HOW TO IMPLEMENT SPEECH-TO-TEXT INTERPRETING IN EDUCATIONAL SETTINGS

Guidelines for making university lectures accessible

ILSA Interlingual Live Subtitling for Access

This project has been funded with support from the European Commission.

This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

How to cite this document

Dutka, Ł., Szczygielska, M., Szarkowska, A., Figiel, W., Romero-Fresco, P., Pöchhacker, F., Tampir, M., Robert, I., Schrijver, I., Haverhals, V. and Meiriño, J. (2020) *How to implement speech-to-text interpreting (live subtitling) in educational settings. Guidelines on making university lectures accessible to hard of hearing, deaf and foreign students.* ILSA Project. **http://ka2-ilsa.webs.uvigo.es/guidelines**

INTRODUCTION

University lectures can be translated and made accessible thanks to speech-to-text interpreting, also referred to as live (sub)titling. There are different methods for producing live (sub)titles, for example with keyboards or with different degrees of automation (automatic speech recognition, machine translation). The ILSA project and these guidelines focus on intralingual and especially **interlingual live (sub)titling** or **speech-to-text interpreting through respeaking**, where a speech-to-text interpreter known as an interlingual respeaker (or transpeaker) translates what is being said (dictating also punctuation marks) to a speech recognition software, which displays the translated words as (sub)titles on a screen.

This document is addressed to accessibility managers at higher education institutions, university decision makers, lecturers and anyone who wants higher education to become more accessible. It aims to help implement speech-to-text interpreting at higher education institutions.

Different terms are used to talk about the provision of live subtitling and speech-to-text services in different countries and markets. In the ILSA project, we attempt to clear up the current terminological confusion. The table below provides the major terms and their definitions, which will be used throughout this document and in all the ILSA guidelines.

DEFINITIONS

BOOTH – sound-proof space (as used, with standardized specifications, for simultaneous interpreting)

INTERLINGUAL RESPEAKING – see 'Transpeaking'

INTRALINGUAL RESPEAKING – see 'Respeaking'

LIVE SUBTITLES – live titles displayed on a screen underneath or as part of the image

LIVE TITLES – written text produced by speech-to-text interpreting

LIVE TITLING – see 'Speech-to-text interpreting'

RESPEAKING – a method of creating live titles using speech recognition technology, whereby a person known as a respeaker repeats and/or paraphrases what is being said, dictating also punctuation marks (respeaking is also known as voice writing)

SPEECH-TO-TEXT INTERPRETING (STTI) – the production of a written version of a spoken message while it is being delivered. It is mainly performed intralingually (in the same language), using keyboards or respeaking, for the benefit of people with hearing loss, but can also be done interlingually. In different countries and contexts, speech-to-text interpreting is also referred to as live subtitling or captioning.

STENOMASK – sound-insulating cover around a microphone used by respeakers to muffle speech sounds (see Fig. 5)

TRANSCRIPT – written text representing a spoken message

TRANSPEAKING – a method of creating interlingual live titles using speech recognition technology, whereby a person known as a interlingual respeaker (or transpeaker) translates what is being said, also dictating punctuation marks

USERS

Speech-to-text interpreting in lectures will benefit many different types of users, in particular:

- **students with hearing loss**, who can participate in classes together with hearing students;
- **international students** and visitors who do not know the language of the lecture and can benefit from interlingual or intralingual speech-to-text interpreting delivered to them in real time;
- other people who may experience difficulties accessing the content of academic lectures, such as people with autistic spectrum disorders or people with dyslexia.

For intralingual live titling, the main target group are people who are deaf or hard of hearing. Interlingual live titling is addressed to foreign students or visitors who do not know the language of the lecture as well as those who are deaf or hard of hearing.

Speech-to-text interpreting can also be used in online lectures and talks, benefitting people with hearing loss, those who do not understand the language of the talk, viewers in noisy environments or those who cannot watch the lecture with the sound on.

