CESSDA Work Plan 2020

Ontology Management System

D3: ELSST Translation guidelines

Document info

Dissemination Level	PU
Due Date of Deliverable	30/11/2020
Actual Submission Date	V1: 03/11/2020 V1.2: 01/12/2020 (resubmitted after CESSDA MO review) V1.3 01/03/2021 (resubmitted after CESSDA WGL review)
Type	Report
Approval Status	Approved by CESSDA Tools & Services Working Group leader Mari Kleemola and CESSDA Technical Working Group leader John Shepherdson
Version	V1.3
Number of Pages	p.1 – p.35
DOI	10.5281/zenodo.4727779

Version history

Version	Date	Comment	Revised by
0.1	20.10.2020	First draft produced by UKDS partner in Ontology Management System team	
0.2	28.10.2020	Comments received from FSD partner in Ontology Management System and UKDS team members	UKDS
1.0	02.11.2020	First version produced by UKDS partner in Ontology Management System team	
1.1	27.11.2020	Review by CESSDA MO	
1.2	01.12.2020	Edits made following CESSDA MO review	UKDS
1.3	01.03.2021	Edits made following CESSDA WGL review	UKDS

Author List

Organisation	Name	Contact information
UK Data Service	Lorna Balkan	balka@essex.ac.uk
Finnish Social Science Data Archive	Taina Jääskeläinen	taina.jaaskelainen@tuni.fi

Peer-review

Organisation	Name	Contact information
UK Data Service	Sharon Bolton	sharonb@essex.ac.uk
UK Data Service	Jeannine Beeken	jeannine.beeken@essex.ac.uk

Contents

Executive Summary

This guide describes the translation procedures for the European Language Social Science Thesaurus (ELSST). It is intended for registered translators or prospective translators of ELSST. The document describes how ELSST translation work is organised and the training provided to translators. It includes information on how a thesaurus is constructed and organised and how relevant components such as concepts, alternative terms and associated notes are translated. Advice on translation, such as clarification and disambiguation of thesaurus terms and preserving cross-language equivalence relationships is also covered. ELSST is currently managed using the VocBench vocabulary platform and published via Skosmos. In addition to the translator guide, a bespoke training programme is also provided for each new translator.

Abbreviations and Acronyms

BARTOC	The Basel Register of Thesauri, Ontologies & Classifications
CAB	Commonwealth Agricultural Bureaux
CESSDA	Consortium of European Social Science Data Archives
ECLAS	European Commission Libraries Catalogue thesaurus
ELSST	European Language Social Science Thesaurus
EMIRE	European Employment and Industrial Relations
ETT	European Training Thesaurus
FINTO	Finnish thesaurus and ontology service
FSD	Finnish Social Science Data Archive
GEMET	General Multilingual Environmental Thesaurus
GESIS	GESIS – Leibniz Institute for Social Sciences
GOC	Government of Canada
IATE	InterActive Terminology for Europe
ILO	International Labor Organization
ISCED	International Standard Classification of Education
ISI	International Statistical Institute
ISO	International Standards Organization
LRE	Learning Resources Exchange for schools thesaurus
NSD	Norwegian Centre for Research Data

OECD	Organisation for Economic Co-operation and Development
OWL	Ontology Web Language
POPIN	Population Multilingual Thesaurus
RAMON	Reference And Management Of Nomenclatures
RDF	Resource Description Framework
SKOS	Simple Knowledge Organization System - home page
SKOS-XL	SKOS eXtension for Labels
TESE	Thesaurus for Education Systems in Europe
TMS	Thesaurus Management System
XKOS	SKOS Extension for Statistics
UNBIS	United Nations Bibliographic Information System Thesaurus
UNESCO	United Nations Educational, Scientific and Cultural Organization
UKDS	UK Data Service

1 Introduction

This guide describes the translation procedures for the European Language Social Science Thesaurus (ELSST). ELSST is currently managed using the VocBench vocabulary platform and published via Skosmos. A bespoke training programme is also provided for each new translator.

2 Organisation of the translation work

2.1 Overview

ELSST translation work is carried out by members of the ELSST Thesaurus Management Team, which consists of representatives from each language of ELSST. On behalf of CESSDA, the UK Data Service (UKDS) is responsible for the source language version of the thesaurus (English) as well as for overall co-ordination of the translation work, the training of translators, and for maintaining ELSST in VocBench.

Each target language has a dedicated spokesperson who is responsible for the co-ordination of the translation of ELSST into their language, for entering the translation into VocBench, and for quality checking the translation.

Target language spokespersons may appoint other translators to help with the translation task.

All translation and quality checking must be carried out by native speakers of the target language who have an excellent knowledge of English.

2.2 Preparation and training

- Before translating ELSST, translators must attend translation training. This is organised by the UK Data Service on behalf of CESSDA and covers the content of this document and instruction on how to use VocBench.
- The translator's organisation must also have a valid ELSST licence.
- Once the translator has completed training, they will receive credentials to use VocBench as a translator, which will allow them to edit their language version in VocBench. If translation work is to be shared, more than one person may apply for credentials per language.

