enosema

report

EFR 2528:2026

Terminology management — Quality review of international standard vocabulary — ISO terms and the IEC International Electrotechnical Vocabulary

First edition

Reference number
EFR 2528:2026

© EFS 2026

(c) 2026 Enosema Foundation.

This document is made available under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Unless otherwise specified, or required in the context of its implementation, you are free to use, share, and adapt this document, provided that appropriate credit is given to the Enosema Foundation and that any changes are indicated. The full license text is available at https://creativecommons.org/licenses/by/4.0/.

Enosema Foundation
2885 Sanford Ave Sw Suite 32361
Grandville MI 49418-1342
USA

E-mail: info@enosema.org

Published in the USA

© EFS 2026 – All rights reserved

Contents

Page

Foreword ..... **v**

Introduction ..... **vi**

1 Scope **1**

2 Terms and conventions **1**

2.1 Terms and definitions 1

2.1.1 localized concept entry 1

2.1.2 definition-bearing entry 2

2.1.3 term-only entry 2

2.1.4 definition length 2

2.1.5 defect 2

2.1.6 document era 2

2.2 Conventions 2

3 The corpora **2**

3.1 The ISO Online Browsing Platform corpus 2

3.2 The IEC International Electrotechnical Vocabulary corpus 3

3.3 Record anatomy 3

3.4 Extraction provenance and scale 4

3.5 Third-party content statement 4

3.6 What each corpus can and cannot show 4

4 Methodology **5**

4.1 Sampling design 5

4.2 Measures 5

4.3 Derived profiles 6

4.4 Limitations 6

5 Findings: the ISO corpus **6**

5.1 Headline statistics 6

5.2 Definition length 7

5.3 Structural discipline 7

5.4 Entry element coverage 8

5.5 Era profile 8

5.6 Notes and examples in detail 8

5.7 Length by era 9

5.8 Circularity in detail 9

5.9 Document concentration 10

5.10 Grammar information 10

5.11 What the ISO corpus does not carry 10

6 Findings: the IEV corpus **10**

6.1 Headline statistics 10

6.2 Languages per concept 11

6.3 Language coverage 11

6.4 The two localization models in one corpus 12

6.5 The tier structure of language coverage 12

6.6 The authoritative pair in detail 13

6.7 A term-only entry set, anatomized 13

6.8 Definition length in the authoritative languages 14

6.9 Era profile 14

6.10 What the IEV corpus lacks 14

7 Comparison of the two corpora **14**

7.1 The systems compared 14

7.2 Where the systems agree 15

7.3 Where the systems differ 15

© EFS 2026 – All rights reserved

7.4 The two localization models 15
7.5 A decision table for registers 16
7.6 What each corpus contributes to the ES calibration 16
7.7 What a third system would add 17

8 Anti-pattern census 17
8.1 The census method 17
8.2 The census 17
8.3 What the census shows 18
8.4 The IEV entries against the census 18
8.5 Reading the census as a register operator 18
8.6 Defect interaction 19

9 Implications for the ES Terminology series 19
9.1 For EFS 2522 (concept entry quality) 19
9.2 For EFS 2525 (concept arrangement) 19
9.3 For EFG 2519 (register operation guide) 20
9.4 For the series as a whole 20
9.5 Stewardship of the evidence base 20

Annex A (normative) **ISO corpus statistics tables 21**

Annex B (normative) **IEV corpus statistics tables 24**

Annex C (normative) **Detection rules 27**

Annex D (normative) **Reproduction 30**

Bibliography 32

© EFS 2026 – All rights reserved

Foreword

ES is the publishing imprint of the Enosema Foundation. ES documents are prepared by the technical committees of the Enosema Foundation under the ES Directives (EFD 101 to EFD 105) and are published under a Creative Commons Attribution 4.0 International licence (CC BY 4.0).

This report was prepared by Technical Committee Terminology.

This report is an informative document of the ES Terminology series in the document class **report** (EFD 102): it reports the findings of a study, states the methods of the study, and reaches conclusions. It does not contain requirements; the requirements that the findings inform are stated in the normative documents of the series — EFS 2513 (concept management systems), EFS 2522 (concept entry quality), and EFS 2525 (concept arrangement in concept systems).

The subject of the study is the vocabulary that the international standardization system actually produces and maintains. Two corpora were analysed:

— the terminology published through the ISO Online Browsing Platform (OBP), extracted in machine-readable form and analysed through a sample of 4 662 localized concept entries drawn from 406 898 extracted records; and

— the IEC International Electrotechnical Vocabulary (IEV) as maintained by IEC TC 1 in Glossarist form, analysed through a sample of 2 000 concept records carrying 23 936 language entries.

All statistics in this report derive from those samples. Annex E states the reproduction instructions, including the sample seeds, so that every figure in this report can be regenerated from the corpora.

© EFS 2026 – All rights reserved

Introduction

0.1 Why a quality review

The ES Terminology series rests on a claim: that the requirements in its normative documents describe what good vocabulary practice is, not what one organization happens to prefer. A claim of that kind needs evidence. The strongest evidence available is the vocabulary that the international system itself produces — the hundreds of thousands of concept entries that ISO has published through its Online Browsing Platform and the tens of thousands of concepts that IEC TC 1 maintains as the International Electrotechnical Vocabulary. If the requirements of the ES series are well calibrated, a corpus conforming termbase measures well against the corpus; if the corpus systematically fails somewhere, the requirements are set to address that failure.

This report is the evidence base made public. It analyses the two corpora with a fixed method (clause Clause 4), reports the statistics (clauses Clause 5 and Clause 6), compares the two systems of maintenance (clause Clause 7), counts the defects that the ES series calls anti-patterns (clause Clause 8), and states the implications of each finding for the normative documents of the series (clause Clause 9).

0.2 What the corpora are

The two corpora represent two eras and two architectures of international vocabulary maintenance:

— The **ISO OBP corpus** is the aggregate of terminology from hundreds of ISO documents, extracted from the published platform into Glossarist-form records. Its entries were authored across four decades, under successive editions of the ISO/IEC Directives and of ISO 10241, by hundreds of committees. It measures the system as a whole — including its oldest practices.

— The **IEV corpus** is one vocabulary, maintained continuously since the early twentieth century and converted in its current phase into structured records by IEC TC 1. Every concept carries a multilingual entry set; English and French carry the authoritative definitions and the other languages carry designations. It measures the depth-maintenance model: one authoritative body, one vocabulary, many languages.

The two corpora differ in shape, and the differences are findings. Where they agree, the agreement is the strongest signal this study can produce: two independent systems, two architectures, one behaviour.

0.3 How to read this report

Each clause states its findings first and interprets them second. Statistics appear in the body where they carry an argument and in Annex A (ISO) and Annex B (IEV) in full. Clause Annex C states the operational definition of every measure — the exact rule by which each count was made — because a defect rate is only as honest as its detector. Clause Annex D gives the commands and seeds.

This report makes no recommendations beyond clause Clause 9; it is a study, not a guideline. The documents that convert its findings into requirements are cited throughout, and each finding carries the traceability note that states which requirements it informs.

0.4 Relation to the ES series

The report serves three documents directly:

— **EFS 2522** (concept entry quality): its criteria, weights and anchors were set from the first ISO sample; this report refreshes the numbers, extends the evidence to the IEV corpus, and confirms or adjusts the calibration;

© EFS 2026 – All rights reserved

— **EFS 2525** (concept arrangement): its arrangement requirements — criteria of subdivision, MECE discipline, cross-system membership — respond to the arrangement-level findings of clauses Clause 5 and Clause 7; and

— **EFG 2519** (register operation guide): its cadence and monitoring practice consumes the era and defect-rate profiles as the benchmark that a conforming register is measured against.

The report is also a demonstration of method. The ES series asserts that registers be machine-readable because machine-readable records can be measured; this report is what measurement looks like, and the reproduction instructions of Annex E are the standing offer that the same measurement can be run over any conforming register — including the registers ES itself operates.

0.5 Findings at a glance

The headline rates of both corpora, side by side, with the clause where each finding is developed:

Table 1

Measure	ISO corpus	IEV corpus	Finding
definition presence	98,8 %	100,0 % (eng, fra)	Clause 5, Clause 6
median definition length	16 words	18 / 19 words (eng / fra)	Clause 7
circularity	3,5 %	not measured this cycle	Clause 5
requirements language in definitions	0,35 %	not measured	Clause 7
notes presence	34,2 %	5,4 %	Clause 7
examples presence	2,9 %	0,1 %	Clause 8
sources presence	100,0 %	not carried	Clause 5
term type uniformity	100,0 %	100,0 %	Clause 7
grammar information	absent as extracted	absent in record format	Clause 5
languages per concept	one authoritative language per document	median 12	Clause 6

The table is the review in one view: strong uniformity where process gates exist, weak production where author initiative is the only force, and two maintenance architectures that converge on the same definition norm while diverging on everything they choose to annotate differently.

0.6 How the report was produced

The review was carried out by the Terminology Committee as part of the EFS 2522 calibration cycle. The analysis scripts are single-pass implementations of the rules of Annex C over the extracted corpora; the report text and the statistics were produced in the same cycle, and the figures in this edition were generated on 13 November 2026. The corpus states, the seeds and the protocol of Annex E make the edition reproducible; deviations introduced by future extractions are attributable to the corpus states, not to the method.

