Language and Law / Linguagem e Direito, Vol. *12*(1), 2025, pp. 61-78
DOI: doi.org/10.21747/21833745/lanlaw12_1a3

Can machine translation really help minority languages in Europe? An analysis with value scenarios

Sergi Alvarez-Vidal [1] and Maarit Koponen [2]

[1] *Universitat Autònoma de Barcelona*, sergi.alvarez@uab.cat
[2] *University of Eastern Finland*, maarit.koponen@uef.fi

Abstract
Machine translation (MT) has greatly improved its quality in the last decade and has become nearly omnipresent in all aspects of society. Neural MT (NMT) and, more recently, large language models (LLMs) such as the generative pretrained transformer (GPT) have made translations to many languages easily accessible to all users from any phone or computer. However, most MT models are English-centric and only produce good quality results for those languages with great amounts of data. For minority languages, the challenge is often understood as the scarcity of data. Systemic differences between language communities should be taken into account for MT systems to actually be useful. In this paper, we use value scenarios to imagine the systemic impacts for two languages with differentiated sociolinguistic realities: Catalan and Karelian. The goal is to outline the main challenges and potential harms when considering MT for minority languages and to suggest some general guidelines that can be followed in future research and applications.
Keywords: *machine translation; minority languages; low-resouce languages; value scenarios.*

Resumo
Na última década, a tradução automática (TA) melhorou significativamente a sua qualidade e tornou-se quase omnipresente em todos os aspetos da sociedade. A tradução automática neuronal (TAN) e, mais recentemente, os grandes modelos de linguagem (LLMs), como o transformador pré-treinado generativo (GPT), tornaram as traduções para muitas línguas facilmente acessíveis a todos os utilizadores a partir de qualquer telefone ou computador. No entanto, a maioria dos modelos de TA são centrados no inglês e só produzem resultados de boa qualidade para as línguas com grandes quantidades de dados. Para as línguas minoritárias, o desafio é frequentemente entendido como a escassez de dados.

As diferenças sistémicas entre as comunidades linguísticas devem ser levadas em consideração para que os sistemas de TA sejam realmente úteis. Neste artigo, recorremos a uma análise de cenários para compreender os impactos sistémicos para duas línguas com realidades sociolinguísticas diferenciadas: o catalão e o careliano. O objetivo é delinear os principais desafios e danos potenciais ao considerar a TA para línguas minoritárias e sugerir algumas diretrizes gerais que podem ser seguidas em investigações e aplicações futuras.
Palavras-chave: *Tradução Automática, línguas minoritárias, línguas com recursos limitados, análise de cenários.*

1. Introduction

For many years, especially after World War I, great efforts were invested into "imposing an ideology of state-centered monolingualism" (Gal, 2017, p. 225), especially for those citizens of new nation-states who, until then, had identified with another language. In the mid-1980s, many political movements advocated for the protection of the "indigenous" languages, especially in the framework of international organizations such as the United Nations (UN) (Muehlebach, 2001). More recently, and in response to international pressure, many of the states changed their policies to fund the conservation and recuperation of these now minoritized languages (Laasko et al., 2016). This policy change has been largely propelled by the European Union (EU), which has linguistic identity as one of its constituent characteristics (Arzoz, 2008).

Europe has a multiplicity of languages, cultures, and peoples. Encouraging the use of these languages means promoting and encouraging diversity and different ways of understanding the world, regardless of the status and number of speakers of each language (Rehm & Way, 2023). According to the UN (2023), 40% of the approximately 7,000 languages that are spoken in the world are endangered to some degree. The *Atlas of the World's Languages in Danger* (Moseley, 2010) defines four levels of vulnerability: vulnerable, definitely endangered, severely endangered and critically endangered.

Language serves an instrumental purpose for accessing information and communication, but it also has an intrinsic function as part of minority identity and culture (Mowbray, 2017, p. 40). Recognizing and supporting minority languages is therefore important to support the autonomy, self-realization and equal dignity of speakers of all languages (De Schutter, 2017, pp. 22-24). However, language diversity has been enormously impacted by globalization, which has facilitated the growth of international languages such as English in the political, social and economic domains (Khan, 2023). Technology has also helped English become a language of prestige and status (Crystal, 2003), and has influenced its spread in media and business. This has largely contributed to its status as a lingua franca.

Language technologies, in particular MT, have also had a notable impact on society, especially since MT quality has been dramatically improved with the emergence of neural MT (NMT) (Wang et al., 2022). While these technologies offer potential benefits, their impact on linguistic diversity and equality should be considered carefully

(Rehm & Way, 2023). Even though communication has never been so easy and accessible to users, technological development risks concentrating powerful resources in a limited number of languages, while marginalising smaller ones. The uses of MT should be planned with consideration for the needs and sociolinguistic situation of the different minority languages. The emerging technological initiatives that develop MT models for these languages (Costa-jussà et al., 2022) should also focus on the target communities and users, in order to understand their needs.

We use value scenarios to incorporate the views of minority language stakeholders to imagine the systemic impacts of MT for two European minority languages. These languages possess different sociolinguistic realities and illustrate some of the potential harms both communities might face when implementing new or improved MT systems. The paper is structured as follows: in Section 2, we contextualize and define minority languages; in Section 3, we describe the most common MT systems and their use with minority languages; in Section 4, we introduce value scenarios and present a use case for two minority languages in Europe with differentiated sociolinguistic situations – Catalan and Karelian. We conclude the paper with Section 5, where we present some general guidelines that should be taken into account when developing responsible and useful MT systems for minority languages.

2. Minority, minoritized, low-resourced languages

The monolingual identification of nation-states, and thus, the negative stigmatization of multilingualism, was propelled after World War I, consolidated after World War II, and lived on for most part of the 20th century. This ideology stressed the importance of having one sole language to guarantee the well-being of all citizens of the same nation-state (Kamusella, 2009), ignoring the multilingual societies in which ideologies were implemented. However, the fall of the Berlin Wall in 1989 escalated tensions and conflicts between ethnic groups in Central and Eastern Europe, leading to growing concern across Europe about ethnic instability. As a consequence, in the beginning of the 1990s, there was a considerable increase in laws and treaties meant to protect minorities, such as the European Charter for Regional or Minority Languages (1992), the Framework Convention for the Protection of National Minorities (1995) and the adoption by the General Assembly of the United Nations of the Declaration on the Rights of Persons Belonging to National or Ethnic, Religious and Linguistic Minorities (1992) (Vieytez, 2014).

