

CHAPTER 18

Pluricentric languages worldwide

1. The objectives of this chapter

This chapter provides a list and brief descriptions of all PCLs identified so far. The length of the description varies depending on the available literature. The list is intended to provide an overview about the PCLs and their different language situations. It also addresses the difficulties that exist in some PCLs when it comes to defining their status.

Languages whose status as pluricentric language is unclear are marked with an asterics (*). A short description is given of the reasons for the unclear status.

The data referring to each PCL come from different versions of Wikipedia (en.wikipedia, fr.wikipedia, de.wikipedia), the Encyklopedia Britannica, the literature mentioned at the end of the description of each PCL and there was also help from ChatGPT. However, the information from the various sources sometimes contradicts each other. This means that the information in this chapter should be checked against other sources. We welcome feedback if some of the information proves to be inconsistent or incorrect. We thank you for it!

1. Albanian

A. Name of the PCL:	Albanian
B. Endonyms - Other name(s) of the PCL:	Shqipja (standarde) e Shqipërisë, Shqipja standarde në Shqipëri, Gjuha shqipe në Shqipëri
C. Language family:	Indogermanic
D. Demography:	Albania: 2.347.700 (2023); Kosovo: 1.485.170 (2024 census); North Macedonia: 481 800 (2023); Germany: 597.000 (2023); Italy: 481.800 (2023); Greece: 374.926 (2021); Switzerland: 306.507 (2023); Serbia: 69.570 (2022); Montenegro: 32.715 (2923); Total: ca. 6.483.650
E. Status-OFL in:	<i>Albania</i> (Albanian Albanian, OFL (official language)); <i>Kosovo</i> (Kosovarian Albanian, OFL); <i>Macedonia</i> (Macedonian Albanian, COFL (Co-official language)); <i>Montenegro</i> , (Montegrin Albanian, COFL), <i>North Macedonia</i> (North Macedonian Albanian, COFL)
F. Recognised minority language (RML) in:	Italy, Serbia, Croatia, Romania
G. DV: (Dominant variety)	Albanian Albanian
H. NDVs of the PCL:	<i>Kosovo</i> (Kosovarian Albanian, OFL); <i>Macedonia</i> (Macedonian Albanian, COFL (Co-official language)); <i>Montenegro</i> , (Mon-

	tegrin Albanian, COFL), <i>North Macedonia</i> (North Macedonian Albanian, COFL)
I. Standardisation, regulated by:	Albania: Academy of Sciences of Albania; Kosovo: Academy of Sciences and Arts of Kosovo; The varieties of Albania and Kosovo are codified or in the process of codification (Kosovo).
J. Alphabet / WS (Writing system)	Latin alphabet
K. Standard form:	Standard Albanian: <i>gjuha letrare shqipe</i> or simply <i>shqip</i>
L. Linguistic fragmentation:	The Gheg variety has no standardised form and is only a spoken variety
M. Mutual intelligibility:	Full mutual intelligibility with other Standard Albanian varieties
N. Corpora:	Albanian National Corpus (ANC) 20 million words, https://web-corpora.net/AlbanianCorpus/ AlCo – “Albanian Corpus” suite (Besim Kabashi et al.) – Reference corpus: ~100 million word-tokens, annotated https://www.besim-kabashi.net/resources.html OPUS (Open Parallel Corpus) covering billions of tokens across 100+ languages, including Albanian↔X pairs: https://opus.nlpl.eu/
O. Dictionaries and grammars:	Kostallari, Androkli (ed.) (1980): <i>Fjalor i gjuhës së sotme shqipe</i> . Tiranë: [s.n.]. xxv, 2273 pp. Thomai, Jani; Samara, Miço; Haxhillari, Pavli; et. al. (eds.).(2006): <i>Fjalor i gjuhës shqipe</i> . Tiranë: Akademia e Shkencave e Shqipërisë, Instituti i Gjuhësisë dhe i Letërsisë, 2006. 1252 pp.; ISBN 978-999438175X. Newmark, Leonard. <i>Oxford Albanian-English Dictionary</i> . Oxford: Oxford University Press, 1999. 1056 pp.; ISBN 0198603223 / 9780198603221. Monolingual Albanian online dictionary; http://www.fjalori.shkenca.org/ ; <i>Fjalori Drejtshkrimor i Gjuhës Shqipe</i> (1976) (Orthographic Dictionary of the Albanian Language)[146] <i>Dictionary of Today's Albanian language</i> (Fjalori i Gjuhës së Sotme Shqipe) (1980). *** Camaj, Martin (1984): <i>Albanian Grammar: With Exercises, Chrestomathy and Glossaries</i> . Translated by Leonard Fox. Wiesbaden: Otto Harrassowitz Verlag, 337 pp.; ISBN 3447024674 / 9783447024679. Newmark, Leonard; Hubbard, Philip; Prifti, Peter R. (347): <i>Standard Albanian: A Reference Grammar for Students</i> . Stanford, CA: Stanford University Press. 347 pp.; ISBN 0804711291 / 9780804711296. Buchholz, Oda & Fiedler, Wilfried. <i>Albanische Grammatik</i> . Leipzig: VEB Verlag Enzyklopädie, 1987.
P. Language situation	Strong rejection of the pluricentric character of Albanian by the cultural and linguistic elites of Albania.