© EFS 2026 – All rights reserved

Terminology management — Quality review of international standard vocabulary — ISO terms and the IEC International Electrotechnical Vocabulary

1 Scope

This report reports the findings of a quality review of international standard vocabulary. It covers:

— the composition of the two corpora studied: the ISO Online Browsing Platform terminology extraction and the IEC International Electrotechnical Vocabulary records maintained by IEC TC 1 (clause Clause 3);

— the methodology of the review: the sampling design, the measures, and the operational detection rules from which every statistic derives (clauses Clause 4 and Annex C);

— the findings for each corpus: presence and length statistics, structural discipline, entry element coverage, era profiles, and document and language concentration (clauses Clause 5 and Clause 6);

— the comparison of the two systems of maintenance and the two localization models they embody (clause Clause 7);

— the census of anti-patterns — the recurring defect classes that the ES quality framework names (clause Clause 8); and

— the implications of the findings for the normative documents of the ES Terminology series (clause Clause 9).

This report is informative. It states findings and conclusions; it does not state requirements. The requirement documents that consume its findings — EFS 2513, EFS 2522, EFS 2525, and EFG 2519 — carry the normative text, and clause Clause 9 traces each finding to the requirements it informs.

The review covers the corpora as they stood at extraction. The ISO Online Browsing Platform changes continuously; the IEV records change with TC 1 maintenance. The figures are dated by the extraction dates of clause Clause 3 and are reproducible from the corpora by the instructions of Annex E.

2 Terms and conventions

2.1 Terms and definitions

For the purposes of this report, the following terms and definitions apply. The report uses the terms of the ES Terminology series as defined in EFS 2510, EFS 2513 and EFS 2522; the entries below fix the measurement senses in which this report uses them.

2.1.1 localized concept entry

record of one concept in one language, comprising its designations and, where the entry carries them, its definition, notes, examples, and sources

[NOTE: In the ISO corpus, a localized concept entry is one language block of one Glossarist concept record. In the IEV corpus, it is one language key of one concept file. The unit of analysis throughout is the localized concept entry unless a clause states otherwise.]

© EFS 2026 – All rights reserved

2.1.2 definition-bearing entry

localized concept entry whose definition field carries a non-empty definition

2.1.3 term-only entry

localized concept entry that carries designations and no definition

2.1.4 definition length

count of whitespace-separated tokens in the concatenation of the definition text of a definition-bearing entry

[NOTE: Stem expressions, markup and punctuation are counted as tokens. The measure overstates length for symbolic definitions and understates it for prose; clause Annex C states the rule verbatim.]

2.1.5 defect

property of a localized concept entry that a detection rule of clause Annex C evaluates as present

[NOTE: Defects are measured, not judged: a defect is whatever the stated rule detects, and the rule is part of the finding. Rates are comparable only through the rules.]

2.1.6 document era

assignment of a source document to one of three periods by its latest year of publication: legacy (before 1998), transitional (1998 to 2014), and modern (2015 and later)

[NOTE: The periods track the regime changes of the drafting rules: the pre-10241-1 practice, the Directives Part 2 consolidation, and the current regime.]

2.2 Conventions

— Percentages are rounded to one decimal place unless a clause states otherwise; sums may therefore differ from 100 by rounding.

— The word **corpus** refers to the full extracted collection; the word **sample** refers to the analysed subset drawn from it (clause Clause 4).

— Sample seeds are stated so that every figure regenerates exactly (Annex E).

3 The corpora

3.1 The ISO Online Browsing Platform corpus

The ISO corpus is the terminology content of the ISO Online Browsing Platform (OBP), extracted into Glossarist-form concept records. The extraction comprises 406 898 machine-extracted records covering the published terminology of hundreds of ISO documents — international standards, technical specifications and technical reports that carry terms and definitions clauses — from 1983 to the date of extraction.

The corpus as extracted has known characteristics that the review treats as part of its object of study rather than as noise:

— the records are machine-derived from published documents, and the extraction fidelity varies with the age and format of the source;

— the records carry the entry structure of the source documents: definitions, notes, examples, sources and term designations where the published document carries them; and

— the records carry term-level metadata — term type and normative status — at the uniformity the platform maintains.

© EFS 2026 – All rights reserved

The analysed sample is 4 662 localized concept entries drawn by the sampling design of clause Clause 4.

3.2 The IEC International Electrotechnical Vocabulary corpus

The IEV corpus is the International Electrotechnical Vocabulary as maintained by IEC Technical Committee 1 in Glossarist form: one record per concept, each carrying the concept's language entries. At extraction the corpus comprises 22 229 concept records. The analysed sample is 2 000 concept records drawn by the sampling design of clause Clause 4, carrying 23 936 localized concept entries.

The corpus has the structure of a maintained multilingual vocabulary:

— every concept carries an English entry and, except for a residual set, a French entry, each with a definition — the authoritative languages of the IEV;

— the remaining languages carry designations without definitions: the term-only model of localization (clause Clause 7);

— each concept carries an identification number of the form `<part>-<section>-<item>` that places it in a subject part of the vocabulary; and

— each language entry carries acceptance and review dates that place the concept in time.

3.3 Record anatomy

Both corpora arrive in Glossarist form, and the review's measures run directly over the record fields. The anatomies below show the shapes; the field names are the ones the rules of Annex C read.

An ISO record (abridged):

```
termid: iso-2382-01
localized_concepts:
  eng:
    terms:
      - type: expression
        normative_status: preferred
        designation: central processing unit
    definition:
      - content: "part of a computer that carries out arithmetic
          and logical operations and controls the execution of
          instructions"
    notes:
      - "In some contexts the term is taken to include the primary
          storage."
    examples: []
    source: ISO/IEC 2382:2015, 2126478
    language_code: eng
    entry_status: Standard
    date_accepted: '2018-05-01'
```

Figure 1

The fields the analysis reads: `terms` (type, normative_status, designation — rules R8, R9), `definition` (rules R1 to R5), `notes` and `examples` (rule R6), `source` (rule R7), and the document identity that carries the era assignment (rule R10).

An IEV record (abridged):

```
termid: 102-01-01
term: equality
eng:
  id: 102-01-01
  terms:
    - type: expression
      normative_status: preferred
```

© EFS 2026 – All rights reserved

```
      designation: equality
  definition: "relation between two entities a and b having the
    properties of reflexivity, symmetry and transitivity"
  notes:
    - "The equality of two entities a and b is denoted by a = b."
  language_code: eng
  entry_status: Standard
  date_accepted: '2008-08-01'
fra:
  id: 102-01-01
  terms:
    - type: expression
      normative_status: preferred
      designation: égalité
  definition: "relation entre deux entités a et b..."
  language_code: fra
  date_accepted: '2007-08-01'
pol:
  id: 102-01-01
  terms:
    - type: expression
      normative_status: preferred
      designation: równość
  language_code: pol
  date_accepted: '2009-11-01'
```

Figure 2

The IEV shape differs structurally from the ISO shape: language keys are top-level, the authoritative pair carries the definitions, and the remaining languages carry only the designation block. Rule R12 counts the language keys with designations; the entry-level measures run per language key.

3.4 Extraction provenance and scale

Table 2

Property	ISO extraction	IEV records
record unit	one concept record with localized concept sub-records	one concept record with language keys
extraction origin	published platform content, per document	TC 1 maintenance database export
record count	406 898 concept records	22 229 concept records
language entries (full corpus, approx.)	406 898 English-language entries	approximately 267 000 language entries at 12 per concept
document spread	hundreds of source documents, 1983–2026	one vocabulary, parts by subject, continuous since early 1900s

The scale ratio shapes the sampling design: the ISO corpus is sampled by entry across a document-dominated population; the IEV corpus is sampled by concept across a language-dominated one. The two designs measure the two different things the clause above names, and the designs are part of the findings' conditions.

3.5 Third-party content statement

The corpora are the published vocabulary of ISO and of IEC TC 1 respectively. The ES research environment holds the extractions for analysis; this report publishes statistics, findings and rule definitions — not corpus content beyond the minimal abridged anatomy records above, which serve to document the format. Source documents and full records are not reproduced here, and the reproduction protocol of Annex E presumes access to the corpora under their respective terms.

3.6 What each corpus can and cannot show

The two corpora measure different things well, and the review uses each for what it measures:

© EFS 2026 – All rights reserved

— The ISO corpus measures **the production system**: what the document drafting process yields when hundreds of committees write terminology under a common rulebook. Its findings are findings about authoring practice at scale.

— The IEV corpus measures **the maintenance system**: what one continuously maintained vocabulary looks like when one body keeps definitions authoritative and distributes designations. Its findings are findings about localization architecture and about the long-run stability of an entry base.

Neither corpus measures conformance to any ES document; the ES documents did not exist when the corpora were formed. The review measures the corpora against the measurement rules of clause Annex C and states the implications for the ES requirements, which are calibrated against this evidence rather than derived from it.

