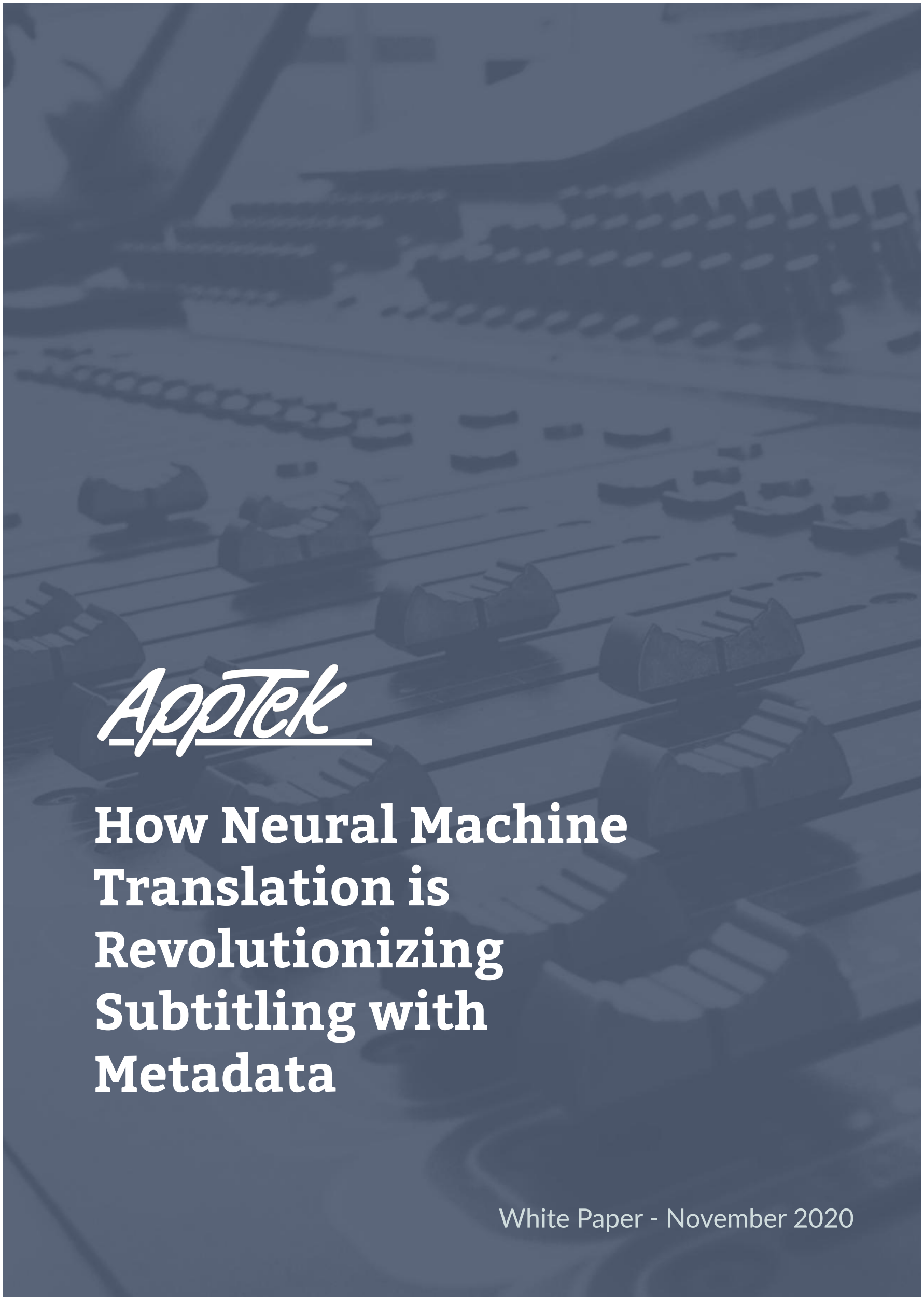


AppTek

How Neural Machine Translation is Revolutionizing Subtitling with Metadata

White Paper - November 2020

Introduction

Advances in neural machine translation (NMT) technology have resulted in media and subtitling companies now relying on it to assist translators in post-editing workflows. While NMT certainly offers a productivity boost as a result of NMT's inherent improvements in fluency, the cost and resource utilization benefits are still constrained by the limits of most popular machine translation (MT) systems.

While the quest for improved MT accuracy continues, this white paper provides an overview of the scenarios that can be addressed by adding metadata in the system training process, with particular focus in the media and entertainment sector. Metadata can be leveraged in the post-editing workflow to increase quality and boost productivity.

