

CECL Papers 3

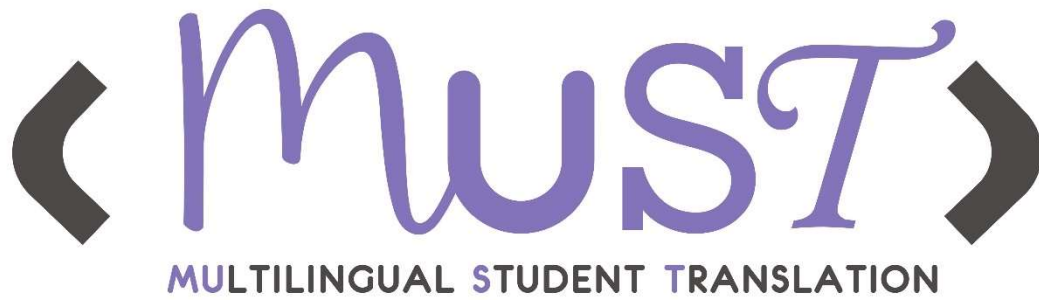
Translation-oriented Annotation System manual

Version 2.0

Sylviane Granger and Marie-Aude Lefer
in collaboration with the MUST partners,
Laura Penha-Marion and Anthonya Delfosse



2021



Translation-oriented Annotation System manual

Version 2.0

Sylviane Granger and Marie-Aude Lefer

in collaboration with the MUST partners,

Laura Penha-Marion and Anthonya Delfosse

October 2021



@ Centre for English Corpus Linguistics – UCLouvain

Table of contents

Introduction.....	6
TAS: general tagging guidelines.....	6
Guideline 1: Tagging discrete errors vs rewriting.....	6
Guideline 2: Describing vs interpreting.....	7
Guideline 3: Correction	7
TAS: error-tagging principles.....	7
Principle 1	7
Principle 2	8
Principle 3	8
Principle 4	8
Principle 5	9
Principle 6	9
Principle 7	9
TAS taxonomy: error categories.....	10
0. OVERVIEW	10
1. CONTENT (CT).....	11
1.1. Distortion (CT-DST).....	11
1.1.1. Incomprehensible (non-sens) (CT-DST-INC).....	11
1.1.2. Incorrect meaning (faux sens) (CT-DST-INM).....	12
1.1.3. Illogical (contresens) (CT-DST-ILL).....	12
1.1.4. Inexact meaning (glissement de sens) (CT-DST-INE)	12
1.2. Omission (CT-OMI).....	12
1.3. Addition (CT-ADD).....	13
1.4. Indecision (CT-IND).....	13
1.5. Non-identification of source-text error/inconsistency (CT-NON).....	13
2. CULTURE (CL)	14
3. TRANSLATION BRIEF (TB).....	14
3.1. Non-compliance with reference materials (TB-REF)	14
3.2. Formatting (TB-FOR).....	14
4. GRAMMAR AND SYNTAX (GR).....	15
4.1. Inflectional morphology (GR-INF)	15
4.2. Tense/aspect (GR-TNS).....	15
4.3. Voice (GR-VOI).....	15
4.4. Word order (GR-WOR).....	15
4.5. Word category confusion (GR-WCO)	16
4.6. Determiner (GR-DET).....	16

4.7.	Preposition (GR-PRE)	17
4.8.	Pronoun (GR-PRO)	17
4.9.	Concord (GR-CCD)	18
4.10.	Conjunction (GR-CJX)	18
4.11.	Complementation (GR-COM)	19
4.11.1.	Adjective (GR-COM-ADJ)	19
4.11.2.	Noun (GR-COM-NOU)	19
4.11.3.	Verb (GR-COM-VRB)	19
4.11.4.	Adverb (GR-COM-ADV)	19
5.	LEXIS AND TERMINOLOGY (LT)	20
5.1.	Word (LT-WRD)	20
5.1.1.	Single word (LT-WRD-SWO)	20
5.1.1.1.	Derivative (LT-WRD-SWO-DER)	21
5.1.1.2.	Cognate (LT-WRD-SWO-COG)	21
5.1.2.	Multiword (LT-WRD-MWO)	21
5.1.2.1.	Compound (LT-WRD-MWO-COP)	21
5.1.2.2.	Collocation (LT-WRD-MWO-COL)	21
5.1.2.3.	Idiom (LT-WRD-MWO-IDI)	22
5.2.	Term (LT-TRM)	22
5.2.1.	Term translated by non-term/non-term translated by term (LT-TRM-TNT)	22
5.2.2.	Terminological inconsistency (LT-TRM-TIN)	22
6.	DISCOURSE AND PRAGMATICS (DP)	23
6.1.	Cohesion (DP-COH)	23
6.1.1.	Linkword (DP-COH-LIN)	23
6.1.2.	Pronoun reference (DP-COH-PRF)	24
6.2.	Information structuring (DP-STR)	24
6.2.1.	Thematic progression (DP-STR-THE)	24
6.2.2.	Split/merge sentences (DP-STR-SPL)	24
6.2.3.	Syntactic preferences (DP-STR-SYN)	25
7.	MECHANICS (ME)	25
7.1.	Spelling (ME-SPE)	25
7.2.	Punctuation (ME-PUN)	26
7.3.	Units, dates, numbers (ME-UDN)	26
8.	REGISTER AND STYLE (REG)	27
8.1.	Degree of (in)formality and register mismatch (REG-FML)	27
8.2.	Heavy (REG-HEA)	27
8.3.	Redundant (REG-RED)	28

9. META-TAGS.....	28
9.1. PLUS.....	28
9.2. Source Language Intrusion (SLI)	28
9.3. Distortion (DIS).....	29
9.4. Error repetition (REP).....	29
Bibliography.....	30

Introduction

The present document is the second version of the *Translation-oriented Annotation System* (TAS) manual designed within the framework of the *Multilingual Student Translation* (MUST) project (Granger & Lefer 2020).

It is important to bear in mind from the outset that all three stages of error annotation – detection, tagging and correction – involve a high degree of subjectivity. The experiment carried out by Granger (2010) showed that error detection is probably the most subjective of the three. The number of errors detected in the same learner text by 10 native speakers of English, all experienced English language and linguistics specialists, ranged from 59 to 106.

Designing a standardized system like TAS is challenging, for several reasons:

- MUST includes a large number of languages, some of which are typologically distant;
- The number of tags to be included in TAS should be manageable, but the annotation system should not be too coarse-grained;
- TAS should ideally be useful for both teaching (lecturers and students) and research;
- It is often difficult to get teachers to adopt an annotation system that differs from their usual annotation practices.

The present manual is divided into three main parts: (1) general tagging guidelines, (2) tagging principles and (3) a detailed description of TAS 2.0 (tag set, definitions, examples, warning boxes).

TAS: general tagging guidelines

In order to address the high degree of subjectivity involved in error annotation, it is important to formulate general guidelines and tagging principles. The guidelines and principles included in the present manual are based partly on the *Louvain Error Tagging Manual* (Granger et al. forthcoming) developed within the framework of the *International Corpus of Learner English* (ICLE) project (Granger et al. 2020). Most of the examples used to illustrate the tagging system are extracted from the MUST corpus. When no example was found, we used examples taken from the *Louvain Error Tagging Manual* with permission from the authors.