2.3 Shared languages

- Where more than one archive share a language, two options are available. One dialect of the language can be chosen as the language for Preferred Terms, and other dialects can be treated as Entry Terms. For example, the German version of ELSST uses German-German as the language for Preferred Terms, with Austrian variants appearing as Entry Terms. Alternatively, different dialects of a language could be treated as separate languages, although there is currently no example of this in ELSST.

The present situation in ELSST is shown in Figure 1:

Language	Language variety of Preferred Terms	Additional language varieties of Entry Terms
Czech	Czech	
Danish	Danish	
Dutch	Dutch Dutch	
English	British English	
Finnish	Finnish	
French	French French	Swiss French (culture-specific terms only (e.g. political institutions)
German	German German	Austrian German

Greek	Greek	
Lithuanian	Lithuanian	
Norwegian	Norwegian Bokmål	
Romanian	Romanian	
Slovenian	Slovenian	
Spanish	Spanish Spanish	
Swedish	Swedish	

Figure 1 Language varieties in ELSST

- Currently different varieties of a language are not distinguished in the ELSST instance of VocBench. For example, German and Austrian terms are both assigned the language tag 'de'. Similarly, although there is an option to tag English terms as British English (en-GB), the more general 'en' is used, which covers all varieties of English.
- Where a language version of ELSST is shared between different countries or archives, translators should organize a system of collaboration of translation and peer review.

3 Thesaurus elements

3.1 Overview

ELSST follows the thesaurus construction principles outlined in the latest international guidelines on thesaurus construction, ISO 25964-1. The previous ELSST Thesaurus Management System (TMS) used names for thesaurus elements based on ISO-25964-1.

ELSST is now modelled as a Simple Knowledge Organization Scheme (SKOS)-(XL) concept scheme in the VocBench management platform. Skosmos uses slightly different terminology again. Since we use the Skosmos platform for publishing ELSST, we adopt the Skosmos terminology for this guide, but present the equivalences between the different terminologies below:

VocBench	Skosmos	Previous ELSST TMS
skos:concept	Concept	Concept
skosxl:prefLabel	Preferred term	Preferred Term
skosxl:altLabel	Entry Term	Use For term/Non-preferred term
skos:broader	Broader concept	Broader term
skos:narrower	Narrower concept	Narrower term
skos:related	Related concept	Related term
skos:definition	Definition	Scope note
skos:scopeNote	Scope note	Use note
skos:editorialNote	Editorial note	Editorial note
skos:historyNote	History note	History note

xkos:additionalContentNote	Definition source	Scope note source

3.2 Concepts

Concepts are 'units of thought' that exist in a domain. Concepts are represented by terms, i.e. the names used to refer to a concept.

3.3 Preferred Terms

A Preferred Term is used to represent a concept when indexing. In ELSST, each concept should have a single Preferred Term in each language. Preferred Terms are also unique to each language version, i.e. a Preferred Term is associated with one and only one concept in each language. Strategies for translating Preferred Terms are discussed in Section 7: Cross-language equivalence in ELSST.

3.4 Entry Terms

- Entry Terms are not assigned to documents but act as pointers to Preferred Terms.
- Entry Terms are unique to each language version of ELSST, i.e. they are associated with one and only one Preferred Term.
- The set of Entry Terms is disjoint with the set of Preferred Terms in each language version of ELSST, i.e. a term cannot be both a Preferred Term and an Entry Term.
- Entry Terms are not necessarily true synonyms. They can also be quasi-synonyms, antonyms, or narrower in meaning than the term with which they have been associated. Their function is to aid information retrieval. Once ELSST is incorporated into the CESSDA Data Catalogue, users will be able to find the same data regardless of whether the Preferred Term or any of its Entry Terms are used as the search term.
- The number of Entry Terms is not necessarily the same in the source language and the target language. This is because Entry Terms may be language-specific; it may be possible to find Entry Terms in the target language even if the source language term has none. Conversely, the source language may have an Entry Term where the target language does not need any. For example, in the source thesaurus the term ADOLESCENTS has an Entry Term TEENAGERS, which has no direct translation in French.