Q. Awareness / Function for identity:	High degree of identity function. Strong linguistic awareness among the elites.
R. Language Policy:	Albania: Ibania: Albanian is enshrined in the constitution as the national language; the rights of minority languages are also guaranteed. No other language policy measures. Kosovo: Albanian is the national language. Law on the Use of Languages (2006/2007) mandates equal treatment of Albanian and Serbian in all public institutions. North Macedonia: The Macedonian language and Albanian are the official languages; in municipalities where at least 20 % of the population belongs to a national minority, that minority language is also official.
S. Note on the pluricentricity of the language	Standard pluricentricity with the language officially recognised language in 3 nations; Reluctance to codify the other NVs than the DV
T. Literature (selection):	Albana Muco (2018): <i>Albanian as a pluricentric language</i> . Lumnije JUSUFI (2018): <i>Pluricentric Developments of Albanian between National Unity and Linguistic Heterogeneity</i> : Lumnije Jusufi (ed.)(2018): <i>The Potentiality of Pluricentrism. Albanian Case Studies and Beyond</i> .

2. Arabic

A. Name of the PCL:	Arabic
B. Endonyms - Other name(s) of the PCL:	al-'arabiyyah in written Arabic العَرَبِيَّة (al-'arabiyyah) Modern Standard Arabic and Classical Arabic is also called al-'arabiyyatu l-fuṣḥā (العَرَبِيَّةُ الْفُصْحَى "the eloquent Arabic") or simply al-fuṣḥā (الْفُصْحَى).
C. Language family and region:	Afro-Asiatic, Semitic language; native to: See point E:
D. Demography	Third most widespread official language with ca. 411 mio. speakers, one of six official languages of the UNO, liturgical language of the Islam.
E. Status-OFL in:	Official language in 28 nations and territories: Algeria, Bahrain, Chad, Comoros, Djibouti, Egypt, Eritrea, Iraq, Jordan, Kuwait, Lebanon, Libya, Mali, Mauritania, Morocco, Oman, Palestine, Qatar, Saudi Arabia, Somalia, Sudan, Syria, Tunisia, United Arab Emirates, Yemen, Zanzibar (Tanzania), Sahrawi Arab Democratic Republic (partially recognized state), Somaliland. International organisations: African Union, Arab League, Organisation of Islamic Cooperation, United Nations
F. Status as minority language (RML) in:	Cyprus, Israel, Mali, Niger, Philippines, Senegal, South Africa
G. DV:	Saudi Arabia, Iraq
H. NDVs of the PCL:	Official language in 28 nations and territories: See Point E.
I. Standardisation / Codification, regulated by:	Done by the 15 language academies: Done by the 15 language academies: Arab Academy of Damascus (Syria); Jordan Academy of Arabic (Jordan); Academy of the Arabic Language in Cairo (Egypt); Iraqi Academy of Sciences (Iraq); Institute for Studies and Research on Arabization (Morocco); Tunisian

	Academy of Sciences, Letters, and Arts (Tunisia); Academy of the Arabic Language in Khartum (Sudan); Palestinian Academy of the Arabic Language (State of Palestine); Supreme Council of the Arabic language in Algeria (Algeria); Mogadishu Institute of Languages (Somalia); Academy of the Arabic Language in Libya (Libya); Academy of the Arabic Language in Israel (Israel); Lebanese Academy of Sciences (Lebanon); Arabic Language Academy in Sharjah (United Arab Emirates); King Salman Global Academy for Arabic Language (Saudi Arabia);
J. Alphabet / WS	Arabic script, except for Maltese Arabic, which uses Latin script.
K. Standard form:	Modern Standard Arabic (MSA), or الفصحى (al-Fuṣḥā) in Arabic.
L. Linguistic fragmentation:	Fragmentation exists geographically (East-West) and socially through strong diglossia between Modern Standard Arabic (MSA) that is taught in school and the national varieties of spoken Arabic that are the primary language of most speakers.
M. Mutual intelligibility:	A minimum level of comprehension exists between all Arabic national varieties, geographic proximity increases it. The written form helps to overcome differences in pronunciation. Despite mutual incomprehensibility among differing spoken varieties, for political reasons, Arabs mostly maintain that they all speak a single language. (Nizar Y. Habash, 2022)
N. Dictionaries and grammars:	Wehr, Hans (1952/1959): Arabisches Wörterbuch für die Schriftsprache der Gegenwart and A Dictionary of Modern Written Arabic (1961) (engl. version ed. by J.M Cowan). Wehr, Hans. A Dictionary of Modern Written Arabic (1994): Editor (English edition): J. Milton Cowan. 4th Edition. Publisher: Spoken Language Services. Year: 1994 (latest widely used edition). ISBN: 9780879500034 Almaany Multilingual Dictionary: Online. https://www.almaany.com Baalbaki, Rohi (2005): Al-Mawrid: A Modern Arabic-English Dictionary. Publisher: Dar El-Ilm Lilmalayin. Year: 2005 (several editions). ISBN: 9789953630 *** Ryding, Karin C. (2005): A Reference Grammar of Modern Standard Arabic. Publisher: Cambridge University Press. ISBN: 9780521771511, 734pp. Alhawary, Mohammad T. (2011): Modern Standard Arabic Grammar: A Learner's Guide. Publisher: Wiley-Blackwell. ISBN: 9781405198433 Brustad, Kristen; Al-Batal, Mahmoud; Al-Tonsi, Abbas (2010): Al-Kitaab fii Ta'allum al-'Arabiyya (Series). Publisher: Georgetown University Press. 3 Volumes, each with textbooks, DVDs/audio, and companion grammar. ISBN (Book 1): 9781589016323
O. Corpora:	Arabic speech corpus (https://en.arabicspeechcorpus.com/), (arWaC) is an Arabic corpus made up of texts from the web (https://www.sketchengine.eu/arabic-web-corpus-wac/); Arabic Speech Corpus at CLARIN: http://hdl.handle.net/