TAS 2.0 relies on the following general principles:

Guideline 1: Tagging discrete errors vs rewriting

Only discrete errors should be tagged. Annotators need to refrain from rewriting.

Example:

When a thrilling program is on, we turn into sponges that imbibe all what is coming.

- Identification of the discrete error: *what* → *that* (grammar: pronoun) (+ possibility of adding a style tag for the whole underlined chunk)
- Rewriting: *When a thrilling program is on, we passively take in everything that comes onto the screen.*
⇒ The pronoun error is overlooked

Guideline 2: Describing vs interpreting

TAS is descriptive in nature, not explanatory or interpretative. In other words, TAS describes the *nature* of errors, not their possible *sources* (e.g. analogy, overgeneralization). One exception to this guideline is the possibility of adding the *Source Language Intrusion* (SLI) meta-tag (see below).

TAS makes no distinction between *errors* and *mistakes* (which are typically due to carelessness), or between *minor* and *critical* errors (no weight is assigned to errors).

Guideline 3: Correction

Specifying the correct form is optional. If you wish to provide a correction, use the 'correction' box in the annotation interface of Hypal4MUST.

TAS: error-tagging principles

Principle 1

Do not tag on the basis of the corrected/targeted word or phrase, but only on the basis of the incorrect word or phrase.

[Example:](#)

They persuaded him to ***abandon** smoking → to **give up** smoking: LT-WRD-SWO (Lexis and Terminology: Word: Single word), [not](#) LT-WRD-MWO (Lexis and Terminology: Word: Multiword)



Exceptions

(i) The inflectional morphology tag (GR-INF) is used to mark inflectional errors, including all cases of confusion between the Saxon genitive and the *of*-genitive.

GR-INF is to be used to indicate cases where the Saxon genitive should not be used:

e.g. Behind the ***Berlin's** wall → Berlin wall

It is also used in cases where the Saxon genitive is called for:

e.g. The ***car of** my sister → My sister's car

⇒ GR-INF is to be used here, and not GRA-PRE (Grammar and Syntax: Preposition), even though *of* is a preposition

(ii) The determiner tag (GR-DET) is used to mark errors involving all types of determiners, including cases of missing determiners and articles.

e.g. not only does ***economy** take over → not only does **the economy** take over

Principle 2

Do not tag errors resulting from the correction of a previous word or phrase.

Example:

***The Belgian** is still active

The Belgian → *Belgians*: plural inflection error, GR-INF (Grammar and Syntax: Inflectional morphology)

is still active: no concord error (concord was correct in the translation)

Principle 3

When there are two (or three) types of error in the same word or phrase, double (or triple) tag the word or phrase.

Example:

Les scandales qui ont secoué l'entreprise → ***Scandals** which ***disrupted** the ***company** → The scandals which have affected the company's reputation

Three discrete errors: GR-DET (Grammar and Syntax: Determiner), GR-TNS (Grammar and Syntax: Tense/aspect), CT-DST (Content: Distortion)



Warning: double/triple tagging

Double tags are different from portmanteau tags (or ambiguity tags), i.e. tags used when the “tagger was unable to determine, with sufficient confidence, which was the correct category, and so left two possibilities for users to disambiguate themselves, if they should wish to do so” (Leech and Smith 2000). In other words, double tagging should not be used when hesitating between two tags, but when a given word or phrase contains two discrete errors.

e.g. highly ***imported** → highly important

⇒ Do not tag as LT-WRD-SWO-DER (Lexis and Terminology: Word: Single word: Derivative) **and** ME-SPE (Mechanics: Spelling); tag as LT-WRD-SWO-DER **only**

It is important to note that in the Hypal4MUST annotation interface, if you wish to double-tag a large segment and a specific word or group of words within that segment, you should annotate the smaller segment first.

Principle 4

Do not tag the same error at two different levels of granularity.

Examples:

(i) to add insult to injury → ***ajouter l'insulte à l'injure** → pour couronner le tout

(ii) schlafen wie ein Stein (lit. 'to sleep like a stone') → ***dormir comme une pierre** → dormir comme une souche ('to sleep like a log')

⇒ Do not double-tag as LT-WRD-MWO (Lexis and Terminology: Word: Multiword) **and** LT-WRD-MWO-IDI (Lexis and Terminology: Word: Multiword: Idiom). Instead, choose at which level of granularity you wish to annotate (here, either third or fourth level).

Principle 5

The length of the highlighted segment should coincide with the tag selected.

Example:

If you wish to tag *in the volume of members* as LT-WRD-SWO (Lexis and Terminology: Word: Single word), do not highlight the full segment. Rather, tag the word *volume*, i.e. in the *volume* of members.

If the lexical error highlighted consists of more than one word, the LT-WRD-MWO (Lexis and Terminology: Word: Multiword) tag should be used.



Warning: discontinuous segments

It is unfortunately not possible to annotate the different elements of a discontinuous segment (e.g. separable verbs in German) as a single error. We therefore suggest that only one element of the discontinuous segment should be tagged and, if necessary, that a note should be added in the comment box on the annotation interface.

Example:

L'avion arrive à 16 heures à Cologne → das Flugzeug **kommt** um 16 Uhr in Köln ***ein** → das Flugzeug **kommt** um 16 Uhr in Köln **an** ('the plane arrives in Cologne at 4 pm')

[bei jdm] um etw einkommen (to apply [to sb] for sth) ≠ ankommen (to arrive)

⇒ In this example, use the tag LT-WRD-SWO (Lexis and Terminology: Word: Single word)

Principle 6

The PLUS meta-tag, which is used to mark good translation solutions, should be used on its own, not in combination with another tag, as tags are used to annotate erroneous/infelicitous translations.

Example:

A particularly felicitous theme-rheme change cannot be tagged DP-STR-THE (Discourse and Pragmatics: Information structuring: Thematic progression) as tags mark errors. Only the PLUS meta-tag should be used (and a comment added in the 'comment' box of the Hypal4MUST annotation interface, if need be).

Principle 7

Use the *Source Language Intrusion* (SLI) meta-tag to mark errors that are suspected of being due to source language influence. The SLI meta-tag can be added to any other tag.

Example:

preziosa donazione → precious donation: collocation error LT-WRD-MWO-COL (Lexis and Terminology: Word: Multiword: Collocation) + SLI (Source Language Intrusion)

TAS taxonomy: error categories

0. OVERVIEW

Here is an overview of the TAS 2.0. taxonomy. The categories are described and illustrated below.