After the lecture, a transcript of the live titles can be used as notes.

BENEFITS

- Speech-to-text interpreting increases overall accessibility and fosters inclusion at your institution.
- By improving participation of the target group, speech-to-text interpreting raises awareness amongst students and academic staff about diverse student groups.
- Most students will benefit from speech-to-text interpreting as it helps them to focus on the content of the lecture. Research shows that subtitling may improve students' performance by fostering comprehension and information retention (see References).
- In the case of language classes, speech-to-text interpreting aids students in understanding speech in a foreign language and helps them learn the correct spelling.
- The by-product of speech-to-text interpreting is a transcript and/or translation of the lecture that can be made available to the lecturer, the university staff or students as needed.
- Speech-to-text interpreting has a positive impact on developing digital competences.
- Introducing speech-to-text interpreting will contribute to the digital transformation of your institution.
- If you record and make your lectures available afterwards, live titles can be re-synced to be used as subtitles for the video.
- Research has shown that students score significantly better on comprehension tests about lecture content when they have been provided with intralingual live subtitling. Moreover, intralingual live subtitling enables L2 students to reach the same performance level as L1 students without subtitles.

WORKFLOWS

Two types of workflow are possible, depending on the location where respeaking is done:

1. **On-site respeaking:** the respeaker works in the same room, preferably in a booth, or using a stenomask.
2. **Remote respeaking:** the respeaker works in a different place, either on-campus or off-campus.

Figures 1 and 2 below show models of on-site and remote respeaking.

