurative</SenseQualifier>
3. <TranslationBlock>
4. <RangeOfApplication>feeling</RangeOfApplication>
5. <Translation>refouler</Translation>
6. <Translation>ravalér</Translation>
7. </TranslationBlock>
8. <TranslationBlock>
9. <RangeOfApplication>words</RangeOfApplication>
10. <Translation>endiguer</Translation>
11. </TranslationBlock>
12. </SenseGroup>

Table 9 — Original data — Block with containers

cleave [kliv:] v (pt cleaved , cleft [kleft], Lit clove [kl.v]; pp cleaved, cleft, Lit cloven ['kl.v(.)n])
--

There are three derived forms for the preterit and three for the past participle of “Cleave”. Each derived form can be refined by its pronunciation or its register.

Table 10 — Schematic representation — Block with containers

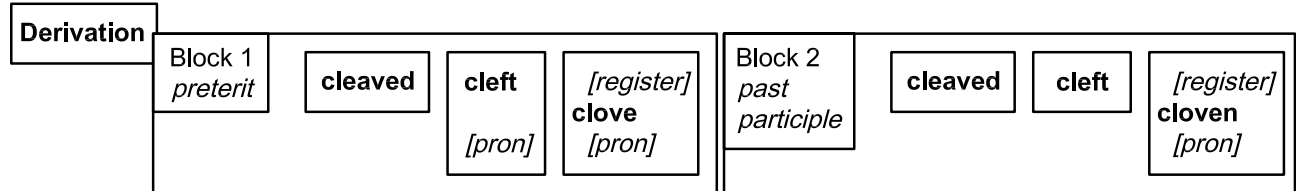


Table 11 — Block encoding — Block with containers

```

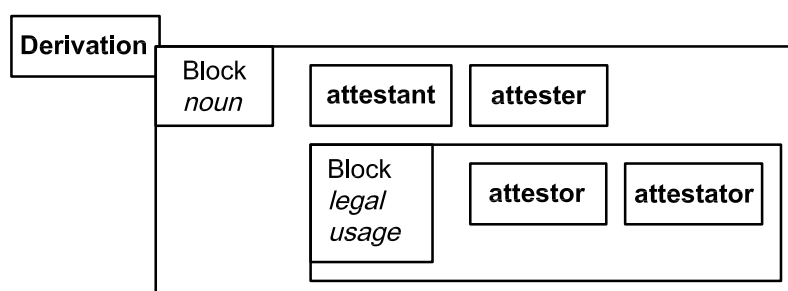
1.<DerivationBlock>
2.  <Tense value = 'preterit'>
3.  <Derivation>cleaved</Derivation>
4.  <DerivationCtn>
5.    <Derivation>cleft</Derivation>
6.    <Pronunciation>kleft</Pronunciation>
7.  </DerivationCtn>
8.  <DerivationCtn>
9.    <Derivation>clove</Derivation>
10.   <Pronunciation>[kl..v]</Pronunciation>
11.   <Register value = 'literary'>
12. </DerivationCtn>
13.</DerivationBlock>
14.<DerivationBlock>
15.  <Tense value = 'pastParticiple'>
16.  <Derivation>cleaved</Derivation>
17.  <Derivation>cleft</Derivation>
18.  <DerivationCtn>
19.    <Derivation>cloven</Derivation>
20.    <Pronunciation>'kl..v(.)n</Pronunciation>
21.    <Register value = 'literary'>
22.  </DerivationCtn>
23.</DerivationBlock>

```

Table 12 — Original data — Derivation block

— at^Éestable *adj.*
— at^testant, at^tester or *esp. in legal usage* at^testor, at^testator *n.*
— attestation (Çæt^ÉestelSⁿ) *n.*

Table 13 — Schematic representation — Derivation block



A <DerivationBlock> is used to associate the <PartOfSpeech> to the four <Derivation>s. An embedded <DerivationBlock> is used to group the two last <Derivation>s with a common <RangeOfApplication>

The X_mLex model defines blocks for seven topics: CompositionalPhrase, Derivation, Headword, Inflection, MultiWordUnit, Synonym and Translation.