Content	CT
Content: Distortion	CT-DST
Content: Distortion: Incomprehensible	CT-DST-INC
Content: Distortion: Incorrect meaning	CT-DST-INM
Content: Distortion: Illogical	CT-DST-ILL
Content: Distortion: Inexact meaning	CT-DST-INE
Content: Omission	CT-OMI
Content: Addition	CT-ADD
Content: Indecision	CT-IND
Content: Non-identification of source-text error/inconsistency	CT-NON
Culture	CL
Culture: Translated untranslatable/untranslated translatable	CL-UNT
Translation brief	TB
Translation brief: Non-compliance with reference materials	TB-REF
Translation brief: Formatting	TB-FOR
Grammar and Syntax	GR
Grammar and Syntax: Inflectional morphology	GR-INF
Grammar and Syntax: Tense/aspect	GR-TNS
Grammar and Syntax: Voice	GR-VOI
Grammar and Syntax: Word order	GR-WOR
Grammar and Syntax: Word category confusion	GR-WCO
Grammar and Syntax: Determiner	GR-DET
Grammar and Syntax: Preposition	GR-PRE
Grammar and Syntax: Pronoun	GR-PRO
Grammar and Syntax: Concord	GR-CCD
Grammar and Syntax: Conjunction	GR-CJX
Grammar and Syntax: Complementation	GR-COM
Grammar and Syntax: Complementation: Adjective	GR-COM-ADJ
Grammar and Syntax: Complementation: Noun	GR-COM-NOU
Grammar and Syntax: Complementation: Verb	GR-COM-VRB
Grammar and Syntax: Complementation: Adverb	GR-COM-ADV
Lexis and Terminology	LT
Lexis and Terminology: Word	LT-WRD
Lexis and Terminology: Word: Single word	LT-WRD-SWO
Lexis and Terminology: Word: Single word: Derivative	LT-WRD-SWO-DER
Lexis and Terminology: Word: Single word: Cognate	LT-WRD-SWO-COG
Lexis and Terminology: Word: Multiword	LT-WRD-MWO
Lexis and Terminology: Word: Multiword: Compound	LT-WRD-MWO-COP
Lexis and Terminology: Word: Multiword: Collocation	LT-WRD-MWO-COL
Lexis and Terminology: Word: Multiword: Idiom	LT-WRD-MWO-IDI
Lexis and Terminology: Term	LT-TRM
Lexis and Terminology: Term: Term translated by non-term/non-term translated by term	LT-TRM-TNT
Lexis and Terminology: Term: Terminological inconsistency	LT-TRM-TIN
Discourse and Pragmatics	DP
Discourse and Pragmatics: Cohesion	DP-COH

Discourse and Pragmatics: Cohesion: Linkword	DP-COH-LIN
Discourse and Pragmatics: Cohesion: Pronoun reference	DP-COH-PRF
Discourse and Pragmatics: Information structuring	DP-STR
Discourse and Pragmatics: Information structuring: Thematic progression	DP-STR-THE
Discourse and Pragmatics: Information structuring: Split/merge sentences	DP-STR-SPL
Discourse and Pragmatics: Information structuring: Syntactic preferences	DP-STR-SYN
Mechanics	ME
Mechanics: Spelling	ME-SPE
Mechanics: Punctuation	ME-PUN
Mechanics: Units, dates, numbers	ME-UDN
Register and Style	REG
Register and Style: Degree of (in)formality and register mismatch	REG-FML
Register and Style: Heavy	REG-HEA
Register and Style: Redundant	REG-RED
PLUS (meta-tag)	PLUS
Source Language Intrusion (meta-tag)	SLI
Distortion (meta-tag)	DIS
Error repetition (meta-tag)	REP

1. CONTENT (CT)

The CT (Content) category groups together all errors arising from a failure to reflect accurately the content of the source text (cf. DFKI 2015).

1.1. Distortion (CT-DST)

CT-DST (Content: Distortion) is used when the source-text content is misrepresented in the translation. A finer-grained distinction can be made between various types of distortion, as discussed below.



Warning: TAS contains both a CT-DST tag and a DIS meta-tag

The CT-DST (Content: Distortion) tag is not to be confused with the Distortion (DIS) meta-tag, meant to signal content distortion caused by other error types (e.g. a lexical, grammatical or discursive error that leads to distortion of source-text meaning) (cf. Meta-tags iii. Distortion (DIS)).

1.1.1. Incomprehensible (non-sens) (CT-DST-INC)

CT-DST-INC (Content: Distortion: Incomprehensible) refers to major distortion errors where the target text is difficult or impossible for the reader to understand, even given the context (Kunilovskaya 2016: 2; UPF taxonomy/Anna Espunya, personal communication). It also includes cases of vague target-text segments.

Example:

ESG research firms → les fournisseurs de services de recherche ESG ('the providers of services of research ESG')

1.1.2. Incorrect meaning (faux sens) (CT-DST-INM)

CT-DST-INM (Content: Distortion: Incorrect meaning) refers to major distortion errors where a word or a phrase in the target text conveys a meaning which at first sight seems to make sense (is plausible) but is in fact incorrect.

Examples:

(i) He **cannot afford to** lose his reputation → Il ***n'a pas les moyens de** perdre sa réputation ('He does not have the means to lose his reputation') → Il **ne peut se permettre de** perdre sa réputation (Delisle et al. 1999: 40)

(ii) Since then the number of monuments to the man branded the "**butcher of the Congo**" **have been taken down** → Depuis, un certain nombre de monuments à la mémoire de l'homme appelé ***« le roi bâtisseur » *ont été démolis** ('Since then the number of monuments to the man branded the "Builder King" have been destroyed') → Depuis, un certain nombre de monuments à la mémoire de l'homme appelé **« boucher du Congo » ont été retirés** ('Since then a number of monuments to the man branded the "butcher of the Congo" have been removed')

1.1.3. Illogical (contresens) (CT-DST-ILL)

CT-DST-ILL (Content: Distortion: Illogical) refers to major distortion errors where a word or a phrase in the target text is given the opposite meaning to the one intended by the source-text author(s) (cf. Dussart 2005: 112).

Examples:

(i) To avoid world wars, world starvation, and world epidemics, **international bodies have been set up** → Um Weltkriege, Hungersnöte und Epidemien zu vermeiden, sind **internationale *militärische Einheiten aufgestellt worden** ('To avoid world wars, world starvation, and world epidemics, international troops have been gathered') → Um Weltkriege, Hungersnöte und Epidemien zu vermeiden, sind **internationale Organisationen gegründet worden** ('To avoid world wars, world starvation, and world epidemics, international organisations have been set up') (Delisle et al. 1999: 389-390)

(ii) Die Nachfrage **aus der Eurozone** nahm um sechs Prozent ab ('Demand from the eurozone fell by 6 per cent') → La demande émanant **de l'extérieur de la zone euro** a reculé de 6 % ('Demand from outside the eurozone fell by 6 per cent') → La demande **en provenance de la zone euro** a reculé de 6 %

1.1.4. Inexact meaning (glissement de sens) (CT-DST-INE)

CT-DST-INE (Content: Distortion: Inexact meaning) refers to minor distortion errors where the meaning of the target text is inaccurate or incomplete with regard to the one intended by the source-text author(s) (cf. Dussart 2005: 116).