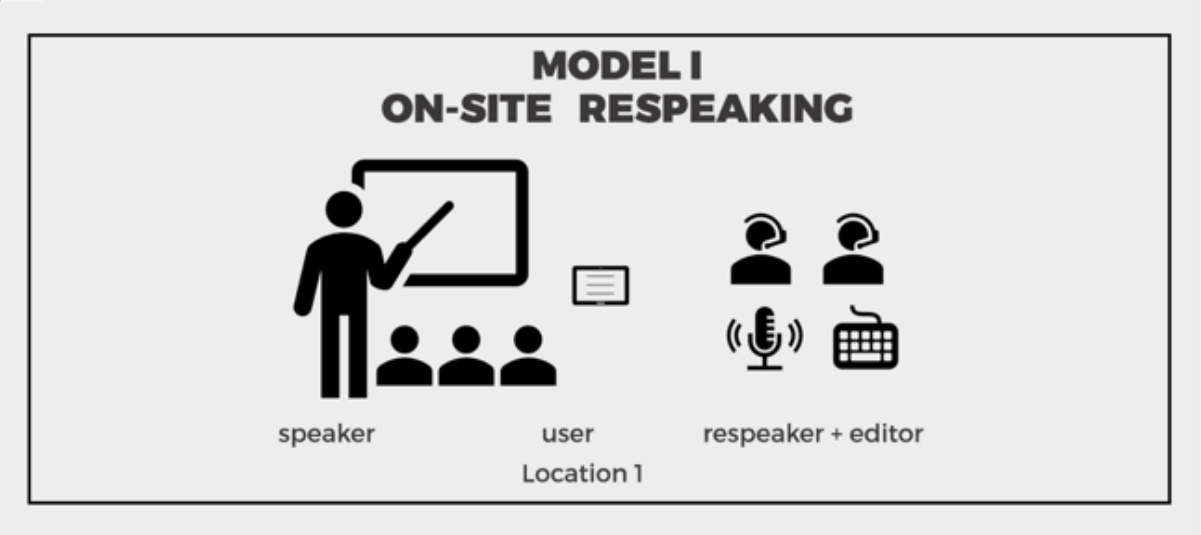


Figure 1. Model of on-site respeaking

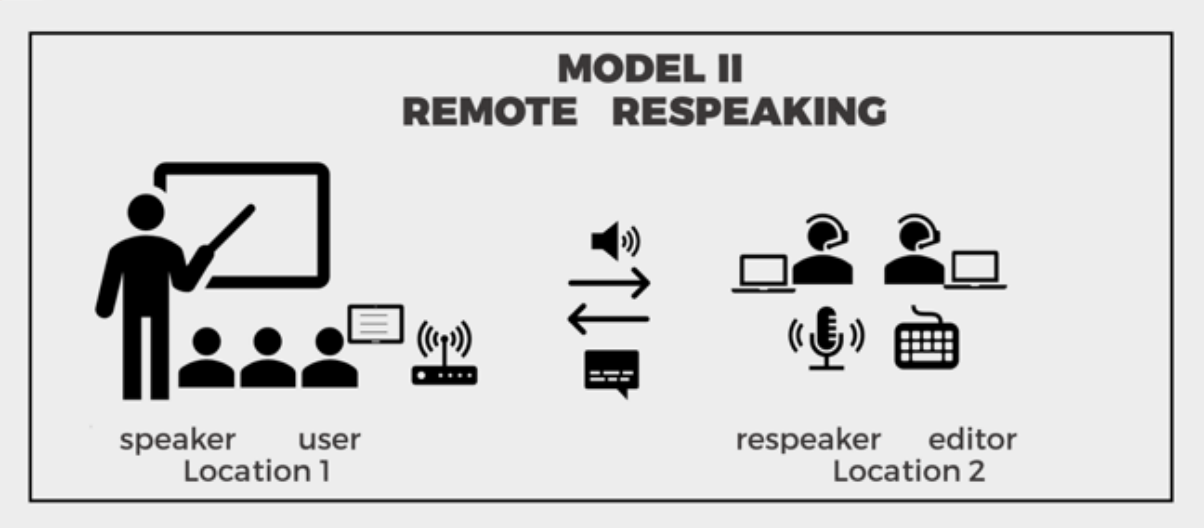


Figure 2. Model of remote respeaking

Although the respeaker can work next to the user in the lecture hall and dictate into the user's document (Fig. 1 and Fig. 3), in practice it is better for the respeaker to work remotely, in a quiet room (Fig. 2 and Fig. 4). In this way, there are fewer distractions and interferences with the speech recognition software.

Figure 3. (© Dostępni.eu)

Figure 4. (© Dostępni.eu)

Figure 5. (© Dostępni.eu)

If the respeaker sits in the same room and not in a booth, it is recommended that a stenomask is used (Fig. 5, Fig. 6). The mask will prevent the respeaker from disturbing the attendees of the lecture, as their voice will be muffled.

Figure 6. On-site respeaking with self-correction (© Sabien Hanoulle)

