

Machine Interpreting: principles, challenges, and future directions

Dr. Claudio Fantinuoli

KUDO Inc., U.S.

University of Mainz, Germany

Kent U.S, April. 2024

Image by Stable Diffusion AI



Outline

- Simultaneous Machine Interpreting: what it is and what are the use cases
- Main (abstract) tasks in a machine interpreting system
- End-to-end and cascading approaches
- Open challenges
- Evaluation
- Ethical issues

Human spoken communication is one the most distinctive and complex characters of our species.

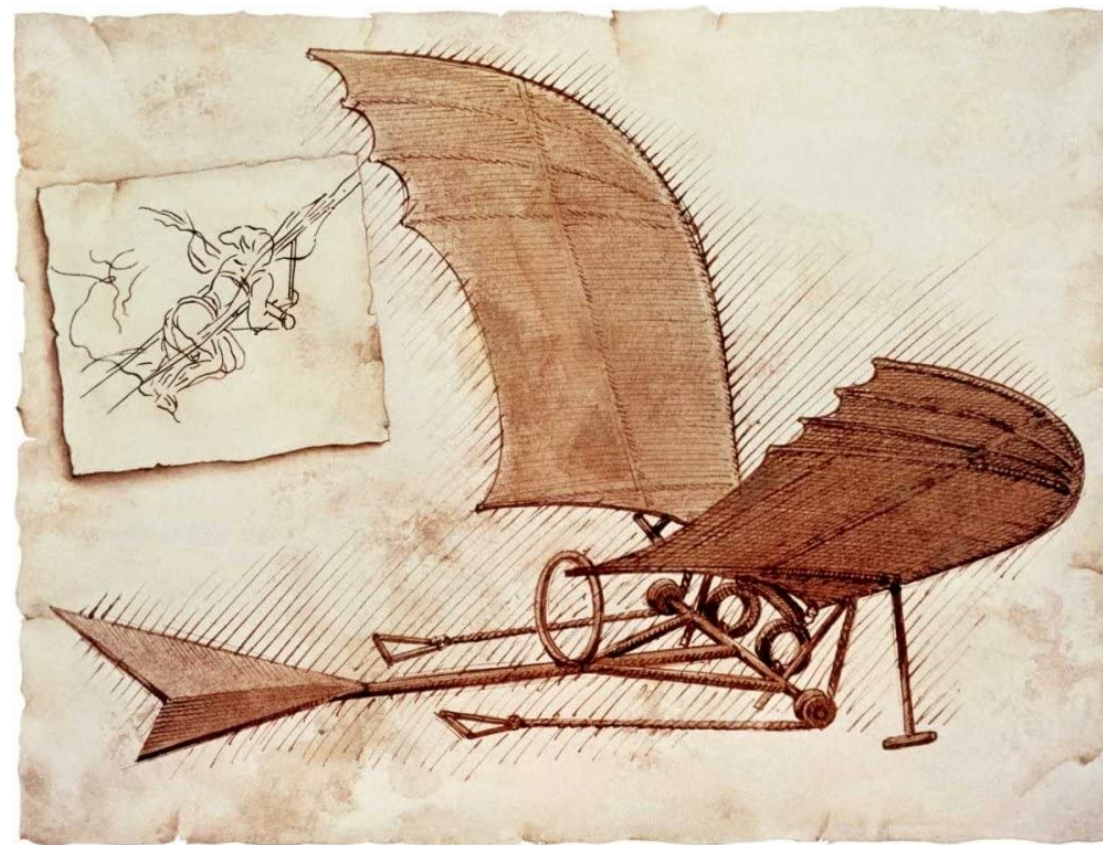
Can a machine – in principle – be able to replicate this communicative agency in a multilingual setting?

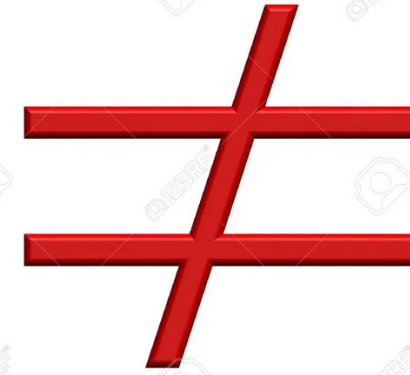
Yes, in a progressive way, because AI does not need to be intelligent to perform things which require intelligence if performed by humans.

nature

A machine does not need to imitate humans to perform similarly or better than them.