Example:

Le langage juridique intransigeant du projet d'accord **pourrait provoquer** un conflit majeur → The uncompromising legal language of the draft agreement ***will provoke** a major row → The uncompromising legal language of the draft agreement **is likely to provoke** a major row

1.2. Omission (CT-OMI)

CT-OMI (Content: Omission) is used when elements present in the source text (titles, headings, words, phrases, sentences, etc.) are missing from the target text and cannot be recovered from the context of the target text.

Examples:

(i) Les trains Eurostar directs reliant Londres et Amsterdam débutent le 4 avril → Direct rail services between London and Amsterdam will begin on 4 April → **Eurostar** direct rail services between London and Amsterdam will begin on 4 April

(ii) A spokeswoman for the party said Wednesday that **up to 27 people** are still believed to be receiving the payments → Une porte-parole du parti a déclaré mercredi que **27 personnes** recevraient probablement encore ces pensions ('A spokeswoman for the party said Wednesday that 27 people are still believed to be receiving the payments')

(ii) Die deutschen Importe fielen im Juli um 0,7 Prozent **zum Vormonat** ('German imports fell by 0.7 per cent in July compared to the previous month') → En juillet, les importations allemandes ont baissé de 0,7 % ('German imports fell by 0.7 per cent in July') → En juillet, les importations allemandes ont baissé de 0,7 % **par rapport au mois précédent**

(iv) Das Risiko für Säuglinge, an **Keuchhusten** zu sterben, liegt auch heute noch in der Größenordnung von 1:100 bis 1:200 der **Erkrankten** ('Today, the risk of infants with whooping cough dying from it is still in the order of 1 in 100 to 1 in 200') → Le risque de décès pour les nourrissons est encore de 1 sur 100 à 1 sur 200 ('Today, the risk of infants dying is still in the order of 1 in 100 to 1 in 200') → Le risque de décès pour les nourrissons **atteints de la coqueluche** est encore de 1 sur 100 à 1 sur 200

1.3. Addition (CT-ADD)

CT-ADD (Content: Addition) is used when information not present in the source text is added to the target text.

Example:

la chanson considérée comme l'hymne national officieux de la Corée → the song regarded as the unofficial national anthem of Korea → the song regarded as the unofficial national anthem **of the two parts** of Korea

1.4. Indecision (CT-IND)

CT-IND (Content: Indecision) is to be used for cases where students add slashes, question marks, asterisks, brackets, etc. to the text to mark indecision or give several options for the same source-text item.

Examples:

(i) This **jeopardized/put in danger** the whole cattle and milk production sector.

(ii) This **jeopardized??** the whole cattle and milk production sector.

1.5. Non-identification of source-text error/inconsistency (CT-NON)

CT-NON (Content: Non-identification of source-text error/inconsistency) is used for cases where students have failed to identify and correct source-text errors or inconsistencies in their translation.

Example:

Barak Obama a deux filles, Malia et Sasha → **Barak** Obama has two daughters, Malia and Sasha → **Barack** Obama...

2. CULTURE (CL)

CL (Culture) refers to errors related to culture-specific references. It covers instances where students fail to include in their translation explanations of realia or cultural concepts which will therefore not be understood by the target audience. The tag also subsumes cases where realia or proper names are translated when they should have been left untranslated, and vice versa.

Examples:

(i) Translation of *cricket* by *cricket* in a context where a culturally equivalent word like *football* would have been more appropriate

(ii) *Apple computers* → *ordinateurs pomme* ('computers apple') instead of *ordinateurs Apple* (cf. DFKI 2015)

3. TRANSLATION BRIEF (TB)

TB (Translation brief) refers to all cases of failure to comply with the specifications of the translation brief.

3.1. Non-compliance with reference materials (TB-REF)

TB-REF (Translation brief: Non-compliance with reference materials) refers to the use of a word, term or expression that is inconsistent with the reference materials provided to the students (these include glossaries, reference termbases, terminology guidelines and style guides).

Example:

Use of *enlargement* when the TB glossary specifies that *expansion* should be used.

3.2. Formatting (TB-FOR)

TB-FOR (Translation brief: Formatting) refers to inappropriate use of layout (e.g. line spacing, font size), paragraph indentation, etc. By default, the tag includes all cases of formatting mismatch between source text and target text.

Examples:

(i) headings should be in bold typeface, but they are not

(ii) line spacing should be 1.15, but is 1.5



Warning: highlighting TB-FOR errors in Hypal4MUST annotation interface

Do not highlight whole paragraphs or whole texts to mark formatting errors, as this will cause serious annotation bugs. Instead, highlight the first word of the erroneously formatted segment and insert the TB-FOR tag.

4. GRAMMAR AND SYNTAX (GR)

GR (Grammar and Syntax) groups together sentence grammar errors and word grammar (i.e. lexico-grammatical) errors. It also includes inflectional morphology errors (derivational morphology errors, by contrast, are part of the LT – Lexis and terminology – category).

4.1. Inflectional morphology (GR-INF)

GR-INF (Grammar and Syntax: Inflectional morphology) involves inflectional errors that result from the misuse of grammatical morphemes (plural, genitive, verb morphology, degree of adjectives, etc.). It also includes all cases of confusion between the Saxon genitive and the *of*-genitive.

Examples:

(i) *watchs → watches

(ii) Behind the *Berlin's wall → Berlin wall

⇒ the Saxon genitive should not be used

(iii) The *car of my sister → My sister's car

⇒ the Saxon genitive should be used

4.2. Tense/aspect (GR-TNS)

GR-TNS (Grammar and Syntax: Tense/aspect) includes all errors related to tense use (past, present and future) or aspect.

Example:

More and more companies *become multi-nationals → More and more companies **are becoming** multi-nationals

4.3. Voice (GR-VOI)

GR-VOI (Grammar and Syntax: Voice) includes all errors related to the passive and active voices (the active used when the passive should be used and vice versa).

Examples:

(i) It is to *hope that → It is to **be hoped** that...

(ii) ... are worth *being underlined → **underlining**

4.4. Word order (GR-WOR)

GR-WOR (Grammar and Syntax: Word order) is used to tag incorrect word order.

Examples:

(i) Some 30 years ago ***was launched the idea of ...** → the idea of ... was launched

(ii) they can therefore ***hand down to him their learning as well as their experience** → **hand down their learning to him, as well as their experience**

(iii) Ich ***bin aufgestanden** um 6 Uhr → Ich **bin** um 6 Uhr **aufgestanden** ('I got up at 6')

4.5. Word category confusion (GR-WCO)

GR-WCO (Grammar and Syntax: Word category confusion) refers to inappropriate use of a grammatical word category, e.g. confusion between an adverb and an adjective or between a noun and an adjective. In other words, "a word is the wrong part of speech" (DFKI 2015).

Example:

Not everybody is ***luckily** → **lucky**

⇒ use of an adverb instead of an adjective



Warning: word category errors vs derivation errors

The GR-WCO (Grammar and Syntax: Word category confusion) tag refers to cases of confusion between grammatical word categories. It should not be used for errors caused by the addition of an affix (be it a prefix or a suffix) resulting in a non-attested word. These should be tagged LT-WRD-SWO-DER (Lexis and Terminology: Word: Single word: Derivative).