- Where different language dialects share the same language version in ELSST (German, for example), Entry Terms may be both language- and country-specific.
- Finding relevant Entry Terms is particularly important for system-specific hierarchies like educational levels, social security etc. where the source Preferred Terms have been based on international, standardised classifications (like International Standard Classification of Education (ISCED)). This makes the terms function cross-nationally but they are not necessarily the terms that information seekers would use for searching, therefore Entry Terms are important. In the EDUCATIONAL INSTITUTIONS hierarchy, for example, there is a term SELECTIVE SCHOOLS. Each language will add its country-specific Entry Terms. English for example has GRAMMAR SCHOOLS, while French has LYCÉES.
- Where translators add an Entry Term to their language that is not an exact synonym of the target language Preferred Term or a translation of a source language Entry Term, they should report this to the ELSST Team. This Entry Term can then be evaluated and suggested to other translators for inclusion in the source language, and for translation into other languages, if appropriate.
- Where the label of a Preferred Term has changed in the source language, the new Preferred Term should be translated, but the translation of the old Preferred Term should normally be retained as an Entry Term, unless there is good reason not to (if, for example it is not politically correct, or if it is broader than the new Preferred Term). Where the translating organisation cannot update its existing metadata to the new target language Preferred Term, due to lack of resources, retaining the old Preferred Term as an Entry Term enables users to find the data using the old Preferred Term, if ELSST is implemented in their search interface. The label change for the Preferred Term should normally be recorded in the History Note (see Section 3.11).

3.5 Broader Concepts

A Broader Concept is a concept that is broader than the one in question.

3.6 Narrower Concepts

A Narrower Concept is a concept that is narrower than the one in question.

3.7 Related Concepts

A Related Concept is a concept that has an associative relationship with the one in question.

3.8 Top Terms

Top Terms are Preferred Terms representing concepts that have no Broader Concept in the thesaurus. In ELSST, a Top Term represents the Broadest Concept in a hierarchy, and is also used as the hierarchy name.

3.9 Definition

A Definition defines or clarifies the semantic boundaries of a concept as it is used in the thesaurus. This meaning is not necessarily the most common meaning associated with the term in the general language. In ELSST, the meaning of terms is closely associated with their meaning in social science research. For example, the source language Definition for HEALTH PROMOTION is:

> ACTION TO MAINTAIN THE BEST POSSIBLE HEALTH AND QUALITY OF LIFE OF THE MEMBERS OF THE COMMUNITY AND TO ENCOURAGE THEIR INVOLVEMENT IN THEIR PERSONAL HEALTH

Note: *In ELSST, a Definition has previously been used to hold many types of information. This information is in the process of being separated into Definition, Scope Note, Definition Source, and History Note. This work has already been carried out in the source language and some target languages. Work on other target languages is still ongoing.*

Definitions should be concise and to the point.

It is ELSST policy to provide as many Definitions as possible in the source language.

It is not mandatory to translate Definitions, but note:

- If the term is not self-evident in your language, it is often useful to translate the Definition as it helps clarify the meaning of the concept in your language.
- Definitions are language-independent, i.e. the target language Definition should express the same meaning as the source language Definition (no more and no less), but it does not need to be a literal translation of it. For example, the Definition for LOCAL GOVERNMENT OFFICERS in English is:
  > THOSE CARRYING OUT A SUPPORT OR ADMINISTRATIVE ROLE WITHIN A LOCAL GOVERNMENT SETTING

  It is perfectly acceptable to translate this in French as:

  > PERSONNES QUI EXERCENT UN RÔLE DE SOUTIEN OU D'ADMINISTRATION DANS UN GOUVERNEMENT LOCAL

(literally: people who carry out a support or administrative role in local government)

- To avoid copyright issues, where possible, use your own words to translate a Definition, and leave the Definition Source field blank. However, if the source language Definition has been taken from a multilingual glossary (e.g. Eurostat), you may use the official translation in that same glossary as your translation. In this case, you must provide the Definition Source, which will be identical to the source language Definition Source.
- All target language Definitions should have a source language equivalent. If you have, or wish to have, a Definition in your language, and there is no source language equivalent, request one from the ELSST Team.
- If the source language Definition does not work in your language, leave the Definition untranslated and inform the ELSST Team. If you have already translated the Preferred Term, you can make a note in the Editorial Note that your translation is incomplete, so that you can find it again easily.

3.10 Scope Note

A Scope Note provides information and guidance on how terms may be used for indexing. It may clarify the usage of the term with respect to other terms in the thesaurus. For example, the source language Scope Note for BIRTH RATE is:

> DO NOT CONFUSE WITH "CHILDBIRTH", "NATALITY", OR "FERTILITY RATE".

It is not mandatory to translate Scope Notes, but note:

- Scope Notes are language-specific, i.e. they may contain different information in each language, including culture-specific information. For example, the Finnish equivalent of BAIL, TAKUUVAPAUTUS, has a Scope Note to say that the bail system is not used in Finland.
- Scope Notes may be required in a target language where none exists in the source language or vice versa. For example, the German term ARBEITSETHIK (WORK ETHIC in English) has the Scope Note:
  > BITTE NICHT MIT DEM BEGRIFF "BERUFSETHIK" VERWECHSELN.

  (literally: DO NOT CONFUSE WITH THE TERM "PROFESSIONAL ETHICS".)

- If you add a Scope Note in your language, you do not need to inform the ELSST Team.