	20.500.14106/2561; International Corpus of Arabic: https://www.bibalex.org/ica/en/about.aspx
P. Awareness / Function for identity:	High degree of identity function. Also in connection with religion as Arabic is the language of the Quran. The Arab League opened a bureau of Arabisation whose task is the Arabisation of the sciences and creating terms of the Arab language.
Q. Language Policy:	Arabisation became a postcolonial policy in Tunisia, Algeria, Morocco and Sudan.
R. Language situation	Strong diglossia and devaluation of all non-dominant (spoken) varieties in the various countries that share the Arabic language. Strong reluctance to codify the (non-dominant) national varieties for fear of jeopardising the unity of the language.
S. Note on the pluricentricity of the language	Standard pluricentricity with the language officially recognised language in 28 nations and territories; Strong diglossia and reluctance to codify NVs and primarily spoken everyday varieties.
T. Literature (selection):	Abd-el-Jawad, Hassab R.S. (1992): Is Arabic a pluricentric language? Aboelezz, Mariam (2016): A History of the Arabic Language and the origin of non-dominant varieties of Arabic. Clyne, Michael / Kipp, Sandra (1999): Pluricentric languages in an immigrant context: Spanish, Arabic and Chinese; Habash, Nizar Y. (2010): Introduction to Arabic Natural Language Processing; Ibrahim, Zeinab (2009): Beyond Lexical Variation in Modern Standard Arabic: Egypt, Lebanon and Morocco; Ibrahim, Zeinab / Karatsolis, Andreas (2013): Excluding Speakers of the Dominant Variety: Two Cases from Greek and Arabic; Marley, Dawn (2012): Competing varieties of French and Arabic in Morocco;

3. Azerbaijani/Azəri

A. Name of the PCL:	Azerbaijani or Azeri
B. Endonyms - Other name(s) of the PCL:	Azərbaycanca, آذربایجانجا, Азәрбајҹанҹа) or Azeri also as Azerbaijani Turkic or Azerbaijani Turkish (Azərbaycan türkcəsi, آذربایجان تۆرکجه‌سی, Азәрбајҹан түркчәси)
C. Language family and region	Turkic language, Oghuz sub-branch
D. Demography and language spread	Ca. 24 mio. speakers in 2022. native to Azerbaijan, Iran, Russia, Turkey, Iraq, Gorgia. Second most spoken language in Iran with about 13 mio. speakers, but no official recognition.
E. OFL in	Azerbaijan, Dagestan (Russia), Organization of Turkic States
F. Recognised minority language (RML) in:	none
G. DV:	Azerbaijan Azerbaijani
H. NDVs of the PCL	Dagestan (Russia)
I. Standardisation, regulated by:	North Azerbaijani: Azerbaijan National Academy of Sciences South Azerbaijani: No regulatory body
J. Alphabet / WS:	Azerbaijan: Latin script (Azerbaijani Latin alphabet)

	Iran: Perso-Arabic script (Azerbaijani Arabic alphabet) Russia: Cyrillic script
K. Standard form:	Shirvani (For North Azerbaijani variety in Republic of Azerbaijan) used in Azerbaijan and Dagestan (Russia) and closely related to modern-day Istanbul Turkish Tabrizi (For South Azerbaijani variety in Iranian Azerbaijan)
L. Linguistic fragmentation:	High amount of fragmentation in regional varieties.
M. Mutual intelligibility:	High degree of mutual intelligibility between North and South Azerbaijani. To a certain extent, speakers of Turkish and Azerbaijani can communicate with each other.
N. Dictionaries and grammars:	Azerbaijani-English dictionary: https://www.lexilogos.com/english/azeri_dictionary.htm Multilingual Turkish Dictionary Azerbaijani to many languages: https://www.adasozluk.com/post/words/list/az/en Ameli (2021). <i>Loghatnâme-ye Torki-ye Āzarbāyġāni</i> (Persian-Azeri dictionary)
O. Corpora:	azWaC: A general-domain web corpus of North Azerbaijani, assembled in January 2012 by crawling Size: 94 million words: sketchengine.eu OPUS Project: A large open-content collection of aligned texts from varied domains (news, subtitles, web-crawl). Coverage: 747 languages, 1 212 corpora, ~58.8 billion sentence pairs overall. OPUS Project. A large open-content collection of aligned texts from varied domains (news, subtitles, web-crawl). SETIMES: Balkan news bitext (Azerbaijani↔English), 17.6 M sentence fragments, 425.9 M tokens total. OpenSubtitles, Global Voices, Tatoeba, JW300, etc. https://opus.nlpl.eu/legacy/ . Access: OPUS website, CLARIN, Sketch Engine
P. Awareness / Function for identity:	High identity function of the term "Azerbaijani", which replaced the language name "Turkish" in 1936-1937 and became the official national language in 1956.
Q. Language Policy:	Language policy that emphasizes the use and promotion of the Azerbaijani language (also known as Azeri) as the official state language. Azerbaijani is the language of instruction in most public schools and universities. Government documents, court proceedings, and official communication are conducted in Azerbaijani. Azerbaijani is the dominant language in TV, radio, and print media. The government formally supports the linguistic rights of minorities, though critics have raised concerns over limited support in practice.
R. Language situation	There are no known language conflicts etc.
S. Note on the pluricentricity of the language	Standard pluricentricity of the officially recognised language in Azerbaijan and a regionally recognised language in Dagestan, Russia
T. Literature (selection):	Ameli, Seyed Hassan (2021). <i>حروف آذربایجانی: جلد ۱</i> (in Persian and Azerbaijani). Mohaghegh Ardabili. Goyushov, Altay: "The Language of Azerbaijan: Turkish or Azer-