Example:

***unconceivable** → inconceivable

⇒ Do not tag GR-WCO (Grammar and Syntax: Word category confusion) but rather LT-WRD-SWO-DER (Lexis and Terminology: Word: Single word: Derivative)

4.6. Determiner (GR-DET)

GR-DET (Grammar and Syntax: Determiner) refers to errors affecting all types of determiners (articles; demonstrative, possessive, indefinite determiners). GR-DET is also used to mark cases of missing determiners (in which case the tag should be attached to the noun modified by the determiner).

Examples:

(i) I think that ***this** → **the** Industrial Revolution should rather be seen as...

(ii) not only does ***economy** take over → not only does **the economy** take over

4.7. Preposition (GR-PRE)

GR-PRE (Grammar and Syntax: Preposition) is used for incorrect, missing, or redundant **independent** prepositions.

[Examples:](#)

(i) ***in** the seaside → **at** the seaside

(ii) we took the boat ***through** the river → we took the boat **across** the river



Warning: dependent vs independent preposition errors

The GR-PRE (Grammar and Syntax: Preposition) tag concerns **independent** prepositions. It should not be used to tag errors relating to **dependent** prepositions, i.e. prepositions that must follow a particular noun, verb, adjective or adverb, which should be tagged GR-COM (Grammar and Syntax: Complementation).

[Example:](#)

Representative ***for** their own identities → **of** their own identities

⇒ Do not tag GR-PRE (Grammar and Syntax: Preposition) but GR-COM-ADJ (Grammar and Syntax: Complementation: Adjective), because the preposition *of* is dependent on the adjective *representative*

4.8. Pronoun (GR-PRO)

GR-PRO (Grammar and Syntax: Pronoun) pertains to errors involving all pronoun categories: demonstrative, interrogative, personal, possessive, reciprocal, reflexive, relative, indefinite, etc.

[Examples:](#)

(i) We could hardly see ***one to another** → **each other** in the fog

(ii) The book was ***her** → **hers**



Warning: unclear pronominal references

The GR-PRO (Grammar and syntax: Pronoun) category does not include cases of unclear pronominal reference. These are cohesion errors that should be tagged DP-COH-PRF (Discourse and Pragmatics: Cohesion: Pronoun reference).

4.9. Concord (GR-CCD)

GR-CCD (Grammar and Syntax: Concord) corresponds to all errors of concord (between subject and verb, verb and object, noun and adjective, determiner and noun, etc.).

[Examples:](#)

Progress ***have** → **has** been made

***This** → **These** elements cannot be separated



Warning: concord errors vs inflectional morphology errors

The GR-CCD (Grammar and Syntax: Concord) tag should not be confused with GR-INF (Grammar and Syntax: Inflectional morphology), which concerns inflectional errors.

[Examples:](#)

***This** → **These** elements cannot be separated

⇒ GR-CCD (Grammar and Syntax: Concord)

***watches** → **watches**

⇒ GR-INF (Grammar and Syntax: Inflectional morphology)

4.10. Conjunction (GR-CJX)

GR-CJX (Grammar and Syntax: Conjunction) includes all errors of missing, redundant or erroneous conjunctions.

[Example:](#)

Wait ***that** → wait **until**



Warning: conjunction errors vs cohesion errors

It should be noted that GR-CJX (Grammar and Syntax: Conjunction) is a local error and is not to be confused with DP-COH-LIN (Discourse and Pragmatics: Cohesion: Linkword). GR-CJX (Grammar and Syntax: Conjunction) refers to grammatical errors, while DP-COH-LIN (Discourse and Pragmatics: Cohesion: Linkword) lies at the discourse level.

4.11. Complementation (GR-COM)

GR-COM (Grammar and Syntax: Complementation) refers to valency errors (grammatical patterning of words), relating in particular to the incorrect use of dependent prepositions.

4.11.1. Adjective (GR-COM-ADJ)

GR-COM-ADJ (Grammar and Syntax: Complementation: Adjective) is for adjective complementation errors.

[Examples:](#)

(i) capable ***to go** → capable **of going**

(ii) representative ***for** → representative **of** their own identities

4.11.2. Noun (GR-COM-NOU)

GR-COM-NOU (Grammar and Syntax: Complementation: Noun) is for noun complementation errors.

[Examples:](#)

(i) students have the possibility ***to leave** → possibility **of leaving**

(ii) it should be a way ***to help** → way **of helping** people to reconcile themselves with society

(iii) there are difficulties ***to establish** → difficulties **in establishing** unity

4.11.3. Verb (GR-COM-VRB)

GR-COM-VRB (Grammar and Syntax: Complementation: Verb) is for verb complementation errors (infinitive instead of gerund and vice versa, *to*-infinitive vs bare infinitive, etc.).

[Examples:](#)

(i) will not help people ***believing** → help people **to believe in...**

(ii) they enjoy ***to gather** → enjoy **gathering** in groups

4.11.4. Adverb (GR-COM-ADV)

GR-COM-ADV (Grammar and Syntax: Complementation: Adverb) is for adverb complementation errors.

[Example:](#)

independently ***on** → independently **of**



Warning: complementation errors vs syntactic preferences

Complementation errors should not be confused with discursive infelicities involving non-finite/finite verb forms, which should be tagged DP-STR-SYN (Discourse and Pragmatics: Information structuring: Syntactic preferences).

A clear distinction is to be made between

- a) preference for finite vs non-finite verb forms (and vice versa) at sentence or clause level; and
- b) word-specific complementation errors, where the form of the verb depends on the noun, verb, adjective or conjunction that precedes the verb form.

Example:

By the way ***estimate** → **estimating** other cultures does not ***mean integrate** → **mean integrating** it to one's own culture.

While the first erroneous verb form, *estimate*, occurs at sentence level and involves the nominal use of the bare infinitive, the second error, *integrate*, is a complementation error which is triggered by the use of the verb *mean*, which must be followed by an *-ing* form.

5. LEXIS AND TERMINOLOGY (LT)

LT (Lexis and Terminology) subsumes problems with single words, multiwords and terms. A term can be defined as “a meaningful lexical unit consisting of a word or a group of words used to univocally designate a concept in a specific subject field. (...) Terms can consist of one word (ex.: *engine*) or more than one word (ex.: *internal combustion engine*)” (Delisle et al. 1999: 185).



Warning: lexical errors vs spelling errors

LT (Lexis and Terminology) does not include spelling errors (e.g. **accomodation*). These should be tagged ME-SPE (Mechanics: Spelling).

5.1. Word (LT-WRD)

LT-WRD (Lexis and Terminology: Word) deals with errors involving one or more words. It does not include terms, which fall within a different category (LT-TRM).

5.1.1. Single word (LT-WRD-SWO)

LT-WRD-SWO (Lexis and Terminology: Word: Single word) refers to cases where an incorrect non-terminological single word is used.

5.1.1.1. Derivative (LT-WRD-SWO-DER)

LT-WRD-SWO-DER (Lexis and Terminology: Word: Single word: Derivative) covers errors due to the use of an affix, either a prefix or a suffix, which results in a non-existent word.

