dpp®

The Future of Localisation

sdvi

Contents

Introduction

There has been a dramatic rise in the demand for content to be localised for different markets and languages. International streaming service launches have created demand from content owners, while audiences have developed a growing appetite for content from other countries and cultures.

As a result, specialist localisation companies are going through a boom, yet still they are struggling to meet the demand. There has not always been enough throughput or enough specialist talent to keep up.

Localisation is experiencing a boom, but providers can't keep up with demand

Media companies need more content localisation, created faster, without dramatically increased costs. And in a world where bad translations become TikTok memes overnight, they cannot afford to lower quality.

Machine learning might be the great hope. Technologies like automated transcription and even synthetic voices are becoming mainstream, while a new generation of generative AI has made headlines in recent months.

So are we on the brink of an automated localisation revolution?

The DPP has spoken with more than 50 experts from content creators, streaming platforms, localisation service providers, and technology vendors, to find out.

Executive Summary

1 Localisation drives both revenue and accessibility

Content owners have increasingly looked to localisation to populate international streaming catalogues, or to maximise ROI in their content. But it is also a critical part of making content accessible to diverse audiences – so much so that many now consider localisation a subset of accessibility services.

2 Audience preferences are changing

Viewers are showing a growing appetite for content from other countries and cultures. This is especially true for English speaking audiences, which have historically been more resistant to subtitling and dubbing. This creates new requirements for localisation providers.

3 Automation is assisting humans

Machine learning is already becoming commonplace to assist humans with tasks such as transcription and translation, while tools such as image analysis and automated audio conform are increasingly prevalent. However, full automation of localisation services remains extremely rare for professional content.

4 AI voices are the next frontier

Synthetic voices are generally still considered unsuitable for use in dubbing high quality content, even though they are improving over time. Meanwhile, voice cloning technology offers exciting new possibilities by applying a machine learning voice model to a human voice actor's performance.

5 We are at an inflection point for automated localisation

Automation will undoubtedly form an increasing part of localisation. Enabling that automation to enhance the work of talented professionals requires a focus on change management, the development of new best practices for intellectual property and artist consent, and investment in training high quality machine learning models.

Contributors

The content for *The Future of Localisation* has been gathered through workshops and interviews with subject matter experts from across the industry. Valuable input has also been provided by our Expert Sponsors, Blu Digital, FPT, SDVI, and Vubiquity, and our Contributing Sponsor, Ateliere.

Although the content of this report has been informed by these discussions, it should not be assumed that every contributor shares all the views presented here.

Peter Abecassis, Grass Valley
Ashley Bailey, Veritone
Andrew Balfour, Netflix
Johanna Björklund, Accurate Video
Philippe Brodeur, Overcast
Tim Burton, 7FiveFive
Lindsay Bywood, University of Westminster
Richard Clarke, Banijay Rights
Emily Corbett, Banijay Rights
Christopher Deas, SDVI
Tom Donoghue, SDVI
Tom Dunning, AdSignal
Ira Dworkin, FPT Software
Michele Edelman, Premiere Digital
Justin Eli, Paramount
Silviu Epure, Blu Digital Group
Nate Frink, Vubiquity
Stefanie Gamberg, Iyuno
Marco Garghentini, Iyuno
Nicky Goldberg, BBC
Garrett Goodman, Papercup Technologies
Jan-Hendrik Hein, A+E EMEA
Magda Jagucka, Deluxe
Isak Jonsson, Vidispine
Eric King, Deluxe
Vanessa Lecomte, BBC Studios
Matt Loreille, Wildmoka
Uisdean Macleod, MG ALBA
Andrey Marchuck, Deluxe
Gustavo Marzolla, Vubiquity
Scott McCarthy, Dreamworks
Thomas Menguy, Wildmoka
Chris Merrill, Grass Valley
Matteo Natale, Vubiquity
Jyothi Nayak, Prime Focus Technologies
Phong Nguyen Xuan, FPT Software
Jeny Nicholson, Deepdub
Lorena Ortega, Take 1 Transcription
Tuncay Pervaz, Fabric
Alan Pimm, MediaSaaS
Chris Reynolds, Deluxe
Jordan Ronaldson, ITV
Scott Rose, Media Globalization
Atul Saxena, Prime Focus Technologies
Jonathan Schiminske, Warner Bros Discovery
Craig Seidel, Pixelogic Media
Amanda Smith, Warner Bros Discovery
Greg Taieb, Deluxe
Eric Toulain, Ateliere
Sarah Turpin, Papercup Technologies
Matt Waldock, Xytech Systems
Stella Yoo, Iyuno

Understanding localisation

KEY INSIGHTS

Localisation means more than 'subs and dubs', also incorporating compliance editing, graphics translation, marketing localisation, and more

Only the highest value content undergoes video edits unless required for compliance, with most localisation being linguistic

Localisation is key to providing access to content: so much so that many organisations now consider it part of accessibility services

There are benefits in greater collaboration between production and localisation teams, with clear deliverables from one to the other

Many aspects of localisation are highly creative, crafting the content with respect for the audience while maintaining the original creator's intent

At a conceptual level, localisation is easy to understand: the process of making adjustments to a piece of content to make it suitable for audiences in a different location. At a practical level, however, it comprises a wide range of processes.

Localisation comprises a huge range of processes

The most obvious are the translation of spoken language, using **subtitles** or **dubbing**. But many other processes are included too.

In some cases, **graphics** must be translated, or other **on-screen text** such as street signs. This is common for certain large budget theatrical releases, though in general it is more usual to rely on subtitles to provide translations of crucial text.

The video itself may be recut or otherwise adapted for different regions. This can include **cultural edits**, such as the famous example of Pixar's *Inside Out*, in which scenes of a child refusing to eat broccoli were replaced with green peppers for the Japanese audience. [*Slate*] However, such edits are not frequently applied to episodic content.

> We approach cultural localisation as much as possible from a linguistic perspective with subtitles and dubbing; it's not very often that video gets re-edited.

Edits for **compliance** reasons – to meet legal and regulatory requirements in different territories – are much more common. These may include removing nudity, violence, smoking, or other content deemed unsuitable for certain audiences.

Video edits are mostly reserved for compliance changes

Specific content types also have their own unique requirements. Sports content commonly receives video edits to address the needs of different audiences, with both live video and highlights packages cut differently to feature athletes and teams from the country in which they will be shown, for example.

