

ISBN: 978-84-1091-017-1 (PDF)

DOI: <https://doi.org/10.14201/0AQ0373337347>

Machine Interpreting to Achieve Social and Economic SDGs? A Use Case for Refugee Communication in Switzerland

¿Interpretación automática para alcanzar los ODS sociales y económicos? Un caso práctico en la comunicación con personas refugiadas en Suiza

Alice DELORME BENITES

ZHAW Zurich University of Applied Sciences
alice.delormebenites@zhaw.ch

Romina SCHAUB-TORSELLO

ZHAW Zurich University of Applied Sciences
romina.schaub-torsello@zhaw.ch

Daniela SAGER

ZHAW Zurich University of Applied Sciences
daniela.sager@zhaw.ch

Mattia TURRA

ZHAW Zurich University of Applied Sciences
mattia.turra@zhaw.ch

ABSTRACT: This project explores how machine interpreting apps can aid communication between employees of institutions and refugees in Switzerland. With around 15,000 refugees arriving annually to the country, language barriers pose significant challenges. Interpreting services are costly, and volunteers for less common languages are scarce. The study assesses the potential and problems of conducting a full conversation between a migrant and a response worker using only Microsoft Translator. We conducted semi-guided interviews to map the needs and expectations of institution employees and used case simulations to evaluate the practicability and the impact of machine interpreting on the conversation. Despite accuracy issues, migrants felt empowered and had a better

conversation control. Yet nonverbal aspects hint towards new emerging turn-taking practices for machine interpreted conversations. This latter point should be further investigated.

KEYWORDS: machine interpreting; refugee communication; community interpreting.

RESUMEN: Este proyecto explora cómo las aplicaciones de interpretación automática pueden facilitar la comunicación entre personal institucional y personas refugiadas en Suiza. Con la llegada anual de unas 15 000 personas, las barreras lingüísticas plantean importantes retos. Los servicios de interpretación son costosos y escasean personas voluntarias que trabajen con lenguas minoritarias. El estudio evalúa el potencial y los problemas de mantener una conversación entre una persona migrante y el personal de respuesta utilizando Microsoft Translator. Se realizaron entrevistas para determinar necesidades y expectativas del personal institucional, y simulaciones para evaluar la viabilidad y los efectos de la interpretación automática en la conversación. Pese a que se detectaron problemas de precisión, las personas migrantes se sintieron empoderadas y tenían mejor control de la conversación. Sin embargo, aspectos no verbales apuntan al advenimiento de nuevas prácticas de toma de turnos en conversaciones con interpretación automática, lo cual debería investigarse más a fondo.

PALABRAS CLAVE: interpretación automática; comunicación con personas refugiadas; interpretación comunitaria.

1. INTRODUCTION

Europe has faced a surge in migration crises in recent years. By the end of 2022, 108.4 million people worldwide were displaced due to various factors, including war, violence, human right violations, and natural disasters (UNHCR 2023). Climate change is anticipated to further exacerbate the situation in future.

The arrival of many foreign-speaking refugees in a country immediately raises communication challenges. Language barriers persist in reception structures, such as asylum centers, healthcare institutions, government offices, courts, and police stations. Interpreters play a crucial role in bridging these gaps by providing foreign speakers with access to public services through interpretation. However, the demand for interpreting services has grown by an average of 25 % in the last three years (Interpret 2022) and has outpaced the availability of trained interpreters in Switzerland. The rising demand and the lack of financial resources to guarantee professional interpreting have led to various strategies to address language barriers, including translation apps, improvised gestures («hands and feet»), family members or co-workers acting as lay interpreters, and even picture cards. Decision-makers often choose from these options based on their knowledge and available (financial and organizational) resources. However, uninformed use of technology, such as relying solely on machine translation apps, can have serious consequences, as evidenced by cases of incorrect asylum decisions in the United States, where asylum hearings were conducted using Google Translate (Bhuiyan 2023). To navigate this complex landscape and avoid an uninformed and, therefore, harmful use of technology, organizations need clear guidelines that consider communication processes (often oral conversations), and raise awareness of the various approaches, their strengths,

and potential risks. As of today, a very limited number of studies tackle this question (e.g., Hudelson and Chappuis 2024; Kapoor et al. 2022; Viera et al. 2022).

