Supervertaler User Guide

Multi-LLM AI-powered Translator & Proofreader

Table of Contents

What is Supervertaler?

Supervertaler is an AI-powered translation and proofreading tool that helps professional translators work more efficiently and accurately. Unlike simple translation tools, it:

- **Uses multiple AI providers** (Claude, Gemini, ChatGPT) for different strengths
- **Considers full document context** instead of translating sentence by sentence
- **Integrates visual context** from images and figures
- **Leverages translation memories** and tracked changes from previous projects
- **Handles both translation and proofreading** workflows

Getting Started

1. Install Required Libraries

Before first use, install the AI libraries you want to use:

```bash
pip install anthropic        # For Claude
pip install google-generativeai  # For Gemini
pip install openai           # For ChatGPT
pip install pillow           # For image support (optional)
```

2. Set Up API Keys

1. Create an `api_keys.txt` file in the same folder as Supervertaler
2. Add your API keys (remove the `#` to uncomment):

```
# API Keys Configuration
google = your_google_api_key_here
claude = your_claude_api_key_here
openai = your_openai_api_key_here
```

3. Get API Keys

- **Claude**: Sign up at console.anthropic.com
- **Gemini**: Get a key from aistudio.google.com
- **ChatGPT**: Create an account at platform.openai.com

Note: You only need ONE API key to get started, but having multiple gives you more options.

Understanding the Interface

The Supervertaler interface has two main areas:

Left Side - Controls

- **Operation Mode**: Choose Translate or Proofread
- **File Selection**: Input and output files
- **Context Sources**: TM, tracked changes, images, custom instructions
- **AI Settings**: Provider, model, language settings
- **Process Controls**: Start, list models, refresh

Right Side - Log

- Shows real-time progress and status messages
- Displays any errors or warnings
- Tracks processing statistics

Operation Modes

Translate Mode

Purpose: Translate source text to target language

Input Format: Text file with one source sentence per line

```
This is the first sentence to translate.
This is the second sentence to translate.
```

Output Format: Tab-separated file

```
Source text    Translated text
This is the first sentence.    Dit is de eerste zin.
```

Best for: Initial translations, new content

Proofread Mode

Purpose: Review and improve existing translations

Input Format: Tab-separated file with source, existing translation, and optional comments

```
Source text    Existing translation    Original comment (optional)
Hello world    Hola mundo    This might need review
```

Output Format: Enhanced with AI improvements and change tracking

```
Source text    Revised translation    Combined comments
Hello world    Hola mundo    PROOFREADER COMMENT (AI): Translation is accurate and appropriate.
```

Best for: Quality assurance, improving existing translations

Context Sources

Context sources provide additional information to improve AI accuracy:

1. TM File (Translation Memory)

What it does: Provides exact matches from previous translations **Formats supported**: .tmx (standard) and .txt (tab-separated) **How it works**:

- Exact matches are used automatically (no AI processing needed)
- Partial matches provide context to the AI
- Most cost-effective way to handle repetitive content

Example TM file (.txt):

```
Welcome to our website    Welkom op onze website
Contact us                Neem contact met ons op
```

2. Tracked-changes

What it does: Shows revision patterns from previous projects **Formats supported**: .docx (from memoQ) and .tsv files **How it works**:

- Load files with tracked changes to see Original → Final patterns
- Use the Browser to search for similar terminology
- Copy useful patterns for reference during translation

Use cases:

- Learn client preferences from previous revisions
- Maintain consistency across projects
- Apply lessons learned from editor feedback

3. Document Images Folder

What it does: Provides visual context for figures and diagrams **Formats supported**: .png, .jpg, .jpeg, .webp **How it works**:

- Name images to match figure references (e.g., "Fig 1A.png")
- AI automatically shows relevant images when translating figure references
- Especially useful for patents, manuals, and technical documents

Naming examples:

```
Figure1.png     → References to "Figure 1"
Fig2A.jpg       → References to "Figure 2A"
Diagram3.png    → References to "Diagram 3"
```

4. Custom Instructions for AI

What it does: Adds specific guidance to the AI system prompt **Use cases**:

- Client-specific terminology preferences
- Industry-specific translation guidelines
- Style and tone requirements
- Special formatting instructions