Table 14 — Block encoding — Embedded derivation block

```

1. <DerivationBlock>
2. <PartOfSpeech value = 'noun' />
3. <Derivation>at<Stress/>testant</Derivation>
4. <Derivation>at<Stress/>tester</Derivation>
5. <DerivationBlock>
6. <RangeOfApplication>esp. in legal usage</RangeOfApplication>
7. <Derivation>at<Stress/>testor</Derivation>
8. <Derivation>at<Stress/>testator</Derivation>
9. </DerivationBlock>
10. </DerivationBlock>

```

4.1.3.3 Groups

A group aggregates independent data elements, containers and blocks.

A group is a compositional element used for representing distinct (and repetitive) sets of information, for instance, multiple senses of a headword within an entry.

Table 15 — Example of groups (from Annex C — excerpts from Example 3)

pneumatophore 1. (Bot) Pneumatophor *n*, Atemwurzel *f*, 2. (Zoo) Pneumatophor *n*, Schwimmglocke *f*, Gasflasche *f* (der Siphonophoren)

Table 16 — Encoding (from Annex C — excerpts from Example 3)

```

1. <SenseGroup>
2. <SubjectField>Bot</SubjectField>
3. <TranslationCtn>
4. <Translation>Pneumatophor</Translation>
5. <PartOfSpeech value = 'noun' />
6. </TranslationCtn>
7. </SenseGroup>
8. <SenseGroup>
9. <SubjectField>Zoo</SubjectField>
10. <TranslationCtn>
11. <Translation>Pneumatophor</Translation>
12. <PartOfSpeech value = 'noun' />
13. </TranslationCtn>
14. </SenseGroup>

```

According to the subject field, the headword has two meanings. Each of them is described in a sense group (<SenseGroup>).

Table 17 — List of Groups

Name	Generic identifier	Explanation	See: Annex Example Line
homograph group	HomographGroup	A grouping element for the description of homographs.	C 1 10
sense group	SenseGroup	A grouping element for the description of one meaning of the headword in an entry. See definition: 3.2.3.	C 1 12

4.2 Formal structure of a dictionary entry

4.2.1 Formal grammar for high-level structures

A dictionary entry is made of

- one or more main lexical units (“headwords”) and their related lexical units (part of speech, pronunciation, orthographical variants, derivations, etc.),
- a description of each meaning of the headwords and related lexical units associated with that meaning (compositional phrases, multiword units, translations, synonyms, etc.).

When a headword has very different meanings, the entry can be split into sub-entries for each homograph or a separate entry (with a homograph number) can be generated for each meaning.

Several entries can be grouped inside a unique “nest” in order to assemble related headwords in a compact dictionary.

The model is expressed according to eBNF (extended Backus-Naur Form) conventions:

- terminals are nouns without delimiters; they are explained before they are used (example: Headword);
- symbols are delimited by <>;
- {} means 0, one or many;
- [] means 0 or 1;
- + means 1 or many.

Note the following.

- The order of the elements within a rule is never meaningful. For instance, within a container, repeatable refining elements may appear before and/or after the refined data element.
- Symbols whose name ends with “Value”, like <GrammaticalGenderValue>, or with “Type”, like <HeadwordType>, are not defined in this International Standard; they refer to permissible values defined either by ISO 12620:1999 or by the TC 37 Data Category Registry.

Values are lists of permissible values allowed for a data element. For instance, <GrammaticalGenderValue> could contain “masculine”, “feminine”, “neuter”, etc.

Types give additional information about the nature of a data element. For instance, <HeadwordType> will enumerate types of headwords (common noun, proper noun, prefix, suffix, geographical name, etc.).

An implementation of this model should always allow the user to add his own types or permissible values according to his needs.

- The description of a citation (CitationCtn) is not defined here. Its components have to be borrowed from ISO 15836 (Dublin Core) or ISO 12615 (Source references).