	<p>bajjani?". Baku Research Institute.</p> <p>Schönig, Claus (1998). "Azerbaijani". <i>The Turkic Languages</i>. London: Routledge.</p> <p>Knüppel, Michael (2010) [2000]. "Turkic languages of Persia: An overview". In Yarshater, Ehsan (ed.). <i>Encyclopædia Iranica</i> (Online ed.). Encyclopædia Iranica Foundation.</p> <p>Rahmati, Nemat (1998). <i>Aserbaidschanisch Lehrbuch : unter Berücksichtigung des Nord- und Südaserbaidschanischen</i>. Korkut M. Buğday. Wiesbaden: Harrassowitz.</p> <p>Ameli, Seyed Hassan (2021). لغتنامه ترکی آذربایجانی: حروف آ (جلد ۱). (in Persian and Azerbaijani). Mohaghegh Ardabili.</p> <p>Goyushov, Altay: "The Language of Azerbaijan: Turkish or Azerbaijani?". Baku Research Institute.</p> <p>Schönig, Claus (1998). "Azerbaijani". <i>The Turkic Languages</i>. London: Routledge.</p> <p>Knüppel, Michael (2010) [2000]. "Turkic languages of Persia: An overview". In Yarshater, Ehsan (ed.). <i>Encyclopædia Iranica</i> (Online ed.). Encyclopædia Iranica Foundation.</p> <p>Rahmati, Nemat (1998). <i>Aserbaidschanisch Lehrbuch : unter Berücksichtigung des Nord- und Südaserbaidschanischen</i>. Korkut M. Buğday. Wiesbaden: Harrassowitz.</p>
--	--

4. Aymara

A. Name of the PCL:	Aymara
B. Endonyms - Other name(s) of the PCL:	Aymar aru (Aymara language)
C. Language family and region:	Aymaran; native to Bolivia, Chile, Peru, Argentina
D. Demography and language spread	Bolivia: ca. 1,5-2,0 million speakers, Peru: ca. 500.000; Chil: ca. 25.000; ca. 2,5 mio. speakers in 2022. native to Bolivia, Peru, Chile, Argentina
E. Status-OFL in:	Bolivia and Peru
F. Status-RML in:	Chile
G. DV:	Bolivian Aymara
H. NDV of the PCL	Peru (official language only in Aymara dominated areas)
I. Alphabet / WS	Latin script, since 1984 Alfabeto Unificado is officially used
J. Standard form:	Bolivian Aymara
K. Standardisation regulated by:	<p>Bolivia: Academia Mayor de la Lengua Aymara (based in La Paz, Bolivia); Plurinational Institute of the Study of Languages and Cultures (IPELC). Oversees the development and preservation of indigenous languages, including Aymara; Ministry of Education (Bolivia): Promotes the use of Aymara in education, public institutions, and media under the country's plurinational framework.</p> <p>Peru: Ministry of Education – Dirección de Educación Intercultural Bilingüe (DEIB). Responsible for standardizing indigenous languages, including Aymara.</p>

L. Mutual intelligibility:	All regional varieties and NVs are mutually intelligible
M. Linguistic fragmentation:	Fragmentation in regional varieties.
N. Corpora:	Corpus del Idioma Aymara (Bolivia): Academia Mayor de la Lengua Aymara (AMLA). Type: Written corpus AILLA (Archive of Indigenous Languages of Latin America). Hosted by: University of Texas at Austin: AILLA.org
O. Dictionaries and grammars:	https://web.archive.org/web/20141219125926/http://www.aymara.org/ http://www.websters-online-dictionary.org/definition/Aymara-english/ Coler, Matt (2014): A Grammar of Muylaq' Aymara: Aymara as spoken in Southern Peru Archived 2014-12-20 at the Wayback Machine. Brill: Leiden.
P. Language situation	There are no known language conflicts etc.
Q. Awareness / Function for identity:	Cultural Significance: For many Aymara speakers, their language and traditions are integral to their identity. The Aymara people have a profound connection to the land they inhabit, which is often passed down through generations. Their language reflects this relationship. Cognitive and Linguistic Features: Aymara has a unique grammatical structure that influences the way its speakers conceptualize the world.
R. Language Policy:	Bolivia: Bilingual Education: The government has implemented a bilingual education system, where Aymara (and other Indigenous languages) are taught alongside Spanish in schools. : The state supports the use of Aymara in public media, with radio stations, television programs, and print media providing content in Aymara, alongside Spanish. Peru: Language in Public Life: Aymara is used in local government activities, official communications, and education within regions where the language is spoken. However, implementation is more uneven compared to Bolivia, and in urban areas, Spanish is still predominantly used. The Peruvian government promotes bilingual education, where students in Aymara-speaking regions are taught in both Aymara and Spanish. However, challenges in resources and teacher training have hindered the full implementation of bilingual education across all areas. However, challenges remain in fully implementing these policies, and ongoing efforts are needed to address issues like resource limitations, teacher training, and the increasing dominance of Spanish in urban settings.
S. Note on the pluricentricity of the language	Standard pluricentricity of an officially recognised co-official language in Bolivia and as a regionally recognised language in Peru. Fulfills all criteria.
T. Literature (selection):	Briggs, Lucy Therina (1976): Dialectal Variation in the Aymara Language of Bolivia and Peru, Dissertation, University of Florida, Gainesville. Hardman, Martha James (1981): The Aymara Language in Its Social and Cultural Context: A Collection Essays on Aspects of Aymara Language and Culture. Gainesville: University Press.

