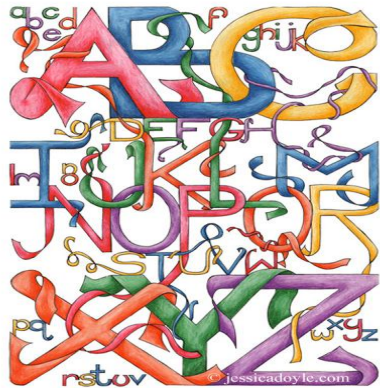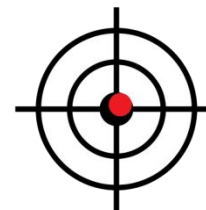Linguistic Systems, Inc.

MT Post-Editing

**Guidance Document for Post-Editors:**

Scope, Objectives and Requirements

Scope

Linguistic Systems Inc. (LSI) sells machine translation (MT) as well as machine translation plus post-editing (MTPE) as low-cost, fast-turnaround alternatives to full human translations. These services are available to select clients via LSI's proprietary **Ai Translate by LSI**^{TM} online ordering system.

The machine translation output is generated by LSI's proprietary MT engines, which may be enhanced with a client's glossary and translation memory. The goal of post-editing by a linguist is to improve the MT output's accuracy and readability, thus bridging the quality gap between MT and a full human translation.

Important: When selling MT or MT plus post-editing, LSI advises its clients that these services are _not_ geared towards creating a 100% accurate translation of top linguistic quality, but that only a full human translation can be expected to achieve those results.
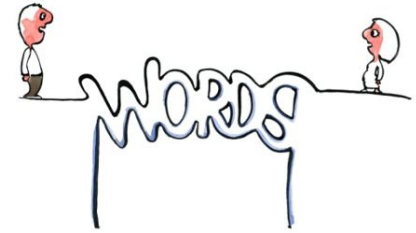
Rationale

While many language professionals have long resisted MT as a viable alternative to human translation, the speed and ease of readily available translation applications, such as Google Translate, have greatly increased the demand for fast-turnaround translation work. MT is here to stay, and it is widely popular and widely used.

LSI has developed its own secure MT engines to meet clients' demands for speed and cost-effectiveness. These demands are especially relevant in the legal field, where large text corpora may need translation, for example, as part of an eDiscovery project. Often, the raw MT output is good enough for "gisting" purposes, helping legal reviewers identify relevant text. At other times, the raw MT output may be too scrambled and flawed to convey a coherent meaning, and the client orders post-editing to increase a text's accuracy and readability.

Data Security Alert: Readily available translation applications do not provide data security and keep all the data for their own purposes. Therefore, using these applications exposes confidential data to unauthorized access. LSI's translators and post-editors are prohibited from using Google Translate and similar public applications when working on LSI translation assignments.

Objectives

LSI sells two levels of MT post-editing:

Light Post-Editing (LPE): The objective of a light MT post-edit is to create a translation in which essential words are translated correctly, so that the reader can capture about 90% of the full meaning of the source text. Improving fluency is subordinated as a tradeoff to the objective of reducing cost. This may result in occasional minor grammar errors and/or awkward style, but the priority of comprehensible correct and unambiguous meaning is preserved. Depending on the quality of the MT output (starting point), a light post-edit should take about 30-40% of the time of a full human translation.

Full Post-Editing (FPE): The objective of a full MT post-edit is to create a translation that is as correct, unambiguous and as fluent as possible, with relatively good syntax. The post-editor is not expected to produce maximum style, but the composition should follow all grammatical rules of proper writing. The accuracy expectation of a full post-edit is 95-100%. Depending on the quality of the MT output (starting point), the full post-edit should take about 65-75% of the time of a full human translation, thereby still shortening turnaround time and cost.
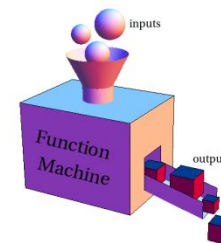
Technical Considerations

The quality of MT output varies greatly and depends on many factors, including source language and file type.

Source language: In general, most Western European languages (for example, German, French and Spanish) yield better MT output than Asian languages (Japanese, Korean and Chinese).

File type: The quality and file type of the source documents can affect the quality of the MT output. LSI's MT engines produce target files in these formats (document layout is retained as much as possible):
DOCX, HTM(L), PDF, PPTX, XLSX, XML, or TXT.