Examples:

(i) ***unpossible** → **impossible**

(ii) ***unconceivable** → **inconceivable**

5.1.1.2. Cognate (LT-WRD-SWO-COG)

LT-WRD-SWO-COG (Lexis and Terminology: Word: Single word: Cognate) deals with errors resulting from the influence of a formally similar word in the source text or in the student's mother tongue which partly or totally diverges in meaning from the source word (false friends).

Examples:

(i) il **prétend** que... → he ***pretends** that... → he claims that...

(ii) **officials** → ***oficiales** → funcionarios (Scott Doyle 2003: 21)

(iii) application [form] → ***aplicación** → **solicitud** (ibid)

(iv) **actualmente** → ***actually** → **presently, currently** (ibid)

5.1.2. Multiword (LT-WRD-MWO)

LT-WRD-MWO (Lexis and Terminology: Word: Multiword) refers to cases where an incorrect non-terminological multiword expression is used.

5.1.2.1. Compound (LT-WRD-MWO-COP)

LT-WRD-MWO-COP (Lexis and Terminology: Word: Multiword: Compound) refers to errors in compounds, i.e. multiword units that constitute one semantic unit although they may be written in one, two or more words, as well as other fully fixed sequences such as complex prepositions, complex adverbs, complex conjunctions and complex verbs, including phrasal verbs.

Examples:

(i) **voiture de société** → ***firm car** → **company car**

(ii) **moorhen** → ***poulet d'eau** → **poule d'eau**

(iii) **d'autre part** → ***on the other side** → **on the other hand**

5.1.2.2. Collocation (LT-WRD-MWO-COL)

LT-WRD-MWO-COL (Lexis and Terminology: Word: Multiword: Collocation) is used to refer to non-compliance with usage or preferences between verb and noun, adjective and noun, adverb and adjective or verb and adverb. In this tagging manual, collocations are regarded as "usage-determined or preferred syntagmatic relations between two

lexemes in a specific syntactic pattern. Both lexemes make an isolable semantic contribution to the word combination but they do not have the same status. Semantically autonomous, the 'base' of a collocation is selected first by a language user for its independent meaning. The second element, i.e. the 'collocate' or 'collocator', is selected by and semantically dependent on the 'base'" (Granger & Paquot 2008).

Examples:

(i) ***important traffic** → heavy traffic

(ii) ***do progress** → make progress

5.1.2.3. Idiom (LT-WRD-MWO-IDI)

LT-WRD-MWO-IDI (Lexis and Terminology: Word: Multiword: Idiom) concerns errors in the use of idioms, i.e. semantically non-compositional (often figurative) combinations of words. In this tagging manual, "idioms are characterized by their semantic non-compositionality, which can be the result of a metaphorical process. Lack of flexibility and marked syntax are further indications of their idiomatic status" (Granger & Paquot 2008). Idioms are usually constructed around a verbal nucleus.

Examples:

(i) jeter l'éponge → ***throw in the sponge** → throw in the towel

(ii) to add insult to injury → ***ajouter l'insulte à l'injure** → pour couronner le tout

(iii) schlafen wie ein Stein (lit. 'to sleep like a stone') → ***dormir comme une pierre** → dormir comme une souche ('to sleep like a log')

5.2. Term (LT-TRM)

LT-TRM (Lexis and Terminology: Term) deals with terminological errors.

5.2.1. Term translated by non-term/non-term translated by term (LT-TRM-TNT)

LT-TRM-TNT (Lexis and Terminology: Term: Term translated by non-term/non-term translated by term) should be used when a source language term that has an established equivalent in the target language is translated by a non-term or when a source language non-term is translated by a term in the target text.

Examples:

(i) monnaie nationale → ***money** → domestic currency

(ii) la communauté européenne des femmes célibataires → the ***European Union** of single women (instead of the 'European community of single women')

5.2.2. Terminological inconsistency (LT-TRM-TIN)

LT-TRM-TIN (Lexis and Terminology: Term: Terminological inconsistency) is for cases where terminology is inconsistent within the target text (DFKI, 2015).

Example:

The terms *durabilité* and *soutenabilité* are used interchangeably to render English *sustainability* within a given translation



Warning: terminological inconsistency vs non-compliance with reference materials

LT-TRM-TIN (Lexis and Terminology: Term: Terminological inconsistency) is to be used for terminology errors. It should not be used for cases where terminology does not comply with the reference materials (e.g. glossaries, termbases). These should be tagged TB-REF (Translation brief: Non-compliance with reference materials).

6. DISCOURSE AND PRAGMATICS (DP)

DP (Discourse and Pragmatics) refers to issues related to cohesion and information structuring, in particular the use of connectors and thematic progression (functional sentence perspective, theme/rheme).

6.1. Cohesion (DP-COH)

DP-COH (Discourse and Pragmatics: Cohesion) subsumes all errors of cohesion, such as inappropriate/incorrect use of cohesive markers and cases of unclear pronoun reference.

6.1.1. Linkword (DP-COH-LIN)

DP-COH-LIN (Discourse and Pragmatics: Cohesion: Linkword) is to be used when source-text connectors are absent in the target text when they should be present, or when connectors are superfluous or convey a different logical link. This category comprises linkwords such as adverbial connectors and conjunctions of subordination or coordination.

Examples:

(i) *Cependant, cette situation n'est pas idéale. En effet, en dépit des droits acquis, l'inégalité des revenus persiste* → However, this situation is not THE ideal one. ~~Indeed~~, in spite of the vested interests, the unequally balanced incomes persist

(ii) **Although, they ended up having three little girls* → **However**, they ended up having three little girls

(iii) ... **i.e.** a way to rate ratings, not just another ratings system → ***par exemple**, une façon de classer les classements, pas simplement un nouveau système de classement ('for example, a way to rate ratings, not just another ratings system') → **c'est-à-dire** ('that is')

(iv) **In der Tat** stottert der chinesische Wachstumsmotor ('Indeed, the Chinese growth engine is sputtering') → Le moteur économique chinois s'enroue ('The Chinese growth engine is sputtering') → **En effet**, le moteur économique chinois s'enroue

6.1.2. Pronoun reference (DP-COH-PRF)

DP-COH-PRF (Discourse and Pragmatics: Cohesion: Pronoun reference) refers to all cases of unclear pronominal reference.

Example:

But there are also imprisoned people waiting for their execution who are innocent. They never had a fair trial and a real chance to get out of ***it** → **jail**

6.2. Information structuring (DP-STR)

DP-STR (Discourse and Pragmatics: Information structuring) is to be used for cases where the source-text structuring of information (i.e. the organization of theme and rheme in a clause with regard to that of the surrounding clauses) is changed in the target text when it should not be, or is not changed when it should be. As stated by House (1997: 44), “any utterance consists of two basic parts which differ in the function they have in carrying information: (a) the theme which refers to facts either taken for granted, universally known, or given from the context, and which therefore do not, or only marginally, contribute to the new information conveyed by the utterance; (b) the rheme containing the main “new” information conveyed by the utterance”. The DP-STR error category also includes cases where sentences have been split or merged when they should not have been, and vice versa.