	Xavier Albó, "Andean People in the Twentieth Century," in <i>The Cambridge History of the native Peoples of the Americas</i> . Vol. III: South America, ed. Frank Salomon and Stuart B. Schwartz (New York: Cambridge University Press, 1999), 765–871.
--	---

5. Basque

A. Name of the PCL:	Basque
B. Endonyms - Other name(s) of the PCL:	Euskara (or Euskera)
C. Language family:	Basque, Language isolate
D. Demography and language spread	Ca. 800.000 speakers in Spain and France, Basque language area called Euskal Herria
E. Status-OFL in:	Spain: (co-offical in:) Basque Autonomous Community and Navarre
F. Status-RML in:	France: Pyrénées-Atlantiques, Nouvelle-Aquitaine
G. DV:	Spanish Basque
H. NDV of the PCL	French Basque
I. Alphabet / WS	Latin script, Basque alphabet
J. Standard form:	Euskara batua
K. Standardisation / Codification, regulated by	Euskaltzaindia - Royal Academy of the Basque Language
L. Mutual intelligibility:	All regional varieties and NVs are mutually intelligible
M. Linguistic fragmentation:	Some fragmentation in regional varieties.
N. Corpora:	Elhuyar Web Corpus (monolingual): automatically collected from 6 202 domains (82 542 documents), totalling 124 million tokens, euWaC (Sketch Engine): a lemmatized, PoS-tagged web-crawled corpus compiled by Igor Leturia in 2012, with ≈100 million words; Pentsamenduaren Klasikoak Corpua "Classics of Thought": 130 KLASIKOAK volumes (1992–2009), 10.7 million tokens; Euskal Klasikoen Corpua (EKC) "Basque Classics Corpus": 496 books (16th c.–1975), 11.9 million tokens; Gero Corpus Historikoa: historical texts from the KLASIKOAK series, normalized into modern Basque; Goenkale Corpua: subtitles/transcripts of the ETB1 TV series "Goenkale" (1994–2014), 13.3 million tokens; Basque Spoken Corpus (ELRA S0123): 42 native-speaker narratives with MP3 recordings and detailed transcripts catalog.elda.org
O. Dictionaries and grammars:	Dictionaries: Euskaltzaindiaren Hiztegia (Unified Dictionary), Literatura Terminoen Hiztegia (dictionary of literary terms), Linguistic Atlas of Euskal Herria: Euskararen Herri Hizkeren At-lasa (EHHA), Idiomatic dictionary Erroibarko eta Esteribarko Hiztegia *** Bendel, Christiane (2006): Baskische Grammatik. Helmut Buske

	<p>Verlag. Hamburg. Euskal Gramatika: Lehen Urratsak (Seven-volume series) (1985-2011): Editor: Gramatika Batzordea, Euskaltzaindia (Royal Academy of the Basque Language). Publisher: Euskaltzaindia Rijk, Rudolf P. G. de (2008): Standard Basque: A Progressive Grammar Series: Current Studies in Linguistics. Publisher: MIT Press José Ignacio Hualde & Jon Ortiz de Urbina (eds.) (2003): A Grammar of Basque. Series: Mouton Grammar Library, Vol. 26, Publisher: Walter de Gruyter, 946 pp.</p>
P. Awareness / Function for identity:	Promotion of Euskara is favoured by 55.2% in the whole Basque area, 17.1% are against. The percentage of people in favour of promoting Euskara is growing, from 47.5% in 1991 to 55.2% in 2011.
Q. Language Policy:	Introduction of Basque in the educational system and extensive codification efforts.
R. Language situation	There are no known language conflicts etc.
S. Note on the pluricentricity of the language	Pluricentricity of a regionally recognised co-official language in Spain with occurrence as minority language in France
T. Literature (selection):	<p>Gerhard Edelmann (2016): Euskara / Basque: The importance of status for the development of a pluricentric language. In: Rudolf Muhr, Kelen Ernesta Fonyuy, Zeinab Ibrahim, Corey Miller (eds.) (2016): Pluricentric Languages and non-dominant Varieties worldwide. Volume 1: Pluricentric Languages across continents - Features and usage. Wien et. al., Peter Lang Verlag. p. 83-98.</p> <p>Arzo, Xabier (2008): Políticas lingüísticas actuales en las tierras del euskara. In: Doppelbauer, Max / Cichon, Peter (eds.): La España multilingüe. Lenguas y políticas lingüísticas de España. Praesens Verlag: Wien.</p> <p>Barreña, Andoni / Ortega, Ana / Amorrortu, Esti (2013): The Basque Language today, Achievements and Challenges. In: Martínez-Areta, Mikel (ed.) Basque and proto-Basque: language-internal and typological approaches to linguistic reconstruction. Peter Lang Verlag: Frankfurt am Main. 11-29.</p> <p>Martínez-Areta, Mikel (2013): Basque Dialects. In: Martínez-Areta, Mikel (ed.): Basque and proto-Basque: language-internal and typological approaches to linguistic reconstruction. Peter Lang Verlag: Frankfurt am Main. 31-87.</p>

6. Bengali

A. Name of the PCL:	Bengali
B. Other name(s) of the PCL	Bangla
C. Language family:	Indo-Aryan, Eastern
D. Demography and language spread	L1: 242 million native speakers, L2: 43 million of 2023; native to Bangladesh and India