Moving Beyond the Status Quo in Machine Translation

Off-the-shelf MT systems are used rather like a black box—you input data and the system outputs the translation in running text format. In the case of subtitling, translators need to turn this text into subtitles by incorporating as-needed corrections and appropriate text segmentation.

Fewer errors in the MT output and faster turnaround times can be expected when such off-the-shelf systems are replaced by those customized to the media and entertainment domain, using available in-domain data from this industry vertical or customer-owned data for bespoke system training. Such customized systems may also include components for automatically segmenting text semantically and syntactically, and outputting it in custom subtitle format.

Customization has become the norm for significantly increasing the quality of the MT output, producing texts that are more like the ones that need translating. With the higher quality output that is now possible with customized NMT, integration of MT in localization workflows for the creative industries has become feasible.

But from a language service provider's (LSP) point of view, if you need high-quality MT for ten different domains or genres, you would need to train, deploy and maintain ten systems —even though they may be idle some of the time. The result is a high environmental footprint and increased costs. There is also a risk of "over-fitting" the training, making it so specific to a particular domain that its performance is worse with different data than it otherwise would be.

So what do you do when you need to subtitle materials as varied as a documentary about planet Earth versus the latest Tarantino movie? Clearly the language used in either will be very different, so would be best translated by a different MT model. But what if instead the register of the language required in each case could be taken care of by the same model that has a switch incorporated to toggle between formal and informal grammar and vocabulary choices?

Adding Metadata Transcends Translation Limitations

The latest advances in NMT from AppTek offer a more productive and cost effective approach to solving the customization issue. The technology has reached not only much higher quality, especially for in-demand, high-resourced language pairs, but also offers flexibility not possible in previous MT generations.

A single NMT system can now be easily customized by inputting minimal additional metadata relevant to each unique domain and scenario that a business requires. The concept is similar to an acoustic mixer that allows you to modify the same sound in various ways. With NMT, the user simply adds an extra (pre-coded) parameter value in an API call that generates a desired translation, for example style=formal or length=short.

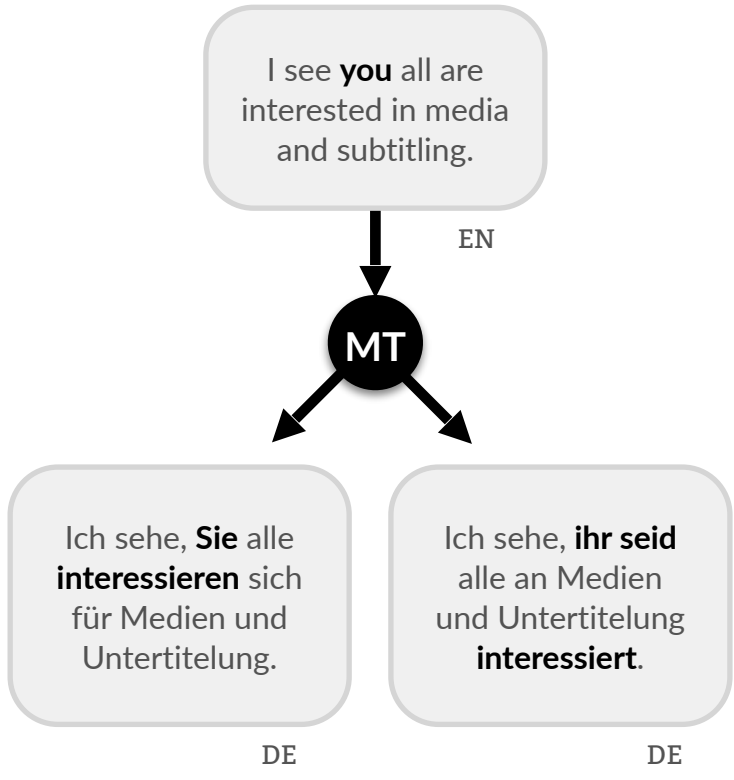
Specific metadata attributes that can be customized are outlined in the following pages.

“A single NMT system can now be easily customized by inputting minimal additional metadata relevant to each unique domain and scenario that a business requires. “

1. Style

such as formal vs. informal register, which is often context dependent

An especially tricky translation to this point has been the English “you” (both singular and plural), a word that is more uniquely distinguished in other languages. Advanced NMT can now choose the right translation, formal or informal, and adjust sentence structure if necessary. Experimentation with the formality feature has shown that such style adaptation goes beyond grammar, affects lexical choice and improves meaning disambiguation.
The vocabulary choices that are output in this manner seem to also correlate with the usage of second-person pronouns and result in formal or informal constructions. It goes without saying that movie dialogue is typically best served with informal MT output, whereas the formal style would probably work better in documentaries.