(see for example Suskind 2021)



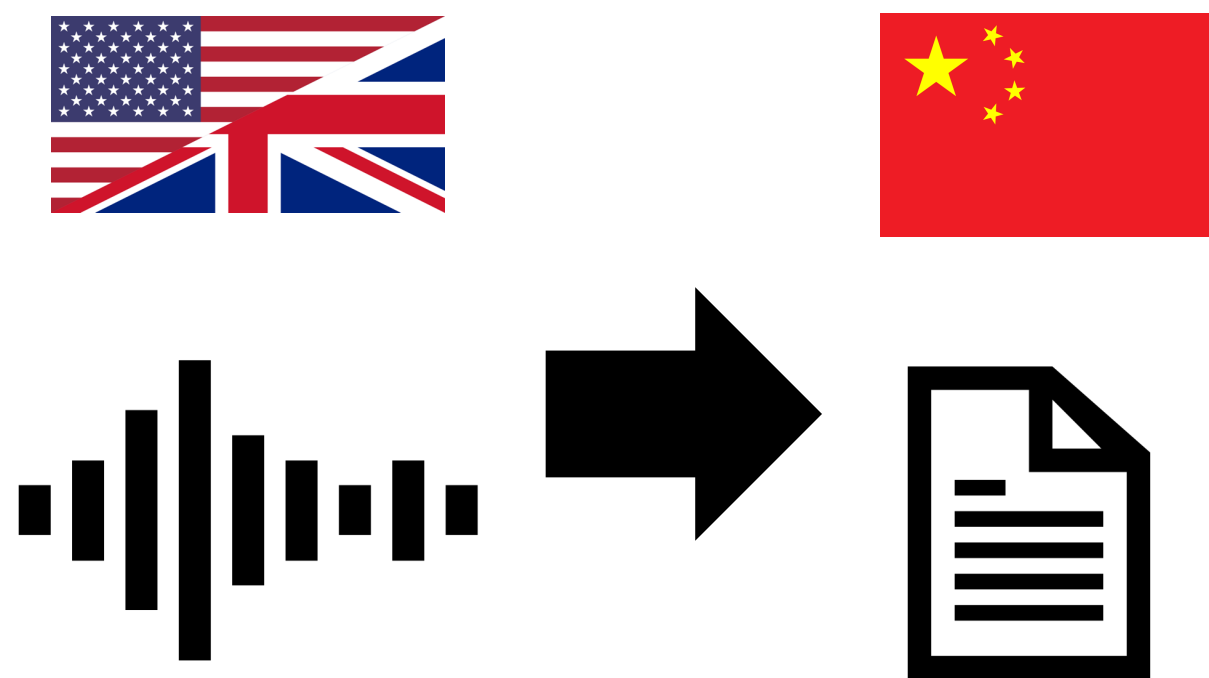
Agency  **Intelligence**

We have detached the ability to solve problems - agency - from the need of being intelligent.