E. Status-OFL in:	Bangladesh, India (West Bengal, Tripura, Assam, Jarkhand)
F. Status-RML in:	None
G. DV:	Bangladesh Bengali
H. NDV of the PCL	Indian Bengali
I. Alphabet / WS	Bengali-Assamese script (abugida), script with letters for consonants, with diacritics for vowels
J. Standard form:	Sadhu bhasha (সাধু ভাষা "upright language") (formal written language, use uncommon now) and Chôlito bhasha (চলিত ভাষা "running language") the official standard now, Standard Colloquial Bengali (SCB)
K. Standardisation / Codification, regulated by	Bangla Academy (Bangladesh) Paschimbanga Bangla Akademi (West Bengal)
L. Mutual intelligibility:	Regional varieties are not always mutually intelligible
M. Linguistic fragmentation:	Some fragmentation in regional varieties.
N. Language situation	High degree of diglossia, literary and standard form differs strongly from the everyday speech of the regions that identify with the language
O. Awareness / Function for identity:	Strong linguistic consciousness of the Bengalis. There was an ethnolinguistic movement in the 1950s that lead to the 21. February 1952 to be commemorated for 5 students who died for their right to speak their language. UNESCO commemorates this day as International Mother Language Day.
P. Corpora:	bnTenTen21 (Bengali Web Corpus 2021): 470 million words (541.6 M tokens, 40.0 M sentences, 1.74 M pages) EMILLE/CIIL Monolingual Bengali (2021): Written component: ≈ 9 million words for Bengali; Spoken component: part of the 2.627 M words of transcribed speech Bengali Raw Speech Corpus (2019): Duration: 128 h 46 m 59 s; 81.2 GB, Speakers: 476 (236 F, 240 M); Domains: news, creative text, sentences, dates, commands, names, phonetically balanced. Provider: CIIL Mysore (2019)
Q. Dictionaries and grammars:	Roy, M. N. (1997): Oxford Bengali-English Dictionary" Publisher: Oxford University Press. Karim, Abdul (1969): Modern Bengali Dictionary" by Abdul Karim. Publisher: Bangla Academy, Dhaka. *** McLeod, W. H. (1969): : A Descriptive Grammar of Bengali" Publisher: Cambridge University Press Iqbal, Zafar (2000): Bangla Grammar. Publisher: National Book Trust, India. Bonazzi, E (2008). Grammatica Bengali. Bologna: Libreria Bonomo Editrice. ISBN 978-88-6071-017-8.
R. Language Policy:	Bangladesh: Bengali is the sole national language in Bangladesh, it is used in government administration, education, media, and all official communication and also the medium of instruction in schools and universities. India: In West Bengal, Bengali is the medium of instruction in

	schools, and it is mandatory for students to study it as part of the curriculum. In some schools in the northeastern states of India (like Tripura and Assam), Bengali is also used as the primary language of instruction. In Bangladesh, the right to use Bengali is enshrined in the Constitution, while in India, the language rights are more varied, depending on the state and regional language policies.
S. Note on the pluricentricity of the language	Standard pluricentricity with occurrence in 2 sovereign nations and two 2 NVs. Contiguous language area.
T. Literature (selection):	Asiatic Society of Bangladesh (2003). Banglapedia, the national encyclopedia of Bangladesh. Dhaka: Asiatic Society of Bangladesh. Bhattacharya, T (2000). "Bengali". In Gary, J. and Rubino. C. (ed.). Encyclopedia of World's Languages: Past and Present (Facts About the World's Languages). WW Wilson, New York. ISBN 978-0-8242-0970-4. Masica, Colin P. (1991). The Indo-Aryan Languages. Cambridge Univ. Press. ISBN 978-0-521-23420-7. Ray, P; Hai, MA; Ray, L (1966). Bengali language handbook. Center for Applied Linguistics, Washington. ASIN B000B9G89C. Thompson, Hanne-Ruth (2012). Bengali. Volume 18 of London Oriental and African Language Library. John Benjamins Publishing. ISBN 9027273138.

7. British English

A. Name of the PCL:	British English (Britain)
B. Other name(s) of the PCL	English, British English
C. Language family:	Germanic, West Germanic
D. Demography and language spread	67.596 mio. speakers, language spread globally via media and commerce
E. Status-OFL in:	United Kingdom (Great Britain)
F. Status-RML in:	None
G. DV:	English English
H. NDV of the PCL	Scottish English, Welsh English, and Northern Irish English on the territory of the UK; Exterritorial varieties: Bermudian English, Falkland Islands English, Cayman Islands English, Gibraltar English; Manx English, and Channel Island English
I. Alphabet / WS	Latin script
J. Standard form:	Standard English, Welsh English, Scottish English, Northern Ireland English
K. Standardisation / regulated by	University of Oxford via the OED
L. Mutual intelligibility:	NDVs varieties are mutually intelligible, regional varieties are not always mutually intelligible
M. Linguistic fragmentation:	Some fragmentation in regional varieties.