3.11 History Note

A History Note may record the date of introduction of a term or concept, or it may give more complex advice on how to search for the concept in earlier or later times. For example, the source language History Note for FOOD SAFETY is:

> TERM RELABELLED APRIL 2018. PREVIOUSLY THE TERM "FOOD CONTROL" WOULD HAVE BEEN USED.

It is not mandatory to translate History Notes, but note:

- History Notes are language-specific, as a target language term may be added at a later date than the source language term. Also, changes can be made to a source language term but not the target language term, or vice versa.
- It is useful to translate History Notes if you have used ELSST to create metadata.
- If you add a History Note in your language, you do not need to inform the ELSST Team.

3.12 Definition Source

The Definition Source holds the bibliographic reference to the Definition.

- Definition Sources are generally only provided for source language terms, since the translations of most Definitions are produced by translators in their own words. The exception is a translation taken from an official multilingual source, where the same Definition Source must be provided in both the source and target language (see Section 3.9 above)
- In order to provide fuller references to Definition Sources, a bibliography of Definition Sources has been provided in the ELSST online guide. For example, the source language Definition Source for HEALTH PROMOTION is 'OXFORD DICTIONARY OF NURSING', abbreviated to 'DN' in the thesaurus.

3.13 Editorial Note

The Editorial Note is a place to record any notes you wish to make during the translation process. This field is visible to all editors of VocBench, including ELSST translators of other languages, but is not published on the Skosmos version of ELSST.

4 Translation status

In ELSST a Preferred Term can have one of four translation statuses:

1. Not Started

2. Incomplete
3. Untranslatable
4. Complete

In the previous ELSST TMS the 'Not started' and 'Untranslatable' statuses were marked syntactically, as follows:

- 'Not Started' terms were of the form XX_TERM, where XX is the 2-letter ISO 3166 country code of the target language, and TERM is the source language Preferred Term. For example, ES_ENERGY PRICES was the untranslated Spanish version of the source language term ENERGY PRICES.
- 'Untranslatable' terms had '_UN' added to them, for example, RIGHT OF WAY_UN

VocBench distinguishes only between active and deprecated terms. 'Not started' terms are left blank, but there is a mechanism in VocBench to find all untranslated Preferred Terms in a language. For the other term statuses, we recommend the following:

- 'Incomplete':
  - It is useful to mark a translation as 'incomplete' if you need to return to it, or if you are collaborating with other translators and wish their input.
  - Make a note in the Editorial Note that a term is 'incomplete'. This way it can be found easily by searching the note fields.
  - Remember to remove 'incomplete' from the Editorial Note once you are happy with your translation.
- 'Untranslatable':
  - In rare cases you may not be able to find an equivalent of a source language Preferred Term in your language (See Section 7.5).
  - Use the source language term as the translation, but make a note in the Editorial Note that the term is 'untranslatable'.
- 'Complete':
  - There is no formal definition of 'complete' in ELSST. Translators may not have the time or resources to translate all elements of the source language concepts. The most important element to translate is the Preferred Term, followed by the Definition (if appropriate), and Entry Terms.
  - Before a new version of the thesaurus is released, you should ensure that all Preferred Terms have been translated, and that you have resolved all 'incomplete' translations.

5 Form of terms

5.1 Part of speech

A Preferred Term should preferably consist of a noun or noun phrase. Verbal nouns are acceptable. Use the form customary in the language involved such as gerunds in languages such as English, infinitives in languages like Dutch or nouns derived from verbs in languages like Romanian.

5.2 Singular or plural forms

Different conventions exist in different languages concerning the use of singulars or plurals in thesauri. In ELSST, each language is free to follow the thesaurus conventions for their language as far as choosing singular or plural forms of nouns is concerned.

In English, the choice of whether to use a singular or plural form of a term is based on whether a particular term is a count noun or a non-count noun.

- Count nouns are names of countable entities that are subject to the question 'How many?' but not 'How much?' For example:
    - DOMESTIC APPLIANCES
    - BIRDS
- Non-count nouns are names of concepts such as materials or substances that are subject to the question 'How much?' but not 'How many?' For example:
    - COAL
    - EQUIPMENT

Note that:

- Some nouns can have both count and non-count versions. The following examples are from English:
    - HEARING (non-count noun)
    - Meaning = 'perception by the ear or auditory sense' (OED)
    - HEARING (count noun)
    - Meaning = 'an official meeting which is held in order to collect facts about an incident or problem.' (Collins online dictionary)
- Both the count and non-count versions of the same noun may appear within the same language version of ELSST
- In English, the source language of ELSST, count nouns are expressed as plurals, and non-count nouns usually as singulars. All target languages except Slovenian have chosen to represent count nouns in their plural form. In Slovenian, the singular form

is preferred, since Slovenian has three values for grammatical number: singular, dual, and plural.

