Strategies to strengthen European linguistic capital in a globalised world

MultiLX Training Meeting 2

MultiLX

MultiLX
Multilingual Europe

UNIVERSITY OF WARWICK

UOC Universitat Oberta de Catalunya

universität freiburg

Oilthigh Ghlaschu

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

UNIVERSITY OF LIMERICK
OLLSCOIL LUIMNIGH

PH LUZERN
PÄDAGOGISCHE HOCHSCHULE

UiO : Universitetet i Oslo

UNIVERSITÀ DI TORINO

MultiLX Training Meeting 2

MultiLX 2025

A report on the second training meeting of Horizon Europe project, 'Strategies to strengthen European linguistic capital in a globalised world (MultiLX)' (101177191)

Please reference as:
MultiLX 2025. *MultiLX Training Meeting 2.* Hosted virtually by University of Limerick and University of Jyväskylä.

Executive Summary

The second MultiLX Training Meeting was held virtually on 26 September 2025 and was jointly hosted by the University of Limerick and the University of Jyväskylä. It centred on the theme "Doing multilingual digital research / Doing research multilingually: Challenges, opportunities, principles and practices." The meeting, organised by Work Package 4 (WP4), addressed both practical and theoretical dimensions of multilingual research in a digital and AI-mediated context, with particular focus on developing and localising research tools for studying young people's engagement with artificial intelligence.

The programme comprised three core sessions. The first explored ethical and methodological challenges associated with digital multilingual research, especially the implications of using Large Language Models (LLMs). Issues such as linguistic bias, inclusivity for low-resource languages, privacy, environmental impact, and the balance between human agency and automation were discussed. Further discussions emphasised the need for conscious and ethical use of AI tools, particularly when working with minority language communities, and for ensuring equitable access to technological resources across languages.

The second session focused on establishing a language policy for WP4 research tools. The group considered how to reconcile an inclusive language ideology with the practical constraints of digital platforms and questionnaire-based data collection. Participants discussed the criteria for selecting languages, the challenges and possibilities of representing mixed or translanguaging practices, and the uneven distribution of linguistic labour within multilingual teams. The outcome reaffirmed a flexible, context-sensitive approach to language policy, underpinned by ongoing reflection and transparent communication.

The final session concentrated on the localisation of WP4 research instruments using digital translation tools and LLMs. Breakout groups reviewed multiple language versions of the project's questionnaire, assessing translation quality, cultural appropriateness, and tone. The comparative exercise underscored the limitations of automated translation – particularly for smaller or less-standardised languages – and highlighted the essential role of human expertise in ensuring linguistic and cultural fidelity.

Overall, the meeting advanced shared understanding of how digital technologies intersect with multilingual research practices. It reinforced MultiLX's commitment to ethical, inclusive, and reflective methodologies, and provided concrete strategies for supporting multilingualism in a technologically complex research environment.

Table of Contents

1. Introduction

The second MultiLX Training Meeting was hosted by Work Package 4 with the theme of 'Doing multilingual digital research / Doing research multilingually: Challenges, opportunities, principles & practices.' The aim of the workshop was twofold: on the one hand to focus on decisions around Work Package 4, one of the cross-cutting WPs in the project, which looks at young people's usage of and experience with AI tools; on the other hand, we wanted to explore these issues more widely across the project.

The meeting was held entirely online on 26th September and hosted jointly by University of Limerick and Jyväskylä, who co-lead WP4. The day consisted of three sessions. The first focussed on ethical and methodological considerations; the second aimed to develop a language policy for the localization of the research instruments in WP4; and the final session focussed on the actual localization with a more hands-on session and a report on the localization and the challenges and opportunities involved.

The training meeting resulted in a number of important decisions in relation to our policies and practices about the use of AI tools in our research, supporting and resourcing multilingual research necessary for WP4 and more widely across the project, and also more widely on using technology and other tools / practices to support our own multilingual working.

2. Ethical and Methodological Challenges and Opportunities

This first session was led by Helen Kelly-Holmes and Sari Pietikäinen and opened by setting the agenda for the day in terms of issues that we wanted to explore as a group, namely:

- What are the differences between / are there differences between – particularly in the current era of Large Language Models?
  - Doing multilingual research digitally
  - Doing research on multilingual digital data
  - Researching digital multilingualism
- What is the experience of discussing our project and being online for the training workshop in terms of actually learning, sharing and discussing our research in a digital space?
- Thinking about how we work with these different types of data and how we use technology to work digitally

- What are the challenges and opportunities?
- And, finally, what principles and practices might emerge, keeping in mind our commitments in relation to AI in the original application?