Diverse labels are used to refer to languages from minority groups, mainly stressing some of their characteristics: minoritized, low-resourced. In this paper, we use the term "minority language", following the definition stated by Cenoz and Gorter (2023). First, a minority is a group that is smaller in relation to another group. It is not a question of absolute numbers, but rather the numerical strength of the group in relation to other groups. Second, the use of "minority language" implies that there is a "majority language", that is, there is an asymmetric relation between both languages. Minority languages are always minoritized languages in power and status compared to the majority language because they have been disempowered by dominant groups in society. Third, the term "minority language" is widely used in many international organizations, such as the UN and the EU, as well as in international law.

According to the European Charter for Regional or Minority Languages, "regional or minority languages" means languages that are: (a) traditionally used within a given territory of a state by nationals of that state, who form a group numerically smaller than the rest of the state's population; and (b) different from the official language(s) of that state – this does not include either dialects of the official languages of the state or languages of migrants (Council of Europe, 1992). In Europe, there are a minimum of 69 minority languages that represent 89% of the linguistic diversity within the EU (Videsott, 2023). These include languages such as the Sámi languages, Aragonese, Karelian, Basque, Breton, Catalan, Corsican, Frisian, Friulian, Galician, Ladin, Occitan, Sardinian, and Welsh, among others.

Even though minority languages include the aforementioned characteristics, they also experience different situations, depending on their specific geographical location and social status. We can differentiate between minority languages that are "unique" and those that are also spoken in a kin-state (Extra & Gorter, 2008). Unique minority languages refer to languages spoken in one or more states, but which are not the dominant language in any state. This is the case of Basque, which is spoken in Spain and France, and also of Catalan, which is spoken in France, Spain, Andorra and the island of Sardinia (Italy). However, neither of these two languages has a dominant status in any of these countries. This also includes languages with very few speakers or that are on the verge of extinction, such as Ume Sámi in Sweden, where the youngest speakers are over 60 years of age (Gorter & Cenoz, 2011).

Kin-state minority languages are languages that are minority languages in one state, but are also the official dominant language in another state. This is the case, for example, of Swedish in Finland, where it is a minority language, even though it is the dominant language in Sweden (Potinkara, 2024). Another example is German, the official language in Germany and Austria, yet it is a minority language in countries such as Belgium, the Czech Republic, Denmark, France and Italy (Gorter and Cenoz, 2011). There are also a few minority languages that are official state languages, even though they function to great lengths in the same way as other minority languages. This is the case of Luxembourgish, which is official in Luxembourg even though the dominant languages in this country are French and German. The same happens with Catalan in Andorra, where it is the only official language. However, much of the linguistic space in this country is occupied by Spanish (Pla, 2000).

During the 20th century, there has also been a standardization process for many minority languages, so they could be used in education, the media, and various institutions (Costa, Korne, & Lane, 2017). In this process, one variety of a language becomes widely accepted throughout the linguistic community as a supradialectal norm. Even though this can facilitate the use of this language among all language users and contribute to the social prestige of the language. Recognizing the spoken and written communication practices of certain speakers as standard has placed many other speakers in a position of inferiority (Gal, 2006). These standards for minority languages can be perceived by the speakers of these languages as lacking authority as well as authenticity, and not being able to represent the speakers implied (Woolard, 2008).

3. Machine translation and minority languages

In recent years, MT quality has been greatly improved with the emergence of neural machine translation (NMT) and large language models (LLMs). This has changed the way people perceive multilingual communications. Users interact with MT daily and use it in very different contexts. When it is used for making sense of a text in a language that users do not understand, we talk about assimilation, which is obviously the most frequent use. When MT is used as part of a more complex process to produce a translation that will be later published, we talk about dissemination purposes (Way, 2018). This is usually used by professional translators, but can also be employed by other professionals like journalists (Fité Labaila, 2001, 2007) as one step in the whole translation process.

Early approaches to MT used handcrafted grammatical and lexical rules to translate into the target language (España-Bonet et. al, 2011). However, there is a high human cost in formalizing all the linguistic rules. Moreover, rules are difficult to transfer from one domain to another, and from one language to another (Wang et al., 2022). With the availability of bilingual corpora, methods using these corpora as the base for MT training became the dominant approach from the mid-2000s onward (Kenny, 2018). These statistical MT approaches often improved translation quality, but lacked fluency, and it had important reordering problems for distant language pairs (Wang et al., 2022). However, rule-based MT (RBMT) systems continue to be used for minority languages that have less available training data and produce poor results when using corpus-based approaches (Bayatli et al., 2018; Islam, Anik, & Islam, 2022; Kasthuri & Kumar, 2014; F. Pirinen & Wiechetek, 2022; T. A. Pirinen, 2019; Sghaier & Zrigui, 2020). Well-known examples of RBMT are Lucy LT1 (Alonso & Thurmair, 2003) and the open-source Apertium (Forcada et al., 2011; Khanna et al., 2021).

NMT is a corpus-based MT, which is trained on huge amounts of corpora, usually formed by pairs of source and target language segments (Forcada, 2017). NMT needs large quantities of data and uses deep neural networks. The first NMT models consist of an encoder and a decoder which are jointly trained (Sutskever, Vinyals, & Le, 2014). The encoder encodes the source language sentences into a sequence of vectors, which are the hidden representations of the source tokens, and also a meaning vector. Then the decoder generates the sequence of tokens for the target language, considering the meaning vector from the encoder. The transformer architecture, which is the one mostly used in current systems, follows mainly the encoder-decoder model with attention passed from encoder to decoder (Vaswani et al., 2017). This employs a self-attention mechanism that allows the encoder and decoder to account for every word included in the entire input sequence. As it yields high quality results, both in automatic and human evaluations (Bentivogli et al., 2016; Castillo et al., 2017; Klubička, Toral, & Sánchez-Cartagena, 2017), with more fluent outputs (Bentivogli et al., 2018), it has been quickly adopted by all industry sectors.