In ELSST, it is possible to find cases where a term in the singular in the source language has an equivalent in the plural form in a target language, or vice versa, because a target language equivalent with the same count as the source language is either unavailable or deemed inappropriate. The following examples are from Romanian:

- Source language count → Target language non-count
e.g. FACILITIES (plural) → INFRASTRUCTURĂ (singular)
- Source language non-count → Target language count
e.g. OVERTIME (singular) → ORE SUPLIMENTARE (plural)

Translators of new languages should contact the ELSST Team before translating to discuss what form of terms to use. In all cases, rules must be followed consistently within a language version.

5.3 Qualifiers

Qualifiers, shown as whole words in round brackets after a term, are used to disambiguate homographs (words that have the same spelling but different meanings) within the same language in the thesaurus. A qualifier should be added to each homographic term. For example:

- LABOUR (WORK)
- LABOUR (BIRTH)

Even if there is only one homographic term within the thesaurus, a qualifier may be added to promote interoperability. For example:

- COFFEE (CROP)
- AIDS (DISEASE)

In any language that contains both the singular and plural form of a term that refer to different concepts, the distinction between them may be reinforced by adding a Definition (in the case of Preferred Terms) and if possible a qualifier. If a Definition is needed, the ELSST Team should be informed so that an equivalent source language Definition can be created first.

For example, the French term STATISTIQUE (meaning 'statistics' in English) may be distinguished from STATISTQUES (meaning 'data' in English) by adding the following qualifiers:

- STATISTIQUE (SCIENCE)

- STATISTIQUES (DONNÉES)

Qualifiers should not, however, be overused, as they make terms slightly cumbersome to apply and some electronic systems have difficulties in applying them.

Qualifiers are language-specific and may not require translation into the other language. Conversely, qualifiers may be required in the target language term even if they do not exist in the source language term.

5.4 Diacritics

VocBench supports all accents and diacritics using UTF-8 encoding. Letters containing diacritics may form part of the alphabet, in some languages. Other diacritics or accents may be added to existing letters. Below is a table of the alphabets and accents found in ELSST.

Language	Alphabet used in ELSST	Additional accents which are not part of the alphabet and which are used in ELSST
Czech	Czech alphabet (42 letters)	none
Danish	Danish/Norwegian alphabet (29 letters)	none
Dutch	Basic Latin alphabet (26 letters)	none
English	Basic Latin alphabet (26 letters)	none
Finnish	Finnish alphabet (29 letters)	none
French	basic Latin alphabet (26 letters)	acute, grave, circumflex, diaeresis and cedilla
German	basic Latin alphabet (26 letters)	all umlauts (Ä, Ö, Ü), plus AE, OE, UE variants in Non-Preferred Terms
Greek	Greek alphabet (24 letters)	diaeresis and acute

Lithuanian	Lithuanian alphabet (32 letters)	none
Norwegian	Danish/Norwegian alphabet (29 letters)	none
Romanian	Romanian alphabet (31 letters)	none
Slovenian	Slovenian alphabet (25 letters)	none
Spanish	Spanish alphabet (27 letters)	none
Swedish	Swedish alphabet (29 letters)	none

Figure 2 Alphabets and accents in ELSST

Translators should remember to inform the ELSST Team of the language specific conventions (e.g. spelling conventions) for their local language.

5.6 Gender

Where a language marks a term for gender (e.g. German distinguishes between male teachers (LEHRER) and female teachers (LEHRERINNEN)), the following strategies for choosing a Preferred Term can be adopted:

- Use a general, neutral umbrella term, for example:

    LEHRPERSONAL in German, meaning TEACHING PERSONNEL
- Use the most common form, be it either masculine or feminine, for example:

    MIDWIVES in English
- Use a compound of the form 'X AND Y', or its equivalent in the target language, where both X and Y are introduced as Entry Terms. For example:

    LEHRER UND LEHRERINNEN

    Entry Terms:

    LEHRER

    LEHERINNEN

The preferred option depends on the language and term concerned, and should be decided on a case-by-case basis.

5.7 Spelling conventions

To promote consistency, a single authoritative source should be used for the spelling of Preferred Terms. Alternative spellings can be added as Entry Terms.

In ELSST, the sources used for the spelling of Preferred Terms in each language are as follows:

Language	Spelling source
Czech	Slovník spisovného jazyka českého/Slovenski pravopis
Danish	Dansk Sprognævn
Dutch	Van Dale dictionary
English	Oxford English Dictionary
Finnish	MOT Kielitoimiston sanakirja
French	Dictionnaire de français Larousse
German	Duden Online-Wörterbuch
Greek	Λεξικό της κοινής νεοελληνικής
Lithuanian	Lietuvių kalbos žodynas
Norwegian	Bokmålsordboka
Romanian	Dicționarul Explicativ al Limbi Române (DEX)
Slovenian	Slovenski pravopis

Spanish	Diccionario de la lengua española
Swedish	Svenska Akademiens ordlista

Figure 3 Spelling conventions in ELSST

6 Clarification and disambiguation of thesaurus terms

In ordinary discourse, a single term might have more than one meaning. In a thesaurus, however, terms in both the source and target languages should be as unambiguous as possible. The intended scope of a term is provided by the context provided by the hierarchy of broader and narrower concepts linked to the concept in question. If this is insufficient to clarify the intended scope, additional measures should be applied:

- The thesaurus terms (especially Preferred Terms) should be expressed as unambiguously as possible. For example, the multi-meaning term DEPRESSION could be reformulated as ECONOMIC DEPRESSION or METEOROLOGICAL DEPRESSION as appropriate.
- A qualifier may be added to the term (see Section 5.3)
- An explicit Definition should be used (see Section 3.9).

