

ÚFAL



ELITR Demo

at ECSPM Symposium

Ondřej Bojar; April 8, 2022

Demo throughout the Symposium



To see the live transcription, go to this page:
<https://elitr.eu/subtitling-at-ecspm-symposium/>

Two useful links there:

- ▶ Open **ELITR demo now**: <https://quest.ms.mff.../demo>.
- ▶ Later, help us by filling out our **feedback form**.

H2020 RIA ELITR (2019–2021 +3 months)



EU research and innovation project ELITR aims:

- ▶ Highly multi-lingual machine and speech translation.
- ▶ Document-level machine translation.
- ▶ Automatic meeting summarization, “Minuting”.



CHARLES
UNIVERSITY



THE UNIVERSITY
of EDINBURGH



PerVoice **alfaview**[®]
VIDEO CONFERENCING SYSTEMS

... and the Supreme Audit Office of the Czech Republic as affiliated user partner.

Main ELITR event: Interpreting at EUROSAT Congress (Prague, May 2021).

H2020 RIA ELITR (2019–2021 +3 months)



EU **research** and **innovation** project ELITR aims:

- ▶ Highly multi-lingual machine and speech translation.
- ▶ Document-level machine translation.
- ▶ **Automatic meeting summarization**, “Minuting”.



CHARLES
UNIVERSITY



THE UNIVERSITY
of EDINBURGH

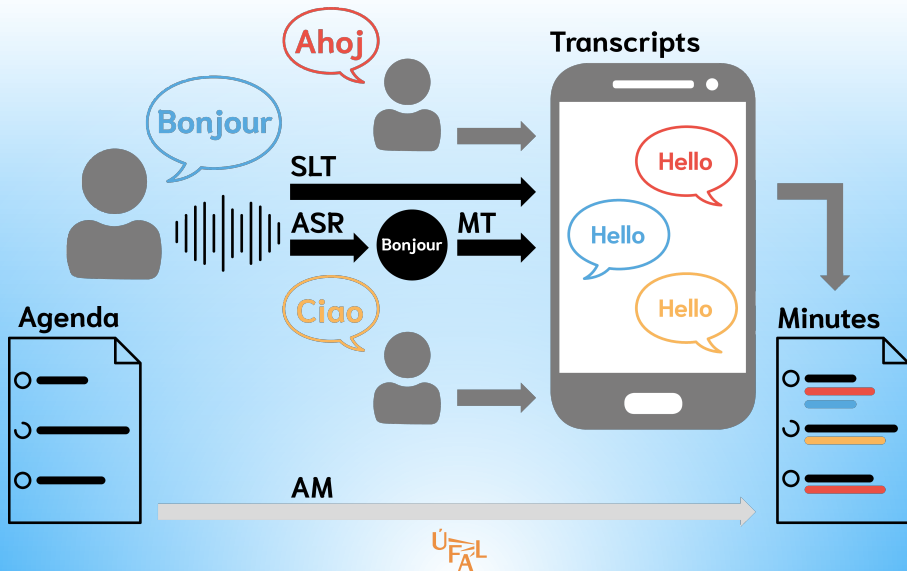


PerVoice **alfaview**[®]
VIDEO CONFERENCING SYSTEMS

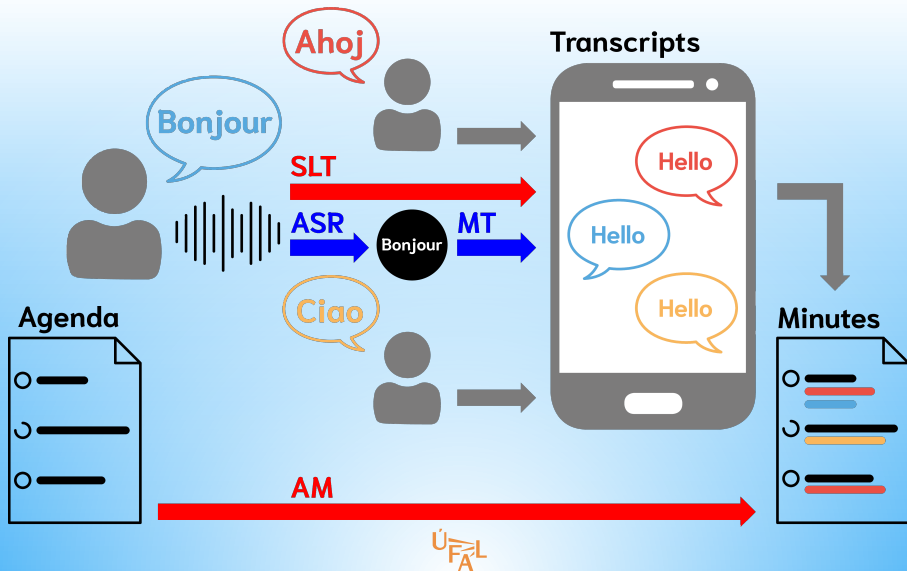
... and the Supreme Audit Office of the Czech Republic as affiliated user partner.