In recent years, the improvements in LLMs have brought new developments in MT and many other natural language processing tasks (Kim et al., 2021). Generative pre-trained transformers (GPTs) are decoder-only LLMs pretrained on large-scale unsupervised text data and are fine-tuned for different natural language processing tasks, which include MT (Brown et al., 2020). GPT models are also based on transformer architec-

ture (as NMTs), but with some differences. They are decoder-only models that use the same parameters to process the context and source as a single input for generating the next output. GPT models are basically trained on large quantities of monolingual data, mostly obtained by scraping texts from the Internet. Since the main language of online digital texts is English, this establishes a bias towards English in the dataset. These high data requirements are an important barrier for most of the smaller language communities, because many of these state-of-the-art tools cannot be applied to them due to the lack of data (Joshi et al., 2019). In fact, even though there is linguistic diversity around the world that includes more than 7,000 languages, few of them are represented in the current language technologies (Haddow et al., 2022; Joshi et al., 2020).

4. Value scenarios

The impact of new technologies is shaped not only by technical design, but also by how individuals and communities engage with them (Friedman, Kahn, Borning, & Huldtgren, 2013). In the case of minority languages, this engagement is deeply embedded in sociolinguistic, cultural, and political realities. MT is not merely a translation tool; it is part of a broader socio-technical system that can reinforce or disrupt language ideologies, power relations, and community practices.

To explore this complexity, we apply the value scenarios method, a speculative and ethical design approach that combines narrative-based scenario building with an explicit focus on human values (Nathan, Klasnja, & Friedman, 2007). Rather than predicting the future, value scenarios offer a structured way to explore how technologies might interact with complex social realities, including unintended consequences and ethical tensions. They help uncover latent assumptions in design and policy choices, as well as encourage reflection on how values such as fairness, representation, and autonomy are affected across different communities. They build on scenario-based design (Carrol, 1999), incorporate imaginative and subversive elements from design noir (Dunne & Raby, 2001), and anchor values such as linguistic justice, identity, and agency. Value scenarios help structure reflections along five key dimensions: stakeholders (both direct users and those indirectly affected), pervasiveness (how widespread and embedded the technology becomes), time (typically near- to mid-term), systemic effects (cascading consequences on language practices and social structures), and value implications (tensions between community needs and the goals or assumptions of technological design).

This approach has been previously applied by Haroutunian (2022), who constructed value scenarios based on stakeholder interviews for the Armenian-English MT context. His work highlighted the importance of participatory development and the need to assess which language technologies might best serve the interests of low-resourced communities beyond MT alone. Inspired by this precedent, our study expands the approach to a comparative perspective by focusing on two minority languages in Europe with contrasting sociolinguistic contexts: Catalan and Karelian.

To ground the analysis, we designed a speculative example of MT technology called MIGHTYMT: a fictional, advanced language technology envisioned to support real-time translation for minority languages. MIGHTYMT is a LLM-based assistant integrated across personal and public devices. It is capable of translating both written and spoken interactions in real time, using voice, text, or multimodal inputs. It operates seamlessly

across messaging apps, search engines, browsers, and voice assistants. Users can query it conversationally or use it passively to translate content displayed on screens, or spoken aloud in physical spaces.

MIGHTYMT supports high-quality, bidirectional translation between minority languages (such as Catalan and Karelian) and major world languages. It uses context-aware translation modules adapted to dialectal variation and allows for user feedback to train personal language profiles. It is designed to be intuitive, customizable, and accessible, with offline functionalities for rural and under-connected regions. Unlike traditional MT systems, it includes modules to explain translation decisions and detect sociolinguistic sensitivities (for example, formality, regional variation, gendered language), making it more transparent and adaptable for speakers of diverse linguistic backgrounds.

This imagined tool serves as the basis for our two value scenarios—one for Catalan and one for Karelian—which explore both the promises and challenges associated with integrating such a system into daily life. These scenarios examine functional, social, and ethical dimensions to better understand how advanced language technologies may interact with complex language ecologies and community values.

4.1. The Catalan Value Scenario

Catalan is a standardized language with around 10 million speakers (Generalitat de Catalunya, 2024), yet it continues to face marginalization in key domains. It has a marginal status in the Balearic Islands and Aragon, is co-official in Catalonia, and is the sole official language in Andorra. It is also spoken in the Sardinian city of Alghero, in some villages in the South of France, and in frontier areas of Murcia. Despite this wide geographical spread, Catalan remains a minoritized language, particularly in the justice and health systems as well as in state-level services, where Spanish is often the dominant language of operation (Pla, 2000).

Since the early 1980s, the education system and public media have played a crucial role in maintaining and promoting Catalan, while civic organizations such as Softcatalà, Òmnium Cultural, and Plataforma per la Llengua have been instrumental in advocating for language rights. In the field of MT, various systems now support Catalan. Google Translate incorporated it in 2008, and Softcatalà maintains several MT engines covering different language combinations to and from Catalan (Briva Iglesias, 2022). Catalan is also increasingly included in LLMs by default, primarily due to its structural similarity to other Romance languages such as Spanish, French, and Italian. However, the resulting outputs are not always genuine or contextually appropriate, often reflecting limited training data or linguistic bias.

In addition to these tools, several research projects are dedicated to the development of language technologies for Catalan. One example is TAN-IBE, which focuses on building NMT models and resources for the Iberian Romance languages (Oliver et al., 2023). Another is Projecte AINA (Montserrat, 2023), a publicly funded initiative coordinated by the Barcelona Supercomputing Center (BSC) in collaboration with the Catalan government. AINA develops large-scale language resources and open-source AI tools—including curated speech and text corpora and Catalan-adapted LLMs—with the overarching goal of ensuring the language's digital viability.

In our value scenario, we consider the deployment of MIGHTYMT and its potential impact on different stakeholder groups, particularly young speakers and healthcare professionals. For young Catalan speakers who are increasingly shifting toward Spanish in digital spaces (Idescat, 2023), MIGHTYMT could lower linguistic barriers and promote the use of Catalan in informal online communication. This may reinforce linguistic confidence and help normalize Catalan in everyday digital practices. However, overreliance on automated output could contribute to linguistic deskilling, limiting opportunities for active language production. If the system consistently reproduces standardized forms, speakers from regions such as the Balearic Islands or Alghero might feel alienated, reinforcing perceptions of exclusion.

