Chapter 8
Minority and diaspora languages as pluricentric languages -An overview

1. Definition of minority languages

> A minority is a "group numerically inferior to the rest of the population of a State, in a non-dominant position, whose members – being nationals of the State – possess ethnic, religious or linguistic characteristics differing from the rest of the population and show, if only implicitly, a sense of solidarity, directed towards preserving their culture, traditions, religion or language". [1] [Capotori, United Nations].

This definition uses three key criteria: (1) A group is numerically small compared to the other (linguistic) groups in a political entity (country, region etc.); (2) it is distinct from the official language(s) of that state concerned[2] and (3) the group is marked by specific (linguistic) features and is willing to preserve them. The ECRML also notes that this definition "does not include either dialects of the official language(s) of the state, nor the languages of migrants."[3] Minority languages may be a small community in a single country, in two or more countries or a linguistic minority in one country, but the majority in another country.

2. The minority languages in Europe

In March 2025, the European Charter for Regional or Minority Languages (ECRML) covers 88 individual minority or regional languages that are the languages of 214 political entities:

1. Albanian	13. Bunjevac	25. Finnish
2. Amazigh	14. Burgenland Croatian	26. Franco-Provençal
3. Aragonese	15. Catalan	27. French
4. Aranese (Occitan)	16. Cornish	28. Frisian
5. Armenian	17. Crimean Tatar	29. Gagauz
6. Assyrian	18. Croatian	30. Galician
7. Asturian	19. Cypriot Maronite Arabic	31. German
8. Basque	20. Czech	32. Greek
9. Belarusian	21. Danish	33. Hungarian
10. Bosnian	22. Darija	34. Inari Sami
11. Boyash Romanian	23. Extremaduran	35. Irish
12. Bulgarian	24. Fala	36. Istro-Romanian
37. Italian	55. Meänkieli	73. Skolt Sami
38. Karaim	56. Moldovan	74. Slovak
39. Karelian	57. Moravian Croatian	75. Slovene
40. Kashub	58. North Frisian	76. South Sami
41. Krimchak	59. North Sami	77. Swedish

42. Kurdish	60. Papiamento	78. Tatar
43. Kven/Finnish	61. Polish	79. Turkish
44. Ladino	62. Portuguese	80. Ukrainian
45. Lemko	63. Romanes	81. Ulster-Scots
46. Leonese	64. Romani	82. Upper Sorbian
47. Limburgish	65. Romanian	83. Valencian
48. Lithuanian	66. Romansh	84. Vlach
49. Low German	67. Russian	85. Welsh
50. Low Saxon	68. Rusyn (as Ruthenian)	86. Yenish
51. Lower Sorbian	69. Saterland Frisian	87. Yezidi
52. Lule Sami	70. Scots	88. Yiddish
53. Macedonian	71. Scottish Gaelic	
54. Manx Gaelic	72. Serbian	

3. Pluricentricity and minority languages

In general, the concept of pluricentricity cannot be applied to minority languages.[4] This is due to the small number of speakers, the isolation from the "mother language" and quite often the inability to form a linguistic centre, which is the prerequisite for the development of a distinct variety and its spread throughout the community. Minority languages mostly use the written variety of the mother variety for teaching and public communication.

However, there are two exceptions to this rule. Scetti (2016) was able to show that the development of a distinct migrant variety is possible, if the number of speakers is large enough. This is the case with the migrant community in Montreal, Canada.

The second exception is when a regional language with many speakers is denied the correct status. Examples for this are Hungarian in Romania, Slovakia and Serbia. Languages whose exterritorial (regional) varieties have a large number of speakers and who are denied the correct status can be considered as regional languages and are therefore pluricentric.