And in genres such as news, video is often recut to match the cadence of out-of-vision speech and to reduce the risk of mismatches between narration and video. The same information conveyed in different languages can be shorter or longer, requiring video to be adjusted to fit.

But localisation isn't limited to the content itself. For most shows and movies, marketing materials such as **trailers**, **artwork** and **metadata** must also be localised. In many cases, these assets are highly differentiated for different regions and cultures.

> Marketing localisation is never literal. It should be something that makes sense for the audience that you are aiming it at.

For example, the genres used to categorise content in programme guides and streaming interfaces may not be directly comparable in different countries, while synopses and taglines might address local cultural references.

> Metadata is often localised by the marketing department. And they're trying to promote the content, so they want to localise specifically for that purpose. The metadata is not there to describe the movie, it's there to sell the movie.

Metadata isn't there to describe the content, it's there to sell it

ACCESSIBILITY

It might be less intuitive to consider accessibility services such as **audio description** as part of localisation. However, they are highly related disciplines, and often managed by the same teams who oversee localisation.

> Audio description isn't technically 'localisation', but it has almost the same workflow and similar technology requirements and therefore fits well.

Indeed, when speaking to the wide range of experts who contributed to the report, it became clear that they felt that localisation is a crucial part of making content accessible to new audiences.

> I think localisation is part of accessibility and not the other way around. When you are localising, you're making content accessible to people from different cultures.

Localisation is a part of accessibility

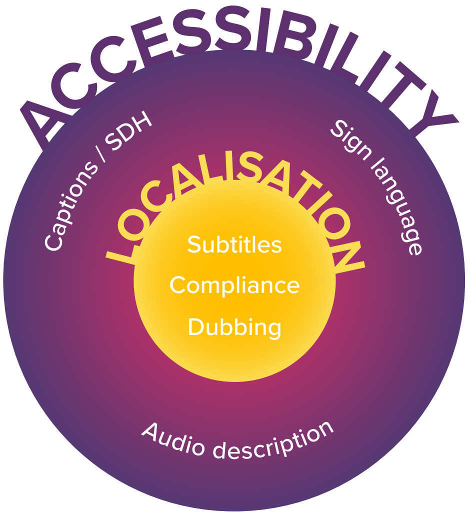
This sentiment was expressed by many individuals from a range of different companies. For some, the relationship between the two disciplines has evolved over time.

> We used to think about access services as part of audio-visual translation, and now there's a movement to flip that. Localisation provides access, so we think of localisation as part of access services.

Accessibility includes other services too, such as **sign language**, and **closed captions**.

Closed captions (CC), or subtitles for the deaf and hard of hearing (SDH), can often be considerably different from subtitles created purely for language translation. They include other audio cues, speaker differentiation, and so on.

Whether formally considered part of access services or not, localisation is certainly key to making content available to new audiences. It can help to maximise the audience for a piece of content, and therefore the revenue it generates. And in an increasingly global content market, it's becoming ever more important.



THE POOR RELATION

When discussing localisation with industry experts, there is a consistent sense that historically, localisation has been somewhat of an afterthought.

> Localisation is seen as the poor relation to broadcasting and content production. It's not thought of as something that will generate revenue directly, even though selling content without appropriate localisation is going to be impossible in many cases.

Many felt that localisation has often been considered a cost of business, rather than an essential part of the content creation process. This is problematic, because localised assets are a key representation of a piece of content.

> Localisation creates the manifestation of our brand in a particular market. If we do it wrong, it damages our brand.

Localisation creates the manifestation of our brand in a particular market

Despite this, it is rare for localisation to be considered during production. Many contributors felt that the process of localisation would be simpler, and yield better results, when given more consideration during production.

> **If localisation and accessibility were considered earlier in the process then the end product would be much better.**

Often, all that is required is a little more attention to creating the right outputs from the production process to make localisation easier. These include accurate scripts, separate dialogue audio tracks, and so on.

> **They don't give us dialogue tracks, because they don't think of us as part of the filmmaking process.**

But with the rising importance of localised content, this is starting to change. One major international broadcaster described how their localisation team now provides feedback to production teams when they find problems with scripts and other deliverables, in an attempt to improve understanding between them.

Others expressed similar aspirations of improving collaboration between producers, distributors, and localisers. And some localisation service providers are starting to feel able to be more open with their customers about how they can help make the process better.

> **I think there's an effort now to address what's been missing, and that is collaboration. I don't think there's enough collaboration between the content creators and post production, and the localisation teams. There's a gap, and there's much more that we could do to bridge that gap.**

There isn't enough collaboration between content creators and localisers

THE ART OF LOCALISATION

One reason localisers would like to build closer collaboration with production teams is that they feel a sense of responsibility to maintain the integrity of the original editorial.

> **The golden rule in localisation is to recreate the director's editorial intent.**

This can be more complex than a simple translation, requiring a creative hand. For example, humour can be particularly demanding. Many jokes don't work when directly translated, needing the localiser to rewrite them while respecting the original creative.

> **I want to be able to go back to the content owners and say, 'Hey, here's how I'm going to adapt it, does that line up with your vision?'. Because I'm altering their product.**

Another challenging area can be sensitive or offensive language. For example, some slurs (such as racial, sexist, or homophobic terms) may be translated very differently based on context – whether used in anger in a historical drama, or factually without passion in a documentary, or with humour by someone reclaiming an insult. This delicacy requires a great deal of intelligence and empathy.

> **A dubbing script writer is just that. A script writer.**

It is striking when speaking to localisation professionals the gravity with which they consider decisions such as these. A very high importance is placed on understanding the cultural context of the viewer they are creating for, as well as that in which the content was created.

Great emphasis is placed on cultural respect for both the creative and the audience

Indeed, the concept of 'cultural respect' is a recurring theme. And it will only become more important as more content is localised for more audiences around the world.

Evolving demand

KEY INSIGHTS

There has been a surge in demand for content localisation, as streaming services have launched large libraries of content internationally

Longer term demand is expected to remain high, due to a growing viewer appetite for international content

Localisation is seen as a key way for content providers to maximise the return on their investment in content

Viewer preferences are changing, with many younger audiences more amenable to subtitling, while some countries are beginning to embrace dubbing

Translation of non-English originals is likely to be the next major growth area for Localisation Service Providers

It is widely reported that there has been a growth in demand for content localisation, and constraints on the supply of localisation services. So we must ask: How has demand changed? What are the causes? And how will it continue to evolve?

