

# THE ROLE OF LANGUAGE TECHNOLOGIES IN PROMOTING THE PARTICIPATION OF LINGUISTIC MINORITIES IN SOCIAL, POLITICAL AND ECONOMIC LIFE

*Petra Lea Láncos*

<https://doi.org/10.47706/KKIFPR.2021.2.73-87>

*“Every new technology drags behind it the inequalities of the world,  
and usually contributes to them in ways nobody thought to foresee.”  
(Sayers et al., 2021)*

**Abstract** While the recognition of language rights is slow to progress, with the incremental development of language technologies, an increasing number of solutions makes the enforcement of fundamental rights of members of linguistic feasible. Although these developments are to be welcomed, such technologies are inherently 'biased' in the sense that these are developed primarily for 'larger' or more powerful minorities. This situation opens new cleavages besides already existing divisions between majorities and minorities, producing different categories of 'privileged' and disenfranchised minorities. The present paper provides an overview of the development of language technologies that may be harnessed for the enforcement of rights. Mapping the different linguistic minorities affected by these developments, the paper seeks to elucidate how new technologies reshuffle power and interest representation opportunities between language groups. Finally, the paper takes a brief look at the challenges of assimilation of minority languages and cultural appropriation.

**Keywords** language rights, development of language technologies, assimilation of minority languages

## Introduction

The establishment of official or quasi-official languages<sup>1</sup> reinforced or privileged linguistic majorities and led to the emergence of some near mono-lingual states and regions. These developments clearly impacted on the social, political and economic status of linguistic majorities and minorities, respectively. Appointing and enforcing an official language for a political community has several advantages. It ensures the efficient communication of political messages (Mill, 1972□ 392), streamlines administration (Spolsky, 2009□47), promotes the security and development of trade (Ammon, 2007□ 322). Namely, political mobilization across linguistically diverse groups is very costly and often unsuccessful□in order to remain effective, administrations must rationalize internal and external communication (Spolsky op. cit. 170)□finally, the volume of trade within a language group outweighs that between language groups due to the costs of linguistic mediation and the low degree of mutual trust (Foreman-Peck, 2007□Fidrmuc et al., 2006□5). All these factors substantiate the benefit of establishing official languages, while at the same time it creates an uneven playing field for the speakers of the different languages within the same political community. The exclusive position of official languages also makes them 'highly resourced languages', with funds and resources channelled into standardizing, documenting, processing, researching and teaching these languages (Ombui & Muchemi, 2015). Members of linguistic minorities may suffer multiple disadvantages in the context of the social, political and economic system underpinned by the official language□in addition to their native tongue they must learn the official language, an effort that may take away capacities and time from investing in further studies.<sup>2</sup> This, coupled with a possible accent will make it harder for them to compete with the linguistic majority for jobs and political positions, or to promote the political interests of their respective group. In the case of haptic or signed languages, information is readily available in their official spoken and written modality. However, making languages available in the signed or haptic modality is more expensive and cumbersome. This makes progress in these areas slow and patchy, disenfranchising linguistic minorities communicating in these language modalities. The factors mentioned above contribute to the 'secondary' social, political

and economic status of linguistic minorities, leading to possible instances of discrimination in the framework of education, health care, access to justice etc.

While the initial progress in the recognition and protection of language rights (themselves vehicles for the enforcement of other rights) on an international level has gradually lost impetus,<sup>3</sup> the emergence of language technologies has had the intended or side effect of helping linguistic minorities overcome some of the disadvantages outlined above. Starting from hearing aids and braille translators, different language and speech communities have benefited from the emergence of language technologies. The development of these technologies has been incremental, gaining particular impetus in the new millennium. While many of these language technologies are initially or primarily developed for military, intelligence or humanitarian purposes, they gradually find their way into civilian uses, benefitting language communities and the economy at large (Hardach, 2021).

Language technologies facilitate communication between speech communities (regional, bilingual, learner, professional etc.), across languages and language modalities (written, spoken, signed, haptic etc.). Today, language learning with a virtual teacher, speech to text programs, including automatic subtitling, computer assisted and machine translation solutions and automated interpreting devices are already a reality (Sayers et al., op. cit. 7). Future developments point toward augmented reality software, integrating virtual visual and auditory objects into our experience, incorporating solutions for the automated processing, translation and interpretation of 'foreign' speech and text, opening up new opportunities for hitherto disenfranchised linguistic minorities.