Main ELITR event: Interpreting at EUROSAT Congress (Prague, May 2021).

ELITR Technologies: ASR, MT, SLT, Minuting



ELITR Technologies: ASR, MT, SLT, Minuting





Languages to be supported by the ELITR project:

Technology	Primary Focus	Covered	Experimental
ASR	En, De	Fr, Sp, It, Ru	Cs
SLT & MT	{ En, De } → { En, De, Cs }	all EU languages → all EU languages	all EUROSAL langs → all EUROSAL langs
Summarization	English, Czech	–	–

24 EU languages: Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Irish, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Slovak, Slovene, Spanish, and Swedish.

Further 19 EUROSAL languages: Albanian, Arabic, Armenian, Azerbaijani, Belorussian, Bosnian, Georgian, Hebrew, Icelandic, Kazakh, Luxembourgish, Macedonian, Moldovan, Montenegrin, Norwegian, Russian, Serbian, Turkish, and Ukrainian.

Combining Two Super-Human Technologies



ARTICLE



<https://doi.org/10.1038/s41467-020-18073-9>


OPEN

Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals


Martin Popel^{1,5}, Marketa Tomkova^{2,5}, Jakub Tomek^{3,5}, Łukasz Kaiser⁴, Jakob Uszkoreit⁴, Ondřej Bojar¹ & Zdeněk Žabokrtský¹

Combining Two Super-Human Technologies





nature
COMMUNICATIONS



Cornell University

We gratefully acknowledge support from the Simons Foundation and member institutions.

arXiv.org > cs > arXiv:2010.03449

Search... All fields Search

Help | Advanced Search

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 7 Oct 2020 (v1), last revised 10 Feb 2021 (this version, v3)]

Super-Human Performance in Online Low-latency Recognition of Conversational Speech

Thai-Son Nguyen, Sebastian Stueker, Alex Waibel

Achieving super-human performance in recognizing human speech has been a goal for several decades, as researchers have worked on increasingly challenging tasks. In the 1990's it was discovered, that

Download:

- PDF
- Other formats (license)

Current browse context: **cs.CV**

< prev | next >

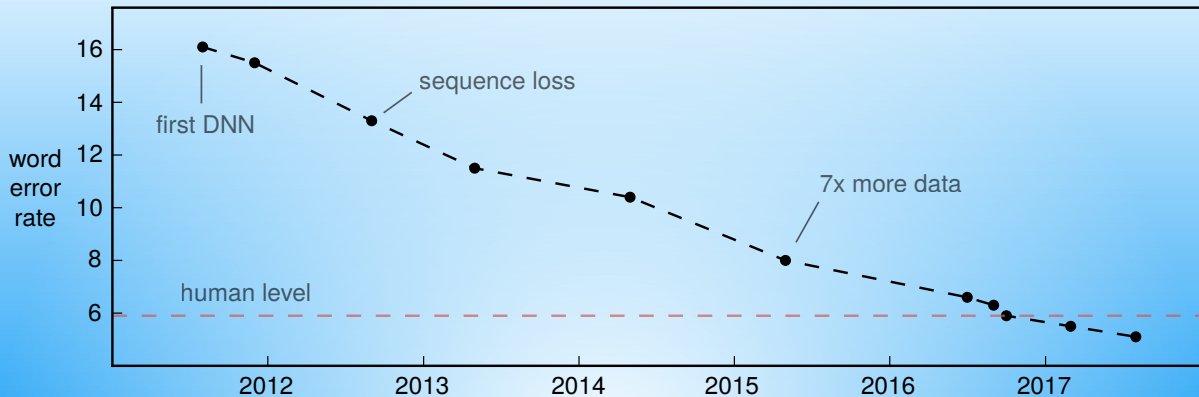
new | recent | 2010

Change to browse by: **cs**

ASR Surpassing Humans



- ▶ Switchboard conversational speech benchmark (2000).
- ▶ 40 phone calls between two random native English speakers.



MT Surpassing Humans?