DEMAND FOR LOCALISED CONTENT

There has certainly been an increase in the volume of content being localised, and there are a number of reasons for this.

The last two years have seen the launch of a number of high profile international streaming services, largely from US media giants. In order to credibly launch in multiple markets, these services require a large catalogue of content for each market. As such, there has been a significant spike in demand to localise library content.

Streaming service launches have created huge demand to localise library content

The shift towards subscription streaming services in particular has accentuated the need to localise large volumes of content, rather than just flagship releases.

> A lot of this was the switch from TVOD to SVOD. In TVOD each title has to generate revenue on its own merits. In SVOD all you're trying to do is keep people from cancelling your service. And all you need is one reason per month for each subscriber.

This demand has been compounded by the pandemic, during which viewing increased overall. Media companies have looked for cost effective ways to meet audiences' seemingly insatiable appetite for content, and have found localisation to be an effective tool.

Once services are launched, the ability to create new content for an international audience rather than just for one country helps content producers to maximise their return on investment.

> From a cost standpoint, if you create a show and it does fantastically in one country, it's much much cheaper to dub it really well than it is to create a new show in a different country.

Localisation helps content owners maximise their ROI

Meanwhile, other new forms of content monetisation, such as FAST channels (Free, Ad-supported Streaming Television, which is explored further in ***Streaming at Scale***), are also adding to demand.

> One of our challenges is the creation of closed captions, even in English. We need cheaper ways to do this, because of the new ways we're exploiting content through FAST channels, self publishing, and so on. Often the cost of subtitling is a barrier to entering markets.

But it is not just the content owners who are looking to localise more content in order to fill catalogues and channels. Audiences have also shown a growing appetite for international content that has been localised, especially in markets with historically low demand, such as English speaking countries.

Some of this has been driven by the growing availability of translation services. Users have grown to expect access to content with subtitles, as platforms such as YouTube have made automated subtitling available across a huge range of video. And high profile international hits have opened audiences' eyes to the content that is available to them.

High profile international hits have opened audiences' eyes

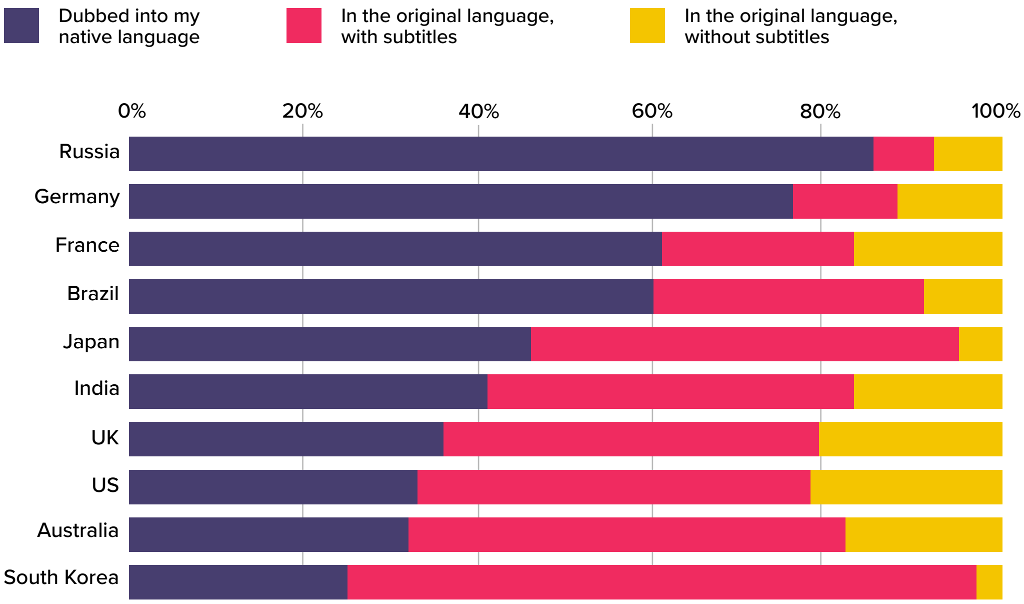
As Bong Joon Ho, the director of Parasite put it in his Golden Globes acceptance speech:

> Once you overcome the 1-inch-tall barrier of subtitles, you will be introduced to so many more amazing films.

LOCALISATION PREFERENCES

Historically, different markets have had different preferences for localisation. In the graph below, we see how audiences in some countries significantly prefer subtitling, while others dramatically favour dubbing.

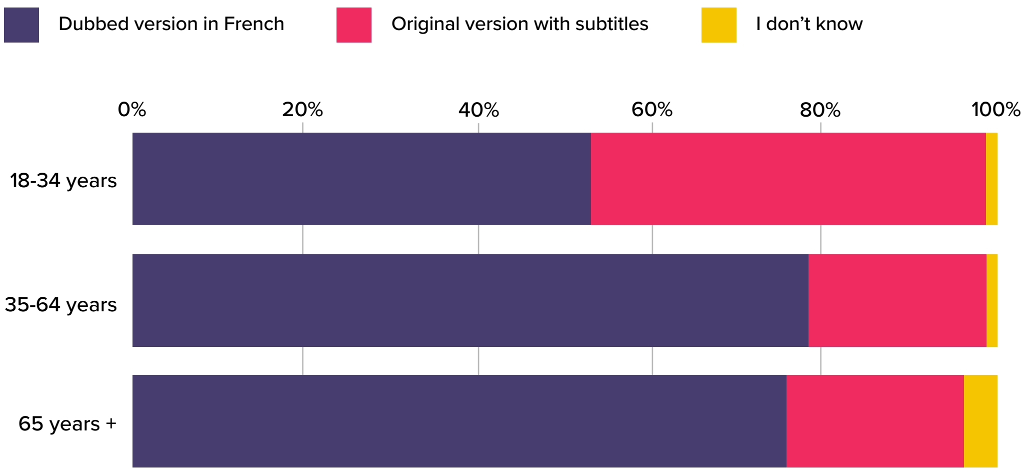
Preferable ways to consume non-native video streaming content in selected countries worldwide as of March 2022



Source: Morning Consult ©Statista 2022
Additional information: Worldwide; March 3 to 8, 2022; 999-2,211; 18 years and older; Online survey

In some countries however, attitudes are changing. The 'one inch barrier' is being broken down, with younger audiences increasingly gravitating towards subtitles. This example from France shows that, even back in 2015, Millennials had a markedly different preference compared with Generation X and Baby Boomers.