N. Language situation	In Britain, most people speak with a regional accent or regional variety. About 2% of speak with an accent called Received Pronunciation (Kings English) that is dominant in Foreign Language teaching. Estuary English and Standard Southern British English have been gaining prominence recently. BBC English is neutral and regarded regionless.
O. Awareness / Function for identity:	Strong identity function of the DV and the NDVs
P. Corpora:	<p>British National Corpus (BNC) (1991-1994): Compiler: BNC Consortium (Oxford Univ. Press lead; Longman; W.&R. Chambers; Oxford & Lancaster Univ.; British Library). – Size: 100 million words (90 % written, 10 % spoken);</p> <p>Spoken British National Corpus 2014 (Spoken BNC2014): Compiler: CASS & Cambridge University Press. Size: 10.4 million words from 1,251 multi-speaker conversations (672 speakers). Annotation: Orthographic transcription;</p> <p>Cambridge English Corpus (CEC): Compiler: Cambridge University Press & Cambridge English Language Assessment. Continuously updated (1960s–present), Size: > 2 billion words (spoken & written);</p> <p>International Corpus of English – Great Britain (ICE-GB): Compiler: Survey of English Usage, UCL (Dir. Sidney Greenbaum; Bas Aarts), Dates: data collected ca. 1990; released 1998 (Release 2 in 2006). Size: 1 million words (60 % spoken [600 000 w.]; 40 % written [400 000 w.]). Annotation: POS-tagged & fully parsed with Quirk et al. grammar; searchable via ICECU</p>
Q. Dictionaries and grammars:	Oxford English Dictionary, Longman Dictionary of Contemporary English, Chambers Dictionary, and the Collins Dictionary;
R. Language Policy:	Strong language spread via foreign language teaching, which generates large economic revenues
S. Note on the pluricentricity of the language	Special case of pluricentricity with 4 NVs in a single country because of the constitutional structure of the UK, which consists of 4 sovereign nations.
T. Literature (selection):	<p>British English; Hiberno-English. In: Oxford English Dictionary (2 ed.). Oxford, England: Oxford University Press. 1989.</p> <p>McArthur, Tom (2002). Oxford Guide to World English. Oxford: Oxford University Press. ISBN 0-19-866248-3 hardback, ISBN 0-19-860771-7 paperback.</p> <p>Bragg, Melvyn (2004). The Adventure of English, London: Sceptre. ISBN 0-340-82993-1</p> <p>Lindsey, Geoff (2019). English After RP: Standard British Pronunciation Today. Switzerland: Springer Nature. ISBN 978-3030043568.</p> <p>Peters, Pam (2004). The Cambridge Guide to English Usage. Cambridge: Cambridge University Press. ISBN 0-521-62181-X.</p> <p>Strycharczuk, P., López-Ibáñez, M., Brown, G., & Leemann, A. (2020). "General Northern English. Exploring regional variation in the North of England with machine learning". <i>Frontiers in Artificial Intelligence</i>, 3, 545883.</p>

8. Catalan

A. Name of the PCL:	Català - Catalan
B. Endonyms - Other name(s) of the PCL	Valencià, Valencian
C. Language family:	Western Romance, Gallo-Romance
D. Demography and language spread	ca. 12 mio. native speakers; native to Spain (Catalonia, Valencia), Andorra, France, and Italy.
E. Status-OFL in:	Spain: Catalonia, Valencian Community (as Valencian), Balearic Islands; Andorra; Italian city of Alghero, Sardinia
F. Status-RML in:	France: French sub-region of: Northern Catalonia (Roussillon), part of Occitania Spanish sub-regions of: La Franja, part of the community of Aragon Carce, part of the Region of Murcia (as Valencian)
G. DV:	Central Catalan
H. NDVs of the PCL	Valencian, Andorra, Balearic Islands, French Catalan varieties
I. Alphabet / WS	Latin script
J. Standardisation / Codification, regulated by	Institut d'Estudis Catalans (IEC), Acadèmia Valenciana de la Llengua (AVL), Institut Aragonès del Català; Consorci per a la Normalització Lingüística
K. Standard form:	Central Catalan, Valencian and Balearic Catalan
L. Linguistic fragmentation:	Western Catalan: North-Western Catalan and Valencian; Eastern Catalan has four varieties: Central Catalan, Balearic, Rossellonese, and Alguerese.
M. Mutual intelligibility:	NDVs varieties are mutually intelligible, feature a relative uniformity; Catalan is one of the most uniform languages of Romance-speaking Europe.
N. Corpora:	Corpus textual informatitzat de la llengua catalana
O. Dictionaries and grammars:	Catalan Language Dictionary of the Institut d'Estudis Catalans; Valencian Normative Dictionary of the Valencian Academy of the Language; Diccionari català-valencià-balear; Diccionari del català contemporani, the Diccionari descriptiu de la llengua catalana (DDL), Gramàtica de la Llengua Catalana; Gramàtica Normativa Valenciana, Enciclopèdia Catalana *** Solà, Joan et. al. (2008): Gramàtica del català contemporani. Publisher: Editorial Empúries, 2008 (4th definitive edition). 3 volumes. Acadèmia Valenciana de la Llengua (AVL): Gramàtica normativa valenciana. Language: Catalan (Valencian variant); Royal Academy of Valencian Culture (RACV) (1979): Norms of El Puig Language: Catalan (Valencian variant) Dols, Nicolau, MansellCatalan, Richard (2017): An Essential Grammar Publisher: Routledge.
P. Language situation	There are conflicts between the Northern and the Valencian variety
Q. Awareness / Function for identity:	Strong identity function of the DV and the NDVs in the struggle to achieve independence from Spain.

R. Language Policy:	Strong codification activities by the regulation bodies. Catalan is used in the education system as language of instruction.
S. Note on the pluricentricity of the language	Special case of pluricentricity with the status as a national language in Andorra and a regional language in Spain that spreads to the neighbouring countries France and Italy with the status of a RML.
T. Literature (selection):	<p>Edelmann, Gerhard (2019): Conflict between Valencian and Catalan: Is Valencian a language of its own or a variety of Catalan? In: Rudolf Muhr, et. al. (eds.) (2019): European Pluricentric Languages in Contact and Conflict: Wien et. al., Peter Lang Verlag. p. 107-120.</p> <p>Edelmann, Gerhard (2015): Catalan as a pluricentric language. In: Muhr, Rudolf Marley, Dawn (eds.), Pluricentric Languages: New Perspectives in Theory and Description. Peter Lang GmbH. Frankfurt am Main, p. 155-168.</p> <p>Mas Castells, Josep-Àngel (2019): Identity and use of a pluricentric language in conflict: Catalan in Spain. In: Rudolf Muhr, et. al. (eds.) (2019): European Pluricentric Languages in Contact and Conflict. p.91-106.</p> <p>Mas, Josep-Àngel (2012): Catalan as a pluricentric language: the Valencian case. In: Non-Dominant Varieties of Pluricentric Languages. Getting the Picture: In Memory of Michael Clyne. Ed. by R. Muhr et. al. Vienna. Peter Lang, p. 263-300.</p> <p>Ferrando, A. / Nicolás, M. (2011). Història de la llengua catalana. Barcelona. UOC.</p> <p>Herling, Sandra (2008): Katalanisch und Kastilisch auf den Balearen. Heidelberg. Winter.</p>