- ▶ MT evaluation is considerably harder than ASR evaluation.
- ▶ Bilingual manual assessment of translation quality in WMT18:

1/10 blocks, 10 items left in block

WMT18SrcDA #1505:Segment #1487

English → Czech (čeština)

Costs are mounting in the case, with hundreds of pages of affidavits, emails and reports by companies including Deloitte, Pitcher Partners and Charter Keck Cramer filed and top barristers including Allan Myers, QC, and senior solicitors retained by both sides.

— Reference text

V tomto případě rostou Chile má deset členů a koncem srpna by měl společností, včetně společností Deloitte, Pitcher Partners a Charty Keck Cramer, a špičkových obhájců včetně Allana Myerse, QC a vyšších právních zástupců, které si ponechaly obě strany.

— Candidate translation



— How accurately does the above candidate text convey the original semantics of the source text? Slider ranges from Not at all (left) to Perfectly (right).

Reset

Submit

MT Surpassing Humans: 2018



- WMT 2018 English-to-Czech news translation results:

	Ave. %	Ave. z	System
1	84.4	0.667	CUNI-Transformer
2	79.8	0.521	UEDIN
	78.6	0.483	Professional Translation
4	68.1	0.128	ONLINE-B
5	59.4	-0.178	ONLINE-A
6	54.1	-0.354	ONLINE-G

MT Surpassing Humans: 2018



- ▶ WMT 2018 English-to-Czech news translation results:

	Ave. %	Ave. z	System
1	84.4	0.667	CUNI-Transformer
2	79.8	0.521	UEDIN
	78.6	0.483	Professional Translation
4	68.1	0.128	ONLINE-B
5	59.4	-0.178	ONLINE-A
6	54.1	-0.354	ONLINE-G

Caveats:

- ▶ Humans translated **whole documents**, MT individual segments.
 - ▶ Evaluation was done for **individual segments**.

U
F
A
L

MT Surpassing Humans: 2019



English→Czech

	Ave.	Ave. z	System
1	91.2	0.642	Professional Translators
2	86.0	0.402	CUNI-Doctransformer-T2T
	86.9	0.401	CUNI-Transformer-T2T-2018
	85.4	0.388	CUNI-Transformer-T2T-2019
5	81.3	0.223	CUNI-Doctransformer-Marian
	80.5	0.206	UdE
7	70.8	-0.156	ONLINE-Y
	71.4	-0.195	TartuNLP-C
9	67.8	-0.300	ONLINE-G
10	68.0	-0.336	ONLINE-B
11	60.9	-0.594	ONLINE-A
12	59.3	-0.651	ONLINE-X

MT Surpassing Humans: 2019



English→Czech

	Ave.	Ave. z	System
1	91.2	0.642	Professional Translators
2	86.0	0.402	CUNI-DoCTransformer-T2T
	86.9	0.401	CUNI-Transformer-T2T-2018
	85.4	0.388	CUNI-Transformer-T2T-2019
5	81.3	0.223	CUNI-DoCTransformer-Marian
	80.5	0.206	UEDIN
7	70.8	-0.156	ONLINE-Y
	71.4	-0.195	TARTuNLP-C
9	67.8	-0.300	ONLINE-G
10	68.0	-0.336	ONLINE-B
11	60.9	-0.594	ONLINE-A
12	59.3	-0.651	ONLINE-X

English→German

Ave.	Ave. z	System
90.3	0.347	Facebook-FAIR
93.0	0.311	Microsoft-WMT19-sent-doc
92.6	0.296	Microsoft-WMT19-doc-level
90.3	0.240	Professional Translation
87.6	0.214	MSRA-MADL
88.7	0.213	UCAM
89.6	0.208	NEU
87.5	0.189	MLLP-UPV
87.5	0.130	eTranslation
86.8	0.119	dfki-nmt
84.2	0.094	online-B
... 10 more systems here ...		
76.3	-0.400	online-X
43.3	-1.769	en-de-task