2. Speaker Gender

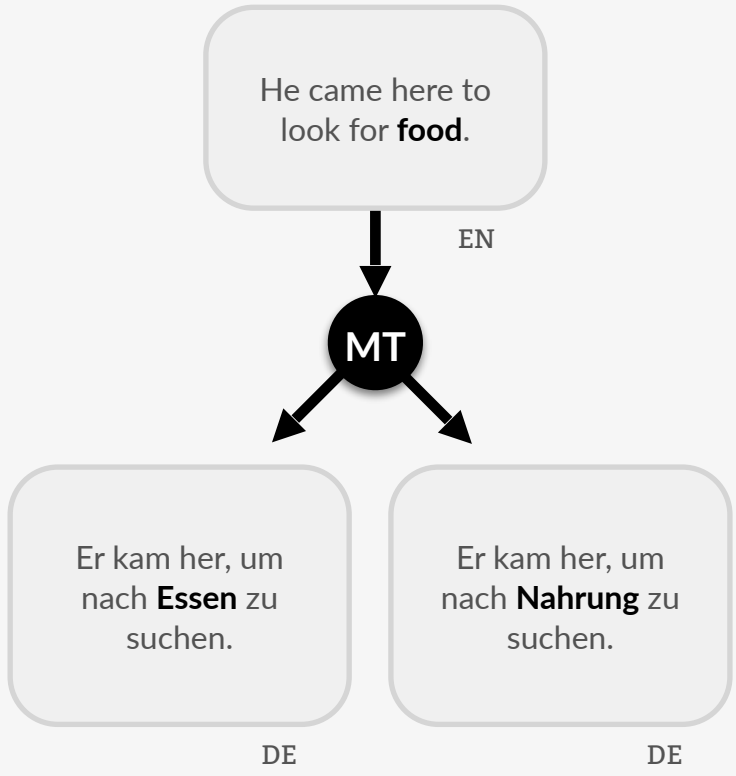
which makes a difference in the morphology (form) of the MT output, e.g. in fusional languages, like Czech, where certain parts of speech are dependent on the gender of the speaker

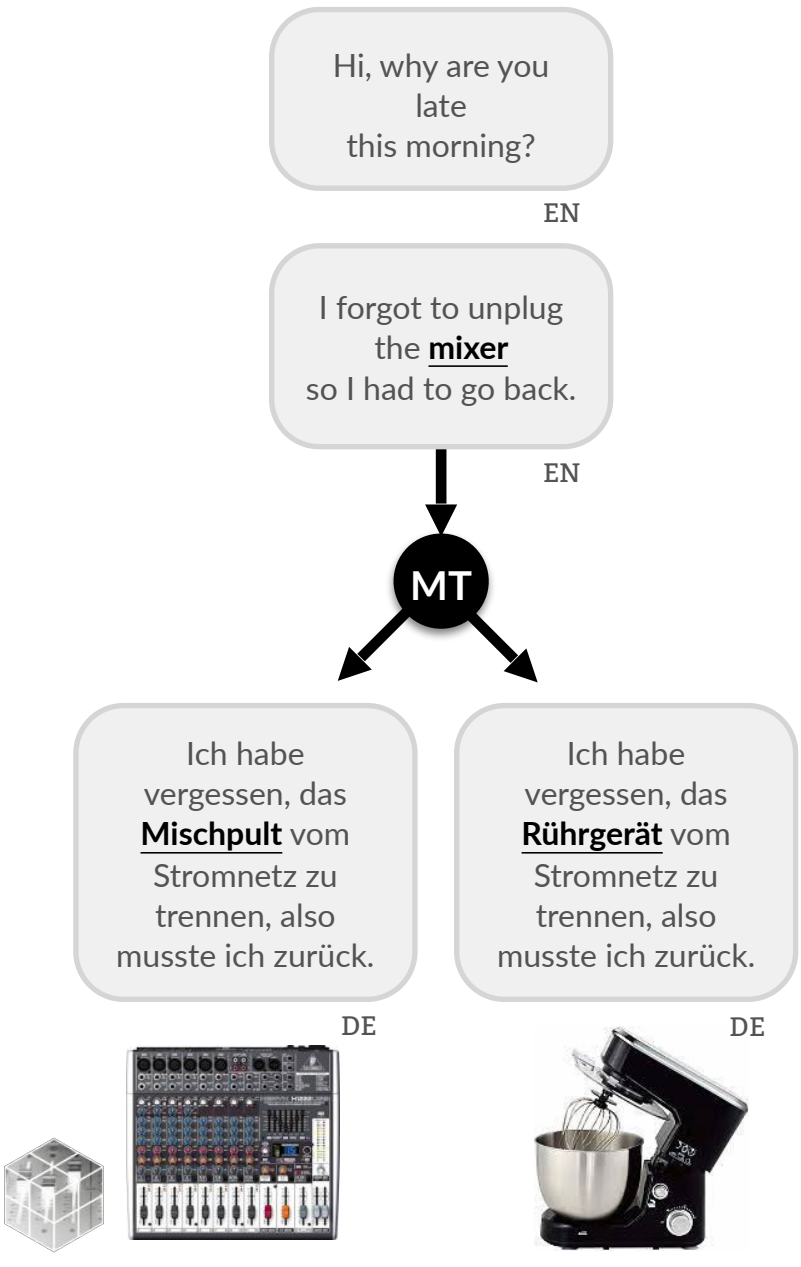
When this information is provided in the metadata of the source text, speaker gender labels can be inferred automatically and result in grammatically correct MT output. In cases of upstream speech recognition tasks, speaker gender information can be inferred straight from the audio as well.