This exploratory study aims to provide an overarching view of machine-translation assisted conversations. First, we mapped out the challenges that conversations with displaced persons pose to government employees with and without human interpreters. Second, we simulated realistic use cases of machine-translation assisted conversations to explore the impact of computer-assistance on the communication situation.

2. CHALLENGES AND OPPORTUNITIES OF INTERPRETATION IN WORKING WITH REFUGEES

In November 2022, we conducted four semi-guided interviews with non-linguists whose day-to-day work involves communicating with refugees: a job coach in charge of helping refugees who have been granted asylum to find work or a vocational training, a team manager at the migration office, a school social worker and a midwife in a hospital. They were asked to share their thoughts and experiences on the following topics:

- a. use of translation technologies;
- b. cooperation with human interpreters;
- c. satisfaction and limits (both for human and machine interpreters);
- d. measurability of success regarding the conversation;
- e. wishes and fears regarding machine translation technologies.

The answers of the interviewees helped us summarize their perception of the challenges and opportunities related to working with human interpreters and with machine translation devices.

2.1. *Considerations regarding Human Interpreters*

In general, the interviewees expressed high satisfaction with human interpreters, both remote and on-site. Yet booking an interpreter in advance is often impossible, particularly in the healthcare sector, where night services (e.g. childbirth) frequently operate without interpreters, as there is no night interpreting service, not even by telephone. Moreover, the respondents reported last-minute cancellations by booked interpreters, so that even carefully planned meetings often had to be postponed or cancelled. Overall, scheduling important conversations is much more complex and uncertain when a human interpreter is required. Further, while the interviewees acknowledged the advantage of having a human interpreter who shares a common cultural background with the refugee, cultural differences between interpreters and refugees also pose challenges, as factors such as age and gender have varying underlying implications in different cultures (e.g. an older man interpreting for a younger woman in a gynecological consult). In most cases, the interpreters are well received, but their personal background remains a variable that can add to the complexity of the conversation. Trust is a critical issue; some migrants fear that interpreters may share information with their country of origin (due to espionage). Recent reports (Rohner 2023) show that the threat

of spying interpreters is real. Institutions need to exercise caution and avoid blind trust. However, the longer the refugee works with an institution, the more trust they build.

2.2. *Considerations regarding the Use of Machine Translation Tools*

When asked about digital technologies, the respondents pointed out that migrants themselves are introducing these technologies. Particularly, school workers report that children increasingly use smartphone apps, especially Google translate and DeepL, as a translation aid for school-related conversations. This new tendency aligns with the recent arrival of Ukrainian asylum seekers, amongst which many children speak the same language. When asked about an ideal machine translation solution, the respondents emphasized the importance of a voice function during conversations, because it would make the setting comparable to telephone interpreting, with which they have good experiences. They would welcome a speech recognition feature for long conversations, to avoid time-consuming manual transcription for the app to translate it. In addition, the reading skills of female refugees are not always very well developed, so that voice output would be necessary as well. However, interviewees expressed concerns about full machine-translated conversations, fearing a loss of interpersonal relationship. They feel like the interhuman connection would be lost, and therefore prefer to rely on human interpreters whenever possible. Yet in certain situations, refugees appreciate being able to communicate more independently, without relying on another person, especially in the case of female refugees: technology could allow workers to speak to them directly without first asking the person accompanying them, who may have a better knowledge of German, but who also often ends up speaking and deciding for them. Finally, the respondents consider data protection as an important factor, as they work with a target group that is very careful to choose the right words and the right thing to say. What they say and how they say it could have serious consequences for them, should this information reach their country of origin. The use of translation applications connected to the Internet presents an additional risk in this context.