WMT: Annual “Competitions” in MT



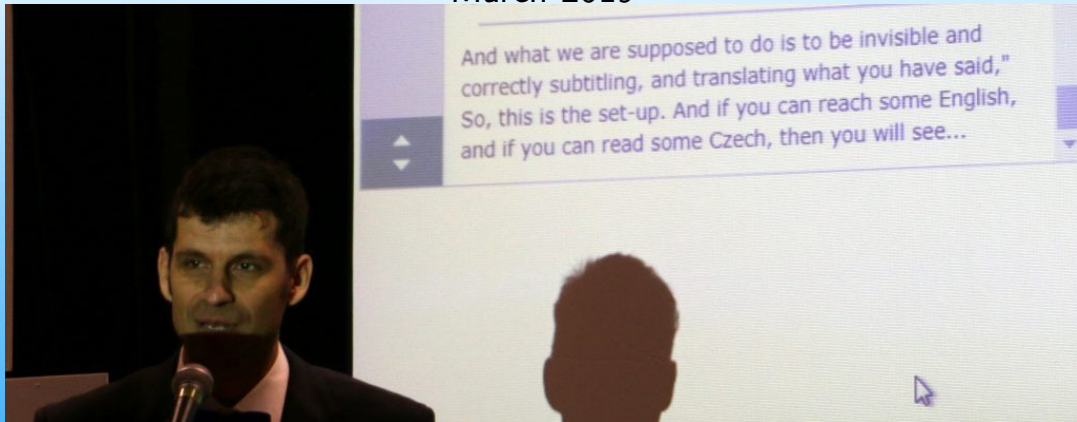
Language Pair	'06	'07	'08	'09	'10	'11	'12	'13	'14	'15	'16	'17	'18	'19
Chinese ↔ English												•	•	•
Czech ↔ English		•	•	•	•	•	•	•	•	•	•	•	•	•
Estonian ↔ English													•	
Finnish ↔ English										•	•	•	•	•
French ↔ English	•	•	•	•	•	•	•	•	•	•				
German ↔ English	•	•	•	•	•	•	•	•	•	•	•	•	•	•
German ↔ Czech														•
German ↔ French														•
German ↔ Spanish			•											
Gujarati ↔ English														•
Haitian Creole → English						•								
Hindi ↔ English									•					
Hungarian ↔ English			•	•										
Kazakh ↔ English														•
Latvian ↔ English												•		
Lithuanian ↔ English														•
Romanian ↔ English											•			
Russian ↔ English								•	•	•	•	•	•	•
Spanish ↔ English	•	•	•	•	•	•	•	•						
Turkish ↔ English											•	•	•	•

Language pairs in the main translation task. <http://www.statmt.org/wmt19/>

Reality Check: A Fair of Student Firms



<https://elitr.eu/clearest-voice-at-trade-fair/>
March 2019



ASR of Non-Natives in Noisy Environment



- ▶ Human error level: 4–6% WER (word error rate).
- ▶ Best neural nets are reportedly there, too.
- ▶ Our test of 90-second speeches of high-school students:
 - ▶ Average WER: 40–50% KIT, 80–90% Google, UEDIN.

The best recognized segment:

Manual	Google	UEDIN	KIT
why do you wear those high heels, if you would wear some sneakers ? I know one really good store, that deals with the sale of freetime footwear	why do Ø where does high heels if you would wear some sneakers I no one re- ally good star that deals with the sale of freedown food to our	're ready where tells us if you would Ø sneaker us I know won the really good store that deals with the sale of freedown food	why do are those highs heels if you would where some sneakers i no one re- ally good story that deals with the sale of freetime food to our

ASR of Non-Natives in Noisy Environment



- ▶ Human error level: **4–6% WER** (word error rate).
- ▶ Best neural nets are reportedly there, too.
- ▶ Our test of 90-second speeches of high-school students:
 - ▶ Average **WER: 40–50% KIT**, 80–90% Google, UEDIN.

The best recognized segment:

Manual	Google	UEDIN	KIT
why do you wear those high heels, if you would wear some sneakers ? I know one really good store, that deals with the sale of freetime footwear	why do Ø where does high heels if you would wear some sneakers I no one re- ally good star that deals with the sale of freedown food to our	're ready where tells us if you would Ø sneaker us I know won the really good store that deals with the sale of freedown food	why do are those highs heels if you would where some sneakers i no one re- ally good story that deals with the sale of freetime food to our



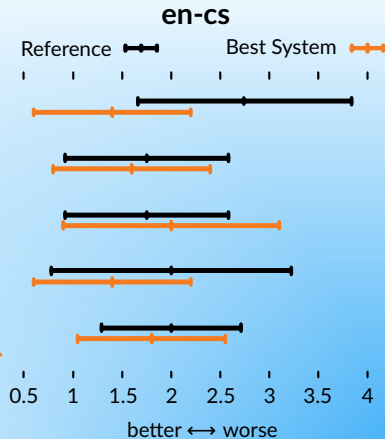
MT of Full Documents: Audit Reports



Manual evaluation by domain experts, scoring in categories:

1. Language Resources – Spelling and Morphology
2. Vocabulary – Adequacy of Terms Used
3. Vocabulary – Clarity of the Text in Terms of Used Words
4. Syntax and Word Order
5. Coherence and Overall Understanding of the Text

plotted as average rank for better comparability



MT of Full Documents: Agreements Input



Supplement No. 1 to the agreement on the sublease the apartment, of 13th May 2016
On the day, month and year written below Marta Burešová, pers. no. 695604/3017
Address: Radimova 8, Prague 6, 169 00 as the tenant on the one hand (Hereinafter referred to as "the tenant") and Karolína Černá, pers. no. 136205/891 Address: Alfrédova 13, Praha 4, 142 00 As a lessee on the other (Hereinafter referred to as "the lessee") collectively also referred to as "the Contracting parties" have agreed on this Supplement No. 1 to the Agreement on the sublease the apartment, of 13th May 2016 (hereinafter referred to as the "Supplement No. 1")

I. Introductory Provisions

On 13th May 2016, the tenant and the lessee closed the Agreement on the sublease of the apartment, under which the tenant let the lessee use the apartment No. 4 (area 49 m²) of size 1+1/L in the ground floor of the house in Prague 4, Alfrédova 13, ...

MT of Full Documents: Agreements Output



Dodatek č. 1 ke smlouvě o podnájmu bytu ze dne 13. května 2016
V den, měsíc a rok níže napsané Marta Burešová, pers. no. 695604/3017
Adresa: Radimova 8, Praha 6, 169 00 jako nájemce na jedné straně (dále jen “nájemce”) a Karolína Černá, pers. no. 136205/891 Adresa: Alfrédova 13, Praha 4, 142 00 jako nájemce na straně druhé (dále jen “nájemce”) společně označované také jako “smluvní strany” se dohodly na tomto dodatku č. 1 ke smlouvě o podnájmu, dále jen “nájemní smlouva”, dále jen “13. května 2016”).

I. Úvodní ustanovení

Dne 13. května 2016 nájemce a nájemce uzavřeli smlouvu o dalším pronájmu bytu, podle níž nájemce pronajímá nájemci byt č. 4 (plocha 49 m²) o velikosti 1+1/I v přízemí domu v Praze 4, Alfrédova 13, ...

MT of Full Documents: Agreements Output



Dodatek č. 1 ke smlouvě o podnájmu bytu ze dne 13. května 2016
V den, měsíc a rok níže napsané Marta Burešová, pers. no. 695604/3017
Adresa: Radimova 8, Praha 6, 169 00 jako **nájemce na jedné straně** (dále jen "**nájemce**") a Karolína Černá, pers. no. 136205/891 Adresa: Alfrédova 13, Praha 4, 142 00 jako **nájemce na straně druhé** (dále jen "**nájemce**") společně označované také jako "smluvní strany" se dohodly na tomto dodatku č. 1 ke smlouvě o podnájmu, dále jen "**nájemní smlouva**", dále jen "13. května 2016").

I. Úvodní ustanovení

Dne 13. května 2016 **nájemce a nájemce** uzavřeli smlouvu o dalším pronájmu bytu, podle níž **nájemce pronajímá nájemci** byt č. 4 (plocha 49 m²) o velikosti 1+1/I v přízemí domu v Praze 4, Alfrédova 13, ...

Some of ELITR Speech Translation Challenges



Input quality:

- ▶ Sound (microphone, background, ...), speaker specifics, dialects, ...
- ▶ ASR:
 - ▶ Training data, training data, training data, training data, training data, ...
 - ▶ Jargon, named entities, abbreviations, domain-specific terms, ...

Spoken Language Translation:

- ▶ Segmentation, adding punctuation.
- ▶ Latency trade-off: translate sooner with less context or better wait?
- ▶ All MT problems: Ambiguity, context, rare words, multilinguality, ...

Delivery:

- ▶ Wifi failing in crowded areas. Font too small.
- ▶ People cannot follow slides, speaker and subtitles on handheld device.

U
F
A
L

Speech Translation from Source or Interpreter?



Image source: https://ec.europa.eu/education/knowledge-centre-interpretation/conference-interpreting/conference-interpreting-explained_en

Speech Translation from Source or Interpreter?



Image source: https://ec.europa.eu/education/knowledge-centre-interpretation/conference-interpreting/conference-interpreting-explained_en

Speech Translation from Source or Interpreter?



Image source: https://ec.europa.eu/education/knowledge-centre-interpretation/conference-interpreting/conference-interpreting-explained_en

Speech Translation from Source or Interpreter?