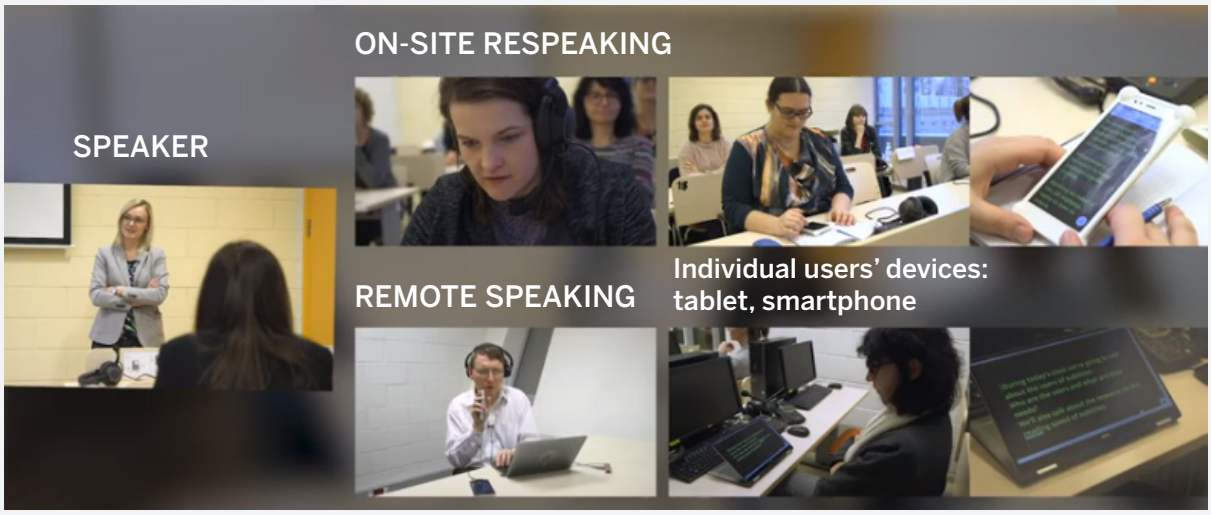


Figure 7. On-site & remote respeaking (© Dostępni.eu)

If respeaking is done on-site, in the lecture room, consider the following elements:

- if a stenomask is used, the respeaker should sit in the first row so that they are close to the lecturer and can hear him or her well;
- if an interpreting booth is used, respeakers need to be able to see the lecturer, the presentation and the room. They should also be able to test the microphone and the headset before the class;

- when an interpreting booth is used, the respeaker will only hear those who speak to the microphone, so it is essential that the lecturer makes sure to always speak to the microphone. If there are other people speaking during the class, they should also be able to speak to the microphone. If this is not possible, the lecturer should repeat the students' questions out loud.

Respeaking can be done intra- and interlingually, as shown by Figures 8-11.

Figure 8. Model of intralingual respeaking

In the case of interlingual respeaking, three workflows are possible:

Figure 9. Live titling – interlingual workflow I with interpreter & intralingual respeaker

The workflow shown in Figure 9 requires the interpreter to hear, understand and translate first (oral language A into oral language B) before the respeaker can start respeaking (oral language B into written language B). This can create more delay but, on the other hand, the respeaker does not need to know the original language.

Figure 10. Live titling - interlingual workflow II with interlingual respeaker

Using interlingual respeakers (Fig. 10) minimizes the delay with which the live titles are displayed to users. However, it might be challening to find interlingual respeakers in some language combinations.

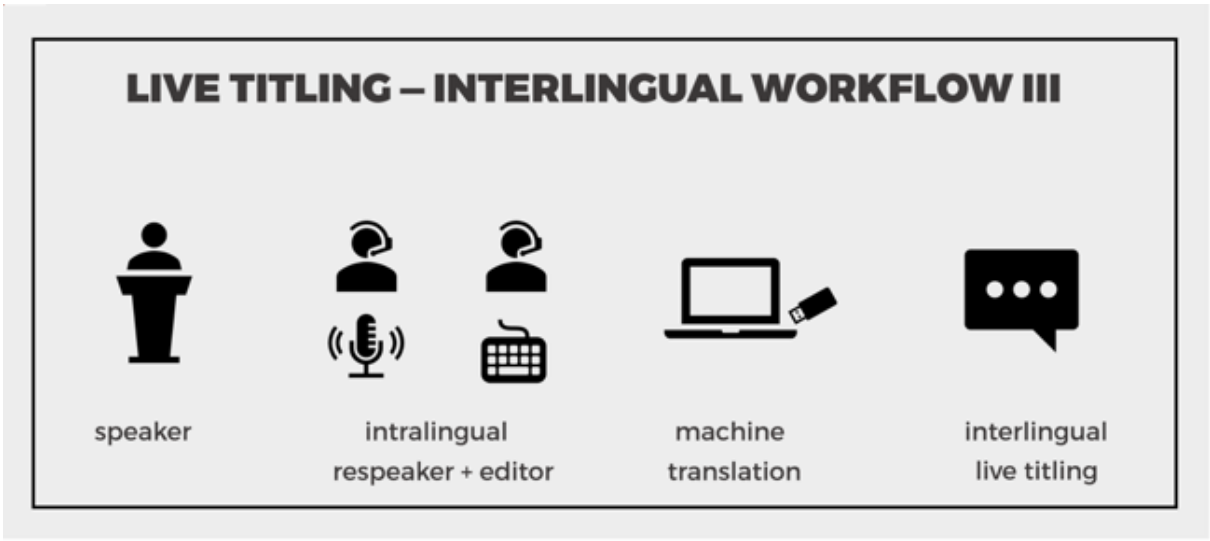


Figure 11. Live titling - interlingual workflow III with intralingual respeaker & machine translation