For healthcare professionals, MIGHTYMT could support communication with Catalan-speaking patients, particularly older adults or individuals with cognitive or speech-related impairments, who may find it easier to express themselves in their first language. This could promote more inclusive and patient-centred care. However, the use of automated tools in clinical contexts raises legal and ethical concerns. If MIGHTYMT processes sensitive health data without adequate safeguards, it may violate patients' privacy rights as defined under the General Data Protection Regulation (GDPR), designating health data as a special category requiring explicit consent and enhanced protection (European Union, 2016, Art. 9). Article 32 further mandates that data controllers implement technical and organisational measures to ensure the confidentiality, integrity, and security of such data, especially in high-risk domains like healthcare.

Systemic effects of deploying MIGHTYMT could include increased visibility and usability of Catalan across sectors, enhanced digital literacy, and a stronger presence in multilingual public communication, reinforcing existing efforts such as the parallel publication of press content in Catalan and Spanish (Fité Labaila, 2007). However, the tool could also entrench dominant language norms, reduce active language engagement, and obscure the need for structural reforms. Public institutions may interpret the presence of MT systems as a substitute for guaranteeing linguistic rights through human-mediated services, thus shifting responsibility for language access onto users and technology.

If developed with community input, MIGHTYMT could also address the longstanding invisibilization of dialectal varieties—such as those spoken in the Balearic Islands or the Valencian Country—which are often flattened in mainstream MT systems. Allowing user-controlled variation and transparent dialect tagging could promote a more inclusive representation of Catalan's internal diversity.

From a legal perspective, integrating MIGHTYMT into public services entails obligations regarding linguistic equality as established by regional autonomy statutes. Inaccurate or biased outputs—particularly in sensitive domains like healthcare or justice—could constitute violations of linguistic rights and require formal accountability mechanisms. When third-party providers are involved, public administrations must ensure compliance with data protection, transparency, and quality standards. In the absence of a clear regulatory framework for AI-based multilingual communication, these risks may go unaddressed. This scenario underscores the tension between enhanced access and the preservation of linguistic agency. While tools like MIGHTYMT may contribute to the normalization of Catalan in digital environments, their design and

governance must be inclusive and community-driven to avoid reproducing structural inequalities.

4.2. The Karelian Value Scenario

Karelian is a critically endangered minority language spoken in Finland and Russia. It is not recognised as an official language in either country, although some laws defining its status exist in the Republic of Karelia in the Russian Federation. The Finnish government included provisions in its language policy programme in 2022 to strengthen the legal status of Karelian in Finland (Valtioneuvosto, 2022). The number of speakers is currently decreasing even though there have been different attempts to revitalize the language, and a planned language revitalization has been in place since the end of the 1980s (Moshnikov, 2022).

Historically, Karelian was a territorial language. As a result of World War II, the majority of speakers of Karelian lost their traditional homelands and were resettled in various parts of Finland. Linguistically, Karelian belongs to the Finnic branch of the Uralic language family, or more precisely, to its eastern Finnic subgroup, which can be distinguished geographically from the southern and western groups, and on the basis of their common history. Karelian does not have a standard variety and is divided into two main dialects: Olonets (or Livvi) Karelian, and Karelian Proper. The latter consists of Viena (North) Karelian and South Karelian (Uusitupa, Koivisto, & Palander, 2017). Based on census data, the number of Karelian speakers in Russia was about 25,000 in 2010 (Moshnikov, 2022, p. 278). In Finland, the number of fluent speakers is estimated to be around 11,000, with a further 20,000 who speak or understand Karelian at least to some extent (Sarhimaa, 2016, p. 3). Some schools in Russian Karelia teach Karelian, and university-level courses are offered at Petrozavodsk State University in Russia and at the University of Eastern Finland, in Finland (Moshnikov, 2022). According to a recent survey conducted by Moshnikov (2022), only 17% of the 161 respondents reported they speak Karelian daily, 18% at least once a week, and 16% more rarely. Additionally, 40% of the surveyed users reported that while they do not speak Karelian, they can understand written or spoken Karelian (34%), or write in Karelian (6%). In the digital sphere, 40% of the respondents search for information in Karelian daily, and an additional 46% at least once a week (Moshnikov, 2022, p. 285). An earlier survey by Salonen (2017, p. 9) found that 73.7% of the nearly 2,000 respondents used Karelian online, mainly for reading, and to a lesser extent also writing content on websites, blogs or discussion forums.

None of the Karelian varieties are currently covered by commercial MT systems such as Google Translate or large-scale initiatives such as Meta's 'No Language Left Behind' (Costa-jussà et al., 2022). Some work has been done on Livvi Karelian as part of the SMUGRI-MT NMT system for low-resource Finnic languages (Purason et al., 2024; Yankovskaya et al., 2023), and a rule-based approach (T. A. Pirinen, 2019). OpenAI's ChatGPT and Anthropic's Claude appear to generate translations that the models claim to be to and from Karelian. However, establishing the output quality these tools produce is outside of the scope of this article.

For our MIGHTYMT value scenario, stakeholders firstly include members of the Karelian speaker community. The tool could help them search for information or answer questions in their language. Particularly for speakers who do not currently use Karelian in an active way, this could give more opportunities to engage with the language.

Another group of stakeholders would be people learning Karelian as a new language, or hoping to improve their skills. This chatbot offers the possibility of interacting in Karelian without having the time constraints of formal learning environments. Speech translation capacities could be useful particularly for those who currently indicate that they do not actively speak although they understand written or spoken Karelian. They could ask for information or talk about current issues to practice their Karelian, and even ask for corrections if they think their level of language is not good enough. It could also enable Karelian speakers to use their own language in interactions with the majority language speakers.

MIGHTYMT could also offer support for translation activities that have been identified as important for language revitalization, such as creating vocabulary and educational materials (cf. Kuusi, Kolehmainen, & Riionheimo, 2017). One important source for Karelian language content is the Finnish national broadcaster Yle, which currently produces news only in one dialect (Livvi Karelian), although some other newspapers publish articles in different variants (Moshnikov, 2022). As a tool for journalists — and others producing materials — MIGHTYMT could also help produce more content in the other varieties, if these are included in the tool.