Do you prefer to watch foreign movies in their original version or dubbed in French?



Source: BVA ©Statista 2022
Additional information: France; BVA; March 10 to 11, 2015; Online survey

Millennials are demonstrating very different preferences to previous generations

Contributors to this research reported similar trends in some other markets, but audience preferences are far from universal.

Certain countries have highly specific local preferences, such as in Poland, where 'lektoring' is used. This is a form of voice over translation, in which a single voice provides a neutral translation over the top of the original programme audio. Although traditional dubbing is increasing in popularity in Poland, lektoring remains dominant.

One of the most notable changes is the shift in demand from English speaking audiences. Viewers in the USA (as well as the UK and elsewhere) have historically been largely uninterested in foreign language originated content.

> Anglophone audiences didn’t need to have content from anywhere else because they had enough of their own.

In the cases where non-English content was available, audiences generally preferred subtitles. In turn, this meant that services to provide English language dubbing were rare, and sometimes poor quality. However, this is starting to change.

> The public in the US never had a preference for dubbing. They just wanted content in English. So what happened is that the quality of dubbing in the US was often very poor. But now with a lot more international content, there has been a big improvement in English dubbing, and you're seeing a lot of the audience accepting it more, and choosing dubbing.

While it is hard to be certain of the reasons behind these changes in attitude, it is likely that the availability of more international content online has helped to change audience behaviours.

> It has been tried in the past but it never took off. Now though, it’s taking off. English speaking audiences now not only tolerate but welcome and want content from other markets.

English speaking audiences now welcome and demand content from other markets

This may prove beneficial not only for English speakers – who are now exposed to a wealth of additional high quality content – but also to viewers who speak other languages, as English speaking content creators’ understanding of the value of good localisation grows.

> The abundance of content that's not originating within the US has caused me to look at this differently, and realise how bad an experience poor localisation can be. I watched a well known imported German show, and the English dubbing was terrible. So now I watch it in German with English subtitles. That was the first time I thought, ‘now I understand why this is so important’.

DEMAND FOR LOCALISATION SERVICES

This increase in demand from both viewers and content providers has had a significant impact on the market for localisation services.

One Localisation Service Provider (LSP) interviewed for this project reported an increase in volume of 75% between 2021 and 2022. More broadly, the market for language services more than doubled between 2009 and 2019 [*CSA research, via Statista*]. And it is forecast to further grow by more than 50% between 2021 and 2027 [*ResearchAndMarkets*].

The language services market doubled between 2009 and 2019

The *Cloud Localization Blueprint* reports that “one of the most pressing concerns is the lack of adequate vendor capacity to service the huge and increasing level of demand for localization coming from content creators and distributors”. And this was certainly echoed in the DPP’s research.

This capacity constraint varies by market, with the German language being cited by multiple contributors as having been particularly constrained, alongside French and some others. In recent years and months, the constraints have ebbed and flowed particularly as major streaming service launches approach in different markets.

The particular dynamics of the localisation market make it especially hard to predict and manage supply. Most of the work of translation, dubbing, and so on is performed by freelancers in-market, and those freelancers will often be on the roster of many LSPs. So it is not possible to understand the total market capacity by summing the theoretical capacity of each LSP, as their capacities may overlap.

The increase in demand has, in some cases, caused rates to increase. But the larger challenge has often been a significant increase in lead times. This can especially impact smaller content providers.

> “Localisers can also pick and choose. They can choose to work on the latest big Netflix or BBC show, and smaller content owners with less famous or less widely promoted content go to the back of the queue.

The next growth area is being driven by the rise in demand for imported content in English speaking markets. Demand for subtitling into English is rising, while dubbing into English is growing fast from a very low base.

Translation of non-English originals is expected to be the largest growth area in the near future

> English dubbing is in its infancy. We're still establishing boundaries and templates of what we should give the audience; what the experience should be. But it's growing.

As localised content becomes more and more central to streaming services' and content producers' outputs, there is also a growing expectation of quality.

> There's a lot more attention to quality now than there was before.

The ways in which quality is assessed and measured are complex. In dubbing, for example, there's the accuracy of the translation, the preservation of original creative intent, the performance of the voice actor, the precision of the lip syncing, the quality of the audio mixing, and so on.

> There is a high demand for quality, and that's increasing, not decreasing. People are asking, how can we make this better? How can we make the experience as natural as possible?

The demand for quality is increasing, not decreasing

However, as we'll explore later in this report, equal quality levels may not be applied to all types of content.

Automating localisation

KEY INSIGHTS

There is great potential for efficiency and automation in the localisation supply chain, and best practices can be learned from other areas of the media supply chain

Machine learning can now automate each of the components of localisation, but the results today are rarely of high enough quality to be used without human intervention

Speech to text and automatic translation tools are increasingly used as aids for human localisers, enabling them to process more content more quickly

Speech synthesis has large potential and is being applied in some areas, but so far it lacks the ability to fully replicate a wide range of human performance

Voice cloning is generating a great deal of excitement, enabling the performance of a professional dubbing actor to be transformed to sound like another actor's voice

Many media companies and service providers are considering automation as a way of driving efficiencies in the process of localisation. With recent developments in machine learning and synthetic media, some even hope for a fully automated future for localisation.

In this chapter, we review how automation and AI are being applied to different parts of the localisation process.

PROCESS AUTOMATION

Localisation involves a complex supply chain, with many media assets, data elements, systems, people, and companies. One of the most salient takeaways from discussion with our contributors was the extent to which the operation of that supply chain could still be improved.

> A lot of the work in localisation supply chains isn't directly involved in localisation itself. There's a lot that has to be done just to get content to the right place, to make sure that it's project managed, and that sort of thing.

The operation of the localisation supply chain has great potential for efficiency improvements

Some companies are more automated than others in this regard, but many report a large amount of manual work still involved.

The opportunity to improve efficiency is garnering new focus as the volume and importance of localisation increase.