7 Cross-language equivalence in ELSST

7.1 Overview

A cross-language equivalence relationship holds between Preferred Terms that represent the same concept in different languages of the thesaurus. In the ideal situation, these terms are exact equivalents in normal discourse. In reality, this is not always the case, so strategies have to be found for how to preserve one-to-one cross-language equivalence relationships between Preferred Terms in the thesaurus.

Following ISO 25964-1, we recognise four degrees of equivalence that may be found between source and target language terms in normal discourse, and discuss ELSST strategies for how to translate these terms where this equivalence is not exact:

7.2 Exact equivalence

Exact equivalence is where "Preferred Terms are semantically and culturally equivalent" (ISO 25964-1, p. 51). In this case, there is no problem in establishing cross-language equivalence between them.

7.3 Inexact or near-equivalence

Inexact or near-equivalence is where "the terms found in different languages carry a small difference in scope, perhaps due to differences of culture, connotation or appreciation (similar to the case of quasi-synonyms within one language.) If, however, the terms are accepted as being close enough to represent the same concept when used in the thesaurus, then a Preferred Term for each language should be selected, equivalence between them established, and the scope of the concept understood as including all the cultural variations." (ISO 25964-1, p. 51).

In ELSST, inexact equivalents can be treated as if they were exact equivalents.

7.4 Partial equivalence

Partial equivalence, also known as 'broader/narrower equivalence' arises when "one language has no term that accurately represents the concept for which a label is sought, but does have a term that is normally considered to represent a broader or narrower concept. If the scope of the concept represented by one term falls completely within the scope of the concept represented by the other, the terms are said to be partially equivalent. If the difference in scope is small enough, it might be acceptable to admit the terms to the thesaurus, treating the two as equivalents that both represent the same concept (similar to the case of specific terms subsumed in a broader concept, within one language.)" (ISO 25964-1, p. 51). An example in ELSST is:

> WISSENSCHAFT in German → SCIENCE in English.

Although the German term is broader than the English term, they are treated as equivalent.

Where the target language term chosen is broader/more ambiguous than the source language term, it may be possible to add a qualifier to narrow its meaning. An example in ELSST is CONSEIL (AVIS), the French term for the English ADVICE.

ISO 25964-1, p. 51 defines a special case of partial equivalence called "**compound equivalence**" where "one of the languages offers two or more partial equivalents that, in combination, represent the whole of the concept represented by one term in the source language."

In ELSST, this type of equivalence is handled by making a combined term in the target language that consists of the two target language terms that represent the concept linked by 'and' (or more precisely the target language translation of 'and'). This is more useful for information retrieval than creating an entirely new term, as information seekers would not know the new term and would thus not use it as a search term. An example in ELSST is:

English term SINGLE-SEX SCHOOLS → German term JUNGENSCHULEN UND MÄDCHENSCHULEN

(literally: boys' schools and girls' schools)

The individual parts of the combined term (i.e. JUNGENSCHULEN and MÄDCHENSCHULEN in the above example) may be added as Entry Terms.

7.5 Non-equivalence

Non-equivalence occurs when "no term in one language can be found that even partially or inexactly represents a concept that is needed in the thesaurus" (ISO 25964-1, p. 52). This may arise because either (a) the concept is totally unknown in the target language or (b) the concept is known in the target culture but is not lexicalised.

We try to exclude from ELSST very culture-specific terms as Preferred Terms. However, in rare cases non-equivalence occurs. Translators have several options:

- Import the source language Preferred Term into the target language as a loan term, i.e. use the source language term as the target language term. The translator may mark the target language term as 'Untranslatable' in the Editorial Note to indicate that no equivalent of the source language term exists in the target language (see Section 4).
- Create an equivalent (coined) term in the target language.
- Very occasionally, the translation of a source language Entry Term can act as the target language equivalent. For example, the term HOMELESSNESS has no direct equivalent in French and Spanish, whereas its Entry Term HOMELESS PERSONS does. Thus the translation of HOMELESS PERSONS is used as the target language term in French and Spanish.

7.6 Summary

Translators should report cases of partial or non-equivalence where they are unable to find a suitable term in their language, to the ELSST Team. It may be possible to modify the scope of the concept so that it can be presented by a suitable term in each of the thesaurus languages. Alternatively, as we have seen, an inexact or partial equivalent may be accepted in one language, or a gap in one of the languages may be filled with a loan term or a coined term.