3. Domain or Genre

such as news, patents, talks, entertainment, etc.

By taking into account domain or genre, the machine will acquire a semblance of the ‘world context’ that it typically lacks when outputting a target sentence. In the case of subtitling, subtitles themselves are often viewed as a ‘domain’, given that the language used in them has to comply with specific characteristics regarding length, use of punctuation, spelling of numbers etc. The broader genre distinction in terms of media would probably have to be in relation to other language qualities, such as whether the language is scripted or more spontaneous, whether it is more stylized or contains a lot of slang, and so on. As such, live broadcasts (e.g. news and talk shows) could probably be grouped together and distinguished from offline entertainment such as films, series, sitcoms, etc.





4. Topic

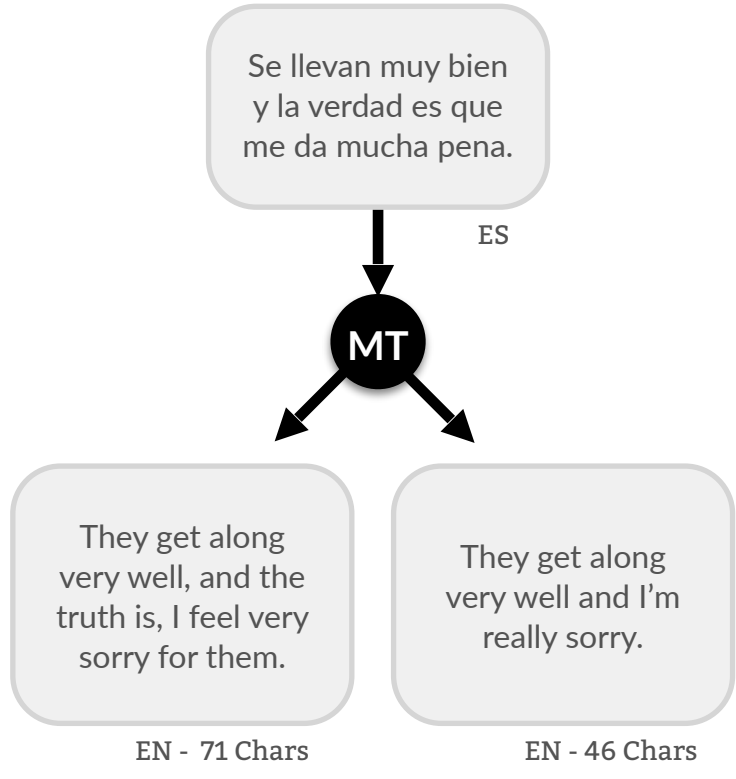
catering to more specific document-level style and terminology differences

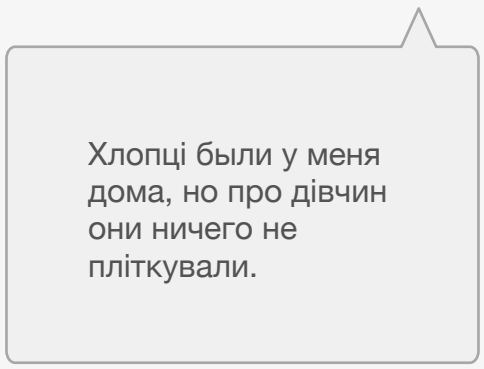
Topic is more specific that domain or genre and more fine-grained topic taxonomies can be defined, which impact the choice of vocabulary in the MT output. An example of topic would be ‘politics’ vs. ‘sports’ vs. ‘weather’ vs. ‘environment’ in the case of live broadcasts, or ‘medical’ vs. ‘military’ vs. ‘cooking show’ etc. in the case of offline entertainment.

