

The Future is Automated

The landscape ahead for M+E looks uncertain. Automation can ease the journey.

WORKFLOWS AND THE CLOUD

From cloud-based productions to unprecedented storage demands, the supply chain has never been more challenging

SMART CONTENT

The questions around AI, metadata, and analytics are endless. But we do have many of the answers already

SECURITY SOLUTIONS:

Cybersecurity and content protection in M&E is always on defense. Can automation can change that?

23,01

DON'T NEGLECT GLOSSARIES WHEN AUTOMATING LOCALIZATION

Reduce linguistic ambiguity, enable information extraction, and enhance the accuracy of automated translations with automated glossaries



ABSTRACT: Glossaries are essential in localization, but often neglected in automation. Automating glossary creation and management offers efficiency and consistency benefits, reduces linguistic ambiguity, enables information extraction, and enhances quality and accuracy of automated translations in fields with specialized terminology.

By Bartosz Adamczewski, Director, Research, Innovation, and Allan Dembry, Chief Technology Officer, Iyuno

Speech synthesis is a technology that has been in the forefront of many discussions recently in the media and entertainment industry, with a strong media presence and funding spree. It is already a large part of our lives in the form of Siri, Alexa, and other digital assistants, found in call centers, embedded in modern automobiles, or as assistive technology for the blind.

Automating glossary creation and management offers efficiency and consistency benefits, reduces linguistic ambiguity, enables information extraction, and enhances the quality and accuracy of automated translations in fields with specialized terminology.

A glossary is a document that contains a list of terms and phrases and their definitions, often used in a particular brand, show, or field.

Glossaries serve as a reference guide for key terms and phrases, ensuring that everyone involved in the project is on the same page and

	A	B	C	D
0	EN	PL	DE	
1	Tom Hardy	Tom Hardy	Tom Hardy	
2	Tomas Nook	Tomasz Nook	Tomas Nook	
3	Poland	Polska	Polen	
4	Warsaw	Warszawa	Warschau	
5	Supervillain	Super Złoczyńca	Superschurke	
6	Palace of Culture and Science	Pałac kultury i nauki	Palast der Kultur und Wissenschaft	
7	Old Town	Stare Miasto	Altstadtstraße	

Figure 1: Glossary.

understands the terminology being used. In the field of automation, glossaries are essential in streamlining processes, improving accuracy, minimizing errors, and enabling suggestions in subtitling and translation systems. They enhance and improve the quality of all downstream processes that must do any processing with text and dialogue.

Of course, one of the primary benefits of using a glossary is that it can help to eliminate ambiguity. If different people use different terminology or have different definitions of the same term, this can lead to confusion and errors in the translation. By using a glossary, everyone involved in the project can refer to the same set of terms and definitions, eliminating confusion and ensuring that everyone uses the same language, thus making it essential in providing consistency in the localization process.

Another key benefit of using a glossary is that it can improve the accuracy of the script translation process, and this is true of both traditional translation and machine translation. When working with scripts, it's easy to make

spelling and form mistakes, and clients often require specific terms to be used in a translation or editing process. However, if everyone involved in the project uses the same set of terms and definitions, this can help reduce the risk of errors and rejections. Enforcing a glossary and providing validations and suggestions can in turn be used to return greater accuracy and more effective automated translation. Historically machine translation engines had a lot of trouble with translating terminology correctly; however, an NMT engine can use a glossary such that when doing the translation process, it uses the correct term spelling and translation from the glossary.

By using tools that integrate the glossary into the script review and translation process, you can gain the benefit of automatically highlighting all inflected terms in the script by using a set of algorithms and machine learning models. Other types of suggestions then become possible whereby the system will inform the user that a better or more suitable term or phrase exists and should be used instead, and that a term should be applied in a specific piece of dialogue but was not. This not only accelerates the process but delivers greater accuracy at the same time. Glossaries can also have metadata; this provides insights so that if the user is unsure whether to use a specific phrase or term in the text, the system can also provide rich metadata and context-specific descriptions of the term or phrase for research purposes as illustrated in figure 2.

Glossaries are also an integral part of the script approval process; when clients want specific terms and phrases to be used, the glossary and highlights terms that were used can



Figure 2: Terms matching, typos, and suggestions.



Allan Dembry is the chief technology officer of Iyuno. He oversees all aspects of the company's information technology. Since joining Iyuno in 2019, Dembry has been focused on developing industry-leading information technology practices and work streams that will underpin the continued growth at Iyuno, the largest and leading media localization partner with the widest range of technology-driven, end-to-end solutions. allan.dembry@iyuno.com @iyunohq



Bartosz Adamczewski is the director of research and innovation of Iyuno. He is responsible for leading and directing research and innovation initiatives at Iyuno. With more than a decade of experience writing fast and scalable big data systems, and natural language technology, Bartosz is primarily focused on driving innovation and applying the latest advancements in artificial intelligence to provide new and improved solutions for media localization and content creation. bartoszadamczewski@gmail.com @iyunohq

IF DIFFERENT PEOPLE USE DIFFERENT TERMINOLOGY or have different definitions of the same term, this can lead to confusion and errors in the translation. By using a glossary, everyone involved in the project can refer to the same set of terms and definitions.