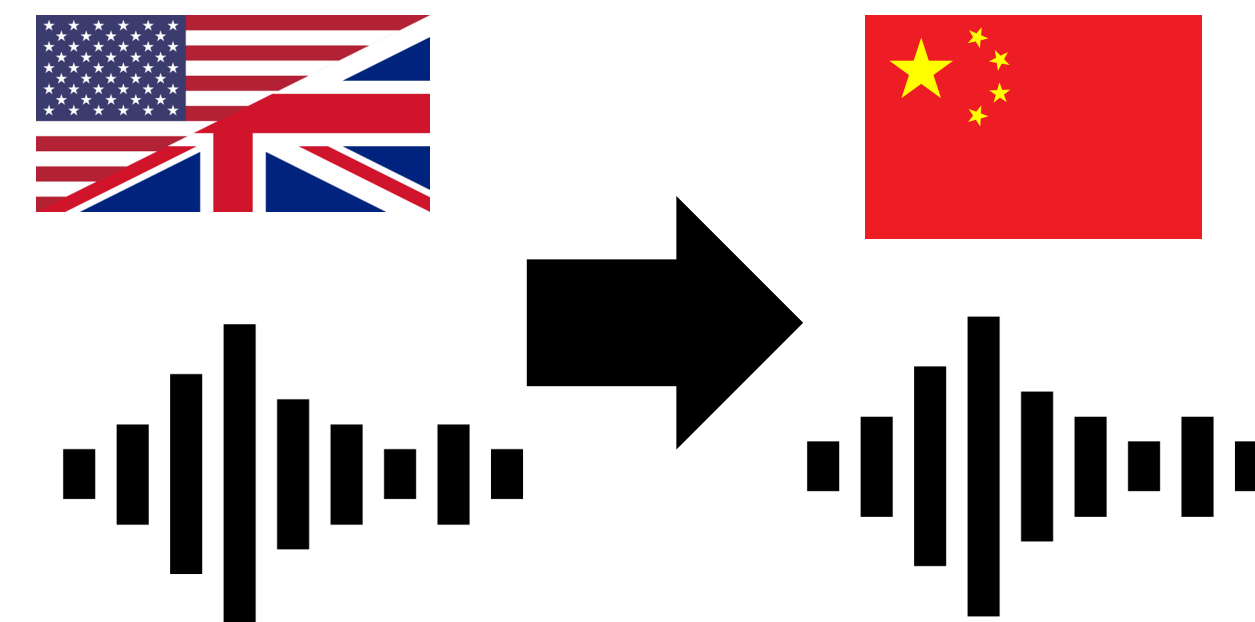
(Floridi 2018)

Spoken Language Translation

Speech-to-Text



Speech-to-Speech

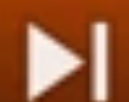


- **Offline:** for later use (audio/video recordings), possibility of editing
- **Real-time:** translation is produced for immediate consumption
 - **Sequential:** translation is produced knowing the whole spoken text
 - **Simultaneous:** produced without knowing the entire spoken text

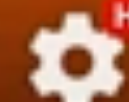
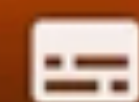
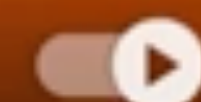
Simultaneous Machine Interpreting

Automated translation in which spoken content is translated from speech to speech in:

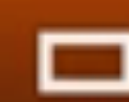
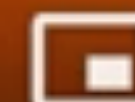
- **real-time**
- **continuously** (streaming)
- with **low latency**
- exposed only to **partial input**
- content is for **immediate consumption** (no or almost no editing)



0:00 / 59:52



HD



Use cases and why it matters

- A large proportion of international communication is conducted in English
- Human interpretation is only able to overcome language barriers in a limited number of cases
- The promise: Increasing accessibility - ubiquity, affordability, democratization - to any live event by means of overcoming the exclusivity of the professions in delivering the service (see Suskind and Suskind 2022)