4 Methodology

4.1 Sampling design

Both corpora are too large to analyse entry by entry in this review, and the review draws probability samples with fixed seeds:

— **ISO corpus**: a simple random sample of 4 662 localized concept entries, drawn from the 406 898 extracted records with seed 2522. The sample size was set by the analysis capacity of the review cycle and yields rate estimates with a width of approximately ±1,4 percentage points at the 95 % level for rates near 50 %, and tighter for the extreme rates that dominate the findings.

— **IEV corpus**: a simple random sample of 2 000 concept records, drawn from the 22 229 records with seed 2522, analysed exhaustively across all language entries of each sampled concept. The concept is the sampling unit and the language entry the analysis unit, which preserves the multilingual structure in the sample.

Both seeds are recorded in Annex E. Re-running the analysis with the same seeds reproduces every figure in this report; re-running with different seeds shifts rates within the sampling error stated above.

The sample sizes are sized to the decisions that the review supports. The ES requirements are calibrated to differences between eras, between corpora and between element rates — differences of five percentage points and more — and a sample that resolves one-point differences is more than the calibration needs. The ISO sample of 4 662 entries resolves half-point differences at the headline rates (Clause A.6), which bounds the sampling error well below the smallest difference any finding turns on; the IEV sample of 2 000 concepts resolves one-point differences at the concept level and yielded language-level rates with ample depth, since every sampled concept contributes its full language set. Growing the samples would sharpen rates that no finding consumes; the sizes in force match evidence to decision.

4.2 Measures

The review measures each localized concept entry on the following variables. Clause Annex C states each rule verbatim.

Table 3

Variable	Meaning	Applies to
definition presence	the entry carries a non-empty definition	both corpora
definition length	token count of the definition text	definition-bearing entries
circularity	the designation (or a distinctive part of it) appears in its own definition	definition-bearing entries
requirements language	the definition contains shall, should or must	definition-bearing entries
article head	the definition begins with the indefinite or definite article	definition-bearing entries
notes presence	the entry carries one or more notes	both corpora
examples presence	the entry carries one or more examples	both corpora
sources presence	the entry carries source citations	ISO corpus

© EFS 2026 – All rights reserved

(continued)

Variable	Meaning	Applies to
grammar information	the designation carries part-of-speech and inflection data	both corpora
term type	the designation's ISO 12620 type	both corpora
normative status	the designation's status value	IEV corpus

4.3 Derived profiles

Three derived profiles organize the findings:

— **the length distribution**: definition lengths bucketed at fixed cut points (1–5, 6–10, 11–16, 17–22, 23–32, 33–50, 51 and more words), with median, mean and the 10th and 90th percentiles;

— **the era profile**: each variable aggregated by document era as defined in clause Clause 2, which shows how practice moved with the drafting-rule regimes; and

— **the concentration profile**: the share of entries and of defects carried by the largest source documents (ISO) and the share of entries carried by each language (IEV).

4.4 Limitations

The review states four limitations that bound its conclusions:

— **extraction fidelity**. The corpora are machine-derived; the extraction inherits the defects of the source formats. The rates reported here are rates in the records as extracted, which for the ISO corpus may differ from the published documents in formatting detail.

— **rule sensitivity**. Each defect rate is conditional on its detection rule; a stricter circularity rule lowers the rate, a looser one raises it. The rules are stated verbatim so that this sensitivity is visible rather than hidden.

— **sample size**. The rates carry sampling error; the review quotes the error bounds where a finding rests on a small difference.

— **era attribution**. The ISO era profile assigns each entry the era of its source document's latest year; a document revised across regimes is assigned one era. The attribution is coarse and the era findings are directional, not exact.

None of the limitations weakens the findings that rest on large differences between corpora or between eras; they bound the precision, not the direction, of the results.

5 Findings: the ISO corpus

5.1 Headline statistics

The sample of 4 662 localized concept entries yields the following headline rates:

Table 4

Measure	Rate	Reading
definition presence	98,8 %	the system publishes definitions almost universally
empty definitions	1,2 %	concentrated in the legacy era (clause [iso-era])
notes presence	34,2 %	one entry in three carries notes
examples presence	2,9 %	examples are the scarcest entry element
sources presence	100,0 %	every entry traces to its source document
term type recorded	100,0 % (`expression`)	uniform type annotation

© EFS 2026 – All rights reserved

(continued)

Measure	Rate	Reading
grammar information	0,0 %	absent in this extraction (clause [iso-grammar])

The system that the OBP corpus represents is strong where the drafting rules have teeth: definitions are present, sources are recorded, term types are uniform. It is weak where the rules leave discretion: examples are rare, and the structural discipline measures below show the residue of discretionary practice.

5.2 Definition length

The 4 606 definition-bearing entries carry definitions with the following distribution:

Table 5

Statistic	Value	Notes
median	16 words	the working norm of the system
mean	18,3 words	right-skewed by long definitions
10th percentile	8 words	very short definitions are rare
90th percentile	31 words	the practical upper working range

The full histogram appears in Annex A. The shape of the distribution carries three findings:

— **the working norm is tight.** Half of the definitions fall between roughly 11 and 22 words; the modal bucket is 11–16 words. The system has an implicit length norm despite no rule fixing a length, and the norm sits close to the 15-to-20-word guidance that terminology practice recommends.

— **very short definitions persist.** 214 definitions (4,6 %) are five words or fewer, and 803 (17,4 %) are ten words or fewer. Short definitions are legitimate where the genus plus one delimiter suffices; the review's circularity analysis (below) shows that the shortest bucket carries a disproportionate share of the circular definitions.

— **a long tail exists.** 385 definitions (8,4 %) exceed 32 words, and 73 exceed 50. The long definitions are concentrated in documents that define complex systems; they are not defective by length alone, and the review treats length as context rather than as a defect.

5.3 Structural discipline

The structural measures test whether definitions follow the genus-differentia discipline:

Table 6

Measure	Rate	Reading
definitional circularity	3,5 %	the designation or its head appears in the definition
requirements language in definitions	0,35 %	definitions almost never prescribe
article-headed definitions	4,3 %	a small residue of prose-style openings

The circularity rate is the headline structural finding. At 3,5 % of definition-bearing entries, roughly one definition in twenty-nine restates its own designation instead of replacing it — the defect that the substitution test of EFS 2522 exists to catch. Clause [iso-era] shows the rate falling by era: the discipline is being learned.

The requirements-language rate is the system's quiet success: definitions describe, and the near-zero rate of shall, should and must inside definitions shows that the boundary between normative clauses and terminology clauses holds across the corpus.

© EFS 2026 – All rights reserved

The article-head rate marks the residue of prose-style definitions — "A device that…" openings that bury the genus under the article. At 4,3 % the residue is small but persistent, and it does not fall by era (clause [iso-era]), which marks it as a drafting habit rather than a regime effect.

5.4 Entry element coverage

Table 7

Element	Rate	Finding
notes	34,2 %	well established; the entry system supports notes at scale
examples	2,9 %	the weakest element; EFS 2522 treats example provision as a scored completeness criterion precisely because the system under-produces them
sources	100,0 %	universal; the platform enforces provenance, and this is the single strongest uniform behaviour in either corpus

The contrast between notes at a third of entries and examples at under three per cent is itself a finding: the drafting culture produces explanatory notes far more readily than illustrative examples. Examples require the author to construct an instance, which costs more than annotating an existing definition — and the rate shows the cost is usually not paid.

5.5 Era profile

Table 8

Era	Entries	Definition present	Empty	Circular
legacy (1983–1997)	405	92,3 %	7,7 %	4,5 %
transitional (1998–2014)	1 667	98,9 %	1,1 %	4,8 %
modern (2015–2026)	2 590	99,8 %	0,2 %	2,6 %

The era profile is the clearest evidence in the review that the drafting-rule regimes work:

— **empty definitions collapse.** From 7,7 % in the legacy era to 0,2 % in the modern era — a 38-fold reduction. The legacy residue is the population of terms published before definitions were enforced; the modern regime effectively eliminates the defect.

— **circularity halves.** From 4,5–4,8 % in the legacy and transitional eras to 2,6 % in the modern era. The improvement is real but incomplete: even under the current regime, one definition in thirty-eight is circular.

— **the article head persists.** The rate holds roughly steady across eras, confirming it as a habit rather than a regime effect.

The modern-era residual defects — 0,2 % empty, 2,6 % circular — are the floor of the current system. They are the numbers that EFS 2522 is calibrated to beat: a conforming termbase should hold empty definitions near zero and circularity below the modern-era rate.

5.6 Notes and examples in detail

The notes rate (34,2 %) rewards a closer look, because notes are the element the drafting culture produces most readily after the definition itself:

— **notes cluster in the vocabulary standards.** The concentration table of Annex A shows vocabulary-bearing documents — the ISO/IEC 2382 series, the 24765 software vocabulary editions, the 5598 and 5593 standards — carrying notes on most entries (68 of 68, 43 of 61, 24 of 24). These documents maintain notes as part of their editorial method.

© EFS 2026 – All rights reserved

— **notes are rare in domain standards.** The long tail of documents carrying fewer than twenty entries shows note rates near zero: a document with twelve definitions typically annotates none of them. Notes are a practice of vocabulary-maintaining committees, not of general drafting.