> **For years, there's been a need to look at our processes, but the drive to do it has never been strong enough because the volumes weren't high enough. But now, the demand is high enough that we need to pay attention to this.**

The challenges mostly correlate into two areas: work management, and asset management.

The communication between organisations and teams is often manual, people intensive, and spreadsheet driven. This is beginning to change, as localisation service providers look to provide APIs for managing work orders and status updates, for example.

> **There's a lot of spreadsheets flying around; there's no single way to send work orders, to send working materials, and to send the deliverables. So you're really trying to build something that helps clients and vendors to exchange information in a single way, via APIs, ideally.**

In the case of asset management, it is important to track the media assets relating to a specific title or job, and to manage versions of each of them. When these processes are too manual, a great deal of time and effort can be wasted.

> **We spend more time looking for stuff – finding the right version, or figuring out what is the version we have – than doing the real work.**

We spend more time looking for stuff than doing the real work

These challenges bear much in common with other parts of the media industry: the need to create software defined, cloud-led supply chains (as discussed in ***Next Gen Supply Chain***), with seamless data flows between systems and organisations (as explored in ***Delivering Seamless Integration***).

More effort has been invested into this area recently, such as the IBC ***Cloud Localization Blueprint***, which proposes a reference model for managing many of these processes using an event based architecture.

Once the basics of data and media management are in place, automation can begin to be applied. Jobs can be automatically assigned to teams or individuals, status updates can be automatically sent, and outputs automatically delivered, for example.

Well designed systems also allow the collection of data on supply chain operations, enabling problem areas and bottlenecks to be identified. This can help drive further process and technology improvements in a targeted way, to drive better ROI.

Beyond these foundations, however, lies a tantalising possibility to automate – or partially automate – the generation of multilingual assets using machine learning.

> **Of course workflow efficiency is important, but it's not unique to localisation. A lot of companies are making good strides on this. The challenges that are unique to localisation are the availability of translators to manage the huge influx of content that now need to be localised. So we do need machine driven or machine aided translation. We need to translate more content faster with the same number of operators.**

We need to translate more content faster with the same number of operators

In the following sections we examine the application of automation to a number of different processes involved in localised asset generation.

SPEECH TO TEXT

Speech to text is the process of generating a text transcription from speech audio. It can be used to generate scripts for translation and dubbing, or as the basis of captions or subtitles.

> When we talk about subs and dubs, they start the same. They both begin with figuring out what's said or otherwise articulated, and converting it into some kind of script.

Speech to text is one of the most widely used forms of automation for accessibility services and localisation today. In some cases, fully automated solutions are deployed, such as in the automated captioning offered on platforms such as YouTube and Facebook.

Speech to text is one of the most widely used forms of automation

A number of organisations which contributed to this report already use automated speech to text, and find it a useful tool. However, the consensus was that it is currently used as an aid to human localisers, rather than as a fully autonomous process. It can help subtitlers to work faster, but it is not commonly being used to replace them altogether.

Humans may be required because the speech to text is not yet 100% accurate, and because of complexity of the subtitling task, which goes beyond transcription. The timing of each word must be correct, speaker tags or colours must be assigned, formatting and layout rules applied, and so on. Automated solutions do now exist for the generation of fully formatted and timed subtitles, though the outputs are often still reviewed by a human.

Sometimes more judgement or creative input must be applied, and generally this is only entrusted to humans. Where spoken text is complicated and fast, it may be necessary to summarise. And in the case of closed captions / SDH, appropriate sound cues and non-verbal information must be added.

Of course, the results generated by automated solutions are highly dependent on the input data. One expert pointed out that improvements in the deliverables from production will dramatically improve the ability to automate subtitle generation.

> The speech to text technology doesn’t work well with a full mix. It doesn't even work well with people talking over each other. But if we get the original dialogue tracks, we can do great speech to text.

If we get original dialogue tracks, we can do great speech to text

As an input into translation processes, pure speech to text may not be sufficient. For example, some languages use different words depending on the formality of the context; English doesn't have these same structures. So it is necessary to interpret the context and include this information, or the translation may be inappropriate.

There are also other use cases for speech to text beyond translation. One example is production tools which are able to use automated transcription to enable operators to quickly select segments of a live sports programme, in order to generate highlights for different markets.

But in general, the most common use for speech to text today is to create a first draft transcription or script, which humans then review and refine before creating subtitles and dubbing scripts. In this capacity it is already very useful, and is improving rapidly. Our contributors had mixed views, but were overall rather positive about the impact of automated text to speech now and in the coming years.

The most common use for speech to text is to create a draft which a human refines

TRANSLATION

Automated translation tools are familiar to most of us, and are freely available on the internet and our mobile phones. They are highly useful, though still far from perfect.

When it comes to professional media applications, the results are highly dependent on the languages involved and the context of the content. Because of this variability, there is even less willingness to use fully automated translation than text to speech.

> There needs to be a monumental change in quality before machine translation will be usable. It's not a 20% increase.

There is little willingness to use fully automated translation for professional content

Problems were reported around the translation engines' ability to maintain context between sentences, to deal with humour, and to understand nuance and sentiment.

However, once again the automated outputs are beginning to prove very useful as input for human translators.

> It's definitely helping us with efficiency and capacity. It's a tool for the creative staff like translators, and it's helping them to get to a draft version of the translation really quickly. It definitely helps them to do things faster.

Users reported rapid progress in recent years, resulting in improved acceptance of the tools by professional translators.

> It's gotten much better based on my experiences and based on the feedback that we're getting from the translators. Four or five years ago, there was huge resistance from the translator pool, that they were never going to use machine translation. They saw the technology as an enemy. But recently I got this awesome long email from one of our country managers saying that it's now a completely different level of quality. That he was blown away by this technology.

Translators now see machine translation as a useful tool, not an enemy

In general, there was low expectation from our experts that fully automated translation will be used for premium content in the short or medium term. However, they recognised the added value that automated translation can provide to other content (such as on YouTube) and as an aid to professional translators.

TEXT TO SPEECH

Synthetic speech generation, or text to speech, is the process of automatically generating spoken word from a written input.

The overall impression held by our experts today is that current synthetic voices lack the nuance, emotion, and overall performance to be used for automated dubbing. It would be fair to say that there was a great deal of scepticism that machine learning could replace dubbing actors.

> In our experience, the application of text to speech is very narrow. I don't see that it is anywhere close to where it needs to be for an actual performance.