As there are varying levels of expertise in and experience of using LLMs across the project, we started with a basic introduction to what these technologies are and what they can and cannot do. In terms of the challenges and opportunities, we discussed the cultural and linguistic biases in LLMs and the recognised WEIRDness of AI systems. For minority language speakers, there is always a dilemma in terms of usage of such systems in terms of balancing the need not to be 'left behind' by such systems versus the inevitable work involved for speakers of languages with lower levels of digital resources who want to be included in these systems. Also of relevance are privacy and security issues and the impact that our interactions with digital technology, in the latest iteration with LLMs, which 'learn' through human feedback and reinforcement have on the type of linguistic and cultural make-up of these systems. One issue we also wanted to bring to the fore of our minds when carrying out research using AI systems is that of the environment. While there is a tendency to reach immediately for these easy and quick tools to carry out research and other project tasks for us, they are much more environmentally costly to use than other specific tools. An interesting finding from our own work in WP4 with preparing draft versions of the questionnaire in different languages using an AI tool was that the production of the versions for lower resource minority languages took more time and many more iterations than those for bigger resource languages.

We also emphasized AI as a complex assemblage of overlapping systems, processes, and innovations with significant environmental, social, and economic implications. Environmentally, AI entails a large footprint through energy use, carbon emissions, and mineral extraction, but can also offer innovations for more sustainable governance. Socially, it may reconfigure needed skills and connections, often reinforcing inequalities, while also enabling reskilling and new forms of knowledge. Economically, AI can be used to accelerate certain type of decision-making and predictions but this raises concerns about dependency on a few powerful actors and related security risks. Importantly, AI does not work the same for everyone, everywhere, or at all times. Viewed through a posthumanist lens, it marks a shift towards decentering the human and reframing questions of power, knowledge, governance, ethics, and creativity in an evolving human–machine relationship.

Having outlined these issues, participants were then divided into breakout groups to discuss the following questions:

- (How) do we want to use / not use AI in our research practices in sites and Work Packages?

- (How) do we want to use AI in our work across Work Packagess (e.g. collaboration, dissemination, management)?
- What should we stop / keep doing / start doing?

Conclusions from these break-out sessions and the plenary discussion can be summarized as follows:

- AI tools can offer many opportunities for lesser-used / small languages, for example, by creating visibility and a space for 'stateless' languages to be seen
- AI tools could help us to create different modes for dissemination of results which could help us to reach a wider and more diverse audience.
- Ethical considerations seem to be particularly relevant when working with minority language speakers in relation to issues of privacy and not exploiting speakers in terms of providing free resources. We also need to respect research participants' wishes in relation to using or not wanting to use AI tools.
- We should try to facilitate interactions between speakers of different languages without needing a lingua franca.
- We should be aiming to choose tools that offer better protections in terms of privacy for research participants.
- Many tools, e.g. automatic transcription, offer time-saving tools for researching language. However, many of the larger platforms are not as effective or accurate as smaller more niche applications. We should aim to use the latter as much as possible in our work.
- Keeping mindful of what we and our research participants are gaining and losing
- Aim for conscious use and non-use despite ambience of the technology
- Create a space where we can share tips, particularly for researching small languages

3. Language Policy for Work Package 4 Research Tools: Researching Multilingual Digital Practices

Work Package 4 is one of the cross-cutting work packages in MultiLX. In other words, it is not site-based and is aiming to collect data on young minority language speakers' usage of and experience with AI tools from across the different research sites in Work Package 3, as well as across Europe generally. For that reason, there is a need for a language policy in relation to the research tools for this work package. An English language version of the questionnaire had already been created and was being

distributed. The questionnaire was primarily designed for and tested on non-native speakers of English, since these would be vast majority of participants. Following this, we wanted to create a number of versions in the main languages of Europe as well as in the project's languages. This represents a challenge, since the number of languages is potentially limitless, and so we wanted to make some decisions about the languages during this part of the workshop.

While we have a very open and inclusive language ideological stance in our Language Policy document, the political economy of the technology that we are using the create and distribute the questionnaire dictates that we select particular languages for participants to use. The questionnaire also clashes with the very localized – linguistically and culturally – approach of the ethnographies in Work Package 3 across a range of sites, and the need to recognise plurilingual and translanguaging practices which are the reality for our research participants. We also reminded ourselves of our commitment in the project application that "the questionnaire for WP4 will be translated into relevant languages to ensure accessibility for participants in different regions". Following an outline of these issues, team members were once again divided into break-out groups to discuss the following questions:

- What are our final decisions around LP for particular sites in relation to translating the questionnaire into 'relevant languages' for participants?
- How will we treat mixed practices in a crude instrument like a questionnaire?
- How will we deal with responses in different languages and mixed language practices?
- What about environmental and labour dimensions?
- Can we find compromises?