Lectures

Events

Town halls

Podcasts

Services

Town halls

How Simultaneous Machine Interpreting works



Simultaneous Machine Interpreting

Many implicit or explicit tasks



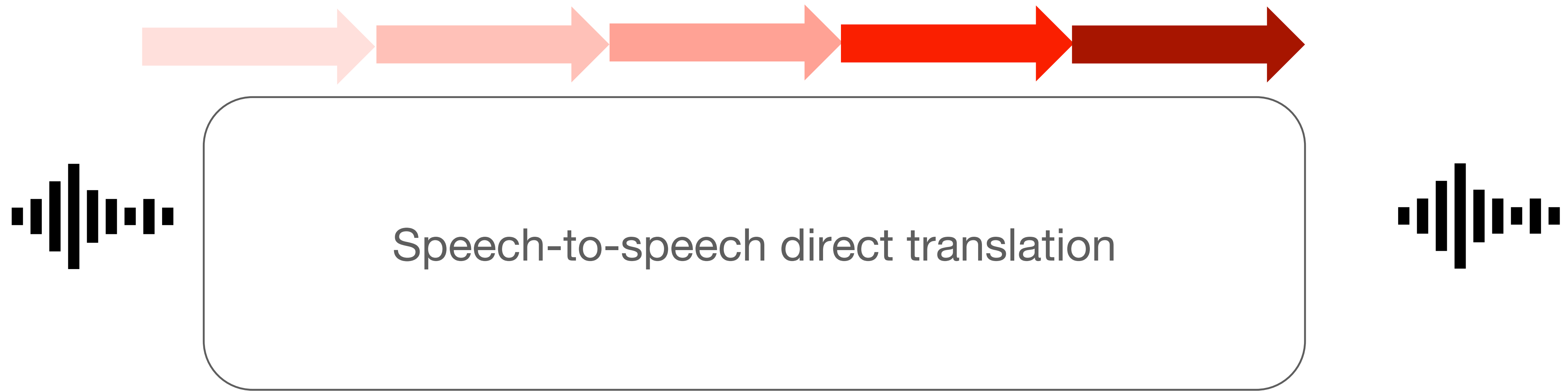
1. Receive a streaming audio as input
2. Analyze the incoming audio in real-time
3. Take decisions about what and when to translate an information, minimizing latency
4. Transform speech (control register/terminology, remove redundancies, solving co-references, disambiguate, etc.)
5. Translate from language A to language B
6. Speak aloud the translation with a natural sounding voice

NO SINGLE SOLUTION!

- Different and evolving approaches with everchanging pros and cons
- Different level of complexities driven by technological advancements and goals
- Do not believe academics that sell you a simplified version of speech translation to support the thesis that it can not work

End-to-end approach (direct)

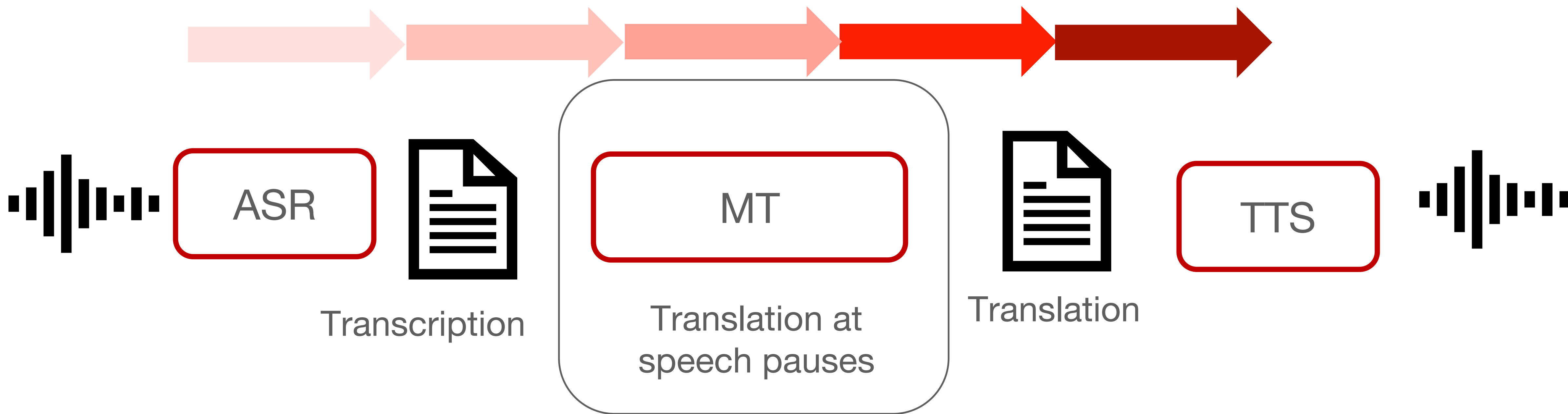
(still not existent for simultaneous)



See for example the project Translatotron (Jia et al. 2021)