There is a great deal of scepticism about fully automated dubbing

Others were more positive, but still feel that the quality of output is lacking compared to what is required for high value television or movies.

> There's so much complexity to dubbing from a linguistic standpoint, but also from an artistic standpoint. You will maybe be able to match somebody's tone of voice using synthetic manipulation of audio tracks, but to actually be able to produce the performance of 10 different characters, with different age groups, and different ways of expressing themselves? That is far from being a reality in everything that I've seen.

However, that is not to say that synthetic voice is without application, even today. Audio Description generally requires less emotion and performance, making synthetic speech an excellent fit in many cases.

> I already see it happening for audio description, and I think it's very positive. There's a huge amount of content out there now, and a lot of catalogue titles, and if we could make that available with AD by using synthetic voice, I think it's a big gain.

Synthetic audio description is happening, and it's very positive

And beyond movies and TV, the use of text to speech is growing at an impressive rate. Apple recently announced the availability of synthetic audiobooks [*The Guardian*], while brands including Bloomberg and Jamie Oliver are using synthetic dubbing for their YouTube output [*Broadcast*].

In each of these examples, humans still review or adjust the outputs, but development is rapid, and future improvements should not be underestimated.

> **I think we are still early days in terms of getting to that performance. But the range that synthetic voices are convincingly able to cover is getting wider. Documentaries, news; even starting to move towards things like reality TV. There's some very interesting directions that we are moving in.**

SPEECH TO SPEECH

The application of machine learning that is perhaps generating the greatest level of excitement in media today is 'voice cloning', a form of speech to speech generation.

The concept is that a speech performance recorded by a human is altered by a machine learning model to sound different – most commonly to replicate the voice of someone else. This has numerous applications, including creating dubbing which sounds as though it's in the original actor's voice.

The advantage of this approach compared with pure speech synthesis is that a trained voice actor can create the right performance, with appropriate emphasis and drama, while the AI model manages aspects of the pitch and timbre. It therefore creates a more convincing final output.

A high profile recent application of this technology was in Disney's Obi-Wan Kenobi, where a voice model of James Earl Jones was used to create Darth Vader's performance without the 91 year old actor having to record the performances. [*The Guardian*] It's also been used for translating podcasts using the original host's voice [*Yahoo*].

Voice cloning was used in Disney's Obi-Wan Kenobi

Other uses include being able to make audio fixes in post production without returning the actors to a recording studio, recreating child actors' voices as they grow up and their voices change, and creating marketing materials for a programme or film using AI models of the lead actors' voices.

While it is early days for this technology, and human intervention is often still required to create the most convincing outputs, the possibilities created a great deal of excitement with many of our contributors.

> **I truly believe that within the next five years, it's just going to become a normal thing in the film and entertainment industries. Post production editors are just going to have this as part of their workflow.**

The possibilities for voice cloning are creating a great deal of excitement

Of course, it is not a silver bullet. Training of the voice models requires a sizeable set of good quality recordings of the source actor's voice, usually requiring a time investment. And some experts felt that currently the process is too manual and expensive to deliver efficiency for high profile content at scale. Others pointed out that key questions around intellectual property ownership must be answered before use becomes widespread (see the section, *Automation challenges*).

One major challenge today is that high profile actors have recognised dubbing actors in various different countries, so audiences may not want to hear the original actor's voice.

> **Italian viewers don't want to hear Brad Pitt's voice. They want to hear the Italian dubbing actor who dubs Brad Pitt.**

While this is undoubtedly true today, perhaps the next generation of actors will be represented around the world using their own voice from the start.

There is exciting potential for voice cloning to deliver great results and enable brand new use cases. Some of our contributors were particularly bullish on the prospects for the near term future.

> **My prediction is that when we see major Hollywood productions in the future, the actors will not only sign off the rights that photographs of them can be used to promote the movie, but also that they will record voice samples so that we can train models and use them to adapt the performances of local dubbing actors. That is something that they will be selling the rights for, within two years.**

In future, actors will sell rights to their voice as well as their photographs

ADDITIONAL AUTOMATION

We have reviewed some of the most prominent forms of automation beginning to impact the processes of localisation. But of course, there are many more.

Machine learning has been deployed at scale to automatically conform audio and subtitles, as explored in detail in ***Automating Media***. This can remove or reduce an otherwise time consuming process for editors managing compliance and other edits.

AI can be used to assist with some of the supply chain and media management problems, such as automatically identifying differences between two video or audio files, when media management and versioning challenges arise. It can also be applied to identify the language of incoming audio tracks, if they are not appropriately labelled. Or to identify on-screen text which needs to be localised – another use case explored in ***Automating Media***.

Machine learning models are also being applied to assist when the inputs into localisation processes are suboptimal. If separate dialogue stems are not available, it is possible to use such models to separate dialogue from music and effects, or to separate speaker tracks. And video segmentation tools can identify title cards, credits, and other elements which may need to be localised.

Virtual humans can now be used to generate entirely synthetic video

Finally, there are synthetic media applications which go beyond the use cases explored here. While we have only considered the localisation of video, there are now real-world applications of fully synthetic video using virtual humans, including to create news bulletins for broadcasters in Korea (see ***CES 2022: What consumer trends mean for the media industry***). The results may not yet make Hollywood directors or actors fear for their jobs, but they are certainly unlocking new media applications and pushing the boundaries of technology.

Applying automation

KEY INSIGHTS

Change management is critical to any technology development: automation will lead to people's roles changing, and the impact of this should not be underestimated

New processes require new industry best practices to be formed around intellectual property ownership, consent, and editorial review

Media companies and LSPs must strike a balance between relying too heavily on automation too soon, and being left behind by failing to innovate

Different content has different requirements, so automation will initially be used on lower value or lower profile titles, with more human review for premium content

To maximise ROI, content providers will need to accurately predict which markets content will succeed in, but remain agile to respond when audiences surprise them

It is clear that automation and machine learning have great potential to enhance the efficiency and productivity of the localisation process. But it is also clear that content owners do not yet feel it delivers high enough quality to be used for top tier content.

So what are the challenges to be overcome in order to move towards applying such solutions? And how will these technologies be pragmatically applied as they continue to develop?