The value scenario presumes that use of the technology would be widespread. In practice, however, the interest of the speaker community to adopt the technology may be limited. Karelian speakers tend to be bilingual, which means that they also speak a majority language, and commonly resort to that in digital environments, in a phenomenon which has been labeled as digital diglossia (Lozano Sañudo, 2023). Hence, Karelian speakers would need to be convinced this new product can accommodate their needs in Karelian before using it in many contexts. The lack of a standard variety also presents a challenge: users may not identify with the variant used by the chatbot and could respond negatively to MIGHTYMT not using their own language variety, which could lead to a low sense of reliability in the chatbot. As Moshnikov (2022) notes, however, the language community is accustomed to multidialectal content in newspapers, for example, so dialectal variation might not be considered an obstacle by Karelian speakers.

Systemic effects of the use of MIGHTYMT could in this way include support for the revitalization efforts of the community and wider use of Karelian both online and off. More widespread use would also improve the visibility of Karelian among majority language speakers. Currently, even the existence of Karelian is not well-known among Finnish people, and it is commonly perceived as a regional dialect of Finnish. By enabling wider use of Karelian in society, MIGHTYMT could raise awareness of the language and strengthen its status.

However, widespread use of the technology could also lead to adverse systemic effects. Firstly, large quantities of machine-generated outputs could start to direct standardization of the language on the terms of the technology, rather than the language community. Secondly, focusing on technological solutions could end up directing societal resources and support away from the speaker community's revitalisation efforts. Thirdly, quality and reliability of the output is naturally a major concern, particularly in high-stakes scenarios like healthcare, legal settings and other public services (see also Guerberof-Arenas & Moorkens, 2023). Similarly to the Catalan case, implementation in these scenarios could have adverse implications for the Karelian speakers' equal rights.

5. Implications for responsible development of MT for minority languages

The envisioned MIGHTYMT system highlights both the potential and pitfalls of deploying MT for minority languages. While Catalan benefits from a relatively high level of institutional support and digital infrastructure, Karelian exemplifies the challenges faced by critically endangered languages with minimal technical resources. The contrast underscores the need for differentiated strategies when developing and implementing MT for diverse language communities.

While Karelian faces a scarcity of digital resources and lacks stable language norms, Catalan is comparatively well-represented in existing language technologies. However, technological inclusion alone does not ensure linguistic empowerment. In both cases, systems like MIGHTYMT risk displacing human-driven language use. For Catalan, overreliance on automated tools may discourage active written production, particularly among young users, and reinforce a narrow standard at the expense of dialectal variation. For Karelian, the risks are amplified in the context of language endangerment: widespread MT use could impose artificial norms on a language with limited standardization and potentially low-quality outputs due to scarce training data. Moreover, mass generation of machine-produced content may saturate digital spaces with synthetic language, displacing and marginalizing authentic, community-produced discourse and expressions (see Aranberri & Iñurrieta, 2024). This risk is particularly acute for endangered languages like Karelian, where vitality depends on active use, and MT quality is likely to remain low due to data scarcity. Some scholars have therefore argued that MT should not be offered for such languages until its quality and consequences are better understood (Bowker, 2021; F. Pirinen & Wiechetek, 2022).

Community acceptance is another crucial factor shaping the role of MT in minority language contexts. Acceptability depends not only on output quality, but also on the perceived purpose and framing of the technology. Research has shown that users tend to accept MT more readily for informational tasks than for contexts involving linguistic identity or visibility (Bowker, 2009; Bowker & Buitrago Ciro, 2015). In Catalonia, for example, some criticism of MT in public services stems from concerns that it might be used to avoid institutional commitments to Catalan, particularly in sensitive domains where language choice is symbolic of rights and recognition (Baumgarten & Cornellà-Detrell, 2019; Nurminen & Koponen, 2020).

These reflections underscore that MT is never a neutral tool. As Bird (2024) argues, language technologies should be designed to support human agency, not replace it. For minority languages, this means embedding MT within community-led strategies that prioritize speaker autonomy, linguistic diversity, and long-term sustainability. Following this principle, development of MT for minority languages should not be carried out simply for the sake of technological development. Rather, the development should be guided by the needs and priorities of the language communities (cf. Riemland, 2023), for example, through participatory research practices. Tools like MIGHTYMT may support access and visibility, but only if they are transparent, participatory, and sensitive to the values and goals of the communities they aim to serve. The central role of the language community as end-users of the technology is reflected in the discussion in our article. In future work, it would be useful to address also the other parts of the

"MT stakeholder ecosystem": companies, governmental agencies, language professionals, researchers and developers (cf. Guerberof-Arenas & Moorkens, 2023, p. 115).

In conclusion, the value scenarios of Catalan and Karelian demonstrate that MT can contribute to the sustainability of minority languages under specific conditions, but may also deepen existing asymmetries if deployed without care. Ensuring that MT supports rather than undermines these languages requires ongoing community involvement, legal and ethical oversight, and a commitment to addressing not only technological gaps, but the structural conditions in which these languages are used.

Acknowledgment

The work of Maarit Koponen was supported by the Strategic Research Council (SRC) established within the Research Council of Finland; funding agreements 352577 (University of Eastern Finland) and 352557 (DECA consortium coordinator University of Helsinki).

References

Alonso, J. A., & Thurmair, G. (2003). The compendium translator systems. In *Proceedings of the Ninth Machine Translation Summit.*

Aranberri, N., & Iñurrieta, U. (2024). When minoritized languages encounter MT: Perceptions and expectations of the Basque community. *The Journal of Specialised Translation*, *41*, 179–205. doi: 10.26034/cm.jostrans.2024.4718

Arzoz, X. (2008). Preface. In X. Arzoz (Ed.), *Respecting Linguistic Diversity in the European Union* (Vol. 2, pp. vii–viii). Amsterdam: John Benjamins Publishing Company. doi: 10.1075/wlp.2.01arz

Baumgarten, S., & Cornellà-Detrell, J. (2019). Translation and the economies of power. In *Translation and the Global Spaces of Power* (pp. 11–26). Bristol: Multilingual Matters.

