

# PDF Handling

## What is it?

- The process of converting a source PDF file into a CAT tool compatible file format in order to obtain a reasonable similarity and text integrity to the initial source file.

## How does it work?

- Ordered via our secure client portal (Ease / API)
- Offered in addition to any translation project or as a standalone service for:
  - **Low complexity level documents:**
    - PDF exported from Adobe Indesign, Illustrator or MS Word file:
    - Selectable text
    - Simple Layout (mostly text and without translatable images or charts)
    - Latin script and Asian scripts languages
  - **High complexity level images:**
    - Scanned or an image saved as PDF:
    - Text cannot be selected
    - Latin, Cyrillic, Asian scripts
- Assigned to Lingo24 DTP specialists
- Project-related information and specifications to be provided
- The process of PDF Handling usually involves two components:
  - **OCR Conversion/Content extraction:**
    - In the pre-translation step, Optical Character Recognition (OCR) Conversion or other forms of text extraction are performed on the received files to generate a potential translatable source file based on the needs/user-case of the client (i.e. layout preservation vs gisting of content)
    - Based on the output generated, the Lingo24 DTP artist will advise the Lingo24 Project Manager and/or Client on the likely success of the conversion process into the translation workflow.
  - **Translation Input Adjustment:**
    - Depending on the complexity of the source and the OCR output, various adjustments on the generated source file may be required to improve the file translatability in Lingo24's Translation Management System, where applicable. This includes formatting simplification (e.g merging of styles/cells), aggressive tag handling (e.g. merging of Microsoft Office tags) and restructuring of content
    - The output of the PDF conversion is always an editable format, either MS Word (DOCX), Plain text format (TXT) or Microsoft Excel format (XLSX)

- We follow general client requirements and project-specific instructions provided by the client.
- **Translation Environment:**
  - At this stage, depending on the output quality, the Lingo24 DTP artist will provide consultancy on whether the translation stage of the core service could be processed in our Translation Management System or if it must be carried out offline to avoid polluting the Translation Memory with low quality entries.
- Output capacity:
  - For low complexity level, up to 20 pages A4 size per language / hour
  - For high complexity level, up to 10 pages A4 size per language / hour
- Delivery is made via the original ordering portal (Ease / API), with its related notifications
- Default service availability: 24 hours a day/5 days a week

## Service Price

- Based upon a per hour pricing scheme
- Minimum charges per language combination apply

## Adherence to service scope

Project Management oversees implementation of and monitors adherence to project configuration:

- Client-Lingo24 agreement, project-related client information and DTP specifications are documented and communicated
- DTP specialist assigned meet the Lingo24 standard requirements for competences and qualifications
- Assignment to a second Lingo24 DTP specialist for DTP peer review purposes
- Ongoing Query Management and Issue Clarification
- Use of relevant tools and programs: - Adobe Illustrator, Adobe Acrobat DC
- Workflow steps set up
- Lingo24's <Post Layout Checking> service may be added to ensure the highest degree of layout similarity and text integrity with the source file, if supplier validation is required.

## Client Feedback

- Client to confirm receipt of deliverables within 1 business day, if validation is specifically requested either by Lingo24 or by the client
- Explicit service acceptance within 3 business days, otherwise acceptance is implied (unless customer specific service acceptance timeframes and criteria have been requested and confirmed by Lingo24)

- Client feedback assessed by the Quality and Experience team when quality questions have been raised and appropriate corrections and/or corrective actions implemented, if necessary.

