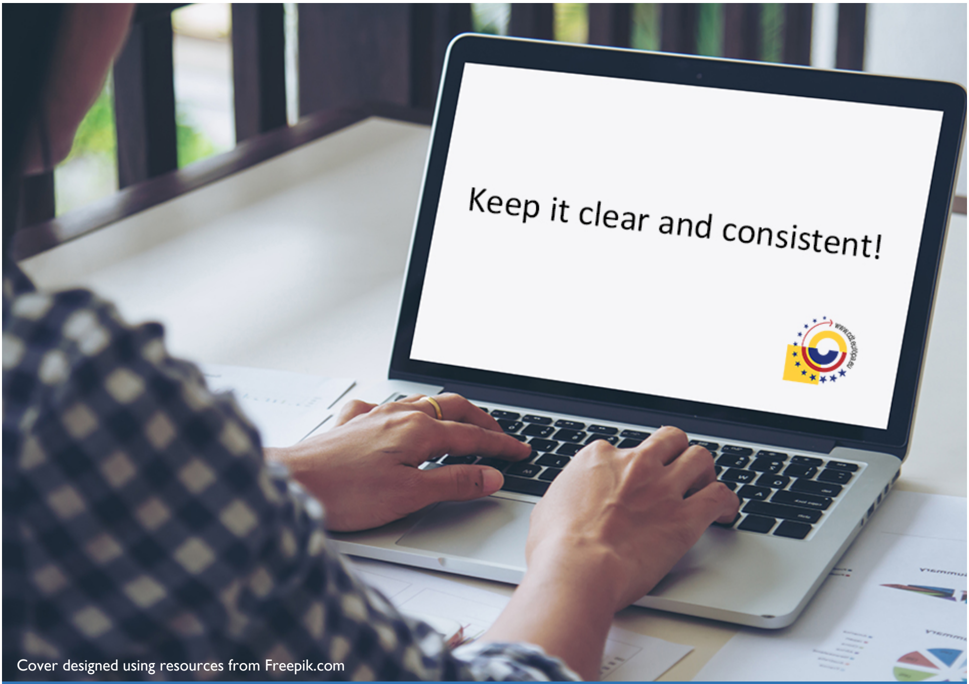
Keep it clear and consistent!

Cover designed using resources from Freepik.com

Writing for machine translation

TRANSLATION CENTRE FOR THE BODIES OF THE EUROPEAN UNION
Bâtiment Technopolis Gasperich ● 12E, rue Guillaume Kroll ● L-1882 Luxembourg
Tel.: +352 42 17 11 1 ● Email: cdt@cdt.europa.eu ● Internet: http://cdt.europa.eu

© Translation Centre for the Bodies of the European Union, 2021.

Last update: 2021-12-09
PDF ISBN 978-92-95212-36-7
doi:10.2817/191981

Contents

INTRODUCTION 3

WHAT IS MACHINE TRANSLATION? 4

WRITING FOR MACHINE TRANSLATION 5

DRAFTING TIPS 6

- 1. DECIDE ON YOUR TERMINOLOGY 6
- 2. USE PRONOUNS SPARINGLY 7
- 3. AVOID ABBREVIATIONS AND ACRONYMS 8
- 4. NOT TOO LONG 9
- 5. NOT TOO SHORT 10
- 6. AVOID LINE BREAKS 11
- 7. PUNCTUATE PROPERLY 12
- 8. AVOID WHOLE SENTENCES IN CAPITAL LETTERS 13
- 9. OTHER TIPS 14

INTRODUCTION

Writers in a multilingual environment need to keep in mind that their texts will be translated, and nowadays this is likely to be by a machine. The machine output is then usually edited by a human (although this is not always the case).

Machine translation (MT) has improved hugely in recent years. When used in the right circumstances and with human supervision it offers many advantages, such as being quicker and more affordable than conventional translation. This allows organisations to leverage multilingualism and reach much larger audiences than before.