AUTOMATION CHALLENGES

We have identified five key challenges to automated localisation:

1. Accuracy

Margins for error are extremely low in professional media, whether in high budget entertainment where brand perception is paramount, or in news where trust is critical. Content providers must find the right mix of human input, review, or oversight to ensure quality.

2. Change management

When new technology is applied, change management is often overlooked. There is fear that automation will lead to job losses; it is likely it will instead be used simply to assist humans. But this will result in changes to people's roles, so the change must be approached and managed carefully.

> **If you're a translator, and you're now asked to work instead as a translation editor, it's a very different application of your skill. You have less creative input. So people might struggle to adapt to that.**

3. Intellectual property

Voice cloning brings a new set of IP and consent questions. With no existing case law, best practices must be defined for issues such as the consent to build a model, ownership of the model, and review and approval of generated media.

> **I actually think it's going to be less of a challenge on the technology side, and more around getting talent consents, and industry best practices on how this is going to be done.**

4. Cost

There is often an expectation that automation will save money, but cutting edge machine learning is costly to develop. There will likely be investment required, and initial costs may be as high as manual processes. As technology matures, costs will likely reduce.

> **It might cut time out, it might improve quality, but technology rarely makes it cheaper.**

5. Moving too slowly

While there are challenges to overcome and risks to relying on automation without appropriate quality checks, there is also risk to not investing in automation. With technology moving so fast, companies who fail to innovate will be left behind by competitors.

A TIERED APPROACH

Applying automated techniques for localisation is not an 'all or nothing' choice of course. Automation can be applied with human oversight, and it can be applied to some content without using it for all.

It is recognised that not all content has equal value, and applying different solutions and processes to different content is nothing new.

> **Even without using AI, there's always been different options at different price points. You can do English dubbing in Africa or in South East Asia, which is more cost effective than doing it in the United States.**

Tiered approaches are not new or unique to machine learning

Each content creator, owner, or platform must make their own value judgments. Those judgments may differ between genres, titles, or regions.

In one of the research conversations for this project, a content provider described children's content as being lower value and hence more suitable for automation, while another saw their children's programming as the most brand critical, and hence the most likely to require manual oversight. Each company is different.

> **Companies with deep back catalogues would have localised their tentpole content. But they're being pulled to make more of the library available in different languages. And they need to do that in a way with a cost structure that will deliver ROI.**

In many cases, the approach taken depends on how closely the viewer is likely to connect the content owner's brand with the localised output.

YouTube, for example, is transparent about offering platform-provided automated captions. Most users understand that these are provided by YouTube not the content creator, and that they are generated using automation. Therefore they may be more accepting of some errors. But Warner Bros Discovery, the BBC, or RTL might reasonably assess that viewers on their flagship streaming platforms would be less accepting of similar inaccuracies.

> Even on YouTube, it depends on the content. For much of the user generated content, automated captions are fine. The viewer isn't going to care if the linguistics are not completely perfect. But there are high profile, high quality channels where audiences are very picky.

For premium content, automation will be blended with top human talent

When it comes to premium content, automation and new technology will undoubtedly play an increasing role. But it is more likely to be blended with top human talent.

> I think there's always going to be tiers of content. There's always going to be the premium tier where George Clooney has a specific actor who does his dubbing. And maybe there's going to be some kind of voice cloning technology to assist with that, but in the end, you need top quality talent to be able to do top quality dubbing. That's never going to go away. At the other end of the spectrum, you might have mass production shows where you have a very high degree of automation. Maybe even 100%. And there will be stuff in between, where humans work with the technology. Fundamentally I think those tiers are permanent.

DATA DRIVEN BUT RESPONSIVE

As the streaming gold rush subsides, demand for localisation is likely to level off. The immediate need to populate large catalogues with localised content in many different countries will give way to a more steady state of localisation demand.

Most content owners and global platforms will return to a 'theatrical model' in which they predict how a title will perform in different markets and languages, and make localisation investments based on those predictions.

Content owners will return to localising content for markets in which it is predicted to perform well

This will be more cost effective than going to the expense of speculatively localising everything into every language.

> Everything used to be done on the basis of research. We knew when a film came out where it would perform well, and what languages to go to first, because the revenue would be there. Now, I feel like it's somewhat of a free for all. Everyone feels they've got to go everywhere. But the costs are high. Is the payback always there?

Content providers will need viewer analytics data to understand how content is performing, and they will use this to improve future predictions. The most sophisticated will create algorithms to identify which languages they should invest in for each piece of content.

They may also perform experiments and tests to learn about changing viewer preferences, such as offering both lektoring and dubbing to Polish audiences, or providing dubbing to Swedish audiences who traditionally prefer subtitles. By seeing which option viewers choose, and whether they stop watching or change audio track, platforms can gain rich understanding of user preferences.

Viewer behaviour data will be used to identify audience preferences

However, no matter how good the predictions, content owners must also remember that audiences are prone to surprising us. They must be ready to respond quickly, providing additional localisation if content performs beyond its predictions. And they should be cautious of the impact that could be caused by lower quality localisation if content reaches a bigger audience than expected.

> Squid Game famously had subtitles that were wrong. And of course, that content was not expected to hit so hard, so they didn't invest big in it. But now it's become the prime example of what could go wrong.

The series was 2021's most in-demand series on Netflix [*Parrot Analytics*], and its popularity ensured that multilingual viewers' analysis of translation errors in the subtitling went viral on social media.

To deliver the optimum blend of quality and cost effectiveness, content providers must be data driven, but still agile.

The future of automation

KEY INSIGHTS

Localisation is close to a tipping point at which automation will become a significant enabler, as machine learning models develop and improve

2022 was the year that generative AI reached the mainstream, with synthetic image and text generation becoming widely available

The next generation of AI developments will deliver better accuracy, support for more languages, and the ability to combine multiple 'modes' such as voice and video

It's early days for interoperability of these technologies, but efforts are underway to build standards for data interchange

There is a palpable sense that the localisation sector is at the start of an exciting time for technology development. For every fear about AI taking jobs or reducing creativity, there is a hope or expectation that it will provide more capacity, better user experience, and greater return on content investment.

> **We're at an apex. We're at the tipping point, where machine learning for localisation is really going to work.**

Some of these developments will come sooner than others, but it's not just the technology experts or vendors who expect them to start making an impact soon. Content owners and LSPs feel optimistic too.