5. Length

generating shorter or longer translations with minimal information loss or distortion

This feature is important in subtitling due to the space constraints available in a subtitle line. It also matters in verticals such as software localization where space is also a hard constraint. In the case of subtitling specifically, reading speed information can be taken into account to inform the decision on the ideal length of the MT output.





6. Language Variety

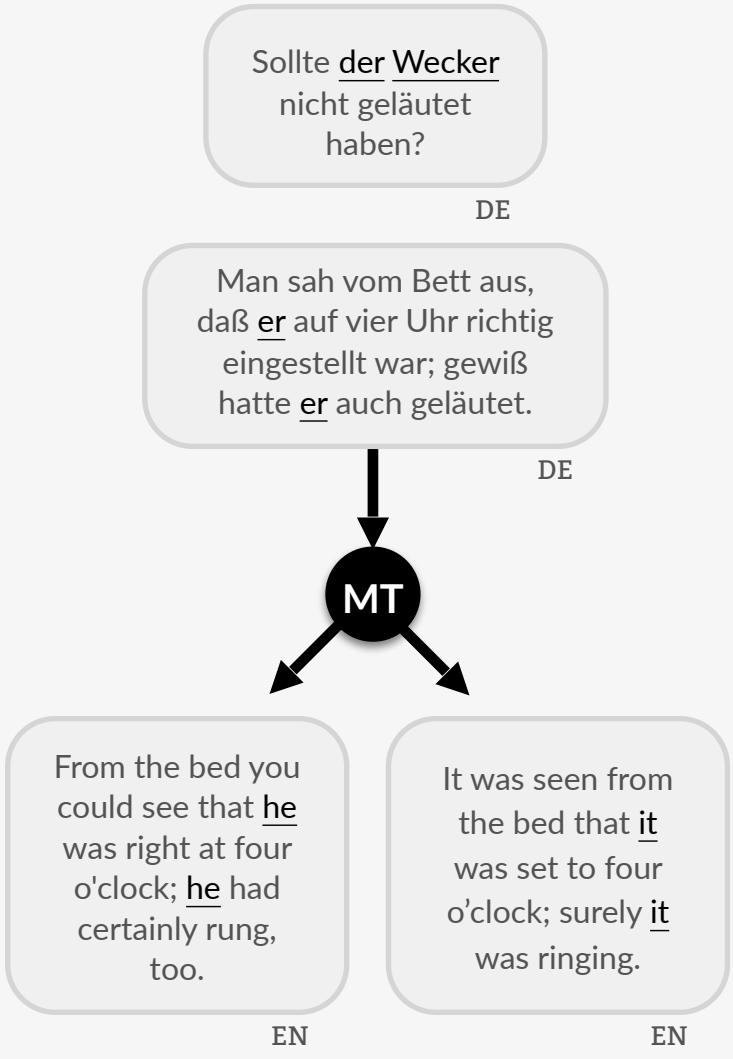
where parallel training data for related languages or dialects can be combined in a single system

Such multilingual or many-to-one systems also have the added advantage of catering to multiple input languages as well as mixed-language input. For example, Ukrainian words used in an otherwise Russian sentence or English and Hindi words mixed, which is frequently the case in Bollywood films. Frequent use cases in the subtitling market would be systems that cater to Castilian versus Latin American Spanish, European versus Brazilian Portuguese, Canadian versus European French, Simplified versus Traditional Mandarin, and others.