Examples:

Use UK English spelling (colour, realise, etc.)
Keep all product names untranslated
Use formal tone throughout
Translate "user" as "gebruiker", not "user"

Basic Workflows

Simple Translation Workflow

1. **Set mode** to "Translate"
2. **Select input file** (one sentence per line)
3. **Choose output location**
4. **Select AI provider and model**
5. **Set source and target languages**
6. **Click "Start Process"**

Enhanced Translation Workflow

1. Follow simple workflow steps 1-3
2. **Load TM file** if available
3. **Load tracked-changes** from previous projects
4. **Set document images folder** if translating visual content
5. **Add custom instructions** for specific requirements
6. Continue with steps 4-6 from simple workflow

Proofreading Workflow

1. **Set mode** to "Proofread"
2. **Select input file** (tab-separated: source, existing translation, comments)
3. **Choose output location**
4. **Load context sources** as needed
5. **Select AI provider and model**
6. **Click "Start Process"**

Advanced Features

Multiple AI Providers

Each AI has different strengths:

Claude (Anthropic):

- Excellent for nuanced, context-aware translations
- Strong with complex sentence structures
- Good following of custom instructions

Gemini (Google):

- Fast processing
- Good with technical content
- Strong multilingual capabilities

ChatGPT (OpenAI):

- Versatile across many domains
- Good balance of quality and speed
- Strong with creative content

Chunking System

- Large documents are automatically split into manageable chunks
- Default chunk size: 100 lines (adjustable)
- Ensures consistent AI performance
- Progress tracking for large files

Model Selection

- **List Models**: See all available models for your chosen provider
- **Refresh Models**: Update the model list
- Different models optimized for different use cases

Tracked Changes Browser

- **Search functionality**: Find specific terms or patterns
- **Full text view**: See complete original and final text (not truncated)
- **Copy to clipboard**: Easy reference during translation
- **Multi-file support**: Accumulate patterns from multiple projects

---

Tips for Best Results

File Preparation

- **Clean input files**: Remove extra line breaks and formatting
- **Consistent encoding**: Use UTF-8 encoding
- **Reasonable chunk size**: 50-200 lines depending on complexity

Context Sources

- **Quality over quantity**: Better to have accurate TM than huge, messy one
- **Relevant tracked changes**: Load changes from similar projects/domains
- **Proper image naming**: Match figure references exactly

AI Settings

- **Choose appropriate models**: Larger models for complex content
- **Experiment with providers**: Different AIs excel at different content types
- **Custom instructions**: Be specific but concise

Language Settings

- **Be specific**: "Dutch (Netherlands)" vs "Dutch (Belgium)"
- **Consistent terminology**: Use same language codes in TM and settings

---

Troubleshooting

Common Issues

"No API keys configured"

- Check that `api_keys.txt` exists in the same folder as Supervertaler
- Ensure you've removed the `#` from the key lines
- Verify your API key is valid

"Model not initialized"

- Check your internet connection
- Verify your API key has sufficient credits
- Try a different model or provider

"No data from input file"

- Check file encoding (should be UTF-8)
- Ensure file has content and proper format
- Remove empty lines and extra formatting

"TM Load Error"

- For TMX files: ensure they contain your source/target language pair
- For TXT files: check tab separation and encoding
- Verify language codes match between TM and GUI settings

Images not loading

- Install Pillow library: `pip install pillow`
- Check image file formats (.png, .jpg, .jpeg, .webp)
- Verify image names match figure references in text

Performance Tips

For large files:

- Increase chunk size to reduce API calls
- Use TM files to reduce AI processing
- Process during off-peak hours for better API response

For cost optimization:

- Use TM files for repetitive content
- Choose smaller, faster models for simple content
- Remove unnecessary context sources

For quality improvement:

- Use larger, more sophisticated models
- Provide comprehensive custom instructions
- Load relevant tracked changes and images
- Review and refine output incrementally

Getting Help

Log Messages: The right panel shows detailed progress and error information **Model Information**: Use "List Models" to see available options **File Format Examples**: Check the format requirements in each section above

Remember: Supervertaler is designed to assist professional translators, not replace them. Always review and refine the output according to your expertise and client requirements.