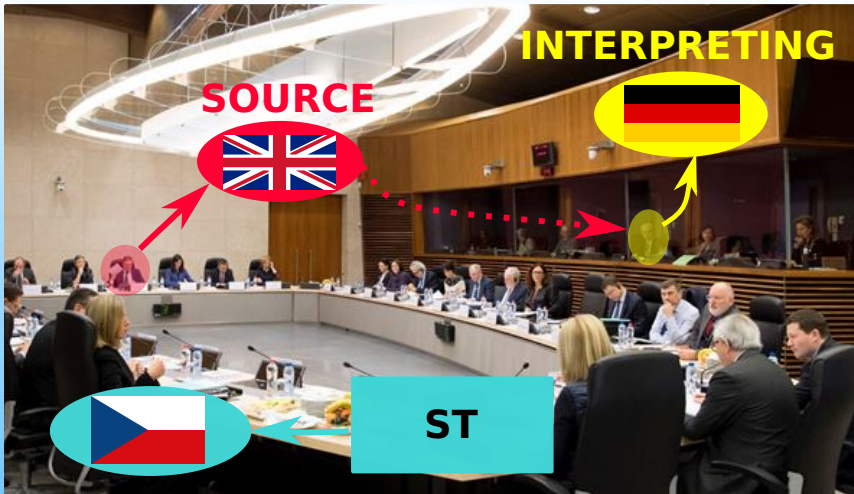


Image source: https://ec.europa.eu/education/knowledge-centre-interpretation/conference-interpreting/conference-interpreting-explained_en

Speech Translation from Source or Interpreter?

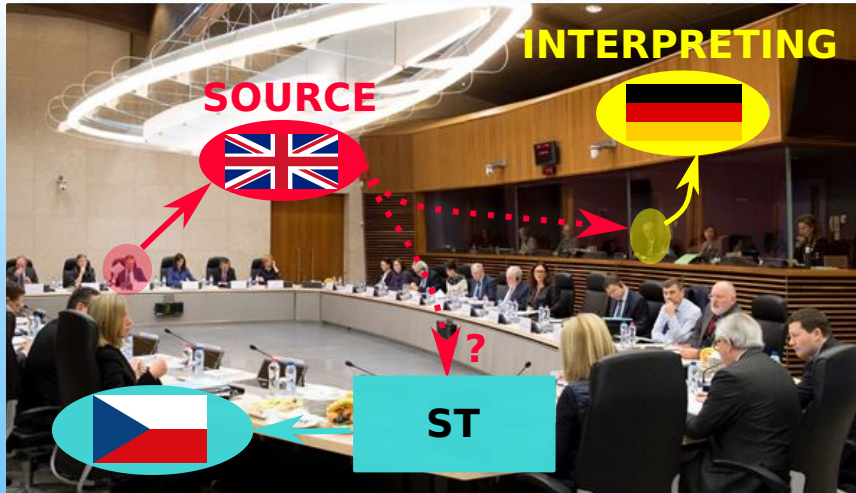


Image source: https://ec.europa.eu/education/knowledge-centre-interpretation/conference-interpreting/conference-interpreting-explained_en

Speech Translation from Source or Interpreter?

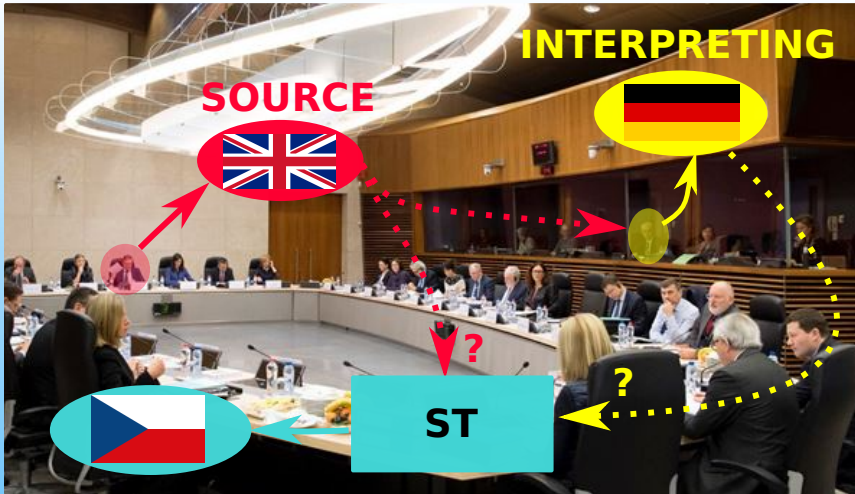
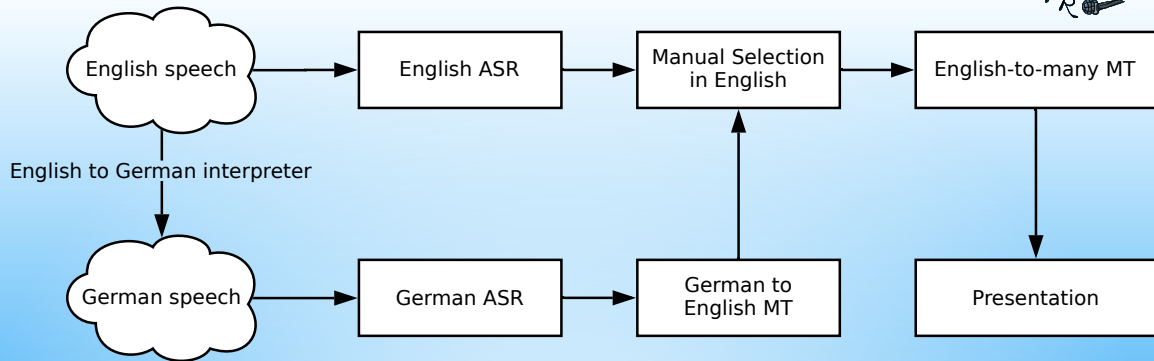