— **the corpus-wide rate is the mixture.** 34,2 % is what the mixture of a few heavy annotators and many non-annotators yields. The figure should be read as "the system carries a notes culture in part of its population", not as "a third of all committees annotate".

The examples rate (2,9 %) decomposes the same way and worse: the 135 sampled entries with examples concentrate almost entirely in the vocabulary standards. Outside those, the example is essentially extinct: the drafting rules carry examples as an optional element, and optional elements without gates go unused.

5.7 Length by era

a) Definition length percentiles by era

Table 9

Era
Median
Mean
10th percentile
90th percentile
legacy (1983–1997)
13
15,7
6
26
transitional (1998–2014)
16
17,9
8
30
modern (2015–2026)
17
19,3
9
32

Definitions lengthened by roughly four words across the regimes. The direction is consistent with the modern rulebook's demand for complete delimiting characteristics — longer definitions are usually more complete definitions — and the percentiles show the whole distribution shifting rather than a tail growing: the practice norm itself moved. For the ES length anchors, the finding sets the reference band at the modern-era profile: median 17, working range 9 to 32.

5.8 Circularity in detail

The 161 circular entries (3,5 %) distribute across forms, and the forms carry different lessons:

— **whole-designation restatement** — the definition begins with or contains the full designation, giving the substitution test nothing: the smallest class, and the classic defect;

— **head-noun echo** — the definition contains the designation's head noun in its defining genus phrase, where the genus is stated and the differentia repeats the term: the largest class, and the one the broad rule R3 over-counts most, since legitimate genus mentions collide with the designation's head; and

© EFS 2026 – All rights reserved

— **stem echo** — a derivational stem of the designation recurs where the concept genuinely differs: the boundary between unavoidable vocabulary overlap and true circularity, and the reason the review states the rule as an upper bound.

The era fall (4,5–4,8 % to 2,6 %) is consistent across the forms; the discipline improved overall rather than one form disappearing. The modern-era residue concentrates in the head-noun echo class, which is the form that review — rather than authoring skill — catches: an author who states the genus honestly collides with the head noun by accident, and the substitution test at review is the net that catches the accident. The placement of the substitution test at review in EFS 2522 is calibrated to exactly this residue.

5.9 Document concentration

Vocabulary production is concentrated. The twenty largest source documents in the sample carry 489 entries — 10,5 % of the sample — and the concentration carries the defects unevenly: five of the twenty documents carry circular definitions, and the largest single carrier of circularity in the sample is a single modern-era systems-integration document. The full table appears in Annex A; the finding is structural: **the corpus is a small number of large vocabularies plus a long tail of small ones**, and quality practice in a handful of committees moves the corpus-wide rate.

5.10 Grammar information

The extraction carries no grammar information: part of speech, gender and inflection are absent from the records as extracted. The finding is bounded: the published platform renders grammar data for entries that carry it in the source documents, and the absence here measures the extraction's field coverage rather than the system's practice. EFS 2522 requires grammar information in conforming termbases; this corpus cannot benchmark it, and the IEV corpus (clause Clause 6) cannot either, for the same reason in its current record format.

5.11 What the ISO corpus does not carry

Three absences bound what the corpus can benchmark:

— **no relation structure.** The extraction carries concept records without the cross-entry relations that the source documents express implicitly by entry order. Arrangement quality is not measurable in this corpus; the evidence for arrangement findings comes from the IEV structure and from the corpus's rake behaviour as published.

— **no grammar fields.** Stated once in clause Clause 5 and repeated here as a coverage bound: rule R8 measures absence, and the absence is the extraction's, not the practice's.

— **no lifecycle history.** The records carry acceptance dates, not revision histories. The era profile is the review's only temporal instrument, and it measures cohorts, not trajectories; a future extraction that carries revision chains would let the review measure movement rather than cohorts.

6 Findings: the IEV corpus

6.1 Headline statistics

The sample of 2 000 concept records carries 23 936 localized concept entries — a mean of 12,0 language entries per concept:

Table 10

Measure	Rate	Reading
concepts with an English entry	100,0 %	English is universal
concepts with a French entry	100,0 %	French is universal
English definitions present	100,0 %	the authoritative language always defines
French definitions present	100,0 %	as does the co-authoritative language

© EFS 2026 – All rights reserved

(continued)

Measure	Rate	Reading
non-authoritative definitions	0,9 % of non-eng/fra entries	other languages carry designations, not definitions
notes presence	5,4 % of entries	notes concentrate in the authoritative entries
examples presence	0,1 % of entries	examples are effectively absent
normative status recorded	100,0 % (`preferred`)	uniform status annotation
grammar information	0,0 %	absent in the current record format

The IEV is a definition-bearing system in exactly two languages and a term-only system in every other language. That is not a defect; it is a declared architecture (clause Clause 7), and its uniformity is the finding: 2 000 of 2 000 sampled concepts carry English and French entries, and the entries that lack definitions lack them by design.

6.2 Languages per concept

Table 11

Statistic	Value	Notes
median languages per concept	12	the working depth of the vocabulary
mean languages per concept	12,0	symmetric around the median
maximum in sample	19	the most translated concepts in the sample

The full distribution appears in Annex B. The distribution is tight: 2 043 of the sampled concepts (the mode region, 10 to 14 languages) form the working core, with thin tails at 7 to 9 and at 15 to 19 languages. A multilingual vocabulary that holds twelve languages per concept at a median is operating at a depth that few terminologies reach, and the regularity of the distribution indicates systematic rather than opportunistic translation.

6.3 Language coverage

The sampled entries distribute across 21 language codes. The table below states the full coverage; Annex B carries the per-language definition analysis.

Table 12

Code	Language	Entries
eng	English	2 000
fra	French	2 000
pol	Polish	1 998
por	Portuguese	1 998
jpn	Japanese	1 991
deu	German	1 920
kor	Korean	1 912
ara	Arabic	1 810
ita	Italian	1 766
spa	Spanish	1 714
swe	Swedish	1 238
zho	Chinese	1 137
fin	Finnish	834
srp	Serbian	619
slv	Slovenian	250
mon	Mongolian	198

© EFS 2026 – All rights reserved

(continued)

Code	Language	Entries
rus	Russian	190
nob	Norwegian Bokmål	158
nno	Norwegian Nynorsk	158
ces	Czech	25
nld	Dutch	20

Three findings emerge from the coverage table:

— **a two-tier structure.** Eleven languages cover at least 85 % of concepts; the remainder trail. The vocabulary grew by tiers, and the tiers are visible in the coverage.

— **near-universal tiers are real.** Polish, Portuguese, Japanese, German and Korean each cover over 95 % of concepts — designation coverage, under the term-only model, across a quarter-million-entry surface when projected over the full corpus.

— **the thin tail is a register of recency and need.** Czech at 25 and Dutch at 20 entries are partial adoptions, not declining ones; their presence at all marks languages being added.

6.4 The two localization models in one corpus

The IEV operates both localization models that EFS 2522 and EFG 2519 name, inside one vocabulary:

— **the definition-bearing model** for English and French: every entry carries a definition, and the two sets of definitions are independently authored — French is not a translation layer but a parallel authority;

— **the term-only model** for all other languages: each language carries its designations against the authoritative definitions, with review concentrated on whether the designation can carry the intension in that language.

The exception proves the model: Russian carries 21 definitions among its 190 entries (11,1 %, median 15 words) — a language that at some point received partial definition authoring and stopped. The residue is detectable precisely because the declared model is term-only for Russian: any definition present in a term-only language is an anomaly, and the corpus carries 21 of them.

6.5 The tier structure of language coverage

Grouping the coverage table by tier shows the vocabulary's expansion history written in its current state:

Table 13

Tier	Languages	Coverage	Reading
authoritative	eng, fra	100,0 %	the definition-bearing pair; the vocabulary's reference point
deep tier	pol, por, jpn, deu, kor	95,6 % to 99,9 %	near-complete designation coverage; the working multilingual core
broad tier	ara, ita, spa, swe, zho, fin	41,7 % to 90,5 %	substantial coverage with gaps; adoption varies by part
partial tier	srp, slv, mon, rus, nob, nno, ces, nld	1,0 % to 31,0 %	recent or limited adoption; growth area

The tiers are not static: the partial tier's presence at all distinguishes a living vocabulary from a frozen export. The coverage gaps within the broad tier are uneven across subject parts — the electrotechnical core parts are covered more deeply than the peripheral ones — which is the pattern a need-driven translation programme leaves, and the pattern that EFG 2519's placement-need records make explicit when a register operates deliberately.

© EFS 2026 – All rights reserved

6.6 The authoritative pair in detail

The English–French pair is the corpus's definition-bearing heart, and its statistics describe the model at its best:

— **presence**: 2 000 of 2 000 concepts carry each — zero gaps in the sample, at every era, in every subject part;

— **length**: English median 18 words, French median 19 — one word apart, and both one to three words above the ISO corpus median, which is the convergence finding of clause Clause 7;

— **notes**: the overwhelming majority of the corpus's 1 285 notes attach to the authoritative entries; the note layer serves the definitions, and the definitions live in two languages only; and

— **independence**: the French definitions are authored against the French usage of electrotechnical vocabulary, not derived by translation default — the corpus gives no mechanical translation signal, and the near-equal length distributions are consistent with parallel authoring under a common norm.