4. Diaspora languages and pluricentricity

4.1. Definition of diaspora languages

Unlike minority languages, diaspora languages usually do not have a clearly defined territory, as their speakers either live scattered in large urban agglomerations or in small communities with little communication between them. The languages of this category usually are often characterised by a lack of mutual intelligibility and for this reason do not meet the criterion (7), which is essential for the definition of PCLs. This situation hinders codification, standardization and the development of linguistic centres that support the maintenance and development of these languages. Their status as PCLs is

disputable because they are highly fragmented and the difficulty to form a linguistic centre

4.2. The pluricentricity of selected diaspora languages

4.2.1. Kurdish - A suppressed and fragmented language trying to survive and remain visible

There are mixed cases, such as *Kurdish*, which occupies both a contiguous area in Turkey, Iraq and Iran but also has a large diaspora community that is the result of decades of emigration from regions where Kurds have been persecuted. The only way to apply the concept of pluricentricity to Kurdish, is to use the concept of high ethnolinguistic awareness, which "holds the language community together". And indeed, Kurdish communities show a high awareness about their language and their ethnic affiliation.

Kurdish is a lingoid that is spoken in *Armenia, Azerbaijan, Georgia, Iran, Iraq, Lebanon, Syria, Turkey*, and *Turkmenistan*. Northern Iraq is the only territory where it has the status of an official regional language (Kurdish autonomous region), apart from being recognised as minority language in Armenia and an recognised language of provincial identity in five villages of Azerbaijan. It is often portrayed as "the largest language without a state of its own"[5]. Kurdish falls into several categories that are not favourable for language maintenance, language ausbau and status improvement. On the one hand, it is a highly fragmented language with three major varieties – *Kurmanji (*Northern Kurdish) in Turkey and in Northern Iraq, *Sorani* (Central Kurdish) and *Palewani* (Southern Kurdish), which are not or almost not mutually intelligible and are separate languages from a linguistic point of view. To make the situation even more complicated, there is no common orthography and no common codification and language planning. There is, however, a very strong ethno-linguistic awareness that leads the Kurdish population themselves consider as one people. Based on this criterion, Kurdish can be considered a PCL (just as Armenian and Yiddish).

Kurdish is included as there is a large Kurdish-speaking community in Turkey (and in many European countries), which is a member of the Council of Europe (COE). Despite the provisions of the COE in respect to human rights, the Turkish authorities are massively persecuting Kurdish speakers (after a short period of political thaw) and the language is currently (again) banned in many public domains. Researchers dealing with the language planning measures of different Turkish governments call this policy the "*invisibilisation*" of Kurdish (Haig, 2004). The attitude is attributed to Turkish nationalism ever since Turkey was founded in 1921. There are no accurate numbers available on how many Kurdish speakers there are in Turkey as the Turkish government has ceased to make demographic surveys based on the criteria of language and ethnicity in

1965. Estimations speak of 10-12.5 million speakers mainly living in the southeast of the country.[6]

The Turkish constitution (Art. 42.9) stipulates that "Turkish nationals may not be taught and educated in any language other than Turkish in the educational institutions."[7] Despite this, there has been a short period of tolerance – the "Kurdish opening" - between 2006 and the so called "putsch" in 2016. During this period, private television stations were allowed to broadcast programmes in Kurdish and from 2009 onwards, there was even a 24-hour Kurdish channel opened by the state-owned Turkish television. The "opening" was a relative one as the teaching of Kurdish in educational TV-programmes was reduced to 45 minutes a day and four hours per week.[8] However, the letters x, y and w – necessary for writing in Kurdish - were banned until 2013 in public texts as they are not part of the Turkish alphabet. After the failure of the peace negotiations between the PKK and the Turkish government in 2015 and even more so after the "putsch" in 2016 the Turkish government has repealed many of the previous measures. The use of Kurdish in the public domain was again restricted, elected representatives and teachers of Kurdish and other minority languages in Diyarbakir were dismissed[9], the Kurdish Institute in Istanbul that had provided language classes has been shut down.[10]

At present, institutions that provide classes in Kurdish are again criminalized by the government and portrayed as supporters of the forbidden PKK guerilla.[11] As a result, teachers are not allowed to speak any language other than Turkish, even in primary school classes where the students have no knowledge of Turkish and do not understand it because their mother tongue is a minority language like Kurdish.[12] This surely is a highly counterproductive form of pedagogy and – judging from the other measures taken - intended to create a kind of linguistic genocide. Kurdish in Turkey is an extreme example of linguistic and ethnic suppression. Speakers of Kurdish therefore consider the right to speak their language in public without being persecuted as an act of liberation: "Speaking Kurdish is important to me and to almost all Kurdish people because it represents our liberty."[13] The situation of this language is one of the most precarious of all PCLs in Europe and a sad example of an ethnic group that has been denied its rights for a long time.