Cascading pipeline

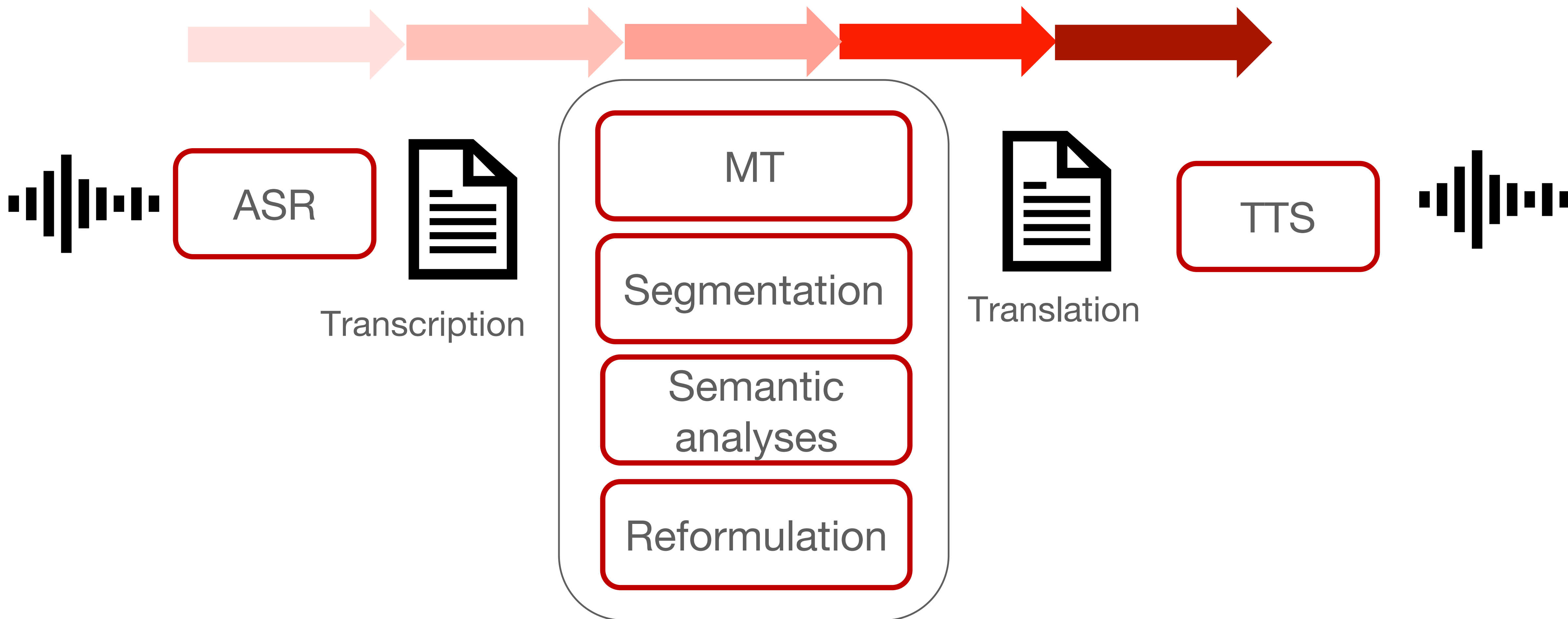
Composite pipeline with high variation architectures



See for example Proceedings of IWSLT (Salesky et al. 2022)

Cascading pipeline

Composite pipeline with high variation architectures



Tendency towards e2e

Classical example is the combination of ASR and MT into a single model



Cascading

- High quality of components
- Propagation error
- Difficult too maintain

End-to-End

- Leveraging speaker's traits from audio
- Only one system to maintain
- Scarcity of data

See Sperber and Paulik (2020)