Using machine translation (Fig. 11) will allow live titles to be translated to multiple languages at once. Depending on the language pair, the quality of machine translation might be inadequate. The results of the machine translation output may improve if the respeakers use simple language and simplify sentence structure.

LIVE CORRECTION

As speech recognition is hardly ever 100% accurate, the text needs to be corrected. Depending on the difficulty of the topic and the number of misrecognized words, correction can be done by the respeaker (Fig. 6) or by a live corrector. In some cases it might also be possible for fellow students to correct the text.

1. self-correction by the respeaker
2. parallel correction by a live corrector
3. parallel correction by users (students)

Parallel correction by a live corrector will produce the best results, especially for fast-paced classes with complex terminology. It requires another professional, which impacts on the cost of the service.

Parallel correction by users (students) is possible when the text of the titles is available in a shared document that users can edit.

Self-correction can be used when classes are slow-paced and there are not too many errors to correct. The respeaker can stop respeaking, correct the error and start respeaking again.

RECOMMENDATIONS

- Whenever possible, speech-to-text interpreting should be done with correction.
- The transcript of speech-to-text interpreting should be corrected by the respeaker after the lecture so that it can be later used as notes.
- As respeaking requires high cognitive effort (and even more so when using a stenomask), it is difficult to sustain focus longer than 20-30 minutes without a break. It is recommended to include breaks for respeakers or have a team of respeakers that can take turns. An intralingual respeaker should not work for longer than 45 minutes (or maximum 1 hour) before a colleague takes over. In interlingual respeaking, individual turns should not be longer than 30 minutes.

Figures 12 and 13. Respeakers usually work in a team (© Sabien Hanoulle)

LIVE TITLING TOOLBOX

Depending on the institution's needs, budget and technical infrastructure, you can choose from a number of solutions.

To start, all you need is a laptop, speech recognition software and a microphone. All the user needs is a device connected to the Internet.

Speech recognition software	
desktop-based solutions	**cloud-based solutions**
–Dragon Professional –Newton Dictate	–Google Speech API –Microsoft Speech API –Amazon Transcribe

Basic tools for providing live titles

- video conferencing apps (Microsoft Teams, Google Meet, ZOOM) to send sound (also the option to send video);
- or instant messaging apps (Messenger, Whatsapp, etc.) to send the sound from the lecturer to the respeaker;
- or a telephone connection to send the sound from the lecturer to the respeaker;
- the text can be displayed in a Google Doc that students can access through a link on their devices.

Professional solutions for providing live titles

- Text on Top – software for producing and displaying live titles, which allows for cooperation between a number of workstations through wireless connection (using USB dongles)
- Various accessibility services providers have their own proprietary platforms or tools, some of which might allow for remote work

In some solutions, students who are users receive dedicated tablets with software and wireless microphone. The users can customise various features of subtitle display, such as font, colour, etc.