7. Extended Context

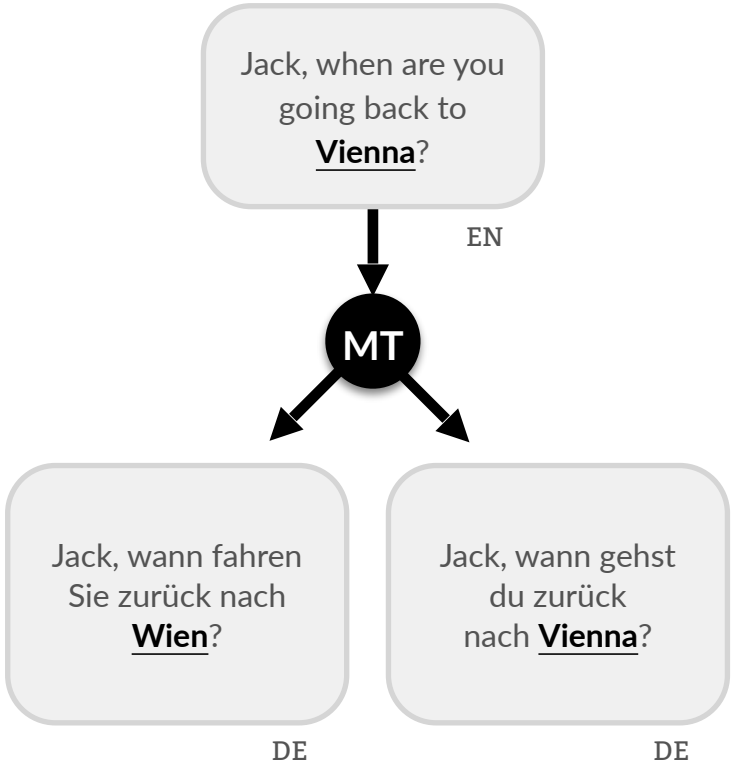
assessing whether or not the context of the previous or the next source sentences should influence the translation of a given sentence

This technique is used for word disambiguation purposes and for pronoun resolution. Agreement type errors in the MT output, e.g. pronoun and noun agreement, which can be quite frustrating for post-editors, are likely to be minimized as a result.



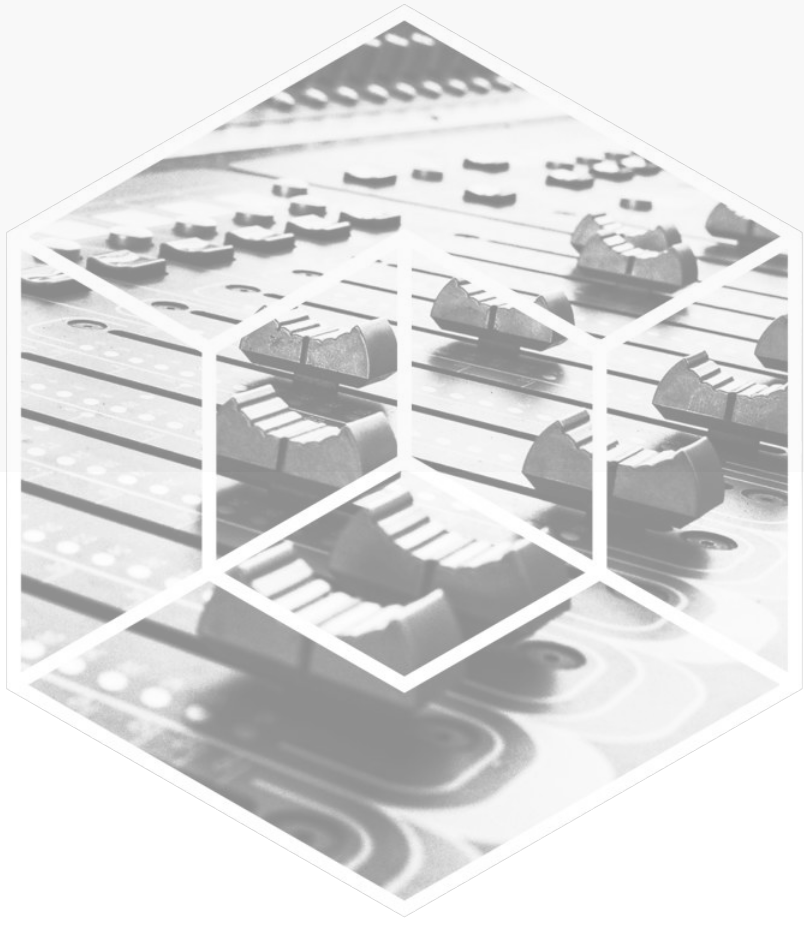
8. Glossary

relating to terms with official or mandatory translations, which the system may otherwise translate differently



The use of custom user dictionaries or glossaries is fast becoming a core requirement in all MT offerings. Users need the ability to import/upload/connect to the internal or end client glossaries and term lists they are required to use. They also need to add new terms and phrases on the go to solve recurrent translation errors in the MT output. The challenge in this case is how to generate the MT for the given term in a grammatically correct form, which can be quite complicated in morphologically rich languages.