Example of composite pipeline



Real-time
Transcription

Unfolding
speech

I am so to say so so happy to be here because uhm I want to share big news

Example of composite pipeline



I am so to say so so happy to be here because uhm I want to share big news

I am so to say so so happy to be here //

Segmentation

Example of composite pipeline



I am so to say so so happy to be here because uhm I want to share big news

I am so to say so so happy to be here //

I am very happy to be here //

Transformation

Example of composite pipeline



I am so to say so so happy to be here because uhm I want to share big news

I am so to say so so happy to be here //

I am very happy to be here //

Estoy muy content? de estar aquí //

Disambiguation

Example of composite pipeline



I am so to say so so happy to be here because uhm I want to share big news

I am so to say so so happy to be here //

I am very happy to be here //

Estoy muy content? de estar aquí //

Estoy muy contenta de estar aquí //

Translation

Example of composite pipeline



I am so to say so so happy to be here because uhm I want to share big news

I am so to say so so happy to be here //

I am very happy to be here //

Estoy muy content? de estar aquí //

Estoy muy contenta de estar aquí //



Adaptive voice generation

Challenges

- Control over terminology, register, gender, context
- Communication goes well beyond the recodification of language structures (pragmatics, etc.)
- Missing multidimensionality of information processing (visual, contextual, etc.)
- Features of real-life spoken language are complex for machines (disfluencies, poorly articulated ideas, etc.)
- Simultaneity is difficult to achieve since it requires progressive processing of speech and its meaning. Simultaneous interpreters are superstars!

Evaluation

- Evaluation of interpretation (also human) is not easy to formalize
- Lots of work going on at IWSLT, first eval of S2S in IWSLT 2022 !!
- Some pilots of user-centered eval (Fantinuoli & Prandi 2021, Javrosky et al. 2022, Korybski et al. 2022)
 - Comparison with human interpretation
 - Based on several metrics: fluency (intelligibility), accuracy (informativeness), naturalness of voices
 - Drawbacks: time consuming, based on transcriptions, difficult to suppress evaluator biases
- Automated metrics (ChF/BLEU) computed between the generated transcript and the human-produced text reference (Anastasopoulos et al. 2022)

Speech-to-text translation

Human interpretation benchmark

Languages	NTR average across speeches	NTR average across speeches	% Change when using the EXPERIMENTAL workflow
	Benchmark workflow	Experimental workflow	
Spanish	98.3%	98.9%	0.6% in favour of EXPERIMENTAL
Italian	98.9%	98.5%	0.4% in favour of BENCHMARK
French	99.5%	99.2%	0.3% in favour of BENCHMARK
Polish	98.7%	98.6%	0.1% in favour of BENCHMARK