1.	<pre> <Dictionary> ::= <generalQualifiers> [sourceLanguage] [targetLanguage] {<DictionaryEntry> <NestEntry> <Insert>} </pre>
2.	<pre> <NestEntry> ::= <generalQualifiers> [sourceLanguage] [targetLanguage] {DictionaryEntry} </pre>
3.	<pre> <DictionaryEntry> ::= [sortKey] [sourceLanguage] [targetLanguage] [homographNumber] <generalQualifiers> (Headword HeadwordCtn HeadwordBlock)+ {HomographGroup} {SenseGroup} {MultiWordUnitCtn CompositionalPhraseCtn {<AdministrativeInformation>} </pre>
4.	<pre> <HomographGroup> ::= <generalQualifiers> [Headword HeadwordCtn HeadwordBlock] {<CoreComments> <SenseGroup>} {<RelatedTopics>} </pre>
5.	<pre> <SenseGroup> ::= <generalQualifiers> [senseNumber] [targetLanguage] {<CoreComments> <RelatedTopics> <SenseGroup>} (See SeeCtn) </pre>

6.	<code><RelatedTopic> ::=</code> <code>{AbbreviatedForm <AbbreviatedFormCtn></code> <code>Analogy <AnalogyCtn> <AnalogyBlock> </code> <code>Antonym <AntonymCtn></code> <code>Citation <CitationCtn> </code> <code>CompanySpecificUsage CompanySpecificUsageCtn </code> <code>CompositionalPhrase <CompositionalPhraseCtn> <CompositionalPhraseBlock> </code> <code>Derivation <DerivationCtn> <DerivationBlock> </code> <code>Example <ExampleCtn> </code> <code>FalseFriend <FalseFriendCtn> </code> <code>FreeTopic <FreeTopicCtn> </code> <code>FullForm <FullFormCtn> </code> <code>Homonym HomonymCtn </code> <code>Inflection <InflectionCtn> <InflectionBlock> </code> <code>InternationalScientificTerm <InternationalScientificTermCtn> </code> <code>MultiWordUnit <MultiWordUnitCtn> <MultiWordUnitBlock> </code> <code>ProprietaryRestriction ProprietaryRestrictionCtn </code> <code>Synonym <SynonymCtn> <SynonymBlock> </code> <code>Translation <TranslationCtn> <TranslationBlock> </code> <code>Variant <VariantCtn>}</code>
7.	<code><HeadwordCtn> ::= <generalQualifiers> Headword</code> <code>{<CoreComments> <RelatedTopics>}</code>
8.	<code><HeadwordBlock> ::= <generalQualifiers> (Headword HeadwordCtn)+</code> <code>{<CoreComments>}</code>
9.	<code><MultiWordUnitCtn> ::= <generalQualifiers> MultiWordUnit+ {<CoreComments> </code> <code><RelatedTopics>} {<SenseGroup>}</code>
10.	<code><CompositionalPhraseCtn> ::= <generalQualifiers> CompositionalPhrase+</code> <code>{<CoreComments> <RelatedTopics>} {<SenseGroup>}</code>
11.	<code><LinguisticComment> ::=</code> <code>Attestation <AttestationCtn> </code> <code>Case <CaseCtn> </code> <code>Complement <ComplementCtn> </code> <code>Display </code> <code>Etymology <EtymologyCtn> </code> <code>Formation <FormationCtn> </code> <code>Formula <FormulaCtn> </code> <code>Frequency <FrequencyCtn> </code> <code>GeographicalUsage <GeographicalUsageCtn> </code> <code>GrammaticalGender <GrammaticalGenderCtn> </code> <code>GrammaticalNumber <GrammaticalNumberCtn> </code> <code>GrammaticalPattern GrammaticalPatternCtn </code> <code>GuidePhrase <GuidePhraseCtn> </code> <code>Mood <MoodCtn> </code> <code>NormativeStatus <NormativeStatusCtn> </code> <code>PartOfSpeech <PartOfSpeechCtn> </code> <code>Person <PersonCtn> </code> <code>Pronunciation <PronunciationCtn> </code> <code>RangeOfApplication <RangeOfApplicationCtn> </code> <code>Register <RegisterCtn> </code> <code>SearchForm <SearchFormCtn> </code> <code>Syllabification <SyllabificationCtn> </code> <code>Symbol <SymbolCtn> </code> <code>TypicalComplement TypicalComplementCtn </code> <code>Tense <TenseCtn> </code> <code>UsageNote <UsageNoteCtn></code>