The glossary to import can also be automatically created by a computer-assisted translation tool that memorizes past user choices and corrections. By using translation memories too, the use of glossary terms is forced to be made in context in the MT output.



“Switches for various captured metadata can be made available through APIs, so that translator interfaces and platforms can leverage such flexibility and offer UI solutions that facilitate respective workflows.”

Extending Subtitling Workflows with Metadata

The metadata needed to train on these attributes can come from a variety of sources. Those include information about the origin of a translated document, directly derived from the data itself, computed from the text data using rules and regular expressions, or predicted through separate machine learning algorithms.

Implementing this approach in MT system training allows you to train single models (e.g., one instead of ten), reducing both environmental footprint and cost at the same time. Switches for various captured metadata can be made available through APIs, so that translator interfaces and platforms can leverage such flexibility and offer UI solutions that facilitate respective workflows.

These NMT-driven advances can deliver significant benefits to media clients, language service providers, and the post-editors and subtitlers who shoulder the high-pressure responsibility of accurate subtitling. An LSP could utilize such customization by setting up some of these parameters in their MT system of choice at the project level, e.g. for attributes such as length or language variation, domain or genre, as such information will be available at the start of any project. Others can be set by default to one setting that generally works for most of the content an LSP handles, e.g. ‘informal’ for film subtitles, with the ability for the user to change this at the subtitle level if needed to facilitate the post-editing process.

Moving Forward

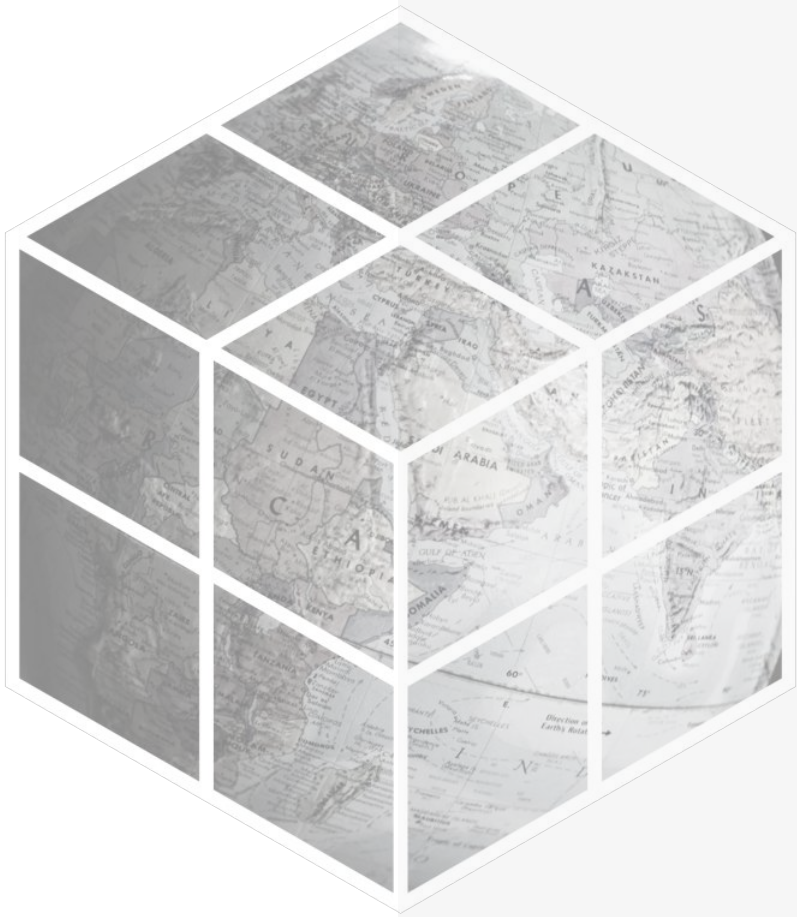
Artificial intelligence for language is developing rapidly. We are already able to reap the benefits that machine learning has made possible via neural network technologies and sequence-to-sequence transformers for machine translation and language generation. While we are still waiting for the next technological breakthrough with machines learning some form of deeper reasoning, customization to the domain and even to the user's own feed is the latest trend in order to benefit from MT output of higher quality.

Customization with metadata is a great example of how to unlock for the user more of the possibilities hidden in the NMT black box. Having proper data regarding style and length, for example, means fewer words and sentences would need to be post-edited. By flipping the switch to the desired metadata, translators can experiment with the technology and get a new tool under their belt to make their daily work easier.

Even so, NMT systems do not (yet) have the world knowledge that human translators have. Inputting metadata will arm a system with at least a small amount of that knowledge, but it takes a combination of human and machine. The real mission is to deliver organizations greater control over whatever subtitling output is given back to them.