For a register choosing the definition-bearing model, the pair is the evidence that the model scales to two languages without drift — provided the parallel authority is genuinely parallel. The model's cost is visible in the pair's exclusivity: no third language sustains definition authoring in the record set, which is the boundary that EFG 2519's model-declaration doctrine expects a register to know about itself.

6.7 A term-only entry set, anatomized

One concept record shows the term-only layer in its ordinary form:

```
# concept 102-01-01, language entries beyond the authoritative pair
pol:
  terms:
    - type: expression
      normative_status: preferred
      designation: równość
  language_code: pol
  entry_status: Standard
  date_accepted: '2009-11-01'
jpn:
  terms:
    - type: expression
      normative_status: preferred
      designation: 等しいこと，等式
  language_code: jpn
  date_accepted: '2011-05-01'
```

Figure 3

Each entry carries exactly what the term-only model requires: the designation set with type and status, the language code, the acceptance date. What it does not carry — a definition, notes, an equivalence note — is what the model assigns to the authoritative layer. The review's measures respect the division: presence rates for the entry elements are computed with the model in view, and the absence of a definition in `pol` is not scored as an empty definition but read as the declared form.

The anatomy also shows the model's audit surface: each term-only entry names its reviewer's question — **can this designation carry the intension of the authoritative definition in this language?** — and carries no field for the answer beyond the designation itself. The equivalence worksheet of EFG 2519 is the instrument that makes the answer explicit when the question is raised, and the corpus shows the question raised rarely: the designations are old, accepted, and stable.

© EFS 2026 – All rights reserved

6.8 Definition length in the authoritative languages

Table 14

Language	Median definition length	Coverage
English	18 words	2 000 of 2 000 entries
French	19 words	2 000 of 2 000 entries

The authoritative definitions sit one word above the ISO corpus median (16 words). The near-equality of independent systems is a strong calibration signal for the length guidance of EFS 2522: two organizations, authoring independently, converge on the same working norm of roughly 16 to 19 words.

6.9 Era profile

The IEV concepts distribute across the era bands by their English acceptance dates:

Table 15

Era	Concepts	Reading
pre-1998	575	the mature core, predating the modern drafting regime
1998–2014	711	the transitional growth band
2015 and later	714	the modern maintenance band

Unlike the ISO corpus, the IEV shows no era gradient in definition presence — English and French definitions are universal at every era. The maintenance model, in which one committee owns the authoritative entries continuously, produces what the ISO document-production model can only approach: uniform presence of the defining elements. The era gradient that the ISO corpus shows for presence is a gradient of the production system, not of multilingual maintenance.

6.10 What the IEV corpus lacks

The uniform strengths have a uniform cost, visible in what the entries lack:

— **examples are absent** (0,1 %) — the term-only model concentrates effort on designations, and no layer of the vocabulary produces examples;

— **notes are thin** (5,4 % of entries) — concentrated in the authoritative languages, where they serve the definitions; the designations of the term-only languages carry almost no notes;

— **grammar information is absent** — the current record format does not carry it; and

— **normative status is uniformly** `preferred` — one value across 23 936 entries, which is uniformity of a form that carries no information: the admission, deprecation and replacement lifecycle that ISO practice records is not exercised in the corpus as extracted.

These lacks are the mirror of the ISO findings: where the production system under-produces (examples), the maintenance system under-produces more; where the production system is uniform (sources, status), the maintenance system matches or exceeds it.

7 Comparison of the two corpora

7.1 The systems compared

The two corpora embody two answers to the same question — how an international vocabulary should be built and kept — and their comparison is the most consequential part of this review:

© EFS 2026 – All rights reserved

Table 16

Dimension	ISO OBP corpus	IEV corpus
production model	document production: committees write terms and definitions clauses	vocabulary maintenance: one committee maintains one vocabulary
authoritative languages	English (per document)	English and French, as parallel authorities
localization model	definition-bearing per document; translation at adoption	definition-bearing (eng, fra) plus term-only (17 languages)
unit of authority	the document	the concept record
uniformity mechanisms	the drafting rules (ISO/IEC Directives Part 2)	the maintenance process of TC 1

7.2 Where the systems agree

Four agreements stand out, and each strengthens the calibration of the ES requirements:

— **the definition length norm.** ISO median 16 words; IEV English median 18, French 19. Two independent systems converge on the same working length. The length guidance of EFS 2522 — genus first, one to two delimiting characteristics, no padding — describes a norm that the international system already lives.

— **definitions describe.** Requirements language inside definitions is 0,35 % (ISO) and effectively zero (IEV). Both systems keep the boundary between prescribing and defining. The rule that a definition states what the concept is, not what shall be done with it, is followed almost perfectly at scale.

— **term type is uniform.** Both corpora record term type at 100 % with `expression` as the working default. Type annotation is a solved problem where the record format carries it.

— **presence of the defining element is achievable.** The IEV holds 100 % definition presence in its authoritative languages across all eras; the ISO modern era approaches it at 99,8 %. Universal definition presence is not an aspiration beyond the practice of international bodies; it is their demonstrated modern state.

7.3 Where the systems differ

— **examples.** ISO 2,9 %; IEV 0,1 %. Neither system produces examples, and the vocabulary-maintenance system produces fewer than the document-production system. The example deficit is systemic — it is the single largest gap between current international practice and the ES entry-quality requirements, which treat examples as a scored element.

— **notes.** ISO 34,2 % of entries; IEV 5,4 %. The document culture annotates; the vocabulary culture concentrates annotation in the authoritative languages. Notes travel poorly across a term-only localization layer, because a note written against a definition has no place in a designation record — the structural reason for the gap.

— **status richness.** ISO records admitted and deprecated statuses in the source documents; the IEV extraction carries a single uniform `preferred`. The lifecycle information that lets a consumer know which designation to avoid is present in one system and absent in the other.

— **structural discipline.** Circularity is measurable only in the ISO corpus (3,5 %, falling by era); the IEV authoritative definitions were not separately scored for circularity in this review cycle, and the review notes the gap in its own coverage (clause Annex D).

7.4 The two localization models

The IEV corpus is the empirical demonstration of the two localization models that the ES series names:

— **the definition-bearing model**: each localization carries its own definition; equivalence is recorded; review covers terminology method as well as language. The IEV's English–French pair is a functioning instance at scale — two parallel authorities, each complete.

© EFS 2026 – All rights reserved

— **the term-only model**: the authoritative definitions serve all localizations; each localization carries designations, status and conversion data. The IEV's seventeen other languages are a functioning instance at scale — median 12 languages per concept without a single localized definition.

The comparison yields the practical finding that EFG 2519 builds on: **both models work at international scale**, and the choice between them is a governance choice — how much definition authoring a language community can sustain — not a correctness choice. The failure modes differ: the definition-bearing model fails through drift between localized definitions and their sources (the equivalence problem), and the term-only model fails through designation fit (the position problem). The register declares its model and reviews for its model's failure mode.

7.5 A decision table for registers

The comparison converts into the choice that a register operator faces, with the corpus evidence for each branch:

a) Localization model selection by register situation

Table 17

Register situation
Model indicated
Corpus evidence
One maintenance committee, definitions authoritative in one or two languages, translation capacity limited
term-only
IEV non-authoritative languages: 17 languages sustained at median 12 per concept without any localized definitions
Two or more language communities able to author definitions independently
definition-bearing, per language
IEV eng–fra pair: 100 % presence in both, parallel length norms, no drift signal in the sample
Localization community with terminology competence and a reviewed equivalence process
definition-bearing, with recorded equivalence
the EFS 2504 apparatus; no corpus instance at scale beyond two languages — the standing gap of this clause
Mixed capacity across languages
declared model per register, never mixed silently
the Russian residue: partial definition authoring under a term-only declaration is detectable but ungoverned

The table's last row is the finding that generalizes: partial adoption of the other model happens whether or not it is declared — the corpus carries 21 live instances — and the governance question is whether the register knows. The ES requirement that a register declare its localization model exists because the alternative is not "no model" but "an undeclared one".

7.6 What each corpus contributes to the ES calibration

a) Calibration inputs and their consumers

Table 18

Calibration input
Contribution
Consumed by
definition length norm
the convergence band 16 to 19 words
EFS 2522 length anchors
presence floor
modern-era presence rates as the achievable floor
EFG 2519 monitoring benchmarks
element deficits
the examples and notes gaps as scored criteria

© EFS 2026 – All rights reserved

(continued)

EFS 2522 completeness criteria
model doctrine
both models functioning at scale with distinct failure modes
EFG 2519 clause 8, EFS 2504
lifecycle evidence
status richness in one corpus, uniformity in the other
EFS 2513 lifecycle requirements
uniformity mechanisms
process gates move rates that manuals do not
EFS 2513 clause 10, EFS 2522 clause 11

7.7 What a third system would add

The comparison is complete for two systems, and the review states what it cannot show: a corpus that operates the definition-bearing model across more than two languages — the practice that EFS 2504 requires equivalence recording for — was not available to this review cycle. The IEV demonstrates the model at two languages; the ES series requires the equivalence apparatus precisely where the model scales past the languages that a maintenance committee can double-author. The gap is a standing item for the next review.