Notwithstanding the clear benefits of providing solutions for members of linguistic minorities to participate in social, political and economic life, the development of these new technologies may open new cleavages between speech communities and language groups. In fact, such technologies will not cater to all linguistic minorities, for lack of economy of scale, market failure, lack of profitable civilian demand etc., resulting in a reshuffling of power and interest representation opportunities between language groups.

Building heavily on the LITHME (Language in the Human Machine Era) Forecast Report, the latest study on the future and effects of language technologies, I describe categories of new and emerging language technologies and the ways in which these may assist members of linguistic minorities in overcoming erstwhile disadvantages suffered in the social, political and economic realm. Next, I turn to the issue of the shifting layers of disenfranchisement between previously disadvantaged linguistic minorities in light of the development of language technologies. Finally, I discuss the question whether language technologies may contribute to the assimilation of linguistic minorities or cultural appropriation, and if so, in what sense. For reasons of space, this paper can only provide a brief overview of the language technology landscape and concomitant threats and opportunities.

In this paper, I shall refer to all speech communities and language groups that are not native speakers of the official language of their state of residence, including those groups whose members share a disability owing to which one or all channels of their communication is impaired (e.g. deafness, deaf-muteness, muteness, blindness, deaf-blindness or conditions otherwise leading to visual and/or hearing impairment) as linguistic minorities. Of course, referring to these diverse groups as linguistic minorities is necessarily reductive, since they are usually at the same time national, ethnic or religious minorities etc. Yet it is their disenfranchisement in communicating with authorities and the wider public that allows us to consider them linguistic minorities for the purposes of this paper.

## Language Technologies and New Frontiers

In line with the definition of language technology advanced by LITHME, for the purposes of this paper, I consider language technology to be any technology that can process language passing between humans, or communicate directly with humans. Or, as LITHME puts it, any technology humans can speak through or to (Sayers et al., op. cit.). The former enables translation, interpreting and facilitates the processing of information, while the latter allows us to interact with technology through communication.

As the Report of the Office of the United Nations High Commissioner for Human Rights entitled *Factors that impede equal political participation and steps to overcome those challenges* (para 2) underlines,

“Political and public participation rights play a crucial role in the promotion of democratic governance, the rule of law, social inclusion and economic development, as well as in the advancement of all human rights. The right to directly and indirectly participate in political and public life is important in empowering individuals and groups, and is one of the core elements of human rights-based approaches aimed at eliminating marginalization and discrimination. Participation rights are inextricably linked to other human rights such as the rights to peaceful assembly and association, freedom of expression and opinion and the rights to education and to information.”

Recommending measures to overcome language barriers to promote participation for all members of society, the Report focuses primarily on political participation and expressly includes “the provision of electoral information and voting papers in a range of accessible formats and languages” as well as the provision of “information and educational materials in accessible formats and languages that present the political process” (ibid. paras 13 and 95).

Expanding the participation perspective to encompass also access to public services, justice, education and health care, language technologies have the potential to improve linguistic minorities’ socio-economic status (Eva, 2014; Joshi et al., 2019). They can promote political participation and increase linguistic minorities’ presence in domestic and international economic life (Thomas et al., 2001:27). As such, language technologies may contribute to social justice, democracy and economic growth in general, while at the same time empowering linguistic minorities and facilitating the enforcement of individual rights (information rights, political rights, social and cultural rights, freedom of enterprise etc.) of their members, in particular.

For linguistic minorities, the main use of language technologies may be to ensure their effective participation in social, political and economic life through providing solutions for inter- and intra-language translation and interpreting, and facilitating language learning. Important technologies

in the realm of inter-lingual translation benefitting among others also members of linguistic minorities are website translators (e.g. WeGlot, Google Translate) and machine translation software (e.g. Google Translate, DeepL, Amazon Translate), promoting the enforcement of language rights by making information in foreign languages accessible. Screen readers (e.g. NVDA, Orca) provide an important service to blind or visually impaired readers of online resources, helping to switch from written into the spoken modality, while smart gloves are the first attempt to translate sign language into written text.

In the realm of text to text machine translation, statistical machine translation (SMT) analysing the source language based on statistical models has given way to neural machine translation (NMT), an AI seeking to incorporate characteristics of human thinking by trying to ‘understand’ meaning. NMT uses so-called vector representations for words, continuously learning such representations via training through millions of sentence pairs. The ongoing training of NMT also means that its neural networks are continuously changing and improving. Newly developed encoders render NMT more ‘context aware’, improving its ability to generate more accurate translations.