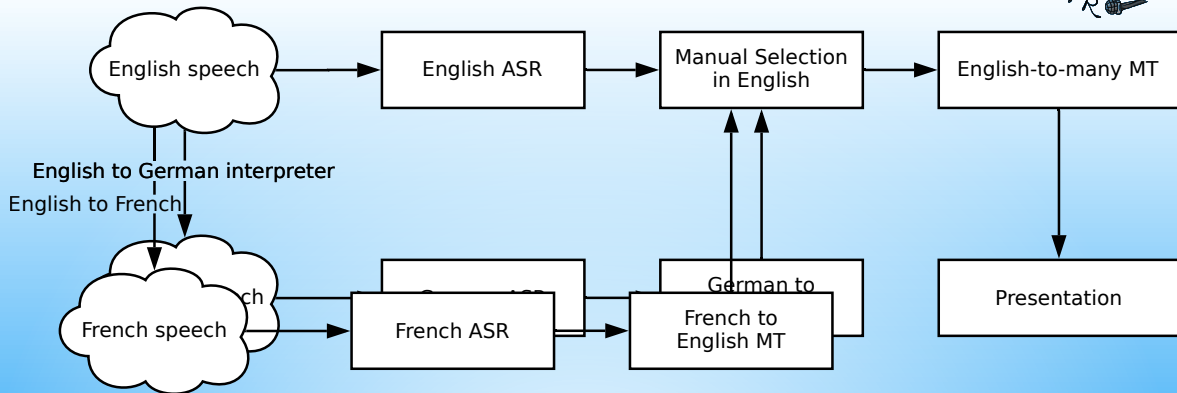
Image source: https://ec.europa.eu/education/knowledge-centre-interpretation/conference-interpreting/conference-interpreting-explained_en

ELITR's Answer: Follow All



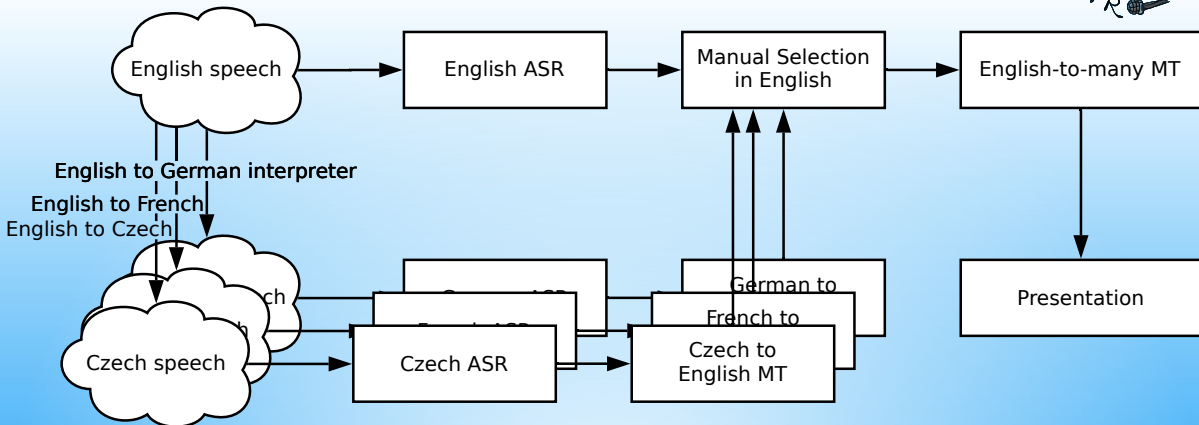
More details in Bojar et al. (2021), **Operating a Complex SLT System with Speakers and Human Interpreters**; <https://aclanthology.org/2021.mtsummit-asltrw.3/>