Bayatli, S., Kurnaz, S., Salimzianov, I., Washington, J. N., & Tyers, F. M. (2018). Rule-based machine translation from Kazakh to Turkish. In *Proceedings of the 21st Annual Conference of the European Association for Machine Translation: 28-30 May 2018* (pp. 49–58). Universitat d'Alacant, Alacant, Spain. Retrieved from https://dialnet.unirioja.es/servlet/articulo?codigo=6474388

Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. (2016). Neural versus phrase-based machine translation quality: A case study. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 257–267). Austin, Texas: Association for Computational Linguistics. Retrieved from http://aclweb.org/anthology/D16-1025 doi: 10.18653/v1/D16-1025

Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. (2018). Neural versus phrase-based MT quality: An in-depth analysis on English–German and English–French. *Computer Speech & Language*, *49*, 52–70. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0885230817301079 doi: 10.1016/j.csl.2017.11.004

Bird, S. (2024). Must NLP be extractive? In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).* Retrieved from https://aclanthology.org/2024.acl-long.797/

Bowker, L. (2009). Can Machine Translation meet the needs of official language minority communities in Canada? A recipient evaluation. *Linguistica Antverpiensia New Series-Themes in Translation Studies*, *8*, 123–155.

Bowker, L. (2021). Translation technology and ethics. In K. Koskinen & N. K. Pokorn (Eds.), *The Routledge Handbook of Translation and Ethics* (1st ed., pp. 262–278). London/New York: Routledge. Retrieved from https://www.taylorfrancis.com/books/9781000288988/chapters/10.4324/9781003127970-20 doi: 10.4324/9781003127970-20

Bowker, L., & Buitrago Ciro, J. (2015). Investigating the usefulness of machine translation for newcomers at the public library. *Translation and Interpreting Studies*, *10*(2), 165–186. doi: 10.1075/tis.10.2.01bow

Briva Iglesias, V. (2022). English-Catalan neural machine translation: State-of-the-art technology, quality, and productivity. *Revista Tradumàtica: Traducció i Tecnologies de La Informació i La Comunicació*, *20*, 149–176.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., … Amodei, D. (2020). *Language Models are Few-Shot Learners.* arXiv. (Version Number: 4) doi: 10.48550/ARXIV.2005.14165

Carrol, J. (1999). Five reasons for scenario-based design. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers* (p. 11). Maui, HI, USA. Retrieved from http://ieeexplore.ieee.org/document/772890/ doi: 10.1109/HICSS.1999.772890

Castillo, S., Moorkens, J., Gaspari, F., Sennrich, R., Sosoni, V., Georgakopoulou, Y., … Gialama, M. (2017, September). A comparative quality evaluation of PBSMT and NMT using professional translators. In *Proceedings of Machine Translation Summit XVI: Research Track.* Retrieved from https://aclanthology.org/2017.mtsummit-papers.10

Cenoz, J., & Gorter, D. (2023). Second language acquisition and minority languages: An introduction. In J. Cenoz & D. Gorter (Eds.), *The Minority Language as a Second Language* (pp. 1–15). New York: Routledge.

Costa, J., Korne, H. D., & Lane, P. (2017). Standardising Minority Languages: Reinventing Peripheral Languages in the 21st Century. In P. Lane, J. Costa, & H. De Korne (Eds.), *Standardizing Minority Languages* (pp. 1–23). New York: Routledge.

Costa-jussà, M. R., Cross, J., Çelebi, O., Elbayad, M., Heafield, K., Heffernan, K., … Wang, J. (2022). *No language left behind: Scaling human-centered Machine Translation.* arXiv. (arXiv:2207.04672 [cs]) doi: 10.48550/arXiv.2207.04672

Council of Europe. (1992). *European Charter for Regional or Minority Languages.* Retrieved from https://rm.coe.int/1680695175

Crystal, D. (2003). *English as a Global Language* (2nd ed.). Cambridge University Press. doi: 10.1017/CBO9780511486999

De Schutter, H. (2017). Translational justice: Between equality and privation. In G. G. Núñez & R. Meylaerts (Eds.), *Translation and Public Policy: Interdisciplinary Perspectives and Case Studies* (1st ed., pp. 15–31). London: Routledge. Retrieved from https://www.taylorfrancis.com/books/9781315521763 doi: 10.4324/9781315521770

Dunne, A., & Raby, F. (2001). *Design Noir: The Secret Life of Electronic Objects.* Springer Science & Business Media.

España-Bonet, C., Labaka, G., Díaz de Ilarraza, A., & Màrquez, L. (2011). Hybrid Machine Translation Guided by a Rule–Based System. In *Proceedings of Machine Translation Summit XIII: Papers.* Xiamen, China. Retrieved from https://aclanthology.org/2011.mtsummit-papers.63

European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). *Official Journal of the European Union, L119*, 1–88. Retrieved from https://eur-lex.europa.eu/eli/reg/2016/679/oj

Extra, G., & Gorter, D. (2008). The constellation of languages in Europe: An inclusive approach. In G. Extra & D. Gorter (Eds.), *Multilingual Europe: Facts and Policies* (pp. 3–61). Berlin/New York: Mouton de Gruyter.

Fité Labaila, R. (2001). La traducció automàtica aplicada a la premsa escrita. El cas d'El Periódico en català. *Treballs de Comunicació*, 21–25.