Automated interpreting devices are also available, such as VERBMOBIL financed by the German Federal Ministry of Research and Technology, developed for the purposes of interpreting between industrial actors in German, English and Japanese. Interpreting demand in warzones has also triggered the development of speech to speech translation software, such as MASTOR, a dialect sensitive solution to mediate between English and Arabic. The Phraselator, a weatherproof hand-held device developed by Applied Data Systems and VoxTec is also used for military purposes and translates English into 40 languages. Finally, Jibbig Translator 2.0 is a free app that translates both speech and text between more than 20 languages (Horváth, 2015). Such devices and software can be of excellent use for members of linguistic minorities, provided that their languages are included in those in which these solutions are made available.

Chatbots (or ‘conversational AI’) represent important opportunities for accessing services, in particular, where such chatbots are available in minority languages (Sayers et al., op. cit. 18). Anticipating questions

and responses of their counterpart, chatbots retrieve pre-compiled responses or carry out actions requested. As such, chatbots may be a cost-effective solution to cater to linguistic minorities in certain areas of public service, but may also be a viable addition to foreign language instruction. Finally, these technologies may be supplemented by speech to text or text to speech functions to accommodate persons with disabilities and impairments, facilitating data input and their translation into the required target language and/or modality. Technology in the area of speech technology is moving towards recognizing not only dialects, but context and emotional nuances of utterances (sarcasm, happiness, sadness etc.) (ibid 20).

The language technologies referred to above promote access to essential public services, such as public service broadcasting, health care, public administration and education. As such, these technologies are an important contribution to the enforcement of fundamental rights of linguistic minorities, in particular, if these technologies are made freely available or are affordable to the members of these communities and are integrated in health care, education and public administration systems. Hence, the development of language technologies may be a vehicle for enforcing citizenship rights, working towards a democracy premised on participation and equal opportunities. At the same time language technologies also facilitate making use of further services provided by private undertakings, including but not limited to assisted living, private health care provision, private and corporate media services and private education etc.

Besides promoting participation of linguistic minorities on a near equal footing in social, political and economic life with majorities, language technologies may also serve to help protect and promote linguistic minorities' culture and language. The documentation of minority languages helps their preservation (Bird, 2020 3505), as well as their processing and research, bolstering efforts at their revitalization through new technologies, such as new media (Eisenlohr, 2004 25). While technologies such as television and radio were long considered triggers of language shift, automated subtitling ensures access in own language. The fact that (majority) language learning is not a must, since mediation takes place,

allows for the possible preservation of language and culture. Conversely, majorities will also have access to these minority cultures, helping build bridges between neighbouring and remote cultures, possibly enriching cultural life and forging mutual understanding.

## Unequal Opportunities for Linguistic Minorities

While speakers of widely used languages such as English benefit from intensive research and development going into language technologies, 'under resourced' languages will yet again become disenfranchised in the human-machine era. Beyond the financial inequalities in access to new technologies, linguistic minorities will have different opportunities to benefit from emerging language technologies, depending on their headcount, linguistic proximity to 'large' or official languages, available resources and the volume of data sets in these languages.

The number of 'speakers' of a given language matters, since this underscores the relevance of such communities for investment and the potential number of future consumers of the language technology to be developed (Thomas et al., *op. cit.* 24). Linguistic proximity plays a role in the quality of the translation and interpreting tools developed; linguistic distance affects the accuracy of mediation between languages (and cultures) in language technologies available to date (*ibid.* 58-59). The availability of resources, such as funds, time, as well as expertise and relevant data in a given language are also decisive for the successful development of language technologies. Finally, the volume of data sets (such as linguistic corpora, speech databases, electronic dictionaries), and in particular, the availability of translations from and to a given language are key. For example, multilingual jurisdictions such as the European Union or Canada provide ample data sets in their official languages and on the topics covered by their legislations, respectively. Meanwhile, minority languages rarely have sufficient data sets to enable the development of applications (*ibid.*)<sup>4</sup> These factors will determine whether technologies will be developed for these languages and in what quality, resulting in a situation where the great dividing line is no longer between official – non-official languages, but between languages with access to new language technologies and those without. As the Mercator Centre's study



put it “the gap between language technology-rich communities and the rest will widen and widen. Languages in which people cannot interact with computers over the Internet will come to be considered inferior, pre-technological” (Thomas et al., op. cit. 24). For linguistic minorities neglected by language technology development, this will mean a further cementing of their ‘secondary’ status in the social, political and economic fabric of society and a possible push towards language shift (Sayers et al., op. cit. 11; Sayers & Láncos, 2017 42; Thomas et al., op. cit. 25).