It is interesting to note that the considerations and questions uttered by the interviewees hardly concern the purely linguistic aspects of interpreting. For example, respondents did not mention the quality of the translation. On the other hand, there was a clear perception of the issues involved in tripartite communication. We therefore adopted the users' point of view for the second part of our study, focusing our observations on the communication process and the course of the conversation rather than on the quality of the translation.

3. USE CASE SIMULATIONS

3.1. *Objectives and Set-up*

In order to gain a better understanding of how communication takes place between specialist professionals and refugees when an oral translation application is used, we

carried out four different simulations. Each simulation scenario was placed in a different context, where a specific conversation between two participants was staged:

- a. A school social worker informing a refugee parent about a school trip in which their child will be taking part.
- b. A social worker establishing the employability of a refugee interested in starting to work.
- c. A hospital employee conducting a first triaging conversation with a refugee after a fall from a bike.
- d. A social worker advising a refugee who wishes to enroll in a German course.

We deliberately excluded scenarios where the refugees just arrived (e.g. reception at an asylum center, health check or administrative meetings upon arrival). These conversations are usually conducted with the help of a human interpreter and have a high potential for emotional distress, which would be ethically questionable to simulate.

The participants were recruited on voluntary basis. The Swiss participants were four professionals playing their own part. Four Ukrainian refugees were recruited to play the client part and they all had limited German proficiency. Compensation was offered to all participants, although some social workers could not accept it due to employment restrictions.

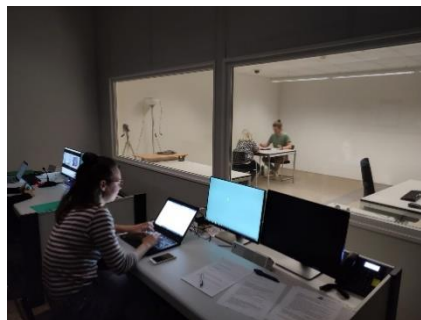


Figure 1. Set-up for the use case simulations

The conversations were prepared in advance with the help of the respective social workers to ensure that they would be as realistic as possible. All participants were instructed on the (fictional) context, their part, and the goal of the conversation beforehand. Further, they were made aware of cameras and microphones recording the interactions and were instructed to use real information (e.g. number of children) while maintaining fictional names and other personal details if necessary. The instructions were given in German and translated by a professional interpreter into Ukrainian.

The participants sat at a table facing each other, with a device (tablet or smartphone) placed on the table between them (see Figure 1). The translation app chosen for the simulations was Microsoft Translator, which featured a dialogue interface allowing simultaneous use by two people (see

Figure 2). The interface had opposing screens, one for each participant. Each screen featured a button to press before speaking. After pressing the button, the person spoke in

their language and the device produced a written transcription, as well as a written translation on the other participant's screen, and also, after the person had finished speaking, an audio output of the translation.

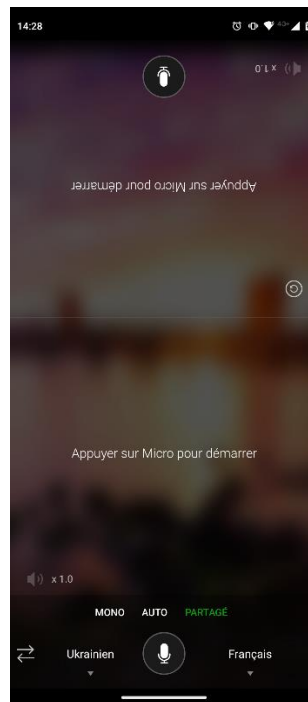


Figure 2. Microsoft Translator shared user interface (Microsoft 2023)

The researchers and the professional interpreters observed the simulation from an adjacent observation room separated by a one-way mirror. The length of the conversations was set to 20 minutes. At the end of each conversation, we conducted a post-task interview with both participants. The Ukrainian participants were assisted by an interpreter.