> **Today if you wanted to fully automate your localisation process, you wouldn't be able to. The impact of significant errors is too great, the risk is too high. But two years from now we'll be in a very different place. It's close to prime time.**

Two years from now we'll be in a different place. It's close to prime time

In the short term, most companies will continue to take a cautious approach, testing new technologies while using humans for many processes, and to assure quality of automated outputs.

> Judicious usage of AI is what's important at this point in time, while we see how the future unfolds. Let the machine do what it is best at, and let humans focus on the creative elements involved in the chain.

It's very common for experts to talk about keeping a 'human in the loop' – and indeed even companies on the forefront of technology development use humans to train models, review outputs, and make adjustments. But leaders across the industry recognise the need to move forward.

> We also need companies that are pushing the boundaries. I don't think we'll end up with completely autonomous solutions, but I will never tell a company that's trying to build that to stop. I need them to push those boundaries, because in reality we'll land somewhere in between.

We need companies that are pushing the boundaries

MACHINE LEARNING DEVELOPMENTS

The world of localisation is benefiting from rapid development of machine learning. In no small part this is because localisation solutions rely on foundational technologies and models which are being developed at extremely large scale for other users and industries.

2022 was the year that generative AI models reached the popular consciousness. First came image generation tools such as ***DALL-E*** and ***Stable Diffusion***, the deep learning model popularised through the Lensa app and its magic avatars. Then came ***ChatGPT***, the publicly accessible chat bot that enables users to converse with the GPT-3 text model.

2022 was the year that generative AI reached popular consciousness

These technologies are a novelty today. But they are beginning to break through to the mainstream. For proof of their future ubiquity, look no further than ***Apple*** optimising their software frameworks to run Stable Diffusion on the chips built into every iPhone. The ability to generate images from text will be in billions of pockets in the coming months and years.

Meanwhile, AI giants are making huge strides. ***Meta*** is investing in direct translation from speech to speech, without intermediate steps of transcription and text translation. And NVIDIA's ***Maxine*** can perform real time video translation. While it may not be ready for premium content just yet, it's a glimpse of the future.

There is an arms race to build larger and better language models. Companies like Google, NVIDIA, Meta, and Open AI, are now approaching models with 1 trillion parameters. [***Omniscien***] And although support for less common languages has so far been a sticking point, that will change as projects like Google's ***Next Thousand Languages*** and Meta's ***No Language Left Behind*** gather pace.

There is an arms race to build better language models

The next generation of developments will include multi-modal models, which use many different inputs to generate an output. For example, using image recognition to understand context which is needed to create better transcription or translation of speech. This could dramatically improve translations that could otherwise be ambiguous or wrong.

Consider translating the sentence "I looked up at the crane" into Spanish. It could be translated to "Miré hacia arriba a la grúa" or "Miré hacia arriba a la grúlla", depending on whether the subject was a piece of construction apparatus or a bird. Multi modal models could tell the difference, if accompanied by the video.

While the development of these fundamental models will underpin the future of localisation automation, we will also need tools that are specifically tuned to professional media use cases. Experts we spoke to for this research recognised their own part in building that future. One localisation service provider explained this clearly:

> **We need to invest in using the tools and providing the feedback. It might even take longer or cost more. But you're building the efficiency for the future.**

And a major content provider agreed:

> It won't evolve unless you test it and allow it to grow and learn and develop. So it's not ready now, but we need to use it now, to see where it can go later.

We need to invest in using the tools, as we're building the efficiency for the future

IMPROVING INTEROPERABILITY

As technology matures, it will also become increasingly important to be able to exchange information between companies.

Many of the key players in the AI space are collaborating, so it is often possible to use one company's speech synthesis model within another's platform, for example. But key assets such as the individual voice models used for voice cloning are proprietary to a given provider.

Some important efforts are underway to improve interoperability, however. Notably, the W3C has created the ***Speech Synthesis Markup Language*** as a standardised format for instructing speech synthesis engines on pronunciation, volume, pitch, rate, etc.

Standardisation efforts are underway in areas such as speech synthesis

Meanwhile, localisation providers and their customers can improve interoperability today by adopting existing standards and common formats such as the ***Language Metadata Table***, the ***Movielabs Digital Distribution Framework***, and the ***Interoperable Master Format***, which is explained further in ***DPP005 IMF Operational Guidance***.

Conclusion

If there is one challenge facing localisation companies and departments today, it's how to balance the need for innovation, efficiency, and scalability with the need to protect the craft of localisation and the experience of users.

> **With all the advancements in technology, we must never lose the creative element to what we do.**

Localisers must balance innovation and efficiency with craft and creativity

By becoming more efficient, and using automation to aid humans, there is the opportunity to both maximise the return on content investment, and to make more great content available to more viewers in different places.

> **We have great stories to tell, and I see great stories coming in from local original productions. Localisation is about being able to bring that to everybody to enjoy it in the same way.**

There is demand for more localisation, and in a challenging economic climate there will be little willingness to pay more for it. So with an industry feeling that the costs of manual localisation cannot be squeezed down further, automation must be an important part of the future.

In a highly unscientific poll conducted at one of our research workshops, we asked attendees to choose whether they need localisation to be faster, better quality, or cheaper – choosing only one option. Almost 40% chose better quality, but remarkably over 60% selected faster, including every respondent from a content owning company. Nobody chose cheaper.

Experts want localisation to be faster above all else

So the future of localisation will undoubtedly require more automation, some of which will be delivered by machine learning. This will be the only practical way to do more, faster, without higher costs. But that automation will be blended with human expertise and creativity from a community that is inspiringly passionate about providing audiences with better access to brilliant content.

> **People get localisation in their blood. It becomes your mission when you wake up in the morning. It’s the combination of storytelling, the creative aspect, the technology, and the global focus where you get to work with cultures around the world.**

The Future of Localisation was authored by **Rowan de Pomerai**. Workshops were hosted by **Edward Qualtrough** and **Rowan de Pomerai**, and organised by **Anh Mao**. The report was designed by **Steph McGonigle**.

About SDVI

SDVI is an Emmy® Award-winning supplier of cloud-based media supply chain technology that empowers organizations to optimize content ingest, processing, packaging, and distribution operations. The company's Rally media supply chain platform helps organizations create a scalable and responsive infrastructure that provides true business agility, operational efficiency, and process intelligence. Learn more about SDVI at ***www.sdvi.com***