8 Anti-pattern census

8.1 The census method

EFS 2522, Annex E, names twelve anti-patterns of concept entry authoring. The census counts their occurrence in the ISO sample where the detection rules of clause Annex C can measure them from the records. Three anti-patterns require context that the records do not carry and are marked not measurable in this cycle; the census therefore reports nine of the twelve.

The census reports rates in the ISO sample and, where the IEV record format exposes the pattern, in the IEV entries. A rate of a pattern in the records as extracted measures the pattern in the extraction; clause Clause 4 states the extraction limitation once and this clause does not repeat it.

8.2 The census

Table 19

ID	Anti-pattern	Rate	Corpus behaviour
AP-01	empty definition	1,2 %	concentrated in the legacy era (7,7 %) and nearly absent in the modern era (0,2 %)
AP-02	definitional circularity	3,5 %	falls from 4,8 % (transitional) to 2,6 % (modern); the short-definition bucket carries a disproportionate share
AP-03	requirements language in definition	0,35 %	the rarest measured pattern; the prescribe-versus-define boundary holds corpus-wide
AP-04	article-headed definition	4,3 %	flat across eras; a drafting habit, not a regime effect
AP-05	missing examples	97,1 %	the complement of the examples rate; the largest measured gap between practice and the ES requirements
AP-06	missing sources	0,0 %	universal provenance; the strongest uniform behaviour found
AP-07	missing grammar information	100,0 % as extracted	not measurable against practice: the extraction and the IEV record format carry no grammar fields

© EFS 2026 – All rights reserved

(continued)

ID	Anti-pattern	Rate	Corpus behaviour
AP-08	missing notes	65,8 %	notes are a minority behaviour in the ISO corpus and thinner in the IEV (94,6 % of entries without)
AP-09	unsupported preferred status	not measurable	requires the full document context of each entry; deferred
AP-10	inconsistent coordinate designation form	not measurable this cycle	requires arrangement-level designation comparison across rakes; deferred to the arrangement review tooling
AP-11	definition longer than the genus requires	8,4 % over 32 words	measured as the length tail rather than judged excess; the tail concentrates in systems documents
AP-12	unmarked modifications of adopted definitions	not measurable this cycle	requires SOURCE-field comparison with the adopted documents; deferred

8.3 What the census shows

Three conclusions follow from the census as a whole:

— **The rare patterns are the ruled ones.** The patterns that the drafting rules enforce hardest — sources, prescribed language in definitions — are at or near zero. The patterns that depend on author initiative — examples, notes — are far from their ES targets. Enforced behaviour holds; discretionary behaviour lapses.

— **The era gradient is the enforcement gradient.** Every pattern with a strong era signal (AP-01, AP-02) tracks a rule that tightened; every pattern without one (AP-04) tracks a habit no rule addresses. The census is evidence for the review's wider proposition: quality follows what the process gates, not what the manuals recommend.

— **The unmeasured third is a tooling agenda.** AP-07, AP-09, AP-10 and AP-12 need arrangement-level or document-context tooling that the current review lacks. They are the worklist for the next cycle, and their absence from the census is stated so that the census cannot be read as all-clear.

8.4 The IEV entries against the census

The IEV record format exposes three census patterns:

— **AP-01 (empty definition)**: by design — the term-only entries are not defects; the declared model decides (clause Clause 7).

— **AP-05 (missing examples)**: 99,9 % — deeper than the ISO corpus, for the structural reason that the term-only layer carries no definition for an example to attach to.

— **AP-08 (missing notes)**: 94,6 % — notes concentrate in the authoritative entries; the term-only layer is nearly bare.

The IEV column of the census therefore reads differently from the ISO column: the same numbers mean different things under different declared models, and the census distinguishes them rather than adding them.

8.5 Reading the census as a register operator

The census rates are not targets to be met individually; they are the profile of the system whose gates the ES series redesigns. A register operator reads the census for three uses:

— **as the baseline**. The rates state what unmanaged practice produces at scale: one entry in twenty-nine circular, one in twenty-nine empty (legacy-weighted), examples on one entry in thirty-five. A register's own rates are meaningful only against this baseline — a 3 % circularity rate is excellent against the transitional era and weak against the modern one, and the baseline choice is a governance statement.

© EFS 2026 – All rights reserved

— **as the checklist input**. The nine measured patterns map onto the review checklist of EFS 2522 Annex A in the same order, so a reviewer can work the checklist knowing how often the corpus fails at each line — the patterns the corpus fails often deserve the reviewer's first attention, and the patterns the corpus never fails (sources) deserve little.

— **as the scoring context**. The EFS 2522 scoring model weights the dimensions; the census rates show where the corpus mass sits relative to those weights. The completeness dimension carries 20 % of the score while the corpus fails it on 97 % of entries — the weighting is a statement about where a conforming termbase should differentiate itself from the corpus, and the census is the quantified statement of that differentiation.

8.6 Defect interaction

The measured patterns are not independent, and two interactions matter for interpretation:

— **short definitions and circularity co-occur.** The 1-to-5-word bucket holds 4,6 % of the definitions but a disproportionate share of the circular entries: a definition that short rarely has room for a full delimiting differentia, and the designation's own words fill it. The interaction says the substitution test matters most where definitions are tightest — the opposite of the intuition that short definitions are safe because simple.

— **notes substitute for examples imperfectly.** Entries with notes carry explanations; entries with examples carry instances. The corpus's heavy-annotator documents raise notes without moving examples, which shows the two elements draw on different authoring efforts and are not interchangeable. A register cannot close the example gap by annotating more, and the scoring model treats them separately for exactly this reason.

9 Implications for the ES Terminology series

This clause states what the findings mean for the normative documents of the series. The statements are informative conclusions; the normative texts carry the requirements.

9.1 For EFS 2522 (concept entry quality)

— **The length anchors hold.** The corpus medians — 16 words (ISO), 18 and 19 (IEV eng/fra) — sit inside the scoring anchors that EFS 2522 sets for definition length. No recalibration is indicated; the report refreshes the anchor evidence.

— **The substitution test is justified.** A 3,5 % circularity rate in the largest vocabulary production system, falling but persistent, is the case for keeping the substitution test as a mandatory-fix criterion rather than a scored nicety.

— **The example criterion is aspirational and should stay so.** At 2,9 % and 0,1 %, current international practice does not produce examples. EFS 2522 scores examples as a completeness criterion; the score weight is correct precisely because the corpus shows the gap — a conforming termbase is meant to out-produce the corpus here, and the scoring gives it credit for doing so.

— **The article-head criterion stays.** The era-flat rate marks a habit; a habit responds to review, and the clarity criteria of EFS 2522 review it.

— **Grammar fields need an evidence source.** Both corpora are silent on grammar in their extracted record formats. EFS 2522 requires grammar information; the requirement rests on ISO 12620-class practice rather than on these corpora, and the next review cycle should seek a corpus that carries grammar fields before adjusting any anchor.

9.2 For EFS 2525 (concept arrangement)

— **Criteria are the gap.** The corpora exhibit rakes whose criteria are implicit; the requirement to record the criterion of subdivision on every membership relation (EFS 2525, A-2, Q-2) is the response to the largest arrangement-level silence the corpora show.

© EFS 2026 – All rights reserved

— **Residual discipline is needed at scale.** The "other"-class node appears across the corpora at every depth. The residual category rule of EFS 2525 (M-6) — declare, record as limitation, prefer positive criteria — is the discipline the corpus practice lacks.

— **Cross-system reality.** The IEV's tiered language coverage and the ISO corpus's overlapping document vocabularies show that concepts live in several systems. The EFS 2525 cross-system requirements (A-9 to A-13) formalize what the corpora do informally.

9.3 For EFG 2519 (register operation guide)

— **The benchmarks are set.** A conforming register compares its defect rates against the era profiles: empty definitions below the modern-era 0,2 %, circularity below 2,6 %, and example provision above the corpus rate are the measurable targets the guide's monitoring practice consumes.

— **The declared-model doctrine is confirmed.** The IEV operates both localization models in one vocabulary and demonstrates the failure mode of each; the guide's requirement to declare one model per register and review for its failure mode is the operational lesson of clause Clause 7.

9.4 For the series as a whole

Three conclusions close the review:

— **Enforcement beats exhortation.** Every corpus behaviour that a process gates (sources, term type, definition presence in the modern era) is at or near uniform; every behaviour left to author initiative (examples, notes) lapses. The ES series gates publication (EFS 2513, clause 10) and scores entries (EFS 2522, clause 11) because gating and scoring move the rates that manuals cannot.

— **Machine-readable records make quality measurable.** This review exists because two vocabularies publish structured records. Every requirement of the series that pushes registers toward published machine-readable records is also a requirement that makes the next edition of this review deeper — the registry becomes the corpus.

— **The review is a cycle.** The census names four patterns this cycle could not measure; the corpora age; the series changes practice and the practice changes the corpora. The next review cycle repeats the method of clause Clause 4 over refreshed extractions and over ES's own registers, and the reproduction instructions of Annex E are written for exactly that reuse.