When all stakeholders in the process – including authors – work together, they can optimise the benefits of machine translation in terms of efficiency, speed and multilingual opportunities.

The source text itself is a major factor in the quality of the machine translation produced. We recommend that authors consider a couple of points when drafting to ensure their writing is clear and will not 'confuse' the machine translation tool. **Authors will also benefit as this will enable their message to be translated accurately and read by more people.**

This guide provides tips on writing 'machine-friendly' texts. These have been drawn up by the Translation Centre's linguists and language technology experts based on their own experience with machine translation, and should not be considered a substitute for academic or scientific research. This guide is also subject to review as technology progresses.

WHAT IS MACHINE TRANSLATION?

Machine translation is the automated translation of text from one natural language to another using a computer application (*ISO 18587:2017 — Post-editing of machine translation output*).

The Translation Centre uses neural machine translation (NMT), which is a state-of-the-art technology based on a deep neural network 'trained' with data. The knowledge model derived from this training phase is used to produce translations from new input data.

NMT generates instant translation output. It also enables customised translation models to be created that can be adapted to a client's particular terminology and style.

However, NMT processes each sentence as a separate unit. It cannot – as yet – infer anything from the text as a whole or take into account the sentence that comes before or after. This can lead to numerous traps, which this guide will help you to avoid.

WRITING FOR MACHINE TRANSLATION

When writing for machine translation, it is important to follow the principles of clear writing in general, above all keeping in mind the purpose of the text and the reader. Please also refer to the Translation Centre's guide *Writing for translation*, which contains tips on how to write concisely and clearly so that your message is conveyed accurately in other languages too.

Along with these more general guidelines, there are some specific points that relate to writing for NMT. Human translators are able to make educated guesses and recognise mistakes in the original text, which they can remedy when they translate. However, machines do not have this awareness, so it is even more important when writing for machine translation to draft clearly and ensure that syntax, spelling and punctuation are correct.

DRAFTING TIPS

1. Decide on your terminology

Use terminology consistently throughout the text. Using synonyms can confuse the machine and result in different translations of the same word, which can be wrong or too general.

Original (EN)	Machine translation (DE)
While in 2014 the module was followed by 311 **asylum officials**, in 2015 the overall number of participants decreased to 102 (-67%).	Während im Jahr 2014 311 **Asylsachbeamte** an dem Modul teilnahmen, ging die Zahl der Teilnehmer im Jahr 2015 sank die Gesamtzahl der Teilnehmer auf 102 (-67 %).
Please note that **one asylum officer** may perform several tasks.	Bitte beachten Sie, dass ein **Asylbeauftragter** mehrere Aufgaben wahrnehmen kann.

Asylum officials' and 'asylum officer' refer to the same role, but the NMT engine does not realise this and so chooses two translations with different meanings in German.

2. Use pronouns sparingly

It is sometimes better to use a noun rather than a pronoun for clarity. NMT engines may generate a pronoun with the wrong gender in the target language, which makes it unclear what is being referred to.

Original (EN)	**Machine translation (FR)**
About the agency	Au sujet de l'agence
It aims to combat counterfeiting.	**Il** entend lutter contre la contrefaçon.
In addition, **it** seeks to support entrepreneurs and creators.	En outre, **il** vise à soutenir les entrepreneurs et les créateurs.

'L'agence' is feminine ('elle'), but the NMT engine has used the masculine pronoun ('il') as it did not know what 'it' referred to.

3. Avoid abbreviations and acronyms

NMT engines may struggle to deal with abbreviations correctly. Sometimes an abbreviation has no official translation and should be left as it is.

Avoid using acronyms and abbreviations in texts as they could be translated incorrectly by the machine. Instead write out the name in full, or use a combination of the full name with the acronym/abbreviation in brackets.

Original (EN)	**Machine translation (FR)**
The **ECS** 2019 captures 'workplace well-being', a concept measured using a composite indicator.	Le **CSE** 2019 met en évidence le «bien-être sur le lieu de travail», un concept mesuré à l'aide d'un indicateur composite.