12.	<code><SemanticComment> ::= Definition <DefinitionCtn> SenseIndicator SenseIndicatorCtn SenseQualifier <SenseQualifierCtn> SubjectField <SubjectFieldCtn></code>
13.	<code><CoreComments> ::= {<LinguisticComments> <SemanticComments> <CrossReferences> <FreeElements> <GeneralElements> <AdministrativeInformations>}</code>
14.	<code><CrossReferences> ::= SeeAlso</code>
15.	<code><AdministrativeInformation> ::= Origination Modification</code>
16.	<code><FreeElements> ::= FreeComment <FreeCommentCtn></code>
17.	<code><GeneralElements> ::= Note <NoteCtn> Source <SourceCtn></code>
18.	<code><ContainerComplement> ::= <FreeElements> <GeneralElements></code>
19.	<code><generalQualifiers> ::= [id] [class] [style] [xml:lang] [documentSize] [display]</code>

4.2.2 Formal grammar for other lexical unit containers

20.	<code><AbbreviatedFormCtn> ::= <generalQualifiers> AbbreviatedForm {<CoreComments> <RelatedTopics>}</code>
21.	<code><AnalogyCtn> ::= <generalQualifiers> Analogy {<CoreComments> <RelatedTopics>}</code>
22.	<code><AntonymCtn> ::= <generalQualifiers> Antonym {<CoreComments> <RelatedTopics>}</code>
23.	<code><DerivationCtn> ::= <generalQualifiers> Derivation {<CoreComments> <RelatedTopics>}</code>
24.	<code><ExampleCtn> ::= <generalQualifiers> Example {<CoreComments> <RelatedTopics>}</code>
25.	<code><FalseFriendCtn> ::= <generalQualifiers> FalseFriend {<CoreComments> <RelatedTopics>}</code>
26.	<code><FreeTopicCtn> ::= <generalQualifiers> FreeTopic {<CoreComments> <RelatedTopics>}</code>
27.	<code><FullFormCtn> ::= <generalQualifiers> FullForm {<CoreComments> <RelatedTopics>}</code>
28.	<code><InflectionCtn> ::= <generalQualifiers> Inflection {<CoreComments> <RelatedTopics>}</code>
29.	<code><InternationalScientificTermCtn> ::= <generalQualifiers> InternationalScientificTerm {<CoreComments> <RelatedTopics>}</code>
30.	<code><SearchFormCtn> ::= <generalQualifiers> SearchForm {<CoreComments> <RelatedTopics>}</code>
31.	<code><SynonymCtn> ::= <generalQualifiers> Synonym {<CoreComments> <RelatedTopics>}</code>
32.	<code><TranslationCtn> ::= <generalQualifiers> Translation {<CoreComments> <RelatedTopics>}</code>
33.	<code><VariantCtn> ::= <generalQualifiers> Variant {<CoreComments> <RelatedTopics>}</code>

4.2.3 Formal grammar for blocks

34.	<code><AnalogyBlock> ::= <generalQualifiers> (Analogy AnalogyCtn <AnalogyBlock>)+ {<CoreComments>}</code>
35.	<code><CompositionalPhraseBlock> ::= <generalQualifiers> (CompositionalPhrase <CompositionalPhraseCtn> <CompositionalPhraseBlock>)+ {<CoreComments>}</code>
36.	<code><DerivationBlock> ::= <generalQualifiers> (Derivation <DerivationCtn> <DerivationBlock>)+ {<CoreComments>}</code>
37.	<code><HeadwordBlock> ::= <generalQualifiers> (Headword <HeadwordCtn> <HeadwordBlock>)+ {<CoreComments>}</code>
38.	<code><InflectionBlock> ::= <generalQualifiers> (Inflection <InflectionCtn> <InflectionionBlock>)+ {<CoreComments>}</code>
39.	<code><MultiWordUnitBlock> ::= <generalQualifiers> (MultiWordUnit <MultiWordUnitCtn> <MultiWordUnitBlock>)+ {<CoreComments>}</code>
40.	<code><SynonymBlock> ::= <generalQualifiers> (Synonym <SynonymCtn> <SynonymBlock>)+ {<CoreComments>}</code>
41.	<code><TranslationBlock> ::= <generalQualifiers> (Translation <TranslationCtn> <TranslationBlock>)+ {<CoreComments>}</code>
42.	<code><VariantBlock> ::= <generalQualifiers> (Variant <VariantCtn> <VariantBlock>)+ {<CoreComments>}</code>