ELITR's Answer: Follow All



More details in Bojar et al. (2021), **Operating a Complex SLT System with Speakers and Human Interpreters**; <https://aclanthology.org/2021.mtsummit-asltrw.3/>

ELITR's Answer: Follow All



More details in Bojar et al. (2021), **Operating a Complex SLT System with Speakers and Human Interpreters**; <https://aclanthology.org/2021.mtsummit-asl-trw.3/>

Various Demo Videos



- ▶ Initial test at 2019 Student Firms Fair:
<https://elitr.eu/the-reverse-babelfish/>
- ▶ CUNI Open Doors Day:
<https://elitr.eu/a-tireless-interpreter/>
- ▶ ASR demo in Full Lecture on Complexity of Our Goals:
<https://lectures.ms.mff.cuni.cz/view.php?rec=455>
- ▶ Unpublished demo of Czech→many subtitling:
<http://ufallab.ms.mff.cuni.cz/~bojar/elitr/720p.mp4>
- ▶ **Coming soon: Popularization of SLT in conference setting:**
<https://www.youtube.com/watch?v=WZy5cGY7JFc>

Recipe to See Miracles Happen



1. Define a task.
2. **Provide training and test data.**
3. Run a “competition”, a shared task.
4. Repeat until satisfied.

Recipe to See Miracles Happen



1. Define a task.
 2. **Provide training and test data.**
 3. Run a “competition”, a shared task.
 4. Repeat until satisfied.
- ▶ WMT does this for **machine translation**.
 - ▶ IWSLT for **spoken language translation**.

Recipe to See Miracles Happen



1. Define a task.
 2. **Provide training and test data.**
 3. Run a “competition”, a shared task.
 4. Repeat until satisfied.
- ▶ WMT does this for **machine translation**.
 - ▶ IWSLT for **spoken language translation**.
 - ▶ ELITR is starting this for **meeting summarization**.
 - ▶ See our AutoMin 2021 shared task:
<https://elitr.github.io/automatic-minuting/>

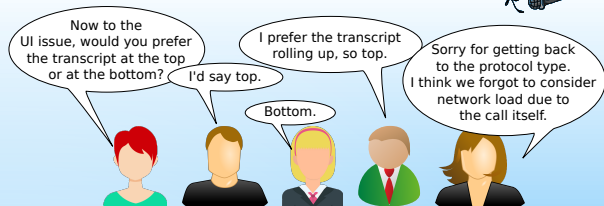
Minuting: We Need Your Data



We need:

- ▶ Audio and/or Transcripts.
- ▶ Agenda
 - ▶ Ideally hierarchical.
- ▶ Minutes
 - ▶ Ideally the Agenda populated with items.
- ▶ Any supplementary files:
 - ▶ Slides, related reports, ...

Please contact us at
info@elitr.eu.



Original agenda as prepared by the organizer beforehand:

- Protocol type: push or pull?
- Layout of the user interface:
 - Transcript grows at the top or bottom of the document?
 - Or in a side pane?

Shared document, everyone allowed to edit.

Starts with the agenda and gets populated by Automatic Minuting (AM)

- Protocol type: push or pull?
 - (AM) > Pull easier to implement.
 - (AM) > Updates can get lost with push in case the user
 - (AM) > Consider network load.
- Layout of the user interface:
 - Transcript grows at the top or bottom of the document?
 - (AM) > Top (AM) > Bottom (AM) > Top, transcript rolling up.
 - Or in a side pane?

Transcript, optionally editable to correct ASR errors:

- 11:03 Sorry for getting back to the protocol type. I think we forgot ...
- 11:02 I prefer the transcript rolling up, so top.
- 11:02 Bottom
- ...

Summary



- ▶ ASR and MT **do surpass** humans, but **in very specific settings only**.
- ▶ Some robustness is achievable through **very large training data**.
- ▶ For ASR, the **input acquisition** is critical, tiny details kill.
- ▶ For MT, handling **context** is now essential, more so in speech.
- ▶ Automatic Minuting is still very much starting.
 - ▶ **We need your work meetings data!**



Thank you!