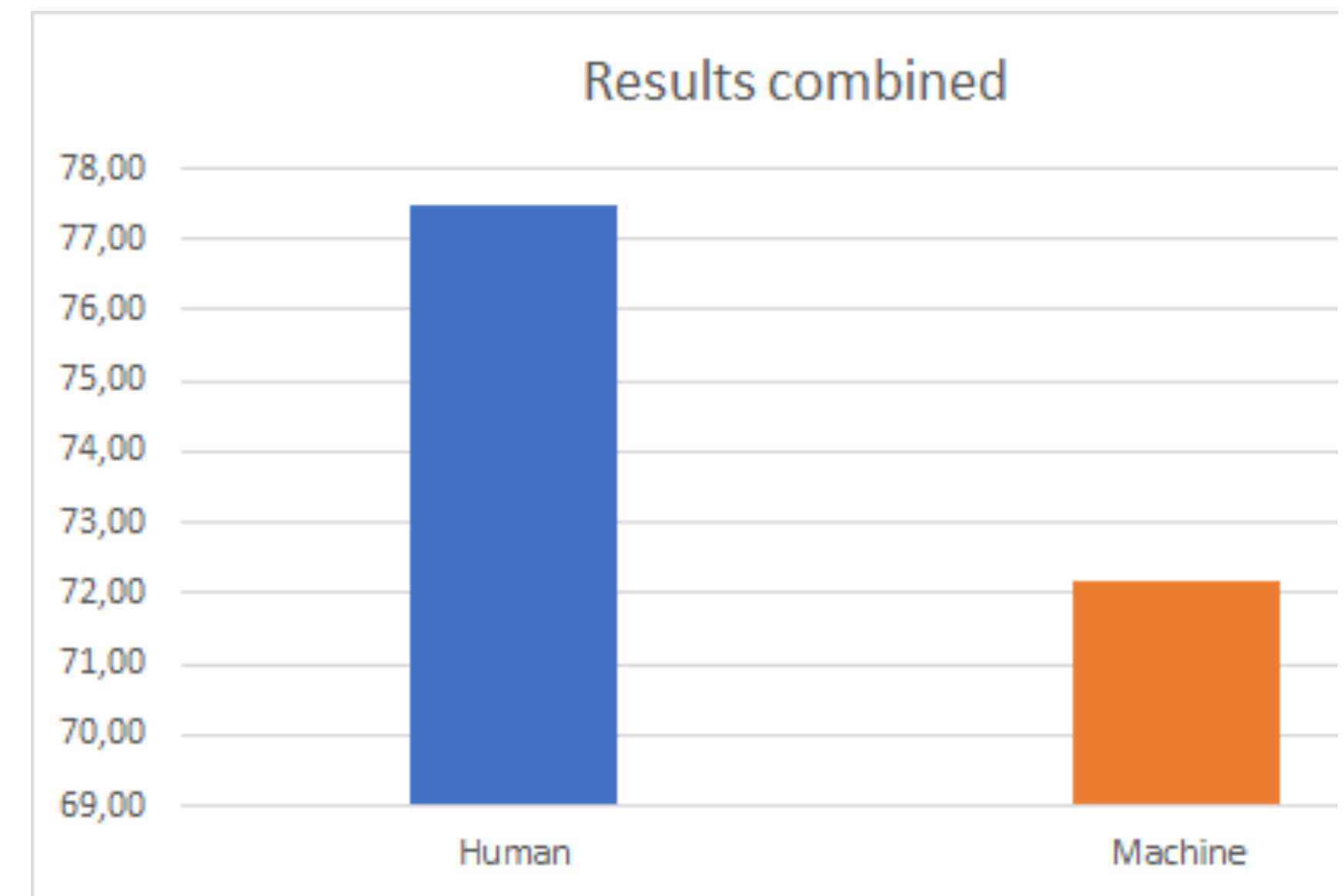
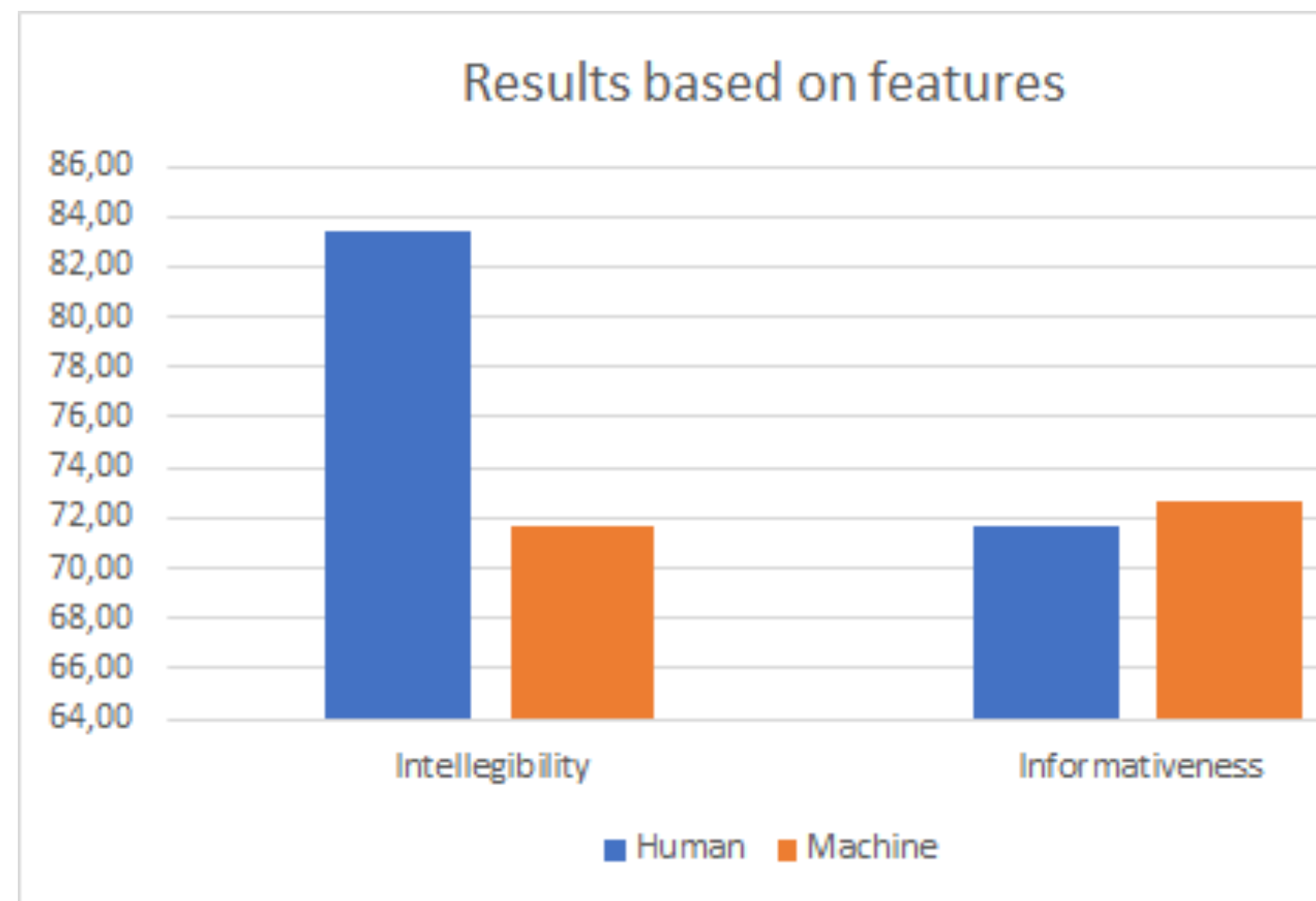
Human interpreters