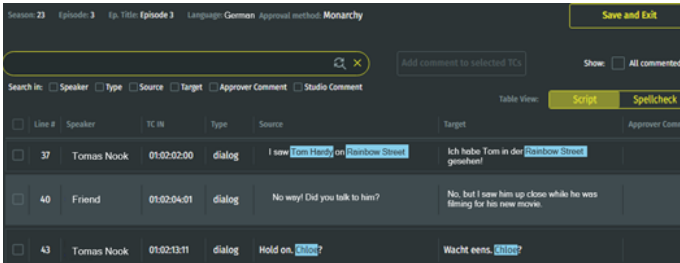


Figure 3: Script approval using a glossary.

be loaded, displayed to the user, and then display warnings about any terms that have not been used or any other potential problems.

Maintaining and keeping glossaries up to date can be a complicated and time-consuming process, and one that is prone to error. Glossary management systems can help with this task, but it's quite a challenge to keep all the terms, phrases, translations, and metadata filled in and up to date; fortunately, again, automation can help here.

Historically, people created glossaries and maintained them in excel spreadsheets and other file types. This presents a challenge with using the correct glossary and correct metadata in the editing or translation systems. Glossary management systems help with this task, but even then, data must be imported into the system. This is where machine learning models can be leveraged to automate glossary imports as well as picking the correct glossary for a given task.

Machine learning models can be used for automatic glossary creation, both for generic glossaries and client and brand-specific glossaries, from a set of scripts and documents; this reduces the time it takes to create a glossary

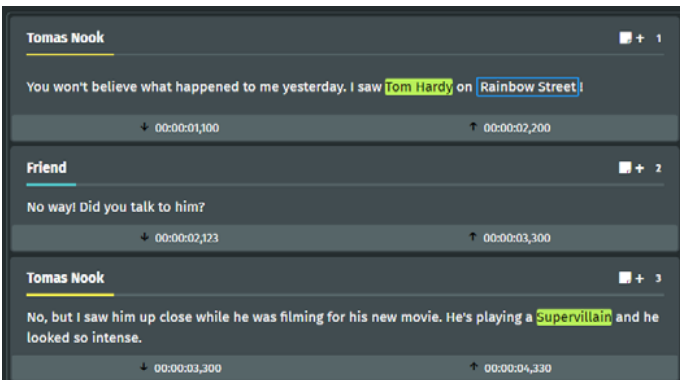


Figure 4: Suggest terms in an Excel glossary.

	A	B	C
0	EN	PL	DE
1	Tom Hardy	Tom Hardy	Tom Hardy
2	Tomasz Nook	Tomasz Nook	Tomasz Nook
3	Poiland	Polska	Polen
4	Warsaw	Warszawa	Warschau
5	Supervillain	Super Zloczyfca	Superschurke
6	Palace of Culture and Science	Pałac kultury I nauki	Palast der Kultur und Wissenschaft

Figure 5: Automatic glossary extraction.

and helps with consistency and quality.

Similarly, relationship extraction models can be applied to extract term metadata and relationships to other terms to enhance the existing glossaries, something that is far more challenging to achieve manually in a timely manner. This automatically generated information can then be used for research when editing or translating.

In summary, the most beneficial aspect of having glossaries is that nearly every piece of automation starts by having an initial glossary, then subsequently building machine learning models and other automation tools based on these data sets. Glossary automation becomes a self-referencing feedback loop of operational and technical improvements and brings even more efficiencies to what can be a long, complex process. ☒

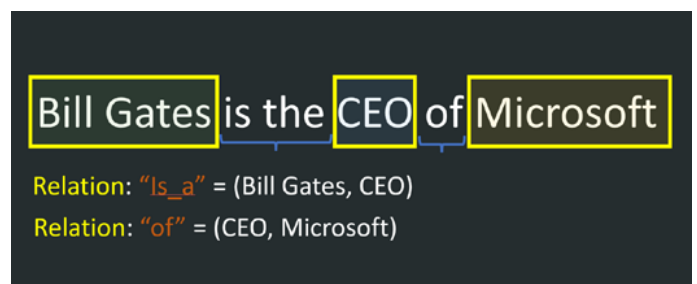


Figure 6: Relationship extraction using existing terms.



Figure 7: Self-enhancing feedback loop.



We. Are. Creative.

Global Storytellers Need a Dedicated Team.

CREATIVE LETTERS | SCRIPT ADAPTATION | CASTING & SUPERVISION
DETAILED MIX REVIEWS | QUALITY CONTROL & CONSISTENCY

CONNECTING CONTENT CONNECTING PEOPLE.

iyuno