How to display live titles

- Live titles can be displayed on an additional, large screen in the lecture hall
- Alternatively, they can be displayed as subtitles, i.e as two lines of text on the main screen, below the presentation
- The most flexible option is for students to display live titles on their own devices such as laptops, tablets or smartphones (this lets them customize the look of the text on their devices)

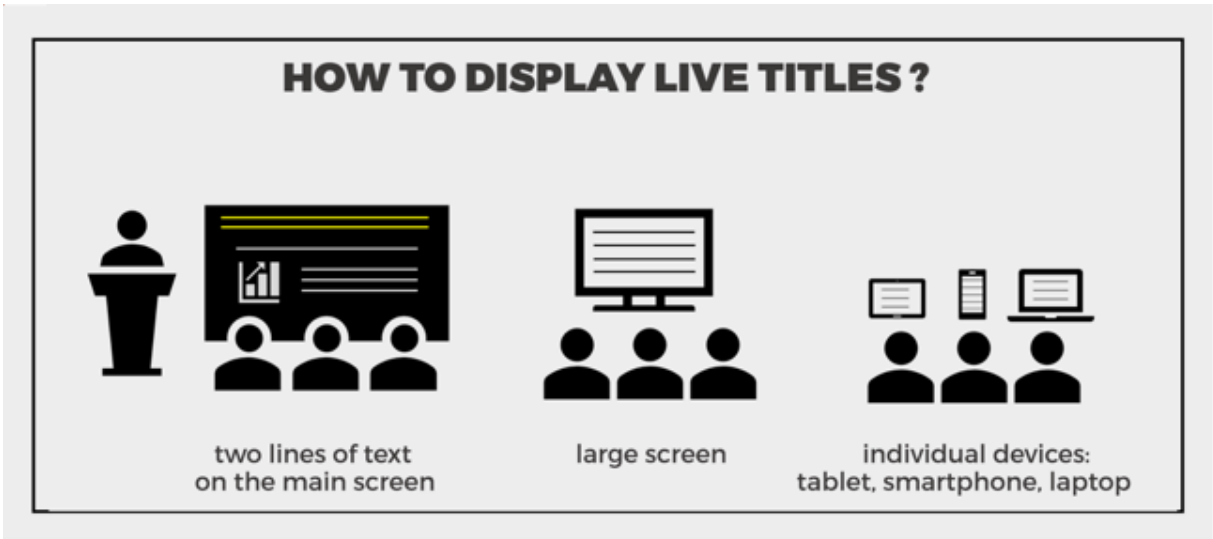
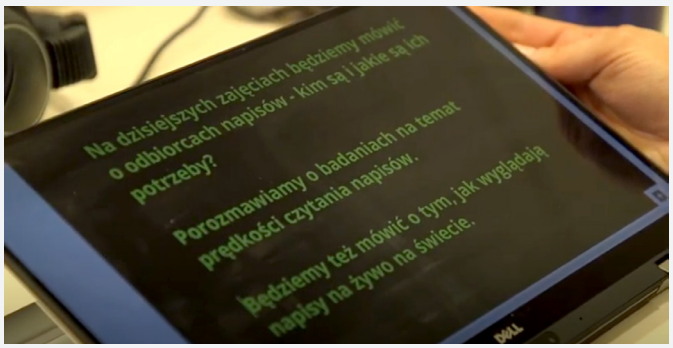


Figure 14. Different ways to display live titles



Figures 15 and 16. Live titles displayed on individual user's devices (© Dostępni.eu)

CHALLENGES AND SOLUTIONS

- **Sound**—if the respeaker works remotely, the sound can be transmitted via a telephone, a messaging app or a video conferencing solution;
- **Display**—titles can be displayed on a website which a user will open on their mobile device, or on a screen in the lecture hall (either the same screen that the lecturer uses or a separate screen);
- **Access to internet**—if there is no WiFi in the venue or the internet connection is slow or overloaded; use telephone connection to send audio. Students can use mobile internet.

DOs and DON'Ts	
DOs	**DON'Ts**
Let the users customize the look of the titles by choosing the font, the size and colour of the text and the colour of the background	Do not use serif fonts to display titles
If users cannot customize the look of titles, make sure the subtitles are displayed in high contrast, left-aligned and the text is big enough so that it is legible	Do not justify text or use italics (as it makes it less legible)
If you have a deaf student, make sure that they are happy with live titles. Some deaf people will prefer a sign language interpreter	Do not make a single respeaker work for the whole day. Like interpreters, for longer events respeakers should take turns/breaks

CHECKLIST FOR THE ACCESSIBILITY MANAGER

The checklist is addressed at the accessibility managers and/or other decision makers at higher education institutions.