'CSE' is not the correct translation of 'ECS'. Using the full form ('European Company Survey') would avoid this error. MT engines recognise data they have previously encountered and 'learnt', so they are unlikely to understand less common abbreviations.

4. Not too long

Write clearly structured sentences as NMT engines have difficulty understanding sentences that are too complicated or long (over 60 words). As a general rule, it is best to convey one idea per sentence to avoid confusion.

Original (EN)	**Machine translation (FR)**
We need to sit dynamically, take frequent breaks, and stretch and move. **The best way to do this is by making sitting a smaller part and moving a bigger part of workplace routine.**	Nous devons nous asseoir de manière dynamique, prendre des pauses fréquentes et nous étirer et nous déplacer. **Le meilleur moyen d'y parvenir consiste à placer une partie plus petite et à déplacer une plus grande partie du lieu de travail.**

This poorly structured sentence ('making [sitting…] and [moving…] of workplace routine') results in a nonsensical translation (back translation: 'The best way to do this is to place a smaller part and move a larger part of the workplace').

5. Not too short

Very short sentences (around seven words or less) may not be understood correctly by machine translation engines as they offer little or no context.

Original (EN)	**Machine translation (RO)**
Polish your presence online.	**Prezența dumneavoastră online în limba poloneză**

🚨 Because the sentence is too short, the NMT engine fails to render the correct meaning of ‘Polish’ (back translation: ‘Your presence online in Polish’).

6. Avoid line breaks

Be careful with formatting and text structure, and do not use line breaks in headings and tables. NMT will treat all parts of a title or heading separately, so elements that go together as a unit of meaning should not be split up.

Original (EN)	**Machine translation (ES)**
PSP Unique Identification Number	**PSP Unique** **Número de identificación**

Because of the line break after ‘Unique’, this table heading has been translated as two separate units of meaning (‘PSP Unique’ and ‘Identification Number’) instead of one.

7. Punctuate properly

NMT engines rely on punctuation to get clues about meaning, so it is important to use it accurately.

Original (EN)	**Machine translation (RO)**
Despite negotiated guarantee measures, the subletting solution is not without risks, magnified by the current **COVID 19 crisis**.	În pofida măsurilor de garantare negociate, soluția de subînchiriere nu este lipsită de riscuri, amplificată de **criza de apă** actuală **din 19**.

The missing hyphen between ‘COVID’ and ‘19’ (‘COVID-19’) confuses the NMT engine, which fails to recognise it as the name of a disease (back translation: ‘the water crisis of 19’).

8. Avoid whole sentences in capital letters

Use normal sentence case when you write full sentences, as NMT is less accurate when processing words in capital letters and may produce strange results. It is also much easier to read lower case letters.

Original (ES)	Machine translation (EN)
EN PARALELO, MI ACTIVIDAD COMO COORDINADOR DE PLANES Y PROGRAMAS DE FORMACIÓN IMPLICA LA DIRECCIÓN DE EQUIPOS REDUCIDOS DE COLABORADORES.	**INPUT, MI ACTIVITY AS A COORDINATOR OF PLANS AND TRAINING PROGRAMMES INVOLVES THE MANAGEMENT OF COLLABORATIVE REDUCED EQUIPMENT.**

The translation makes no sense. Compare the MT output with the actual translation: 'At the same time, my work as the coordinator of training plans and programmes involves managing small teams of staff.'

9. Other tips

- Use **automatic heading styles** in MS Word (this also ensures they will be translated in the same way as the table of contents).
- Use **quotation marks** consistently (stick to the same type throughout the text).
- Use **automatically generated bullet points** rather than creating them manually.
- Use **full words instead of special signs or symbols**. NMT engines may not understand signs or symbols that are not used in the target language. For example, write the conjunction ‘and’ instead of ‘&’.

TF-01-21-517-EN-N

Publications Office
of the European Union