Thus, linguistic minorities will not be affected by the development of language technologies in the same way (Joshi et al. 2020). For example, while a linguistic minority of their state, certain language groups already benefit from the fact that their mother tongue is the official language in their kin-state, e.g. receiving cultural, political or economic opportunities from such state, or benefitting from the EU official language status of their native language (e.g. Hungarian) (ibid 30; Láncos, 2012 95-96).<sup>5</sup> Similarly, such linguistic minorities will see the advantages of the private or public investment made into the development of language technologies for these official languages.

Meanwhile, as early as 2000 the Mercator Centre highlighted with respect to the smaller and minority languages in the EU that

“a larger number of languages which lack the full array of language resources – linguistic corpora, electronic dictionaries etc. – are in danger of being excluded not from the Internet as it is now, but from many of the processes, including machine translation and other language processing functions, that will increasingly be carried out over the Internet. (...) The development of language technology and language resources for all European languages is therefore essential from the point of view of citizenship and equal opportunity in the information society” (Thomas et al., op. cit. 5, 7).

The importance of language technology is echoed by Kaleimamoowahinekapu Galla, who stresses in her study on Hawaiian language revitalization,

“In spite of technological advancements and the proliferation of digital technology, many Indigenous peoples do not have equal and sustained access and infrastructure to digital technology

in comparison to the global world. It is quite difficult to imagine the survival of Indigenous languages without support from digital technologies, with their ability to record, preserve, analyse, manipulate and transmit languages in a myriad of ways” (Kaleimamoowahinekapu Galla, 2018:100).

But as the LITHME forecast report underlines, even official languages may find themselves in a disenfranchised position, where the small number of speakers does not attract investment (e.g. Latvian) (Sayers et al., op. cit.:9). Official languages’ dialects (e.g. Austro-Bavarian) or national varieties (e.g. Moroccan Arabic), or different registers for distinct speech groups or uses may be also disadvantaged. These may have no standard written spelling (e.g. Swiss German, varieties of Romani) (Thomas et al., op. cit.:18) or only relatively small available data sets (e.g. so-called Low Resource African Languages) (K4all.org, 2021). Since language technologies currently rely on Natural Language Processing, most non-standard forms of language (and, as a corollary, most ‘speakers’ of these languages) are disenfranchised, since “language processing with such language as input suffers from low accuracy and high rates of errors” (Sayers et al., op. cit.:8).

Finally, it is not only non-standardized languages that are disadvantaged in the process of language technology development. As of yet, solutions for signed and haptic languages have as of yet yielded poor results. ‘Smart gloves’ for example are still of poor quality, due to the fact that sign language does not merely consist of hand signals, but include facial gestures and body movements and posture (Quer & Steinbach, 2019:2, 5), which such gloves cannot read. In addition, since signed and haptic languages have their own standard national versions (e.g. Slovak Sign Language) or varieties (e.g. depending on the deaf school or club attended) (ibid:4), or are language-specific (e.g. French Braille) or even domain-specific (e.g. literary Braille, pharmaceutical Braille etc.), and the number of their ‘speakers’ is relatively low, developing technologies for these languages seems particularly slow, difficult and underfunded (Sayers et al., op. cit.:13). Meanwhile, these linguistic minorities do not have the opportunity of modality shifting with the consequence of remaining disenfranchised.

## Do Language Technologies Promote Assimilation and Cultural Appropriation?

A pertinent question arising in connection with linguistic minorities' increased participation in social, political and economic life is whether such tendencies may render them more amenable to assimilation and cultural appropriation. At this point however, the diversity of linguistic minorities and its all-embracing concept employed in this paper make it difficult to discuss the issues of assimilation and cultural appropriation from a general perspective.