4.2.2. Assyrian, Neo-Aramaic:

Communities of this language are found in *Turkey, Iran, Iraq, Syria, Azerbaijan, Iraqi Kurdistan*. It is a highly fragmented language, which has official recognition only in Kurdistan in the North of Iraq and in Armenia under the name of "*Assyrian Neo-Aramaic*" and "*Chaldean Neo-Aramaic*" which are linguistically different.[14] Gonzalez (2013) points out that an Assyrian spoken

koine developed in Iraq, and serves "as the main interdialectal means of communication used by most modern Aramaic dialect speakers."

4.2.3. Yiddish and Judaeo-Spanish: The language(s) of Jewish communities in Europe

4.2.3.1. Yiddish:

Yiddish is the historical language of the Ashkenazi Jews, who settled all over Europe but were persecuted throughout history, which ultimately led to the Holocaust in WWII. The language is a typical diaspora language, as the Jewish communities are/were scattered across many regions and countries and developed a Western and Eastern variety. Due to its occurrence, it could be considered as a PCL, but it is only recognized as a European minority language under the Charter of European Minority Languages in seven countries: *Bosnia and Herzegovina, Finland, Netherlands, Poland, Romania, Slovakia, Sweden, and Ukraine*. Despite its recognition as a minority language, the lack of a substantial centre and the lack of support from European nations, it is to be considered a diaspora language and not a PCL.

4.2.3.2. Judaeo-Spanish (Ladino):

This language derives from Spanish of the 15th century when the Jewish (Sephardic) community in Spain was banned from Spain due to the edict of Isabella I in 1492. The community dispersed throughout the Mediterranean countries, especially in the Ottoman Empire with Salonika (Thessaloniki, Greece) and Constantinople (Istanbul, Turkey) as their main centres.[15] Through contact with local languages, several varieties developed in the different countries, which caused a decline in mutual intellibility. Today it is only spoken by small minorities in the Balkans, Greece and Turkey, and it is seriously threatened with extinction. Bosnia and Herzegovina is the only European country where it has the status of a minority language. The only support is found in courses given at some universities in Israel and the US and the announcement by the Spanish Royal Academy (RAE) in 2017 that it would establish a special academy for Judaeo-Spanish in Israel. The language suffers, like all other diaspora languages, from the lack of a central agency and a school system that supports it. Judaeo-Spanish, like Yiddish, is a diaspora language rather than a PCL.

5. European Languages with a high amount of internal linguistic fragmentation that makes their status as a PCL unclear, impedes their standardization and makes the survival of the language difficult

The languages of this category are characterised by a high degree of

internal fragmentation, which reduces mutual intelligibility and obstructs the development of a common standard variety. This in turn often leaves the single varieties at the level of informal spoken languages (vernaculars), whose functionality is reduced to a few domains.

5.1. Occitan:

Occitan, in its *Aranese Occitan* variety, has the status of an officially recognised regional language in Catalonia, Spain. Occitan is also spoken in France and Italy. The language is heavily fragmented as there are six varieties: *Auvergant, Gascon, Provençal, Limousin, Languedocien, Vivaro-Alpin* in France and *Aranese* in Spain. The situation is aggravated by the fact that France signed, but never ratified the European Charter of Minority Languages[16] and leaves Occitan unprotected and unsupported in France, where most speakers live. Carrera (2019) provides a thorough update of the present situation of Occitan and an overview of attempts to create a common orthography.

As with many fragmented languages there is a discussion going on among linguists whether the lingonym is an umbrella term for several single languages or a language covering all varieties (Holtus/Metzelin/Schmitt, 1991). Sumien (2006) deals with this extensively. There is no unifies written standard and within one and the same variety (such as in Béarnais and Gascon)[17] there are even competing spelling norms. There are two main orthographic norms: The *Mistral writing system* (developed by the writer Frederic Mistral in his dictionary of 1879 and the *classical writing system*, which is based on medieval Occitan. An updated form of the classical writing system was adopted by the *Institut d'Estudis Occitans.* However, as Sumien (2012) shows in detail, other spelling systems for individual varieties were introduced and used in parallel, but have also contributed to the development of koiné varieties in some sub-varieties that have helped the language to survive. Efforts to create a standard variety are still going on. Due to its fragmentation and the lack of standardization, the status of Occitan as a PCL is in doubt. Too many protagonists seem to fight for their particular interests instead of following the model of the development of Rumantsch in Switzerland (Schmidt, 1982 / Darms, 1989) where a united written language across seven varieties has successfully been created and is now used throughout the community.