Ultimately, it's all about better business performance and results.

"It takes a combination of human and machine."



AppTek Solutions for NMT

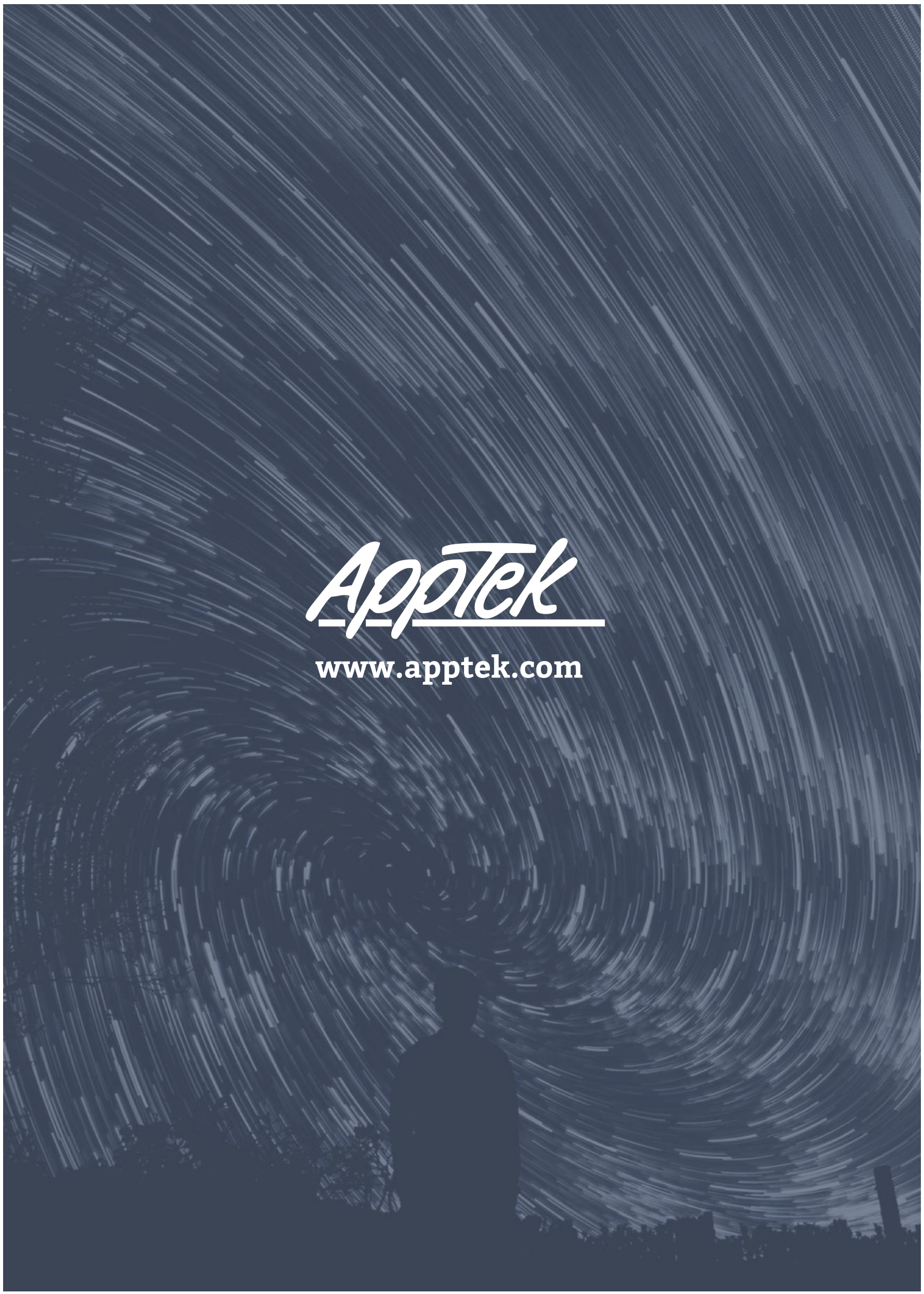
AppTek empowers you to generate highly accurate, customizable and scalable translations across hundreds of language pairs.

AppTek offers enterprise-grade state-of-the-art neural machine translation technology that supports modern architectures.

Our models are trained on large amounts of public and proprietary data, and cover a wide range of data types and domains.

We also offer customized solutions for fast domain adaptation to customer translation memory data.

No matter your translation or localization needs, AppTek can help globalize your application.


AppTek
www.apptek.com