In fact, linguistic minorities whose language is an official or majority language coded in a signed or haptic form are actually members of these majority cultures, possibly with their own particular cultures (e.g. deaf culture). Meanwhile, other linguistic minorities will access new cultures through the vehicle of language technologies, including the possibility of language learning, making cultural and linguistic assimilation and language shift possibly imminent. It is in these cases that policy decisions targeting the revitalization and preservation of under resourced minority languages are well placed to prevent language loss through, among others language technology intervention (Dooly, 2010).

Fears of cultural appropriation in connection with the digitalization of minority languages are also prevalent (Kaleimamoowahinekapu Galla op. cit. 106). Making minority languages accessible through technology also makes the cultures such languages 'encode' accessible and, consequently, vulnerable to appropriation, allowing for the potential exploitation of minority communities. One example would be the traditional knowledge appropriated from minorities through access to their language and culture. As MacPherson explains, local communities hold important knowledge about their immediate environment encoded in their languages. Appropriating this knowledge for industrial and commercial purposes without sharing the benefits with such local communities would be a form of cultural appropriation through the vehicle of language facilitated by language technology. While political participation and access to justice boosted by language technologies may allay these threats, more research is needed on how to protect minority cultures in parallel with the empowerment of their individual members through the development of language technologies.

## References

- Ammon, U. (2006). "Language Conflicts in the European Union". *International Journal of Applied Linguistics*, Vol. 16, No. 3
- Bird, S. (2020). Decolonising speech and language technology. In Proceedings of the 28th International Conference on Computational Linguistics, pages 3504–3519, Barcelona, Spain (Online).
- Cordis.europa.eu. (1997). Language Engineering Preparatory Action For Linguistic Resources Organization for Language Engineering. Retrieved from <https://cordis.europa.eu/project/id/LE24017>
- Dooly, M. (2010). "Empowering Language Minorities through Technology Which Way to Go?" *eLearning Papers*, 5.
- Eisenlohr, P. (2004). "Language Revitalization and New Technologies Cultures of Electronic Mediation and the Refiguring of Communities." *Annual Review of Anthropology*. Vol. 33, No. 1. pp. 21-45.
- Eva, J.C. (2014). "Minority languages fight for survival in the digital age". *Phys.org*, 17 February. Retrieved from <https://phys.org/news/2014-02-minority-languages-survival-digital-age.html>
- Fidrmuc, J., Ginsburgh, V.A., Weber, S. (2006). *Economic challenges of multilingual societies*. Working Paper No.11/04. Retrieved from [www.researchgate.net/publication/228651837](http://www.researchgate.net/publication/228651837) [Accessed 27 September 2021]
- Foreman-Peck, J. (2007). *Costing Babel, The contribution of language skills to Exporting and Productivity in the UK*, Cardiff University, 1. Jan.
- Hardach, S. (2021). "The languages that defy auto-translate." *BBC.com*, 23 March. Retrieved from [https://www.bbc.com/future/article/20210322-the-languages-that-defy-auto-translate?utm\\_content=172507833&utm\\_medium=social&utm\\_source=facebook&hss\\_channel=fbp-117630151602650](https://www.bbc.com/future/article/20210322-the-languages-that-defy-auto-translate?utm_content=172507833&utm_medium=social&utm_source=facebook&hss_channel=fbp-117630151602650)
- Horváth, I. (2015). "Információs és kommunikációs technológiák a tolmácsolásban és gépi tolmácsolás." In I. Horváth. (Ed.). *A modern fordító és tolmács*. Budapest: ELTE Eötvös Kiadó. 164 et seq.
- Joshi, P., Barnes, C., Santy, S., Khanuja, S., Shah, S., Srinivasan, A., Bhattamishra, S., Sitaram, S., Choudhury, M., Bali, K. (2019). "Unsung

