Automated translation into Chinese

It is now time to take automatic translation seriously. The development of AI has required systems to understand language better, as most interfaces to AI are in written or spoken language. In the past, automatic translation of words and sentences has achieved some success. AI has increased success rates not just for words and sentences but also for paragraphs and papers.

Currently (September 2024), internet search engines do not automatically translate your internet content into multiple languages worldwide. This will change, allowing your business to be indexed worldwide in all written languages. You need to review your website content to be ready for the opportunity. Remember, AI systems have already started!

Historically

When reading information, you will assume the author has written the contents and not an automatic translation program.

Although automatic translation is improving, it still makes mistakes, and this will reflect poorly on the website's author.

For many years, browsers have included a feature that translates website content.

Chinese people often use automatic translation software to read foreign content, such as news reports. In doing so, they know that the translation will not be 100% correct, so errors are not attributed to the author.

Chinese engineers have been developing translation software from English into Chinese, the first foreign language in China. We believe Chinese translation software is equal to, if not ahead of, the market in translating from English to Chinese.

Not having your internet content translated into Chinese is not an issue, provided that

- Do not use automatic translation
- Avoid complex language and structure
- Images with text are not translated automatically
- SEO data (currently not translated by search engines)

We recommend not using automated translation in your website content. The Chinese reader will use an automatic translation service, e.g., Baidu Translate, when needed.

Today (September 2024)

What can we do to improve the automatic translation?

Firstly, let us look at why the automatic translation fails.

Why is automatic translation failing?

The automatic translation process often encounters technical, linguistic, and contextual challenges.

1. Complexity of Natural Language

- **Context and Ambiguity**: Human language is inherently ambiguous. Words can have multiple meanings depending on the context. Automatic translation systems often struggle to pick the correct meaning without fully understanding the context, resulting in awkward or incorrect translations, for example,

    In English, "bank" could refer to a financial institution or the side of a river. Automatic translation might choose the wrong meaning based on insufficient context.

- **Idioms and Phrases**: Many languages have idiomatic expressions or colloquialisms that don't translate, for example,

    The English idiom "it's raining cats and dogs" would be confusing in a literal translation. Automatic systems may translate it directly instead of recognising it as an idiom and providing an equivalent expression in the target language.

- **Cultural Nuances**: Cultural references, humour, and tone are complex to convey in another language. Automatic systems may translate words correctly but fail to represent their implied meaning or cultural significance.

2. Grammar and Syntax Differences

- **Word Order Variations**: Languages have different grammatical structures and word orders. For instance, the subject-verb-object order in English doesn't apply to all languages. Automatic translation can mess up the word order, leading to awkward phrasing or misunderstood sentences.
- **Tense and Gender Differences**: Some languages have complex tense systems (e.g., Greek, Russian), gendered nouns (e.g., French, Spanish), or other grammatical features (e.g., cases in German or Slavic languages) that don't have direct equivalents in different languages. Automatic translators might oversimplify or get these details wrong.

3. Lack of Contextual Knowledge:

- **Domain-Specific Terminology**: In specialised fields like medicine, law, or technology, automatic translators may struggle with domain-specific jargon. The systems often default to general vocabulary, leading to incorrect or nonsensical translations.
- **Named Entities**: Proper nouns like names, places, brands, or technical terms are sometimes not translated correctly. If an automatic translator doesn't recognise a term as a proper noun, it might incorrectly translate it, which is confusing.

4. Limited Data for Some Languages:

- **Less Data for Low-Resource Languages**: Languages like English, French, or Spanish have vast training data for machine translation models. However, low-resource languages (e.g., Swahili and Basque) may lack sufficient data, leading to lower translation quality. The system might translate to/from English first or use approximations.
- **Abbreviations:** The Automatic translation process often does not understand/misunderstands abbreviations. Try to avoid abbreviations.
- **Dialects and Regional Variations**: Automatic translation systems often don't account for regional dialects, slang, or regional variations within the same language. If the system is trained on only one language variant, this can result in awkward or incorrect translations.

5. Machine Learning Limitations:

- **Neural Network Errors**: Many translation systems use neural machine translation (NMT) based on deep learning models. While NMT is very powerful, it can still make errors, especially in complex sentence structures. Since these models rely on pattern recognition rather than proper understanding, mistakes can happen when the model encounters something it hasn't seen often or at all in its training data.
- **Generalisation Issues: Machine translation models are trained on vast text corpora,** but are only as good as the data they've seen. The system will struggle to generalise to those cases if the training data doesn't cover specific, nuanced phrases, rare words, or complex structures.

6. Lack of Real-World Understanding:

- **No Real-World Context**: Automatic translators don't have real-world understanding or common sense. They can't "know" what the speaker or writer intends beyond the text they are analysing. For instance, understanding the difference between polite and impolite language in social interactions requires social knowledge that automatic systems lack.
- **Subjectivity and Tone**: Translating subjective content or emotional tone accurately is complex because the meaning of certain words or phrases changes based on the emotional intent behind them. Automatic systems may fail to pick up on this nuance.

7. Sentence Fragmentation and Long Texts:

- **Incoherent Sentence Structure**: Sometimes, automatic translation systems translate sentences in fragments, especially if punctuation is inconsistent or the sentence structure is too complex. This can lead to disjointed or confusing translations as the system fails to maintain continuity or handle long sentences properly.
- **Text Segmentation Issues**: Some systems struggle when translating entire paragraphs or long texts because they may treat each sentence independently rather than considering the overall flow and context. This can lead to inconsistent word choices, different translations for the same term, and loss of meaning.

8. User Input Errors:

- **Typos and Slang**: If the source text contains spelling errors, informal language, or abbreviations, automatic translation systems might struggle to interpret them correctly, leading to poor translations.
- **Incomplete Sentences**: If the source text is fragmented or incomplete, the automatic translation system might produce confusing or incomplete output.

How to Improve Automatic Translations

- **Use Simple, Clear Language**: When possible, simplify the input text to reduce the chance of confusion for the machine translation system.
- **Short sentences:** The "Churchill" short-sentence writing style gives more precise information to the automatic translation process.
- **Provide Context**: When available, provide additional context or use multi-sentence input for translations to help the system understand the meaning.
- **Business English**:

Proofread Translations

Firstly, use the multilingual people in your staff, family, and friends. Regardless of language, ask multilingual people to review the automatic translation into their own/second language. When they see errors in the translation, they should review the English text and look for some of the problems listed above. Make any necessary changes to correct the translation.

Then, try a second language. At this stage, the number of translation errors should have been significantly reduced. The automatic translation software can better understand the English you change from the first translation review. Depending on the available resources, you can repeat this step several times in different languages.

Finally, have the translation checked in Chinese. We always believe in using a local final-year college student studying Chinese. This archive;

- Using a person will help them understand better what you are trying to achieve in the English text. A literal translation can be technically correct, but it misses the point of what has been said. A Chinese person often feels uncomfortable suggesting changes to the English text.
- A local person can be educated about your business to understand what you are presenting to the Chinese market.
- It is a good investment in local people who could support your business in the future.

SEO Data

Search engines currently do not automatically translate SEO data into the local language.

- SEO data will always be present on your website in the language of your target market.
- If, in the future, an SEO automatic SEO translation process is developed by search engines, reviewing your website's automatic translation will be even more critical.
- For more information, please see the Chinese engine SEO