- ☐ Find out what are the needs of your users:
  - –Do your deaf students prefer live titles or sign language interpreting?
  - –Do foreign students and visitors need a translation of the lecture into their mother tongue or into English?
- ☐ Contact your IT department to check if the Internet connection is sufficient to provide the service.
- ☐ Choose whether speech-to-text interpreting will be done remotely or on-site.
- ☐ Choose the language of live titles to be provided.
- ☐ Contact the lecturer to discuss the provision of the speech-to-text interpreting service.
- ☐ Ask the lecturer to provide preparatory materials.
- ☐ Share the information about the introduction of the service with the whole community at your institution.
- ☐ Make sure the information about the service is continuously accessible to all students (for instance on the website).
- ☐ Provide a mechanism for quality assurance and feedback from users.
- ☐ Decide if the transcript of the classes will be made available and to whom.
- ☐ If you make a recording of the lecture available later, make sure the subtitles are available with it.

CHECKLIST FOR THE LECTURER

There are a few things that you can do as a speaker to improve the experience for the students who are accessing your talk thanks to speech-to-text interpreting:

- ☐ Make reference materials available to respeakers before the lecture, in particular the presentation, script or notes. (So that the respeakers can get prepared before the class and familiarise themselves with the vocabulary.)
- ☐ Make sure you have not prepared too many slides for your lecture, which might mean that you would have to rush through them.
- ☐ During the presentation, avoid changing slides too quickly as the students will be reading live titles alongside the visual content you are talking about and live titles appear with some delay. Allow extra time for the students to read titles and and then look at the graphics or written content you are sharing with them.
- ☐ Use the microphone throughout the lecture. (Respeakers rely on clear audio.)
- ☐ Make sure that others who speak during the lecture also use the microphone and remind them to do so as needed.
- ☐ If there are comments or questions from audience members who do not have a microphone, repeat them to the microphone. (As respeakers are only able to respeak what they can hear.)
- ☐ Do not speak too fast. (As this makes it more difficult for respeakers as well as all students to understand your lecture)

CHECKLIST FOR THE RESPEAKER

- [ ] Contact the accessibility manager to discuss the workflow.
- [ ] Obtain reference materials in advance.
- [ ] Familiarise yourself with specialised vocabulary.
- [ ] Update the speech recognition software's lexicon so that you are sure that the main terms and names are included in the lexicon and will be recognized correctly.
- [ ] Check the Internet connection.
- [ ] Check if you can hear the speaker and adjust the volume of the sound.
- [ ] At the beginning of the class, remind all participants to speak to the microphone.
- [ ] Stop your work and intervene or ask others to intervene on your behalf if someone is not speaking to the microphone or the audio quality deteriorates significantly.
- [ ] During the class, verify if students can see the titles.
- [ ] After the class, save all the text as a transcript in case it is needed later.

QUALITY IN SPEECH-TO-TEXT INTERPRETING

Although there are different ways to assess the quality of live (sub)titles produced by speech-to-text interpreting, the most common method used is the NER model (Romero-Fresco & Martínez, 2015), which makes a distinction between recognition errors (caused by the interaction between the respeaker and the software) and edition errors (caused by the respeaker's incorrect decisions when omitting or changing information). These errors can in turn be minor, standard or serious depending on how they impact on viewers' comprehension. The NER model is currently being used by governmental regulators, broadcasters and universities in countries such as Spain, the UK, Belgium, Poland, Switzerland, South Africa, Australia, the US and Canada, where it has been included within the national accessibility legislation to assess live subtitling quality. More information about the NER model can be found on the website of the Galician Observatory for Media Accessibility as well as in the ILSA course.