All videorecorded conversations were coded using MaxQDA (VERBI Software, 2021) with two parallel coders for each video. The coding was qualitative and semi-explorative, focusing on «abnormalities» or noticeable behavior on the following aspects: handling of the technology, verbal communication, paraverbal communication, nonverbal communication, turn-taking and overall behavior. The coded videos were then discussed in a workshop to identify the most salient and significant observations made throughout all videos.

3.2. Results

3.2.1. OVERALL SATISFACTION

The post-task interviews showed a high overall satisfaction with the conversation and the translation tool, especially amongst the Ukrainian participants, who were pleasantly surprised by its high quality. Most participants perceived the conversation as very good, with no significant problems. However, the present professional interpreters identified several translation errors when sitting in the observation room during the

simulations, e.g. interruptions, omissions, gender and pronoun errors. Interestingly, the participants did not mention these errors in the post-task interview, either due to lack of notice or immediate recall. Further, both a Ukrainian and a Swiss participant indicated that the translation and voice output conveyed a more neutral, less emotional tone. Both felt positively about it.

3.2.2. PROBLEMS AND SHORTCOMINGS OF THE SET-UP

While the participants were generally satisfied with the tool, the research team noticed a range of recurring or particularly salient issues. First, technical issues frequently occurred during the simulations, such as the app failing to record or translate utterances. These blanks lacked clear explanations, making prevention nearly impossible.

Further, the app prematurely translated and produced vocal output before the participants completed their sentence, especially during slower speech or pauses. As a result, only part of the utterance was translated. Longer utterances were more frequently interrupted by the app's vocal output, especially when the participants spoke slowly. Micro-pauses, such as hesitation, searching for words, brief thinking pause also triggered the automatic translation. Consequently, turn-taking was negatively affected or at least influenced by the device. Therefore, phatic elements and turn-keeping strategies should be handled differently in device-translated conversations. When confronted with these interruptions, participants adapted by speaking faster, splitting their utterances into smaller units, or simply repeating their utterance without pauses. Accordingly, this technical issue had a major impact on the conversation flow. Surprisingly, during the post-task interview almost no participant (and particularly none of the refugees) perceived said interruptions as a problem, likely due to participants developing coping strategies rapidly.

Second, most of the participants read along the transcript of their own utterance while (or even before) hearing the translation output. Reading the transcript or the translation on a smartphone (split) screen may be challenging, especially for elderly individuals.

Third, in the hospital use case, the device's positioning (on a table between two sitting participants facing each other) was inadequate and unrealistic, as most conversations usually occur during physical examinations. This shows the need for further investigation into a more ergonomic hardware solution.

Finally, there were a range of recurring errors that seemed to result from the combination of speech recognition and automatic translation. For example, the German formal pronoun «Sie» being a homonym to the neutral pronoun «sie» (plural, «they»), the tool was not always able to differentiate between a formal address («Haben Sie Kinder?» or «Do you have children?») and a plural form («Haben sie Kinder?» or «Do they have children?»). This led to misunderstandings that had to be resolved during the conversation. Abbreviations, technical terms and realia do not work well either in speech recognition or in machine translation, resulting in non-sensical translations. There is still a strong gender bias, so that e.g. female participants were translated in the masculine form

when talking about themselves. Furthermore, many questions were translated into assertive statements, which directly impacted the turn management.

3.2.3. CONCENTRATION OF THE PARTICIPANTS

At the beginning, the participants (especially the social workers) often spoke explicitly, clearly and slowly into the device. They avoided abbreviations and specialized terminology. Towards the end of the conversation, however, they spoke faster and with less care, using more abbreviations and specialized terms. Moreover, we could observe some signs of exhaustion such as sighs, signs of effort, and various slips of the tongue. This leads to the question whether the mental load required to conduct the conversation using a machine translation device is particularly high and the concentration wanes after a while—in this case, the allocated 20 minutes seem to be already too long. On the other side, this could simply be the result of the gain of experience throughout the conversation. Since the quality of the translation is more at risk with the increased use of abbreviations and specialized terms, and with the incorrect or inaccurate pronunciation, the duration of the conversation and the level of concentration of the participants seem to be important factors for a successful scenario.