Fité Labaila, R. (2007). Cas d'integració de la TA: El Periódico. *Tradumàtica: Traducció i Tecnologies de La Informació i La Comunicació*, *4*. Retrieved from https://raco.cat/index.php/Tradumatica/article/view/56010

Forcada, M. L. (2017). Making sense of neural machine translation. *Translation Spaces*, *6*(2), 291–309. doi: 10.1075/ts.6.2.06for

Forcada, M. L., Ginestí-Rosell, M., Nordfalk, J., O'Regan, J., Ortiz-Rojas, S., Pérez-Ortiz, J. A., … Tyers, F. M. (2011). Apertium: A free/open-source platform for rule-based machine translation. *Machine Translation*, *25*(2), 127–144. doi: 10.1007/s10590-011-9090-0

Friedman, B., Kahn, P. H., Borning, A., & Huldtgren, A. (2013). Value sensitive design and information systems. In N. Doorn, D. Schuurbiers, I. Van De Poel, & M. E. Gorman (Eds.), *Early engagement and new technologies: Opening up the laboratory* (Vol. 16, pp. 55–95). Dordrecht: Springer Netherlands. Retrieved from https://link.springer.com/10.1007/978-94-007-7844-3_4 (Series Title: Philosophy of Engineering and Technology) doi: 10.1007/978-94-007-7844-3_4

Gal, S. (2006). Contradictions of standard language in Europe: Implications for the study of practices and publics*. *Social Anthropology*, *14*(2), 163–181. doi: 10.1111/j.1469-8676.2006.tb00032.x

Gal, S. (2017). Visions and Revisions of Minority Languages: Standardization and Its Dilemmas. In P. Lane, J. Costa, & H. De Korne (Eds.), *Standardizing Minority Languages* (pp. 222–242). New York: Routledge.

Generalitat de Catalunya. (2024). *Situació del català, de l'occità aranès i de la llengua de signes catalana i principals línies d'intervenció en política lingüística durant la legislatura.*

Gorter, D., & Cenoz, J. (2011). Multilingual education for European minority languages: The Basque Country and Friesland. *International Review of Education / Internationale Zeitschrift Für Erziehungswissenschaft / Revue Internationale de l'Education*, *57*(5/6).

Guerberof-Arenas, A., & Moorkens, J. (2023). Ethics and machine translation: The end user perspective. In *Towards Responsible Machine Translation: Ethical and Legal Considerations in Machine Translation* (pp. 113–133). Cham: Springer.

Haddow, B., Bawden, R., Barone, A. V. M., Helcl, J., & Birch, A. (2022). Survey of low-

resource Machine Translation. *Computational Linguistics*, *48*(3), 673–732. doi: 10.1162/coli_a_00446

Haroutunian, L. (2022). Ethical considerations for low-resourced Machine Translation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop* (pp. 44–54). Dublin, Ireland: Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-srw.5

Idescat. (2023). *Enquesta d'usos lingüístics de la població.* Retrieved from https://www.idescat.cat/pub/?id=eulp&n=7195

Islam, M. A., Anik, M. S. H., & Islam, A. B. M. A. A. (2022). An enhanced RBMT: When RBMT outperforms modern data-driven translators. *IETE Technical Review*, *39*(6), 1473–1484. doi: 10.1080/02564602.2022.2026828

Joshi, P., Barnes, C., Santy, S., Khanuja, S., Shah, S., Srinivasan, A., … Bali, K. (2019). Unsung challenges of building and deploying language technologies for low resource language communities. In *Proceedings of the 16th International Conference on Natural Language Processing* (pp. 211–219). NLP Association of India. Retrieved from https://aclanthology.org/2019.icon-1.25

Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic fiversity and inclusion in the NLP world. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 6282–6293). Online: Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.560

Kamusella, T. (2009). *The Politics of Language and Nationalism in Modern Central Europe*. London: Palgrave Macmillan UK. Retrieved from http://link.springer.com/10.1057/9780230583474 doi: 10.1057/9780230583474

Kasthuri, M., & Kumar, S. B. R. (2014). Rule Based Machine Translation System from English to Tamil. In *2014 World Congress on Computing and Communication Technologies* (pp. 158–163). Trichirappalli, India: IEEE. doi: 10.1109/WCCCT.2014.50

Kenny, D. (2018). Machine translation. In P. Rawling & P. Wilson (Eds.), *The Routledge Handbook of Translation and Philosophy* (pp. 428–445). Abingdon/New York: Routledge.

Khan, A. (2023). Language and globalization: A critical study on language, culture, and power dynamics. *International Journal of English Learning & Teaching Skills*, *5*(3), 3412–3423. doi: 10.15864/ijelts.5305

Khanna, T., Washington, J. N., Tyers, F. M., Bayatlı, S., Swanson, D. G., Pirinen, T. A., … Alòs I Font, H. (2021). Recent advances in Apertium, a free/open-source rule-based machine translation platform for low-resource languages. *Machine Translation*, *35*(4), 475–502. doi: 10.1007/s10590-021-09260-6

Kim, Y. J., Awan, A. A., Muzio, A., Salinas, A. F. C., Lu, L., Hendy, A., … Awadalla, H. H. (2021). *Scalable and Efficient MoE Training for Multitask Multilingual Models.* arXiv. (Version Number: 1) doi: 10.48550/ARXIV.2109.10465

Klubička, F., Toral, A., & Sánchez-Cartagena, V. M. (2017). Fine-grained human evaluation of neural versus phrase-based Machine Translation. *The Prague Bulletin of Mathematical Linguistics*, *108*(1), 121–132. doi: 10.1515/pralin-2017-0014

Kuusi, P., Kolehmainen, L., & Riionheimo, H. (2017). Introduction: Multiple roles of translation in the context of minority languages and revitalisation. *Trans-Kom*, *10*(2), 138–163.

Laakso, J., Sarhimaa, A., Åkermark, S. S., & Toivanen, R. (2016). *Towards Openly Multilingual Policies and Practices: Assessing Minority Language Maintenance Across Europe*. Bristol: Multilingual Matters.

Lozano Sañudo, B. (2023). The role of language technologies and MT in fostering multilingualism by contributing to the revitalization of minor and endangered languages. In F. M. Sirignano & R. M. Roig (Eds.), *Educación, investigación e innovación en la red, 2023* (pp. 131–149). Aula Magna. Retrieved from https://dialnet.unirioja.es/servlet/articulo?codigo=9527870

Montserrat, M. V. (2023). El projecte AINA, la IA i les tecnologies del llenguatge. *Terminàlia*, *27*, 80–84.