The NER model is used for the assessment of intralingual live subtitles (those produced in the same language as the original audio). The assessment of interlingual live subtitles (those translating the original audio into another language) may be done with the NTR model (Romero-Fresco & Pöchhacker, 2017).

MORE INFORMATION

On educational benefits of subtitling

De Meulder, A., Robert, I. S., Schrijver, I. (in preparation). Live subtitling for access to education: A pilot study of university students' reception of intralingual live subtitling.

Gernsbacher, M. A. (2015). Video Captions Benefit Everyone. *Policy Insights from the Behavioral and Brain Sciences*, 2(1), 195-202. doi:10.1177/2372732215602130

Kruger, J.-L., Hefer, E., & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. L2. *Journal of Eye Movement Research*, 7(5), 1-15.

Jelinek Lewis, M., & Jackson, D. W. (2001). Television Literacy: Comprehension of Program Content Using Closed Captions for the Deaf. *Journal Of Deaf Studies And Deaf Education*, 6(1), 43-53. doi:10.1093/deafed/6.1.43

Linebarger, D., Piotrowski, J. T., & Greenwood, C. R. (2010). On-screen print: the role of captions as a supplemental literacy tool. *Journal of Research in Reading*, 33(2), 148-167. doi:10.1111/j.1467-9817.2009.01407.x

Vanderplank, R. (2013). 'Effects of' and 'effects with' captions: How exactly does watching a TV programme with same-language subtitles make a difference to language learners? *Language Teaching*, 49(02), 235-250. doi:10.1017/s0261444813000207

Yoon, J.-O., & Kim, M. (2011). The Effects of Captions on Deaf Students' Content Comprehension, Cognitive Load, and Motivation in Online Learning. *American Annals of the Deaf*, 156(3), 283-289. doi:10.1353/aad.2011.0026

On live subtitling through respeaking

Pöchhacker, F., &Remael, A. (2019). New efforts?: A competence-oriented task analysis of interlingual live subtitling. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 18, 130–143
https://lans-tts.uantwerpen.be/index.php/LANS-TTS/article/view/515/471

Robert, I. S., & Remael, A. (2017). Assessing quality in live interlingual subtitling: A new challenge. Linguistica Antverpiensia, New Series: Themes in Translation Studies, 16, 168-195. https://lans-tts.uantwerpen.be/index.php/LANS-TTS/article/view/454

Romero-Fresco, P. (2011). *Subtitling through speech recognition: respeaking*. Manchester: St. Jerome Publishing.

Romero-Fresco, P. (2012). Respeaking in Translator Training Curricula. The Interpreter and Translator Trainer, 6(1), 91-112. doi:10.1080/13556509.2012.10798831

Romero-Fresco, P., & Martínez, J. (2015). Accuracy Rate in Live Subtitling: The NER Model. In J. Díaz-Cintas & R. Baños Piñero (Eds.), Audiovisual Translation in a Global Context. Mapping an Ever-changing Landscape (pp. 28-50). London: Palgrave.

Romero-Fresco, P., & Pöchhacker, F. (2017). Quality assessment in interlingual live subtitling: The NTR Model. *Linguistica Antverpiensia*, *New Series – Themes in Translation Studies*, 16, 149–167.
https://lans-tts.uantwerpen.be/index.php/LANS-TTS/article/view/438

Romero-Fresco, P. (2018). Respeaking. In L. Pérez-González (Ed.), *The Routledge Handbook of Audiovisual Translation* (pp. 96-113).

Romero-Fresco, Pablo and Eugeni, C. (2020). Live subtitling through respeaking. In Łukasz Bogucki and Mikołaj Deckert (Eds.), *Handbook of Audiovisual Translation and Media Accessibility*, Palgrave.

ILSA Interlingual Live Subtitling for Access

This project has been funded with support from the European Commission.

This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

universität wien · Universidad de Vigo · vrt · Universiteit Antwerpen · dostępni.eu · University of Warsaw

With the support of the Erasmus+ Programme of the European Union