- Challenges of Building and Deploying Language Technologies for Low Resource Language Communities.” In D.M. Sharma, Pushpak, B. (Eds.) Proceedings of the 16th International Conference on Natural Language Processing. NLP Association of India. pp. 211-219. Retrieved from <https://aclanthology.org/volumes/2019.icon-1/>
- Joshi, P., Santy, S., Budhiraja, A., Bali, K., Choudhury, M. (2020). The State and Fate of Linguistic Diversity and Inclusion in the NLP World Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Retrieved from <https://aclanthology.org/2020.acl-main.560.pdf>
- K4all.org. (2021). *Cracking the Language Barrier for a Multilingual Africa*. Fellowship to develop datasets and strengthen capacities and innovation potential for Low Resource African Languages. Retrieved from <https://www.k4all.org/project/language-dataset-fellowship/>
- Kaleimamoowahinekapu Galla, C. (2018) “Digital Realities of Indigenous Language Revitalization A Look at Hawaiian Language Technology in the Modern World.” *Language and Literacy*. Volume 20, Issue 3, pp. 100-120
- Láncos, P.L. (2012). “Nyelvpolitika és nyelvi sokszínűség az Európai Unióban”. Doctoral dissertation. Budapest Pázmány Péter Catholic University, Doctoral School of Law and Political Sciences.
- MacPherson, S. (2003). TESOL for Biolinguistic Sustainability The Ecology of English as a Lingua Mundi. *TESL Canada Journal*.
- Mill, J.S. (1861). “Considerations on Representative Government.” In *Utilitarianism, On Liberty and Considerations on Representative Government*. Reprint, London Everyman, 1972.
- Ombui, E.O., Muchemi, L. (2015). “Wiring Kenyan Languages for the Global Virtual Age An audit of the Human Language Technology Resources”. *International Journal of Scientific Research and Innovative Technology* Vol. 2 No. 2., pp. 35-42.
- Quer, J., & Steinbach, M. (2019). “Handling Sign Language Data The Impact of Modality.” *Frontiers in Psychology*, Vol. 10., pp 1-8.
- OHCHR. (2014). Factors that impede equal political participation and steps to overcome those challenges. Report of the Office of the United Nations High Commissioner for Human Rights, A/HRC/27/29.

- Sayers, D., R. Sousa-Silva, S. Höhn et al. (2021). *The Dawn of the Human-Machine Era*. A forecast of new and emerging language technologies. Report for EU COST Action CA19102 'Language In The Human-Machine Era'. Retrieved from [www.lithme.eu](http://www.lithme.eu), 10.
- Sayers, D., Láncoş, P.L. (2017). "(Re)defining linguistic diversity: What is being protected in European language policy?" *SKY Journal of Linguistics*, pp. 35-73, 42.,
- Spolsky, B. (2009). *Language Management*, Cambridge: Cambridge University Press
- Thomas, N., King, A., Jones, E.H.G. (2001). *Machine Translation and the Internet*. Retrieved from [https://www.researchgate.net/publication/37146251\\_Machine\\_Translation\\_and\\_the\\_Internet](https://www.researchgate.net/publication/37146251_Machine_Translation_and_the_Internet)

## Endnotes

- 1 Not all jurisdictions enshrine the official status of the language(s) they use in administration formally in law, e.g. the USA has no federal official language, and only certain States have codified the official status of English.
- 2 In *Ádám and Others v. Romania* (Application no. 81114/17 et al) the European Court of Human Rights considered the applications of students' belonging to the Hungarian minority alleging discrimination in the Romanian education system which required that students studying in their mother tongue take an additional two exams in Romanian language and literature to complete their baccalaureate (school leaving qualifications). The Court found that „the fact remains that pupils in the applicants' situation have to pass two more exams than pupils studying in Romanian. That is however the direct and inevitable consequence of the applicants' conscious and voluntary choice to study in a different language and the State offering them such an opportunity. In this connection, the Court observes that the law recognises a right but does not impose an obligation on pupils belonging to a national minority to study in their mother tongue (para 101).
- 3 1948 Universal Declaration of Human Rights (Article 2)□1966 International Covenant on Civil and Political Rights (Article 27)□1989 ILO Convention No. 169 on Indigenous and Tribal Peoples (Articles 28, 30)□1990 Copenhagen Declaration of the OSCE (points 32-34)□1992 UN General Assembly Declaration on the Rights of Persons Belonging to National or Ethnic, Religious and Linguistic Minorities (Articles 1, 2, 4)□1992 European Charter for Regional or Minority Languages□1995 Framework Convention for the Protection of National Minorities□1998 Oslo Recommendations regarding the Linguistic Rights of National Minorities of the OSCE.
- 4 But even languages with millions of speakers but scarce data sets such as for example the African languages Wolof, Yoruba and Ewe are less amenable to developing language technologies, giving rise to a collaborative project with the ambition of building data sets to serve as resources for language processing tools (K4all.org).
- 5 It is worth mentioning however the early, 1996 LE-PAROLE Project funded by the EU, which aimed at building a large-scale harmonized set of corpora and lexica for EU languages. The project not only included selected official languages of the EU but also Catalan (Cordis.europa.eu).