4.2.4 Formal grammar for other comment containers

43.	<code><AttestationCtn> ::= <generalQualifiers> Attestation {<ContainerComplement>}</code>
44.	<code><CaseCtn> ::= <generalQualifiers> Case { Example ExampleCtn <ContainerComplement>}</code>
45.	<code><DefinitionCtn> ::= <generalQualifiers> Definition {<ContainerComplement>}</code>
46.	<code><EtymologyCtn> ::= <generalQualifiers> Etymology {<ContainerComplement>}</code>
47.	<code><FormulaCtn> ::= <generalQualifiers> Formula [Source] <InflectionCtn> ::= Inflection { <LinguisticComment> }</code>
48.	<code><FreeCommentCtn> ::= <generalQualifiers> FreeComment {<CoreComments>}</code>
49.	<code><FrequencyCtn> ::= <generalQualifiers> Frequency {<ContainerComplement>}</code>
50.	<code><GeographicalUsageCtn> ::= <generalQualifiers> GeographicalUsage {<ContainerComplement>}</code>
51.	<code><GuidePhraseCtn> ::= <generalQualifiers> GuidePhrase {<ContainerComplement>}</code>
52.	<code><GrammaticalGenderCtn> ::= <generalQualifiers> GrammaticalGender {<ContainerComplement>}</code>
53.	<code><GrammaticalNumberCtn> ::= <generalQualifiers> GrammaticalNumber {<ContainerComplement>}</code>
54.	<code><GrammaticalPatternCtn> ::= <generalQualifiers> GrammaticalPattern {<ContainerComplement>}</code>
55.	<code><InsertCtn> ::= <generalQualifiers> Insert {<ContainerComplement>}</code>

56.	<MoodCtn> ::= <generalQualifiers> Mood {Example ExampleCtn <ContainerComplement>}
57.	<NormativeStatusCtn> ::= <generalQualifiers> NormativeStatus {<ContainerComplement>}
58.	<NoteCtn> ::= <generalQualifiers> Note {<ContainerComplement>}
59.	<PartOfSpeechCtn> ::= <generalQualifiers> PartOfSpeech [Subcategorisation SubcategorisationCtn] [GrammaticalGender] [GrammaticalNumber] {<ContainerComplement>}
60.	<PersonCtn> ::= <generalQualifiers> Person { Example ExampleCtn <ContainerComplement>}
61.	<PronunciationCtn> ::= <generalQualifiers> Pronunciation {<ContainerComplement>}
62.	<RegisterCtn> ::= <generalQualifiers> Register {<ContainerComplement>}
63.	<RangeOfApplicationCtn> ::= <generalQualifiers> RangeOfApplication {<ContainerComplement>}
64.	<SeeCtn> ::= <generalQualifiers> See { Corecomments } (HomographNumber) (SenseNumber) {<ContainerComplement>}
65.	<SenseIndicatorCtn> ::= <generalQualifiers> SenseIndicator {<ContainerComplement>}
66.	<SenseQualifierCtn> ::= <generalQualifiers> SenseQualifier {<ContainerComplement>}
67.	<SourceCtn> ::= <generalQualifiers> Source {<ContainerComplement>}
68.	<SubcategorisationCtn> ::= <generalQualifiers> Subcategorisation {<ContainerComplement> [SubcategorisationCtn]}
69.	<SubjectFieldCtn> ::= <generalQualifiers> SubjectField {<ContainerComplement>}
70.	<SyllabificationCtn> ::= <generalQualifiers> Syllabification {<ContainerComplement>}
71.	<SymbolCtn> ::= <generalQualifiers> Symbol { Source } {<ContainerComplement>}
72.	<TenseCtn> ::= <generalQualifiers> Tense { Example ExampleCtn <ContainerComplement>}
73.	<TypicalComplementCtn> ::= <generalQualifiers> TypicalComplement {<ContainerComplement>}
74.	<UsageNoteCtn> ::= <generalQualifiers> UsageNote { <CoreComments>}

4.2.5 Formal grammar for lexical units

75.	<AbbreviatedForm> ::= <generalQualifiers> [AbbreviatedFormType] <CoreComponentContent>
76.	<Analogy> ::= <generalQualifiers> <CoreComponentContent>
77.	<Antonym> ::= <generalQualifiers> [AntonymType] <CoreComponentContent>
78.	<Citation> ::= <generalQualifiers> [CitationType] <CoreComponentContent>