Post-editors are assigned files in the format produced by the MT engines, and they are expected to return files in the same format as received.

Common Errors Found in Raw MT Output

MT post-editors should be aware of these errors commonly found in MT output:

- Inaccurately translated words,
- Words left in the source language,
- Words translated that should not have been translated (i.e. proper names),
- Words omitted,
- Words scrambled (i.e., mix of letters, numbers and symbols),
- Literal translations that create inaccurate or ambiguous meaning,
- Incoherent sentences or paragraphs,
- Grammar mistakes,
- Stylistic errors.

MT Post-Editors: Basic Competence

MT post-editing is closely related to, but not the same as translation. LSI follows the competence requirements for post-editors outlined in the international MT post-editing standard, ISO 18587.

The basic post-editor competence requirements include:

- Translation competence and fluency in source *and* target language,
- Cultural competence in source and target language,
- Linguistic and textual competence to produce coherent texts,
- Domain competence to understand the content to be post-edited,
- Ability to follow instructions.

MT Post-Editors: Technical Competence

MT post-editors should also have the following technical competences:

- Experience in using modern translation tools and IT systems,
- Experience in using online research tools to acquire specialized knowledge,
- A general knowledge of MT technology and common MT errors,
- A general knowledge of how terminology management systems interact with MT,
- Ability to assess whether post-editing an MT output makes sense (time and effort).

LSI may provide assignment-specific guidance and training to post-editors, for example, when a project includes the use of translation memory and glossaries that enhance the MT output to be post-edited.

Assignments

A post-editing assignment from LSI includes the following:

- The source material,
- The raw MT output,
- Translation memory files (if available),
- The type of post-editing required (LPE or FPE),
- The expectation of accuracy to be achieved,
- The estimated time to be spent on the assignment , and/or
- The estimated dollar amount to be paid for the assignment.

Important: Since MT post-editing is cleaning up MT output without the expectation of perfection, the question of *how-clean-is-clean-enough* does arise from time to time. The estimated time and remuneration for the assignment should serve as a rough guide for applying the appropriate level of MT revision.

Review of the Assignments

It is the responsibility of LSI's project management team to analyze the quality of the MT output before assignment, and to provide a ballpark estimate of the post-editing effort required to achieve the specified results.

Nevertheless, the quality of the MT output may not be uniform across assigned files in a large project. For example, MT output may be unreadable in parts of the project, where the MT engine mixed letters with numbers, such as "sc1no0" for "school". In this case, the MT output may not be suitable for MT post-editing, as it may take longer to unscramble the MT than to translate the document from scratch.

Important: It is the responsibility of the post-editor to alert LSI's project manager immediately, if a project – or parts of a project – are not suitable for MT post-editing, or if the effort to clean up the MT is likely to take much more time than estimated.

Post-Editing Process

A client who orders MT post-editing expects to receive an accurate and comprehensible translation quickly and economically. Therefore, time management is key to achieving this goal!

While working on the assignment, the post-editor must keep in mind that a light post-edit (LPE) should take about 30-40% of the time of a full human translation, and a full post-edit should take about 65-75% of that time.

An effective way to manage time is to start with corrections to <u>key words</u>, such as the subject, verb and object of a sentence, before moving on to supporting words such as adjectives and prepositions, and finally to word order and grammar.

✓   **Important**: LSI's post-editors are expected to <u>refer to the source text frequently</u> in order to correct errors made by the MT engine. The post-edited text should be as accurate and comprehensible as possible, given the time constraints placed on the effort.

Feedback and System Improvements

The MT post-editing process has many variables. Therefore, achieving intended outcomes depends on good communication between all project team members as well as the continuous improvement of technical resources.

Feedback is important

Post-editors are strongly encouraged to provide feedback on the performance of the MT system and the post-editing process to their LSI project manager.

Acknowledgement

The review of this presentation is part of LSI's qualification process for MT post-editors. Therefore, LSI needs to have on file your acknowledgement that you have read and understood the content of this presentation, and that you possess the competences required to carry out post-editing assignments.

Please return your acknowledgement via email to our vendor management staff at vendor@linguist.com, subject line "MT-Post-Editing-Training".

Once your qualification process has been completed and you have been assigned an actual MT post-editing project, your LSI project manager will guide you through the assignment and answer any questions you may have.