6.2.1. Thematic progression (DP-STR-THE)

DP-STR-THE (Discourse and Pragmatics: Information structuring: Thematic progression) refers to cases of clumsy “theme-rheme distribution” (House 1997: 44). The DP-STR-THE category also includes topicalization errors.

Example:

Le Musée des Beaux-Arts abrite une collection de renom international. Il contient la plus grande collection du pays → ***A collection of world renown is housed by the Musée des Beaux Arts**. It contains the largest collection in the country → The Musée des Beaux Arts houses a collection of world renown. It contains the largest collection in the country

6.2.2. Split/merge sentences (DP-STR-SPL)

DP-STR-SPL (Discourse and Pragmatics: Information structuring: Split/merge sentences) concerns errors involving the splitting of a sentence that should not have been split or the merging of sentences that should not have been merged (run-on sentences), and vice versa.

Example:

He said he would do the dishes, ***however** he did not → He said he would do the dishes. **However**, he did not

6.2.3. Syntactic preferences (DP-STR-SYN)

DP-STR-SYN (Discourse and Pragmatics: Information structuring: Syntactic preferences) is used for cases where non-finite verb forms are preferred over finite ones (and vice versa) or passive verb forms preferred over active ones (and vice versa).

Example:

People **having made** a difference to the world in a positive way will be remembered → people **who have made** a difference to the world in a positive way will be remembered



Warning: syntactic preferences vs voice errors

The DP-STR-SYN (Discourse and Pragmatics: Information structuring: Syntactic preferences) tag concerns *preferences*, i.e. items that are not wrong *per se* but that one would still want to improve. It should not be confused with GR-VOI (Grammar and Syntax: Voice), which involves *errors* related to the passive and active voice.

Example:

People **having made** → people **who have made** a difference to the world in a positive way will be remembered
⇒ DP-STR-SYN (Discourse and Pragmatics: Information structuring: Syntactic preferences)

vs

It is to ***hope** that ... → It is to **be hoped** that ...
⇒ GR-VOI (Grammar and Syntax: Voice)

7. **MECHANICS (ME)**

ME (Mechanics) groups together all errors related to spelling, punctuation and typography.

7.1. Spelling (ME-SPE)

ME-SPE (Mechanics: Spelling) consists of all spelling errors (including accents, diacritics, within-word hyphens and capitalization).

Examples:

(i) ***ambiguïty** → ambiguity

(ii) ***larousse** → Larousse, ***montag** → Montag

(iii) sondage ***publique** → sondage public

(iv) ministre bruxelloise de la ***mobilité** → ministre bruxelloise de la **Mobilité**

(v) non respect → non-respect

7.2. Punctuation (ME-PUN)

ME-PUN (Mechanics: Punctuation) covers all punctuation errors (full stops, commas, colons, semicolons, quotation marks, question marks, exclamation marks, brackets, en dashes, em dashes, apostrophes, slashes, etc.).

Examples:

(i) John Lennon who was born in Liverpool, met Yoko Ono in 1966 in London. → John Lennon, who was born in Liverpool, met Yoko Ono in 1966 in London.

(ii) Elle se dit contente qu'un nom de femme ait été choisi mais indignée que ce soit celui d'une femme blanche → Elle se dit contente qu'un nom de femme ait été choisi, mais indignée que ce soit celui d'une femme blanche ('She says she is happy that a woman's name has been chosen but bemoans the fact that it is a white woman's')

(iii) "It should have been one of the many brilliant Congolese women who fought the oppression of the evil colonialist ruler" → « Cela aurait dû être une des nombreuses femmes congolaises qui ont lutté contre l'oppression du malfaisant dirigeant colonisateur » → « Cela aurait dû être une des nombreuses femmes congolaises qui ont lutté contre l'oppression du malfaisant dirigeant colonisateur »

⇒ A space is missing before the closing quotation mark in the French translation



Warning: punctuation in numbers

This category does not include cases of punctuation errors in numbers. These should be tagged ME-UDN (Mechanics: Units, dates, numbers).

7.3. Units, dates, numbers (ME-UDN)

ME-UDN (Mechanics: Units, dates, numbers) is used for incorrectly formatted dates, units, or any other numbers in the target language (e.g. telephone numbers and currencies). It includes cardinal and ordinal numbers.

Examples:

(i) (AmE) 17/7/2000 → 7/17/2000

(ii) 2300 people → 2,300 people

(iii) 4th of February → 4 February

(iv) 7,3 million → 7.3 million



Warning

This category does not concern inconsistencies relating to numbers in source and target texts. These are content errors and should be signalled by the CT-DST (Content: Distortion) tag.

Examples:

1620 *instead of* 1610
16th *instead of* 17th century

8. REGISTER AND STYLE (REG)

In this tagging manual, register and style are regarded as “different approaches or perspectives for analyzing text varieties”, i.e. categories of texts that share social or situational characteristics (Biber & Conrad 2009: 15). Both register and style focus on the typical linguistic features of text varieties (e.g. lexical and grammatical characteristics). However, while the *register* perspective “connects those features functionally to the situational context of the variety” (ibid, 16), the *style* perspective associates them “with aesthetic preferences, influenced by the attitudes of the speaker/writer about language” (ibid, 18). In other words, in the register approach, linguistic patterns serve functional purposes; in the style approach, they serve aesthetic purposes. REG (Register and Style) groups together errors resulting from the use of words or structures which do not fit with the communicative function(s) of the translation and clumsy, awkward chunks of text that do not contain any errors as such but that one would want to reformulate. It also encompasses cases of redundant or unnecessary words that would be better left out.

8.1. Degree of (in)formality and register mismatch (REG-FML)

REG-FML (Register and Style: Degree of (in)formality and register mismatch) refers to the use of inappropriate register in the target text. It is used to signal instances of informal words or structures used in contexts where a neutral or formal word or structure would be more appropriate, and vice versa (e.g. *du* vs *Sie* in German, *tu* vs *vous* in French, *vosotros* vs *ustedes* in Spanish). It is also used for cases where the target language requires a more formal or informal tone than the source language (e.g. in marketing).

Examples:

(i) ***kids** → children

(ii) Although **people** in the US and New Zealand report similar levels of life satisfaction New Zealand’s overall HPI is 13 points higher than the US → Même si ***les gens** aux États-Unis et en Nouvelle-Zélande déclarent avoir des niveaux de satisfaction de vie semblables, le HPI général en Nouvelle-Zélande dépasse de 13 points celui des États-Unis (‘Although people in the US and New Zealand report similar levels of life satisfaction, the overall HPI in New Zealand is 13 points higher than in the US’) → même si **les citoyens** américains et néo-zélandais [...] (‘Although US and New Zealand citizens [...]’)

8.2. Heavy (REG-HEA)

REG-HEA (Register and Style: Heavy) is used for chunks of text which are stylistically heavy, clumsy or awkward.