3.2.4. EYE CONTACT

At the beginning of the conversations, participants made very little eye contact, which was unexpected for non-mediated interactions. They greeted each other and asked questions while primarily looking at the translation device. This could be encouraged by the fact that the translation apps transcribe the vocal input and simultaneously produce a written translation before the vocal output. This might have triggered the participants to focus on reading the text that kept appearing on their part of the split screen. Although eye contact increased during the conversation, it remained low. Experts feared this decline in human connection, but it is possible that refugees used the device strategically to avoid vulnerability. Swiss participants, in their professional role, initiated more eye contact. Additionally, participants often unnecessarily leaned toward the device (smartphone), possibly due to its shape and size. One might wonder if this strong tendency is linked to the lack of eye contact: maybe the participants were so absorbed by their interaction with the device that eye contact was neglected, or they possibly used the interaction with the device as a strategy to avoid eye contact altogether. Investigating other device forms could provide insights on this matter.

3.2.5. NONVERBAL COMMUNICATION

The recordings revealed numerous nonverbal communication elements. Participants used gestures to regulate emotions, express frustration (when the device malfunctioned), and encourage each other. Frequent smiles reinforced mutual understanding and explicit nodding, a strategy for active listening, signaled comprehension. Unlike smiling, nodding did not always require eye contact. This can partly be explained by the fact that the participants were often looking at the device to read the translation produced in real time by the app (even before the voice output started),

and then nodded while reading to signal that they had understood the utterance. In that context, nodding seemed to be a major element of the conversation regulation.

Another interesting nonverbal aspect was the various gestures used for turn-taking. In general, the participants did not start talking as long as the device was producing a voice output—with almost no overlap or interruption. Most of the turn-taking was paced by the device, with each participant making their complete utterance, waiting for the translation and then listening to the other party's verbal reaction. This forced pauses in one's speech, where both parties waited for the translation. In general, these technology-induced pauses seemed to be unnatural interruptions in the flow of the conversation that had to be endured in some way. The participants also used finger gestures (thumbs up, hand up, or forefinger up) or the English word «sorry» to negotiate turns. It appears that fingers played a special role in the negotiation of turns: as soon as the participant understood what the other person had said (either through reading or listening to the vocal output), they hovered their finger over the device, signaling their intent to take the turn. Sometimes, both participants would place their finger over the device at the same time, and the turn-taking was then regulated through finger gestures. This new, technology-adapted strategy allowed participants to control the conversation flow to a certain extent.

3.2.6. CONVERSATION AND ROLE MANAGEMENT

All participants showed explicit interest and engagement in the conversation. They all reacted when problems occurred, using different strategies to maintain the conversation flow. Gaining physical control over the translation device (via finger hovering or adjusting speaking pace) emerged as a crucial factor. Interestingly, during the post-task interview, a Swiss participant expressed the need to clarify roles early on when using a translating device. More concretely, they would rather explicitly state that they will ask the questions while the participant will simply answer them. This reaction indicated that the Swiss participant was surprised by the fact that the refugee participant also asked questions and initiated turns. This was possible due to the split interface of the translation app that made the translating button equally accessible to both parties.

4. DISCUSSION

The use case simulations showed how the translation through a device impacted the conversation and, to some extent, the relation between the participants. First, the role of nonverbal communication changed, with elements from active listening (nodding) becoming crucial to the conversation, while others (eye contact) were neglected or avoided. The haptic dimension was enhanced by the necessity to tap a shared device before speaking. The physical affordances of the device seem to play a major role in the emerging nonverbal practices.