9.5 Stewardship of the evidence base

The corpora that made this review possible are maintained by others, and the ES series holds them as an entrusted instrument:

— **the extractions are cited by state.** Every edition of this review names its corpus states — extraction dates, record counts — so that the communities maintaining the platforms can see exactly what was measured and correct the record where the extraction misreads their published content;

— **the analysis gives back.** The detection rules of Annex C and the reproduction protocol of Annex E are offered to the corpus communities: a committee that adopts the rules can measure its own vocabulary with the instrument this review used, and compare against the published baselines on equal terms; and

— **the next cycle is scheduled with the series.** The review repeats when the corpora change materially or when the series publishes requirements whose calibration needs new evidence, whichever comes first; the coverage gaps of this cycle are the agenda of the next.

© EFS 2026 – All rights reserved

Annex A
(normative)

ISO corpus statistics tables

This annex carries the full statistics of the ISO sample (4 662 localized concept entries). Each table restates or extends a finding of clause Clause 5.

A.1 Definition length histogram

Table A.1

Words	Entries	Share
1–5	214	4,6 %
6–10	803	17,2 %
11–16	1 415	30,3 %
17–22	1 005	21,6 %
23–32	784	16,8 %
33–50	312	6,7 %
51 and more	73	1,6 %

A.2 Presence statistics

Table A.2

Measure	Count	Rate
entries	4 662	—
with definition	4 606	98,8 %
without definition	56	1,2 %
with notes	1 594	34,2 %
with examples	135	2,9 %
with sources	4 662	100,0 %
with circularity marker	161	3,5 % of defined entries
requirements language	16	0,35 % of defined entries
article-headed	198	4,3 % of defined entries

A.3 Era profile

Table A.3

Era	Entries	Defined	Empty	Empty %	Circular n	Circular %
legacy (1983–1997)	405	374	31	7,7	17	4,5 %
transitional (1998–2014)	1 667	1 649	18	1,1	79	4,8 %
modern (2015–2026)	2 590	2 583	7	0,2	67	2,6 %

© EFS 2026 – All rights reserved

A.4 Term type distribution

Table A.4

Term type	Count	Share
expression	4 662	100,0 %
NOTE The extraction carries a single term type value. Source documents distinguish abbreviations and symbols in their entry structure; the platform normalization folds them into the expression value in the extracted records.		

A.5 Concentration: the twenty largest source documents

Entries per source document, with the defect counts the review measures. The document identifier is the platform source key.

Table A.5

Source document	Entries	Empty	Circular	Notes
iso-iec.2382.2015	68	0	0	68
iso.10303-2.2024	61	0	5	43
iso-iec-ieee.24765.2017	35	0	1	9
iso.7348.1992	27	0	0	0
iso.5598.2020	27	0	0	6
iso-iec-ieee.24765.2010	24	0	2	3
iso.6707-1.2020	24	0	0	1
iso.472.2013	24	0	0	8
iso.5593.2023	24	1	0	24
iso.4135.2022	18	0	0	3
iso.17769-1.2012	15	0	0	9
iso.1382.2012	15	0	0	3
iso.4618.2014	15	0	0	0
iso-iec.19762.2016	15	0	0	9
iso.772.2011	14	0	0	0
iso.5598.2008	14	0	0	2
iso.9000.2015	12	0	0	8
iso.14040.2006	12	0	0	4
iso.22300.2018	12	0	0	6
iso.5127.2017	12	0	0	5

The twenty documents carry 489 entries (10,5 % of the sample). The circular counts concentrate: `iso.10303-2.2024` alone carries five — the largest single-document circularity count in the sample — and the two vocabulary-integration standards (24765 editions) carry three between them.

A.6 Sampling error

At n = 4 662, the 95 % confidence half-width for a rate p is approximately $1{,}96 \times \sqrt{p(1-p)/n}$. For the headline rates:

Table A.6

Rate	Estimate	95 % half-width
definition presence	98,8 %	±0,10 pp
circularity	3,5 %	±0,53 pp
notes presence	34,2 %	±1,36 pp

© EFS 2026 – All rights reserved

(continued)

Rate	Estimate	95 % half-width
examples presence	2,9 %	±0,48 pp

The IEV rate half-widths at the concept level (n = 2 000) are approximately twice these values.

A.7 Notes and examples by document tier

The concentration finding, decomposed: entries grouped by the size class of their source document, with the element rates per class.

a) Entry elements by source-document size class

Table A.7

Document size class
Entries
Notes %
Examples %
Sources %
large (24 entries and more)
431
62,4
7,7
100,0
medium (12 to 23 entries)
223
30,5
1,3
100,0
small (fewer than 12 entries)
4 008
31,0
2,5
100,0
NOTE The small class aggregates hundreds of documents; its rates are document-mixture rates, not per-document rates. The large class is the vocabulary-standard tier of clause Clause 5.

© EFS 2026 – All rights reserved

Annex B
(normative)

IEV corpus statistics tables

This annex carries the full statistics of the IEV sample (2 000 concept records, 23 936 language entries).

B.1 Languages per concept: full distribution

Table B.1

Languages	Concepts
2	2
7	12
8	1
9	99
10	323
11	560
12	364
13	275
14	140
15	69
16	58
17	71
18	8
19	18

The mode region is 10 to 14 languages (1 662 concepts, 83,1 %). The thin lower tail (7 to 9 languages, 112 concepts) marks partially adopted concepts; the upper tail (15 to 19, 97 concepts) marks the most internationalized concepts.

B.2 Language coverage with definition status

Per language: entries, definition-bearing entries, and the median length of the definitions present.

Table B.2

Code	Language	Entries	Definitions	Median words
eng	English	2 000	2 000 (100,0 %)	18
fra	French	2 000	2 000 (100,0 %)	19
pol	Polish	1 998	0	—
por	Portuguese	1 998	0	—
jpn	Japanese	1 991	0	—
deu	German	1 920	0	—
kor	Korean	1 912	0	—
ara	Arabic	1 810	0	—
ita	Italian	1 766	0	—
spa	Spanish	1 714	0	—
swe	Swedish	1 238	0	—
zho	Chinese	1 137	0	—

© EFS 2026 – All rights reserved

(continued)

Code	Language	Entries	Definitions	Median words
fin	Finnish	834	0	—
srp	Serbian	619	0	—
slv	Slovenian	250	0	—
mon	Mongolian	198	0	—
rus	Russian	190	21 (11,1 %)	15
nob	Norwegian Bokmål	158	0	—
nno	Norwegian Nynorsk	158	0	—
ces	Czech	25	0	—
nld	Dutch	20	0	—

B.3 Entry element coverage

Table B.3

Measure	Count	Rate of entries
entries	23 936	—
with notes	1 285	5,4 %
with examples	24	0,1 %
with normative status `preferred`	23 936	100,0 %
with grammar information	0	0,0 %
with script annotation	0	0,0 %

B.4 Concept acceptance era

Table B.4

Era (English acceptance)	Concepts	Share
pre-1998	575	28,8 %
1998–2014	711	35,5 %
2015 and later	714	35,7 %

The near-uniform distribution across the three bands marks a vocabulary under continuous maintenance rather than one built in a single conversion: a quarter of the sampled concepts predate the modern drafting regimes entirely, and a third arrived within the last decade of the extraction.

B.5 The Russian residue

Russian is the only language whose entries carry definitions besides the authoritative pair: 21 of 190 entries (11,1 %), median 15 words. Under the declared term-only model for Russian, these definitions are anomalies of the record set — most plausibly the residue of a partial definition-authoring effort. The review records them without removal: the corpus is reported as extracted, and the anomaly is a finding about the record set's history.

B.6 Language tier coverage

The tier analysis of clause Clause 6, with exact coverage per tier from the sample.

a) Language coverage by tier

Table B.5

Tier

© EFS 2026 – All rights reserved

(continued)

Languages
Concepts covered (max)
Coverage range
authoritative
eng, fra
2 000
100,0 %
deep
pol, por, jpn, deu, kor
1 998
95,6 % to 99,9 %
broad
ara, ita, spa, swe, zho, fin
1 810
41,7 % to 90,5 %
partial
srp, slv, mon, rus, nob, nno, ces, nld
619
1,0 % to 31,0 %

B.7 Notes distribution across models

Every notes-bearing entry in the sample carries an authoritative language code: 642 English and 643 French entries, 1 285 in total, with zero notes in the term-only layer. The note element belongs to the definition-bearing model in this corpus, entirely: a note written against a definition has no carrier in a designation-only record, and the maintenance system produces none. The finding sharpens the model contrast of clause Clause 7 — under the term-only model the annotation culture of the authoritative languages does not leak outward at all.

© EFS 2026 – All rights reserved

Annex C
(normative)

Detection rules

This annex states the operational definition of every measure. Each rule is stated so that an implementer can reproduce the count exactly. The rules were applied uniformly to every entry of the samples.

C.1 R1: definition presence

An entry is definition-bearing when its definition field exists and, after extraction, carries a non-empty text value. A definition field that is present but empty, or a list field with no content items, makes the entry term-only.