Example:

Why may it be a good idea to incorporate sustainability information into stock or bond selection? → **Pourquoi l’idée d’inclure des renseignements sur la durabilité au moment du choix des stocks et des obligations semble bonne ?** (‘Why does the idea of including sustainability information when selecting stocks and bonds seem good?’)

8.3. Redundant (REG-RED)

REG-RED (Register and Style: Redundant) is used for all redundant or unnecessary words.

Example:

He went to London and **to** New York → He went to London and New York

9. META-TAGS

TAS contains four meta-tags, which are described in this section. While PLUS (positive feedback) should be used on its own, the other three meta-tags can be added to any error tag.

9.1. PLUS

PLUS is used to mark positive features, i.e. particularly good translations of words or larger chunks of text. It should be used on its own, not in combination with other tags (which are used in TAS to mark errors).

Example:

Facebook s'est notamment excusé après la transmission des données personnelles de 87 millions d'utilisateurs à la société Cambridge Analytica → After Cambridge Analytica collected personally identifiable information of 87 million users, **Facebook issued a public apology**

9.2. *Source Language Intrusion (SLI)*

Although TAS is descriptive, rather than explanatory or interpretative, in that it describes the nature of errors and not their possible sources, the SLI meta-tag makes it possible to mark errors that are suspected of being due to source language influence. It can be used in combination with any error tag.

Example:

Last June, a group of BLM protesters climbed onto the imposing statue of King Leopold II **waving** the flag of the Democratic Republic of Congo → En juin 2020, un groupe de protestants du mouvement Black Lives Matter avait grimpé sur l'imposante statue du roi Léopold II **agitant** le drapeau de la République démocratique du Congo

- ⇒ CT-DST (Content: Distortion): in English, *waving* directly follows *King Leopold II*. In French, a literal translation with the same word order and word category implies that it is the statue that is waving the flag
- ⇒ SLI: the English present participle is translated literally into French

9.3. Distortion (DIS)

The DIS meta-tag is used to signal semantic distortion caused by any error (typically, a lexical error). DIS can be used in combination with any error tag.

Example:

[Context: the process of renaming the Leopold II tunnel in Brussels]

Annie Cordy won a fifth of the **ballots** → Annie Cordy a remporté un cinquième des **scrutins** (instead of **voix**)

- ⇒ LT-WRD-SWO (Lexis and Terminology: Word: Single word)
- ⇒ DIS: in French, the noun **scrutin** implies that we are dealing with (political) elections (which is not the case in the present context)
- ⇒ The DIS meta-tag is added to the LT-WRD-SWO (Lexis and Terminology: Word: Single word) tag to signal that the lexical error leads to meaning distortion



Warning

The Distortion (DIS) meta-tag should not be confused with the CT-DST (Content: Distortion) tag. The Distortion (DIS) meta-tag is typically used to signal minor semantic shifts caused by erroneous segments, such as lexical items.

Example:

Die **Werksferien** fielen diesmal vor allem auf den Juli ('This time, plant holidays fell mainly in July') → Cette année, la plupart des **congés annuels** tombaient en juillet ('This time, annual holidays fell mainly in July') → la **fermeture des usines** ...

- ⇒ LT-WRD-MWO (Lexis and Terminology: Word: Multiword): *congés annuels* is not the equivalent of *Werksferien*
- ⇒ DIS: the use of *congés annuels* instead of *fermeture des usines* in French causes a meaning distortion

vs

CT-DST: He **cannot afford to** lose his reputation → Il ***n'a pas les moyens de** perdre sa réputation ('He does not have the means to lose his reputation') → Il **ne peut se permettre de** perdre sa réputation (Delisle et al. 1999: 40)

9.4. Error repetition (REP)

REP is used to highlight cases of repeated errors in a given translation.

Example:

Le nom du **Roi** [ME-SPE] Léopold II retiré du plus grand tunnel de Belgique [...] Mireille Tsheusi-Robert est une fervente militante de l'effacement des traces du **Roi** [ME-SPE, REP] Léopold II dans les rues de Belgique

- ⇒ In French, no capitalization of titles of nobility

Bibliography

- Biber, D. & Conrad, S. (2009). *Register, Genre and Style*. Cambridge: Cambridge University Press.
- Castagnoli, S., Ciobanu, D., Kunz, K., Kübler, N. & Volanschi, A. (2011). Designing a learner translator corpus for training purposes. In Kübler, N. (ed.), *Corpora, Language, Teaching and Resources: From Theory to Practice*, 221–248. Bern: Peter Lang.
- Delisle, J., Lee-Jahnke, H. & Cormier, M.C. (1999). *Terminologie de la traduction*. Amsterdam/Philadelphia: John Benjamins.
- DFKI (2015). Multidimensional Quality Metrics (MQM) Issue types. Retrieved from <http://www.qt21.eu/mqm-definition/issues-list-2015-05-27.html>
- Dussart, A. (2005). Faux sens, contresens, non-sens... un faux débat ? *Meta* 50(1), 107–119.
- Granger, S. (2010). Error annotation of learner corpora. Paper presented at the 4th Brazilian Computational Linguistics School (IV EBRALC), Porto Alegre, 6-9 October 2010.
- Granger, S., Dupont, M., Meunier, F., Naets, H. & Paquot, M. (2020) *The International Corpus of Learner English. Version 3*. Louvain-la-Neuve: Presses universitaires de Louvain. <https://corpora.uclouvain.be/cecl/icle/trial/>
- Granger, S. & Lefer, M.-A. (2020). The Multilingual Student Translation corpus: A resource for translation teaching and research. *Language Resources and Evaluation* 54(4), 1183–1199.
- Granger, S. & Paquot, M. (2008). Disentangling the phraseological web. In Granger, S. & Meunier, F. (eds), *Phraseology: An Interdisciplinary Perspectives*, 27–49. Amsterdam/Philadelphia: John Benjamins.
- Granger, S., Swallow, H. & Thewissen, J. (forthcoming). *The Louvain Error Tagging Manual Version 2.0*. CECL Papers 4. Centre for English Corpus Linguistics, Université catholique de Louvain.
- Halliday, M.A.K. & Matthiessen, C. (2004). *An Introduction to Functional Grammar*. London: Arnold.
- House, J. (1997). *Translation Quality Assessment – A Model Revisited*. Tübingen: Gunter Narr.
- Kunilovskaya, M. (2016). Description of RusLTC translation error typology. Retrieved from https://rus-ltc.org/static/references/description_RusLTC-error-typology2016.pdf
- Leech, G. & Smith, N. (2000). Manual to accompany The British National Corpus (Version 2) with Improved Word-class Tagging. University Centre for Computer Corpus Research on Language, Lancaster University.
- Lommel, A., Görög, A., Melby, A., Uszkoreit, H., Burchardt, B. & Popović, M. (2015). QT 21 – Deliverable 3.1 – Harmonised Metric. Retrieved from <http://www.qt21.eu/wp-content/uploads/2015/11/QT21-D3-1.pdf>
- Scott Doyle, M. (2003). Translation Pedagogy and Assessment: Adopting ATA's Framework for Standard Error Marking. *The ATA Chronicle* 32(11): 21–29.