Conclusions from the break-out sessions and the plenary discussion that followed are summarised here:

- Deciding on the 'relevant' languages of the questionnaire requires a balancing act and reveals a tension between different types of inclusion; issues of literacy in very small languages; different types of speakers; and needing but not wanting to draw a line and make firm decisions.
- Decisions we make about languages could impact on maintaining trust that has been built up in a particular language with our participants and this needs to be weighed against the practicality of these decisions.
- The obstacles to decoding and translating multilingual data need to be kept in mind, particularly working with data in languages we don't speak
- It is important to recognise the uneven burden on certain languages and speakers across the project and to compensate for this in whatever other ways we can.

- The Language Policy for the project is not designed to be a one-size fits all. We should encourage decision-making on LP in small group working and try to facilitate multilingual conversations and different lingua francas

- Although the de facto working language of the project is English, this is of course English as a Lingua Franca. First language English speakers need be mindful of this and adapt appropriate practices in their communication.

- The project has made a commitment in its application to "put in place a language policy to mitigate difficulties [..] of translations and interpretation in working across a range of languages." Could we use technological tools to facilitate team members and support multilingualism across the project, e.g. subtitling for online meetings?

- We should keep having explicit discussions about our language policies and practices and critique and reflect on our practices.

It is worth noting that this session sparked a lively and important discussion about language practices and policies in the wider project, not just in relation to the localization of the questionnaire, and this is something which we will be looking to explore and address in the project.

4. Using Digital Tools and LLMs for Multilingual Research: Localising the WP4 Research Instruments

The research conducted within the remit of Work Package 4 (WP4) relies on two separate, but interconnected, research instruments to generate insights into young people's usage of, and experience with, AI tools. The first of these is an online questionnaire; the second is a task which is carried out using AI and reported on by participants. This third session of the day, led by Mark Ryan, focused on translating and localising the WP4 research instruments using digital tools and Large Language Models (LLMs). The purpose of the session was to share experiences from the translation process, outline the tools used, and collaboratively review the resulting language versions of the online questionnaire in breakout groups.

The workshop began with a presentation outlining the background to the translation and localisation work. Mark Ryan explained the choice of the software, as well as their various pros and cons. Two separate translation tools were chosen to provide participants with varied content to review. Participants were shown how translation workflows were managed using both Translation Tool 1 and Translation Tool 2, which was also integrated into the questionnaire platform. The session also covered practical considerations such as managing formal and informal registers, specifying dialects, and

addressing platform limitations such as formatting inconsistencies and inconsistent translation quality across the project's languages.

After the presentation, participants divided into breakout groups. Each group was assigned one or more language versions of the questionnaire to review, depending on the expertise of the members of the group. Using a digital whiteboard space, they recorded their observations, discussed translation accuracy, register, and cultural appropriateness, and proposed suggestions and amendments. The digital white board proved valuable for consolidating insights from across language teams. The combination of collaborative annotation and comparative discussion helped to surface both linguistic and methodological issues. A further benefit of the digital white board is that it was exported after the workshop, remaining as a visual artefact of the reflections of the day. Participants concluded that while the digital translation tools have considerable potential, effective localisation requires a combination of automated support and human expertise.

Some break-out sessions operated in site-specific and working languages (e.g. Catalan), while others worked across sites using English. Comments and feedback on the different versions of the questionnaire centred around register and tone, difficulties with terms across contexts, lack of consistency and clarity in the translations, the appropriacy of questionnaires for working with certain groups of participants, and issues and possible solutions for pluricentric languages.

Following the session, several follow-up actions were identified to support continued development and refinement of the WP4 research instruments. These steps will ensure that the translations remain accurate, culturally appropriate, and consistent across all language versions.

- Review and Consolidation of Translations: Each language group will review translations relevant to their context and update their respective language versions accordingly. Particular attention should be paid to formality levels, idiomatic phrasing, and clarity in question wording.
- Technical Integration and Testing: Revised translations will be uploaded to the questionnaire platform by the WP4 team.

5. Conclusion and Reflections

The second MultiLX Training Meeting provided a space for reflection, experimentation, and practical collaboration. A central insight emerging from the discussions was the recognition that many platforms that we use for research do not always deliver on multilingual usage, especially when related to minoritized languages. This limitation has implications for projects such as MultiLX, which seek to

foreground lesser-used and minoritised languages. It demands additional labour, creativity, and ethical sensitivity from researchers.

The event highlighted that doing multilingual research is inherently challenging and requires sustained resources, careful planning, flexibility, and ongoing commitment. In this context, digital translation tools were seen as valuable aids that can reduce the burden of manual work, provided their outputs are refined and validated through human expertise.