## Quality and Scope of service

### High-level Error Categories

Accuracy	Text extraction	Fail to extract non editable text from images. Fail to remove not needed content.
	Graphics	Incorrect use of graphics.
	Omission	Fail to localize the country specific elements, such as PNs, phone numbers, addresses etc.
	Truncation/ text expansion	Missing text or text flowing outside text boxes.
	Callouts	The callouts or the leaders do not mirror the source file.
Technical requirements	CAT tool related	Providing improper CAT tool friendly files
	File optimization	Fail to implement file/client optimization standards
	File management	Fail to respect the file management standards.
	Workflow	Fail to respect the standard or client specific DTP workflow.
	Printing specs	Fail to respect the printing instructions received from the client.
	Delivery package	Fail to respect client's standards for the delivery files.
Layout	Mirror the source	Fail to match the general source style and to preserve the overall visual impact of the page.
	Layout Framework	Change the position or dimension of layout/design components.
	Text handling	Text boxes are overlapping with each other.
Design style	Company style	Text formatting done outside the standards set by specific client workflows.
		Fail to use the latest image elements.
		There is a problem relating to design aspects (vs. linguistic aspects) of the content.
	Local formatting	Issues related to local formatting (rather than to overall layout concerns).
Length	There is a significant discrepancy between the source and the target text lengths.	
Other	PDC	Fail to implement all the amends suggested by the reviewer in the post DTP checking stage.
		Any other issues.

## High-level Error Category Weight

Accuracy	Multiplier: 3
Technical requirements	Multiplier: 3
Layout	Multiplier: 2
Design style	Multiplier: 2
Other	Multiplier: 1

## Error Severity Levels

Severity Level	Definition	Multiplier
Minor	Errors that don't lead to loss of meaning layout or design and wouldn't confuse or mislead the user but would be noticed, would decrease stylistic quality, fluency or clarity, or would make the content less appealing.	1
Major	Errors that may confuse or mislead the user due to significant change in meaning, layout or design, or because errors appear in a visible or important part of the content.	5
Critical	Errors that may render the document unusable for its intended purpose, carry health, safety, legal or financial implications, violate geopolitical usage guidelines, damage the company's reputation, cause the application to crash or negatively modify/misrepresent the functionality of a product or service, or which could be seen as offensive.	10

## Quality Threshold

Professional DTP file formats such as the ones provided by Adobe (Illustrator or Photoshop) offer the possibility of having editable text within images hence more control over text allowing for enhanced similarity with the source. The quality pass/fail threshold for these file formats is 98%.

For image formats where the text and the background are merged into one object (PNG, JPG etc), the DTP artist needs to rebuild parts of the image before adding the translated text. This process may lower the similarity with the source down to 75%.

The quality pass/fail threshold is based on:

- Number of confirmed errors per high-level error category
- The error category weight
- Error severity level and multiplier
- Number of pages.

We do not label as confirmed errors, customer revisions or feedback on Lingo24's DTP which are:

- Preferential changes: both the original layout and the revised version have equal merit and are correct; either can be used
- Unsupported changes: changes not in line with the general client requirements and DTP specifications, changes not in line with the source layout, or changes containing errors.

## Impact on Quality

There are various service-specific risk factors that can have a negative impact on the quality of the core service scope:

- The output of the low complexity documents undergoing PDF Handling will have 90% to 100% similarity with the source in terms of layout
- The output of the high complexity documents undergoing PDF Handling will have a minimum of 75% degree of similarity with the source in terms of layout and text integrity.

## Consultancy

Get in touch with your Account Manager / Project Manager for any custom requirements:

- Shorter turnaround times
- Source file format and content not compliant with the service standard
- Revisions or additions to the source content after order confirmation and with service provision in progress
- Developing a brand new layout
- Weekend and 24/7 coverage.

## Disclaimers

- Following assessment of <PDF Handling> requirements, capabilities and capacity, Lingo24 reserves the right to agree one or more of the following:
  - Longer order confirmation times
  - Longer turnaround times than requested by client
  - To decline requests if no agreement can be reached and service provision cannot be guaranteed (incomplete DTP specifications: missing links, missing fonts, ambiguous information about the content to be translated, etc.)
  - To suggest a different service package
- Due to the file construction approaches used by OCR Conversion tools, Translation Memory leverage may be impacted due to increased formatting tags (internal style markers protected from linguists). In such cases, we may recommend the translation stage be carried out outside Lingo24's Translation Management System. In both cases, the client will be advised of the impact and decision taken on how to proceed.