5.2. Frisian:

This language is split into three varieties (*West Frisian, North Frisian, and Saterland Frisian*), which are not mutually intelligible. In addition to their geographic distance, the three varieties are also influenced by different languages: West Frisian is strongly influenced by Dutch, Saterland Frisian and

North Frisian by German, and Low German. The status of Frisian as a single language is therefore disputed by scholars who prefer to speak of the "Frisian *languages*". Its status as a PCL is in doubt. The social visibility of the three varieties/languages is guaranteed via the recognition as regional or minority languages in the Netherlands and Germany.

5.3. Sami:

This lingonym is an umbrella term for, "depending on the nature and terms of division, ten or more Sami languages"[18]. The language area extends across Norway, Sweden, Finland, and Russia with neighbouring varieties mutually intelligible but not with varieties that are further apart. Moreover, Sámi has minority language status of in all countries and very few speakers (Rueter (2019)): Sami languages and number of speakers: *Inari Sami* ~300, *Kildin Sami* ~600, *Lule Sami* ~2.000, *Northern Sami* ~ 26.000, *Pite Sami* ~20; *Skolt Sami* ~420, *South Sami* ~600, *Ter Sami* ~2; *Ume Sami* ~20.

Sami cannot considered as a PCL due to its fragmentation. Social visibility and support are guaranteed via the recognition as regional or minority languages to a greater or lesser degree in the four countries where the language is native. For an insight into the linguistic situation of these languages / varieties, see Rueter (2019).

Footnotes:

[1] https://digitallibrary.un.org/record/10387?ln=en [acc. 28.03.2025]

[2] https://rm.coe.int/1680695175 ECRML, Art. 1

[3] Ibid.

[4] See Langer 2021b for a failed attempt to create pluricentricity for Friesian, which supports this statement.

[5] McDowall (2005) and www.mashallahnews.com/language/languages-stateless-nation.html [acc. 07.03.2025]

[6] See https://en.wikipedia.org/wiki/Languages_of_Turkey [acc. 15.03.2025]

[7] https://global.tbmm.gov.tr/docs/constitution_en.pdf [acc. 07.03.2025]

[8] Skutnabb-Kanga, Tove / Fernandes, Desmond (2008).

[9] https://www.nytimes.com/2017/06/29/world/middleeast/amid-turkeys-purge-a-renewed-attack-on-kurdish-culture.html [acc. 07.03.2019]; See also: https://en.wikipedia.org/wiki/Kurdish%E2%80%93Turkish_conflict_(2015%E2%80%93present) [acc. 07.03.2025];

[10] https://www.thenation.com/article/in-turkey-repression-of-the-kurdish-language-is-back-with-no-end-in-sight/ [acc. 06.03.2019] / https://en.wikipedia.org/wiki/Kurdish%E2%80%93Turkish_conflict_(2015%E2%80%93present) [acc. 06.03.2025]

[11] A guerrilla group that is fighting for the independence of the Kurdish areas and is considered as terrorist group by the Turkish government and by some European governments.

[12] ibid. Fn 50.

[13] www.mashallahnews.com/language/languages-stateless-nation.html [acc. 07.03.2025).

[14] See Beyer (1986) and Khan (2018) for an overview.

[15] See Gabriel / Grünke (2018) for the general situation.

[16] The latest attempt to ratify the Charter in France in the National Assembly failed in 2015 due to the resistance of the political right-wing parties. [https://www.francetvinfo.fr/societe/le-senat-dit-non-a-la-charte-europeenne-des-langues-regionales_1712811.html, acc. 21.92.2025]

[17] Joubert (2015: 164).

[18] https://en.wikipedia.org/wiki/S%C3%A1mi_languages [acc. 25.02.2025]