Respeaker + MT

(Korybski et al. 2022)

Machine Interpreting

Human interpretation benchmark



(Fantinuoli & Prandi 2021)

Ethical issues

- Intrinsic issues:
 - Bias: gender, racial, cultural, etc.
 - Language divide -> AI is not evenly distributed among languages
- Extrinsic issues :
 - Concentration of power -> unbalance between open-source and proprietary
 - AI-divide -> many countries have not the power to keep up
 - Empowering people -> risk to depower professionals

Outlook and Conclusion

Where are we now?

- Utopia of a larger diffusion and democratization of access to real-time information in other languages may be nearer, but challenges should not be underestimated
- Given the current technology, MT has still a lot of potential to improve. Ceiling is not in sight (but there is one)
- Generative models represent the new “hot potato” since they allow – for the first time – some level of speech understanding. Still unexplored
- Multimodality will increase contextualization of the translation process
- New breakthrough can happen at any time
- Humans to be unmatched in many scenarios (high stakes translation, trust, where extreme flexibility is required, where “making sense” is difficult)

Bibliography on Interpreting and Technology

<https://www.claudiofantinuoli.org/site/itb.html>

zotero

Groups

Documentation

Forums

Get Involved

Log In

Q Title, Cr

Group Libraries					
Interpreting Technologies	Title	Creator	Date		
Computer-assisted Interpreting T...	AIIC Guidelines for Distance Interpreting (Version 1.0)	AIIC	2019		
Computer-assisted Interpreting T...	AIIC Position on Distance Interpreting	AIIC	2018		
Remote Interpreting	Have Interpreting and Technology Reached a Tipping Point?...	Allen and Olsen	2015		
	Information and Communication Technologies (ICT) in Inter...	Andres and Falk	2009		
	Community Interpreting-oriented Terminology Management...	Antón	2016		
	BootCaT: Bootstrapping corpora and terms from the web	Baroni and Bernardini	2004		
	Corpora for translator education and translation practice	Bernardini and Castagnoli	2008		
	Printed glossary and electronic glossary in simultaneous int...	Biagini	2016		
	Recommendations for the use of video-mediated interpretin...	Braun	2011		
	Technology and interpreting	Braun	2019		
	Populating a 3D virtual learning environment for interpretin...	Braun and Slater	2014...		
	Video-mediated interpreting: an overview of current practic...	Braun and Taylor	2011		
	'It's like being in bubbles': affordances and challenges of vir...	Braun et al.	2020...		

Write me: fantinuoli@uni-mainz.de or claudio@kudoway.com

Thank you