There is rarely a unique correct solution to dealing with equivalence problems. The best option is the one that best serves information retrieval. What is important is that a source language term and its target language equivalent should allow you to retrieve comparable datasets.

The structural relations between concepts are the same for all language versions, although not all concepts have a Preferred Term in each target language. This happens if translations could not be updated due to external factors. Where the structure is felt to be too culture-specific or language-specific, it should be reported to the ELSST Team as the hierarchy in question may need to be reviewed.

8 Checking the data

Once translation is complete, various quality checks should be performed on the translation before the release of a new version. They include the following:

- Completeness: Make sure all translations are as complete as possible, and remember to remove 'incomplete' from Editorial Notes.
- Translation updates: Are all translations up-to-date? In particular, have they taken into account recent additions/edits/deletions in the source language?
- Qualifiers: Have qualifiers been translated only where necessary, and inserted where appropriate?
- Cross-reference to other terms inside notes: It is easy to change a target language term without checking that all references to it in the Definitions/Scope Notes of other concepts have also been changed. Search inside note fields to find such cases.
- Consistency: Check that you have been consistent in your treatment of accents, plurality, and punctuation.
- Cross-referencing your translation with other language versions: If you know another language well, it is useful to compare terms in that language with yours, but remember that you should always translate from the source language.
- Hierarchical structure: Check whether the hierarchical structure is appropriate for the target language. If not, a new concept may be required – report it to the ELSST Team.
- Peer review of hierarchies: Ideally, you should get your translation peer-reviewed by a subject expert or experts.

References

Bokmålsordboka. Available at
http://ordbok.uib.no/perl/ordbok.cgi?OPP=&bokmaal=+&ordbok=bokmaal , accessed 28
October 2020.

Dansk Sprognævn (Danish Language Council). Available at
https://dsn.dk/om-os/about-the-danish-language-council, accessed 28 October 2020.

Diccionario de la lengua española. Available at http://www.rae.es/, accessed 28 October
2020.

Dicționarul Explicativ al Limbi Române (DEX). Available at https://dexonline.ro/, accessed 28
October 2020.

Dictionnaire de français Larousse. Available at http://www.larousse.fr/dictionnaires/francais,
accessed 28 October 2020.Duden Online-Wörterbuch. Available at
http://www.duden.de/woerterbuch, accessed 28 October 2020.

Eurostat glossary. Available at
https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Category:Glossary,
accessed 28 October 2020.

Extended Knowledge Organization System (XKOS):
https://ddialliance.org/Specification/RDF/XKOS, accessed 28 October 2020.

ISCED 1997: International Standard Classification of Education. Available at
http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-e
ducation-1997-en_0.pdf, accessed 28 October 2020.

ISO 639-1: 2002: Codes for the representation of names of languages – Part 1: Alpha-2
code. Available at https://en.wikipedia.org/wiki/ISO_639-1, accessed 28 October 2020.

ISO 25964-1 (2011): Information and documentation – Thesauri and interoperability with
other vocabularies, Part 1: Thesauri for information retrieval. Available at
http://www.niso.org/schemas/iso25964/#schema, accessed 28 October 2020.

ISO 3166: Codes for the representation of names of countries and their subdivisions – Part
1: country codes. Available at https://www.iso.org/iso-3166-country-codes.html, accessed
28 October 2020.

Λεξικό της κοινής νεοελληνικής (Modern Greek Dictionary). Available at
http://www.greek-language.gr/greekLang/modern_greek/tools/lexica/triantafyllides/,
accessed 28 October 2020.

Lietuvių kalbos žodynas. Available at http://www.lkz.lt, accessed 28 October 2020.

MOT Kielitoimiston sanakirja

Oxford English Dictionary. Available at http://www.oed.com/, accessed 28 October 2020.

Simple Knowledge Organization System (SKOS). Available at:
https://www.w3.org/2004/02/skos/ , accessed 28 October 2020.

Simple Knowledge Organization System eXtension for Labels (SKOS-XL). Available at:
https://www.w3.org/TR/skos-reference/skos-xl.html , accessed 28 October 2020.

Skosmos. Available at http://skosmos.org/, accessed 28 October 2020.

Slovenski pravopis. Available at
http://www.fran.si/iskanje?FilteredDictionaryIds=134&View=1&Query=%2A, accessed 28
October 2020.

Slovník spisovného jazyka českého. Available at http://ssjc.ujc.cas.cz/ , accessed 28
October 2020.

Svenska Akademiens ordlista.

Van Dale dictionary. Available at https://www.vandale.nl/ , accessed 28 October 2020.

VocBench. Available at http://vocbench.uniroma2.it/, accessed 28 October 2020.

Appendix: Reference sources for translation

As well as monolingual dictionaries and sources, the following multilingual resources are useful in looking for language equivalents.

All online resources accessed 28 October 2020.