Moseley, C. (2010). *Atlas of the world's languages in danger*. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000187026

Moshnikov, I. (2022). The use of Karelian language online: Current trends and challenges. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics*, *13*(2). doi: 10.12697/jeful.2022.13.2.09

Mowbray, J. (2017). Translation as marginalisation? International law, translation and the status of linguistic minorities. In G. G. Núñez & R. Meylaerts (Eds.), *Translation and Public Policy: Interdisciplinary Perspectives and Case Studies* (1st ed., pp. 32–57). Routledge. doi: 10.4324/9781315521770

Muehlebach, A. (2001). "Making Place" at the United Nations: Indigenous Cultural Politics at the U.N. Working Group on Indigenous Populations. *Cultural Anthropology*, *16*(3), 415–448. doi: 10.1525/can.2001.16.3.415

Nathan, L. P., Klasnja, P. V., & Friedman, B. (2007). Value scenarios: A technique for envisioning systemic effects of new technologies. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems* (pp. 2585–2590). San Jose CA USA: ACM. doi: 10.1145/1240866.1241046

Nurminen, M., & Koponen, M. (2020). Machine translation and fair access to information. *Translation Spaces*, *9*(1), 150–169. doi: 10.1075/ts.00025.nur

Oliver, A., Vàzquez, M., Coll-Florit, M., Álvarez, S., Suárez, V., Aventín-Boya, C., … Pardos, A. (2023, June). TAN-IBE: Neural Machine Translation for the romance languages of the Iberian Peninsula. In M. Nurminen et al. (Eds.), *Proceedings of the 24th Annual Conference of the European Association for Machine Translation* (pp. 495–496). Tampere, Finland: European Association for Machine Translation. Retrieved from https://aclanthology.org/2023.eamt-1.50/

Pirinen, F., & Wiechetek, L. (2022). Building an Extremely Low Resource Language to High Resource Language Machine Translation System from Scratch. In R. Schaefer, X. Bai, M. Stede, & T. Zesch (Eds.), *Proceedings of the 18th Conference on Natural Language Processing (KONVENS 2022)* (pp. 150–155). Potsdam, Germany: KONVENS 2022 Organizers. Retrieved from https://aclanthology.org/2022.konvens-1.18/

Pirinen, T. A. (2019). Workflows for kickstarting RBMT in virtually No-Resource Situation. In A. Karakanta et al. (Eds.), *Proceedings of the 2nd Workshop on Technologies for MT of Low Resource Languages* (pp. 11–16). Dublin, Ireland: European Association for Machine Translation. Retrieved from https://aclanthology.org/W19-6803/

Pla, J. T. i. (2000). La Llengua catalana a Andorra: Problemes i perspective. *Treballs de*

sociolingüística catalana, 165–167.

Potinkara, N. (2024). Finland-Swedes and the concept of national minorities in Sweden. *Ethnopolitics*, *23*(1), 59–75. doi: 10.1080/17449057.2022.2108596

Purason, T., Ivanov, A., Yankovskaya, L., & Fishel, M. (2024). SMUGRI-MT - Machine Translation System for Low-Resource Finno-Ugric Languages. In C. Scarton et al. (Eds.), *Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 2)* (pp. 31–32). Sheffield, UK: European Association for Machine Translation (EAMT). Retrieved from https://aclanthology.org/2024.eamt-2.16/

Rehm, G., & Way, A. (Eds.). (2023). *European Language Equality: A Strategic Agenda for Digital Language Equality*. Springer.

Salonen, T. (2017). Karelian – a digital language? In *Reports on Digital Language Diversity in Europe.* Retrieved from http://www.dldp.eu/sites/default/files/documents/DLDP_Karelian-Report.pdf

Sarhimaa, A. (2016). *Karelian in Finland: ELDIA Case-Specific Report.* Retrieved from https://phaidra.univie.ac.at/detail/o:471733

Sghaier, M. A., & Zrigui, M. (2020). Rule-based Machine Translation from Tunisian dialect to modern standard Arabic. *Procedia Computer Science*, *176*, 310–319. doi: 10.1016/j.procs.2020.08.033

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). *Sequence to Sequence Learning with Neural Networks.* arXiv. (Version Number: 3) doi: 10.48550/ARXIV.1409.3215

United Nations. (2023). *Why Indigenous languages matter: The international decade on Indigenous languages 2022–2032.* Retrieved from www.un.org/development/desa/dpad/wpcontent/uploads/sites/45/publication/PB151.pdf

Uusitupa, M., Koivisto, V., & Palander, M. (2017). Border Karelian dialects and the terminology of Karelian borderlands' language varieties. , *121*.

Valtioneuvosto. (2022). Kielipoliittinen ohjelma: Valtioneuvoston periaatepäätös (No. 51; Valtioneuvoston julkaisuja). *Valtioneuvosto*. Retrieved from http://urn.fi/URN:ISBN:978-952-383-645-7

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). *Attention Is All You Need.* arXiv. (Version Number: 7) doi: 10.48550/ARXIV.1706.03762

Videsott, P. (2023). Minority and minoritised languages as part of the European linguistic and cultural diversity | Think Tank | European Parliament. *European Parliament*. Retrieved from https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2023)751273

Vieytez, E. J. R. (2014). Cultural Diversities and Human Rights: History, Minorities, Pluralization. *The Age of Human Rights Journal*, *3*(Article 3).

Wang, H., Wu, H., He, Z., Huang, L., & Church, K. W. (2022). Progress in Machine Translation. *Engineering*, *18*, 143–153. doi: 10.1016/j.eng.2021.03.023

Way, A. (2018). Quality Expectations of Machine Translation. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation Quality Assessment* (Vol. 1, pp. 159–178). Cham: Springer International Publishing. Retrieved 2025-12-02, from http://link.springer.com/10.1007/978-3-319-91241-7_8 (Series Title: Machine Translation: Technologies and Applications) doi: 10.1007/978-3-319-91241-7_8

Woolard, K. A. (2008). Language and identity choice in Catalonia: The interplay of

contrasting ideologies of linguistic authority. In K. Süselbeck, U. Mühlschlegel, & P. Masson (Eds.), *Lengua, nación e identidad: La regulación del plurilinguismo en España y América Latina.* Iberoamericana Vervuert. Retrieved from https://dialnet.unirioja.es/servlet/articulo?codigo=5371500

Yankovskaya, L., Tars, M., Tättar, A., & Fishel, M. (2023). Machine Translation for Low-resource Finno-Ugric Languages. In T. Alumäe & M. Fishel (Eds.), *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)* (pp. 762–771). Tórshavn, Faroe Islands: University of Tartu Library. Retrieved from https://aclanthology.org/2023.nodalida-1.77/