The haptic turn-taking rituals also challenge the traditional hierarchy of worker-refugee conversations. By taking control over the device, the refugee can negotiate more speaking turns and is no longer limited to the question-answering role. Along with the (cautious) hypothesis that the low eye-contact frequency might be a protection strategy

for the refugees, this could be the sign that a targeted use of language technology can empower vulnerable populations such as refugees. This, of course, should be investigated more in-depth before stating any general claim.

The simulations also showed a need for more conversation-oriented solutions. The numerous interruptions by the machine showed that neither the diverse speaking paces nor the varying lengths of utterances are accounted for by the current systems. Moreover, the written translation produced during the speech recognition phase (and usually before the vocal output starts) changes with each word spoken, creating a constantly reshaped text that is difficult to read in a linear fashion. Since we observed that most participants read this written translation, there is a clear need for a more reader-friendly solution.

Regarding the current quality of the tools, it seems that it does not really affect the conversation. The potential of errors and the risk of technical issues are still quite high, but the participants were able to overcome most problems. Since Large Language Models are improving quickly, we expect the quality of such apps to improve significantly in the next years—making device-translated conversations even safer.

Nevertheless, a certain degree of adaptation is required from the participants, regardless of the translation quality. As some interviewees anticipated, the interhuman relation changes when using such a device. Fortunately, the use case simulations pointed towards a more complex evolution than a simple loss of interhuman bonds. At this point, we can only assume that the use of oral translation devices opens new avenues for both improvements and problems. However, the empowerment that the refugees have gained through the possibility of tapping for translation is certainly a welcome discovery to be further investigated.

REFERENCES

- Bhuiyan, Johana. September 7, 2023. «Lost in AI Translation: Growing Reliance on Language Apps Jeopardizes Some Asylum Applications». *The Guardian*. Accessed June 11, 2024. https://www.theguardian.com/us-news/2023/sep/07/asylum-seekers-ai-translation-apps?CMP=share_btn_tw.
- Hudelson, Patricia, and François Chappuis. 2024. «Using Voice-to-Voice Machine Translation to Overcome Language Barriers in Clinical Communication: An Exploratory Study». *Journal of General Internal Medicine* 39: 1095-102.
- Interpret. 2022. «Einsatzstatistiken zum interkulturellen Dolmetschen und Vermitteln». Accessed June 11, 2024. <https://www.inter-pret.ch/de/service/statistiken-77.html>.
- Kapoor, Ravish, Corrales, German, Flores, Manuel P., Feng, Liyun, and Juan P. Cata. 2022. «Use of Neural Machine Translation Software for Patients with Limited English Proficiency to Assess Postoperative Pain and Nausea». *JAMA Network Open* 5 (3): e221485. Accessed June 11, 2024. <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2789715>.
- Microsoft. 2023. *Microsoft Translator*. V.3.0. Microsoft. Mobile app.
- Rohner, Raphael. August 26, 2023. «Spione aus Eritrea arbeiten für Schweizer Migrationsbehörde – Bund leitet Untersuchung ein». *Aargauer Zeitung*. Accessed

August 7, 2024. <https://www.aargauerzeitung.ch/schweiz/fluechtlingspolitik-verdacht-auf-spionage-bei-dolmetschern-aus-eritrea-bundesrat-gesteht-fehler-ein-id.2501694?reduced=true>.

United Nations High Commissioner for Refugees (UNHCR). 2023. «Global Trends - Forced Displacement in 2022». Accessed June 11, 2024. <https://www.unhcr.org/global-trends-report-2022>.

VERBI Software. 2021. *MAXQDA*. VERBI Software. Software.

Vieira, Lucas N., Minako O'Hagan, and Carol O'Sullivan. 2021. «Understanding the Societal Impacts of Machine Translation: A Critical Review of the Literature on Medical and Legal Use Cases». *Information, Communication & Society* 24 (11): 1515-32.