Multilingual thesauri

Name of thesaurus	Subject area	Languages	URL
Finnish thesaurus and ontology service (Finto)	general	Finnish, Swedish, English	https://finto.fi/en/
UNESCO	social sciences	English, French, Spanish	http://databases.unesco.org/thesaurus/
OECD	social sciences	English, French, Spanish	http://bibliotecavirtual.clacso.org.ar/ar/oecd-macroth/
EUROVOC Multilingual Thesaurus of the European Union	general	EU languages	https://op.europa.eu/en/web/eu-vocabularies/th-dataset/-/resource/dataset/eurovoc

United Nations Bibliographic Information System Thesaurus (UNBIS)	general	English, French, Spanish, Russian, Chinese, Arabic	http://metadata.un.org/thesaurus/?lang=en
General Multilingual Environmental Thesaurus (GEMET)	environment	37 languages including all current ELSST languages	http://www.eionet.europa.eu/gemet/about
European Health Promotion Thesaurus	health	Dutch, English, French, Italian, Portuguese, Spanish	http://www.vocabularyserver.com/hpmulti/en/index.php

Learning Resources Exchange for schools (LRE) thesaurus	education	available in 19 languages including: Czech, Danish, Dutch, English, Finnish, French, German, Greek, Norwegian, Slovenian, Spanish, Swedish	http://lreforschools.eun.org/web/guest/lre-thesaurus searchable via http://europeanschoolnet-vbe.lexaurus.net/vbe/ (select 'LRE thesaurus' in Advanced search)
Thesaurus for Education Systems in Europe (TESE)	education	available in 15 languages, including: Czech, English, Finnish, French, German, Greek, Lithuanian, Romanian, Spanish	https://www.vocabularyserver.com/eurydice/en/

European Commission Libraries Catalogue thesaurus (ECLAS)	EU topics	English, French	https://op.europa.eu/en/publication-detail/-/publication/5fdfd2a9-3c29-4308-ac45-3c6ee3157b0c
POPIN	population	English, French, Spanish	http://www.cicred.org/OLD2004/thesaurus/INTEGRAL/index.html
ILO Thesaurus	work, employment	English, French, Spanish	https://metadata.ilo.org/thesaurus.html
European Training Thesaurus (ETT)	education	available in 11 languages including: Danish, Dutch, English, Finnish, French, German, Greek, Swedish	Downloadable from http://www.cedefop.europa.eu/en/publications-and-resources/publications/4114

European Education Thesaurus	education, learning	languages include: English, French, German, Spanish	http://vocabularyserver.com/tee/en/
AGROVOC	agriculture, forestry, fishery, food and related (e.g. environment	available in 23 languages including: Czech, English, French, German, Spanish	http://aims.fao.org/standards/agrovoc/functionalities/search
Commonwealth Agricultural Bureaux CAB	life sciences	equivalents for most terms in: English, Dutch, Portuguese, Spanish, plus lesser content in: Danish, Finnish, French, German, Norwegian,	http://www.cabi.org/cabthesaurus/

		Swedish	
Government of Canada (GOC) Core Subject Thesaurus	social sciences	English, French	http://www.thesaurus.gc.ca/recherche-search/thes-eng.html

Multilingual glossaries, terminologies, and vocabulary repositories

Name	Resource type	Subject area	Languages	URL
InterActive Terminology for Europe (IATE)	term bank	general	EU languages	https://iate.europa.eu/home
European Glossary on Education	glossary, also info on institutions by educational level in each country	education	EU languages	Downloadable from https://op.europa.eu/en/publication-detail/-/publication/6dc168d4-7a44-4a90-a247-4300e9769e47
Multilingual glossary of technical and popular medical terms	glossary	medicine	9 EU languages including: Danish, Dutch, English,	http://users.ugent.be/~rvdstich/eugloss/welcome.html

			French, German, Spanish	
European Employment and Industrial Relations (EMIRE) Dictionary	dictionary Explains the national industrial relations systems of the EU member states through their terminology	labour and employment	available in EU language, including all current ELSST languages	http://www.eurofound.europa.eu/emire/emire.htm
International Statistical Institute (ISI) glossary	glossary	statistics	29 languages	http://isi.cbs.nl/glossary/
Reference And Management Of Nomenclatures (RAMON)	metadata server	multiple	multiple	http://ec.europa.eu/eurostat/ramon
The Basel Register of Thesauri, Ontologies & Classifications (BARTOC)	database of Knowledge organization systems	multiple	multiple	http://www.bartoc.org/

EVROTERM	Multilingual terminology database	multiple	multiple	http://www.evroterm.gov.si/

Parallel corpora

Name	Resource type	Subject area	Languages	URL
Reverso	parallel corpora	all	multiple	http://context.reverso.net/translation
Linguee	parallel corpora	all	multiple	www.linguee.com
Eur-lex	parallel corpora	law and related	EU languages	http://eur-lex.europa.eu/advanced-search-form.html