Reflections on the organisational aspects of the virtual meeting also provided valuable learning. From an organisational point of view, running the event online was more stressful in some ways and less in others. While digital delivery reduced travel demands and widened participation, it introduced challenges around spontaneity, control, and engagement. Adapting "in the moment" proved more difficult than in physical settings, and organisers found themselves constrained by the structure and rhythm of the technology. At the same time, this format enabled broader inclusion across sites and time zones, raising interesting questions about the nature and depth of engagement in digital spaces. The organisers observed that the online format required a more methodical, step-by-step approach—a need to "think more logically" to ensure smooth coordination and participant focus. While the process could at times feel tiring or tedious, it also simplified certain logistical aspects, notably by removing the need for travel and reducing associated costs and environmental impact.

Participants underscored that working multilingually demands conscious decision-making. These decisions should remain open to scrutiny and revision throughout the project, supported by transparent communication and reflexive dialogue. Crucially, multilingual research needs to be resourced and supported adequately.

Ultimately, the training meeting reinforced MultiLX's core ambition: to strengthen European linguistic capital in a globalised and technologised world. It demonstrated that advancing this goal requires constant negotiation between ideals of inclusivity and the realities of technological provisions. Through deliberate collaboration, ethical reflection, and pragmatic innovation, the MultiLX consortium continues to explore how research can both navigate and reshape the digital conditions that too often marginalise linguistic diversity.

Appendix 1: MultiLX Participants in WP4 Meeting, September 2025

Özgül Bendes, Universität Freiburg / University of Freiburg

Şilan, Bezaydaş, Universität Freiburg / University of Freiburg

Adrian Blackledge, University of Warwick

Katharina Brizić, Universität Freiburg / University of Freiburg

Angela Creese, University of Warwick

Victor Corona, Universitat Oberta de Catalunya / Open University of Catalonia

Mehdi Jafarzadeh, Universität Freiburg / University of Freiburg

Helen Kelly-Holmes, Ollscoil Luimnigh / University of Limerick

Edina Krompák, Pädagogische Hochschule Luzern / University of Teacher Education Lucerne

Pia Lane, Universitetet i Oslo / University of Oslo

Marta Lupica Spagnolo, Università di Torino / University of Turin

Jai Mackenzie, University of Warwick

Sidar Bayram Orhan, Universität Freiburg / University of Freiburg

Bernadette O'Rourke, Oilthigh Ghlaschu / University of Glasgow

Cecilia Paulsson, University of Warwick

Sari Pietikäinen, Jyväskylän yliopisto / University of Jyväskylä

Maite Puigdevall, Universitat Oberta de Catalunya / Open University of Catalonia

Joan Pujolar Cos, Universitat Oberta de Catalunya / Open University of Catalonia

Mark Ryan, Ollscoil Luimnigh / University of Limerick

M Seppola Simonsen, Universitetet i Oslo / University of Oslo

Paula Teixeria Molans, Oilthigh Ghlaschu / University of Glasgow

Victoria Wasner, Pädagogische Hochschule Luzern / University of Teacher Education Lucerne

Vida Colliander, Universitetet i Oslo / University of Oslo

Appendix 2: Invitees

Erin McNulty

Sofia Nemouchi

Verena Platzgummer

Appendix 3: Schedule of Training Day

MultiLX WP4 Workshop – Doing digital multilingual research / doing multilingual research digitally – challenges, opportunities, principles, and practices

Friday 26th September 2025, 10.00 - 16.00 (Irish Summer Time)

Welcome and Introductions (Helen Kelly-Holmes)

Session 1: Ethical and methodological challenges and opportunities

Agenda setting: (Helen, Sari, Mark)	20 minutes
Breakout and discussion in small groups	30 minutes
Thoughts on how we work with AI across the project (plenary discussion)	20 minutes
BREAK (11.20)	15 minutes
Optional time for reflection on the topic 1 (individual)	10 minutes

Session 2: Language Policy for WP4 research tools – Researching Multilingual Digital Practices (11.55)

Working with the languages of the project in WP4 (Helen)	15 minutes
Breakout and discussion in small groups	20 minutes
Principles of and LP for research tools in WP4 (plenary discussion)	20 minutes
LUNCH BREAK (12.50)	30 minutes
Optional time for reflection on the topic 2 (individual work)	10 minutes

Session 3: Using digital tools and LLMs for multilingual research – localising the WP4 research instruments (13.30)

Overview of working practices to date/tools used/versions	20 minutes
Breakout in site-specific working groups evaluating versions	45 minutes
BREAK (14.45) & Optional time for reflection on the session 3 (individual)	15 minutes
Report back on evaluation and improvements **(15.00)** (plenary session)	30 minutes

Session 4: General discussion and reflection on researching multilingualism digitally and use of AI and digital tools for multilingual research (plenary session)   20 minutes

Day finishes (16.00)