C.2 R2: definition length

The definition text is the concatenation of the content items of the definition field, in order, joined with single spaces. The length is the count of whitespace-separated tokens. Markup tokens — stem expressions in AsciiMath or LaTeX notation, inline formatting — count as tokens; a symbolic definition therefore overstates its word length, and the finding of clause Clause 5 treats length as context, not as a defect measure.

C.3 R3: circularity

An entry is circular when the lowercase form of the preferred designation (or, for multi-word designations, any word of the designation of length four or more) occurs within the lowercase definition text. The rule is deliberately broad: it counts “terrain analysis: analysis of terrain” and ” “drain: to drain” alike. A stricter rule would demand the full designation; the reported rate is an upper bound on true circularity and a lower bound on agreement, and the same rule was applied to every era, which is what the era comparison consumes.

C.4 R4: requirements language

An entry carries requirements language when its definition text contains any of shall, should or must as a word, case-insensitive. The rule does not parse modality; quoted uses count. At the observed rates the distinction is immaterial to the finding.

C.5 R5: article head

An entry is article-headed when its definition text, after leading whitespace, begins with a, an or the followed by a space, case-insensitive.

C.6 R6: notes and examples presence

An entry carries notes when its notes field carries one or more non-empty items; the same rule applies to examples.

C.7 R7: sources presence

An entry carries sources when its source field carries one or more items. The rule does not validate the cited document.

© EFS 2026 – All rights reserved

C.8 R8: grammar information

An entry carries grammar information when its designation record carries a non-empty grammar field (part of speech, or gender or plurality). The rate measures the record set, not the practice that produced the source documents.

C.9 R9: term type and normative status

The term type is read from the designation record's type field; the normative status from the designation record's status field. The values are reported as found, without normalization beyond the extraction's own.

C.10 R10: era assignment (ISO)

The era of an entry is the era of its source document, assigned by the latest four-digit year appearing in the source document key. The assignment is coarse; clause Clause 4 states the limitation.

C.11 R11: era assignment (IEV)

The era of a concept is assigned from the date accepted of its English entry.

C.12 R12: languages per concept (IEV)

The count is the number of language keys of the concept record that carry at least one designation. Empty language keys do not count.

C.13 Worked examples of the rules

The examples show the rules applied, so that an implementer can calibrate against the intended behaviour rather than the prose.

R2 applied: the definition "part of a computer that carries out arithmetic and logical operations and controls the execution of instructions" yields 17 tokens. A definition containing a stem expression yields one token per stem symbol — the measure's known overstatement, stated in the rule.

R3 applied: designation `drain valve`, definition "valve for draining fluid from a vessel". The head noun `valve` (length five) occurs in the definition; rule R3 marks the entry circular, although the definition states a proper genus-plus-differentia form. The rule trades precision for coverage, uniformly; the census and the era profile consume the uniform overcount.

R3 not applied: designation `drain`, definition "to remove fluid by drainage" — the word `drain` does not occur (the definition carries `drainage`, a different token under exact matching), so the rule does not mark it; the stem echo class of clause Clause 5 names this boundary case.

R5 applied: definition "A device that measures voltage" begins with `A ` — article-headed. Definition "Device for measuring voltage" does not — the genus stands first, and the entry passes.

R6 applied: a notes field of one non-empty item counts; an empty list does not; a list of one empty string does not.

R10 applied: source key `iso.10303-2.2024` carries the latest year 2024; the era is modern. Key `iso.7348.1992` carries 1992; the era is legacy.

C.14 Rule change policy

A rule change between editions invalidates the affected trend comparisons, and the review constrains changes accordingly:

© EFS 2026 – All rights reserved

— a rule may be tightened or loosened only in a new edition, with the change stated in that edition and both rates reported for the first edition in which both are computable;

— era comparisons within one edition use one rule set, without exception — the era findings consume uniformity, and a mixed rule set would manufacture or mask gradients; and

— the rules of this annex are the baseline: any future change is stated as a diff against them, not as a replacement text.

The policy exists because the review's authority rests on reproducibility; a rule that changes silently between editions would make every trend line negotiable.

C.15 Rule sensitivity summary

Table C.1

Rule	Sensitivity	Direction of bias
R3 circularity	broad	overcounts true circularity; uniform across eras, which is what the era comparison consumes
R4 requirements language	no parsing	negligible at observed rates
R10 era	coarse	attenuates era differences; observed gradients are lower bounds

© EFS 2026 – All rights reserved

Annex D
(normative)

Reproduction

This annex states how every figure in this report regenerates from the corpora. The analysis is a set of scripts run over the extracted corpora; the seeds and the corpus states are the two things a reproduction holds fixed.

D.1 Corpus states

Table D.1

Corpus	State	Scale
ISO OBP extraction	Glossarist-form concept records extracted from the ISO Online Browsing Platform	406 898 records
IEV maintenance records	Glossarist-form concept records of the IEC International Electrotechnical Vocabulary, as maintained by IEC TC 1	22 229 records

The ISO extraction and the IEV record set are held in the ES research environment. The corpora are third-party content; the extraction and the analysis are internal to ES, and the report publishes statistics only.

D.2 Sample seeds

— ISO: `random.sample(records, 3000)` with `random.seed(2522)`, analysing all localized concept entries in the sampled records — 4 662 entries in this corpus state.

— IEV: `random.sample(sorted(records), 2000)` with `random.seed(2522)` — 2 000 concepts, 23 936 entries.

The seed value 2522 is the document number of EFS 2522, whose first calibration cycle introduced the analysis. Future cycles may change seeds; the seed in force is stated with the edition of this report.

D.3 Measures implementation

Each rule of Annex C is a direct implementation over the parsed records:

— definition presence and length: rules R1 and R2 over the definition field of each language entry;

— circularity, requirements language, article head: rules R3 to R5 over the definition text;

— presence measures: rules R6 to R9 over the corresponding fields; and

— era and language measures: rules R10 to R12 over the record metadata.

The implementations are intentionally single-pass and dependency- free: a reviewer with the corpus directory, a YAML parser and the rules of Annex C reproduces every rate without the original tooling.

a) Analysis passes and their outputs

Table D.2

Pass
Input and rule set
Output
ISO entry pass

© EFS 2026 – All rights reserved

(continued)

sampled records; rules R1 to R10
presence, length, defect and era rates per entry
ISO concentration pass
full sample, grouped by source key
per-document counts of Clause A.5 and Clause A.7
IEV concept pass
sampled records; rules R1 to R9, R11, R12
per-language coverage, definition status and era rates
IEV element pass
sampled records; rules R6, R9
notes and status distributions of Clause B.3 and Clause B.7

D.4 Regeneration protocol

A regeneration for a future edition of this report proceeds:

a) refresh the extractions from the live platforms and record the extraction dates and record counts;

b) re-run the sampling with the seed in force and record the sample sizes;

c) re-apply the measures of Annex C, unchanged unless a rule change is documented in the edition;

d) compare the new rates with the rates of this edition, with the sampling error of Clause A.6 as the yardstick; and

e) state the rule changes, the corpus state changes and the rate changes in the new edition, so that every difference is attributable.

D.5 Known coverage gaps for the next cycle

The reproduction protocol carries the review's own improvement list:

— circularity scoring for the IEV authoritative definitions (clause Clause 7, gap);

— grammar-information measurement from a record format that carries grammar fields (both corpora lack it as extracted);

— anti-patterns AP-09, AP-10 and AP-12 (clause Clause 8), which need document-context and arrangement-level tooling; and

— a definition-bearing multilingual corpus beyond two languages, for the equivalence-model evidence that clause Clause 7 names as missing.

© EFS 2026 – All rights reserved

Bibliography

[1] ISO 704, *Terminology work — Principles and methods*.

[2] ISO 1087-1[1)] , *Terminology work — Vocabulary — Part 1: Theory and application*.

[3] ISO 10241-1, *Terminological entries in standards — Part 1: General requirements and examples of presentation*.

[4] ISO 12620[2)] , *Management of terminology resources — Data category specifications*.

[5] ISO 25964-1, *Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval*.

[6] ISO/IEC Directives, Part 2, *ISO/IEC Directives, Part 2, Principles and rules for the structure and drafting of ISO and IEC documents*.

[7] EFS 2504:2025, *Terminology management — Multilingual terminology*

[8] EFS 2510:2025, *Terminology management — Concept systems and concept relations*

[9] EFS 2513:2026, *Terminology management — Concept management systems — Requirements*

[10] EFS 2516:2026, *Terminology management — Concept data interchange*

[11] EFG 2519:2026, *Concept registers — Operation guide*

[12] EFS 2522:2026, *Terminology management — Concept entry quality — Guidelines, requirements and scoring*

[13] EFS 2525:2026, *Terminology management — Concept arrangement in concept systems — System membership, relationship establishment and qualification, and mutual exclusivity of subdivision*

[14] ISO OBP, *ISO Online Browsing Platform*, International Organization for Standardization, https://www.iso.org/obp/

1) Withdrawn.

2) Withdrawn.

© EFS 2026 – All rights reserved

enosema

Price based on 32 pages

© EFS 2026
All rights reserved

Making meaning meaningful