9. Chewa/Chichewa*

A. Name of the PCL:	Chewa
B. Endonyms - Other name(s) of the PCL	Chichewa, Nyanja; Zambia: Nyanja <i>or</i> Chinyanja
C. Language family and region:	Bantu, South-east Africa, native to Malawi, Zambia, Mozambique
D. Demography and language spread:	ca. 7 mio. in 2007; ca. 10,6 million L1 speakers in Malawi; Zambia: ca. 1 million; Mozambique: ca. 1,35 million. Total: 16 million speakers.
E. Status-OFL in:	Malawi, Zambia (co-official with English)
F. Status-RML in:	unrecognised regional language in Mozambique
G. DV:	Malawi
H. NDVs of the PCL	Zambia
I. Standard form:	Malawian Chichewa
J. Alphabet / WS	Latin script (Chewa alphabet), Mwangwego
K. Standardisation / Codification, regulated by	No regulation at present, Between 1972–1995 the Chichewa Board for language regulation existed
L. Mutual intelligibility:	The various forms of Chichewa across Malawi, Zambia, and Mo-

	zambique are mutually intelligible and considered varieties of the same language.
M. Linguistic fragmentation:	Standard Nyanjy, Town Nyanja (in Zambia);
N. Language situation	Town Nyanja, is the lingua franca of the Zambian capital Lusaka and is widely spoken as a second language throughout Zambia. Standard Nyanja as used in schools differs strongly from the variety spoken in Lusaka. It has been identified as a barrier to the acquisition of literacy among Zambian children.
O. Awareness / Function for identity:	Strong identification with Town Nyanja in Zambia
P. Corpora:	WebCrawl African Parallel Corpus (2022): 695 000 aligned sentences for 74 language pairs between English and 15 African languages, among them Chichewa-English. webcrawl-african-nya-eng.nya (Chichewa) and webcrawl-african-nya-eng.eng (English), facilitating direct alignment. The monolingual Chichewa file webcrawl-african-nya-eng.nya mirrors the parallel data and can serve as a standalone text corpus. Hosted on GitHub under the ANVITA lab. African Dataset on Hugging Face: A standardized parallel corpus for benchmarking African language MT: Dataset: allenai/wmt22_african, containing 248 language pairs, including Chichewa-English. Usage: Ready-to-load via Hugging Face Datasets;
Q. Dictionaries and grammars:	Paas , Steven (2017): Chichewa-English/English-Chichewa Dictionary. Oxford. OUP. 1158 Pages. ISBN: 9780190416591 Hetherwick, Alexander (1907): A Practical Manual of the Nyanja language. *** Watkins, Mark Hanna (1937). A Grammar of Chichewa: A Bantu Language of British Central Africa Archived 2018-12-10 at the Wayback Machine, Language, Vol. 13, No. 2, Language Dissertation No. 24 (Apr.-Jun., 1937), pp. 5–158.
R. Language Policy:	Malawi: : English and Chichewa are the only official languages. The National Language Policy mandates that from Standard 1–4, instruction be delivered in the dominant regional mother tongue—except for English and Chichewa, which are taught in their own language. The Centre for Language Studies (CLS) at the University of Malawi to coordinate research, orthography, and promotion of all Malawian languages, with Chichewa as the flagship. Media use: Chichewa is the only vernacular used on national radio and television, and literature in Chichewa has flourished since its institutionalisation. Zambia: English as sole official language: Three-language instructional model (adopted post-1977 orthography standardisation): Grades 1–2: mother tongue; Grades 3–5: regional African language (e.g. Nyanja); Grade 5 onward: English. Mozambique: National languages” are recognised as part of Mozambique’s cultural and educational heritage. From 2013 onward, bilingual-education pilots in provinces such as Niassa and Tete have included Cinyanja (Chichewa) alongside Portu-

	guese in early primary grade.
S. Note on the pluricentricity	The pluricentricity of Chichewa is not very pronounced, which is mainly due to its low value in Zambia.
T. Literature (selection):	Downing, Laura J.; Al D. Mtenje (2017). <i>The Phonology of Chichewa</i> . Oxford University Press. Gray, Andrew; Lubasi, Brighton; Bwalya, Phallen (2013). <i>Town Nyanja: a learner's guide to Zambia's emerging national language</i> Maxson, Nathaniel (2011). <i>Chicheŵa for English Speakers: A New and Simplified Approach</i> . ISBN 978-99908-979-0-6.

10. Chinese | Mandarin

A. Name of the PCL:	Chinese Mandarin <i>Pu Tong Hua</i> 普通话 lit. 'The Common Language'. (PRC); <i>Guo Yu</i> 國語 lit. 'The National Language' (Taiwan); <i>Hua Yu</i> 華語 lit. 'Language of the ethnic Chinese' (Singapore)
B. Endonyms - Other name(s) of the PCL	漢語 "H'ny"; 華語 Huáy'; 中文 Zh'ongén
C. Language family and region	Sino-Tibetan, Sinitic, located in Central Asia, native to China Taiwan Hong Kong Macau Malaysia, Singapore
D. Demography and language spread	ca. 1,39 billion people, (17% of the global population), speak a variety of Chinese as their L1 native and L2 language; native to a China, Taiwan, Hong Kong Macau, Malaysia and Singapore
E. Status-OFL in:	In China and Taiwan as Mandarin, Co-official as Mandarin in Singapore, as Cantonese in Hong Kong and Macau together with English and Portuguese, UNO (working language);
F. Status-RML in:	None
G. DV:	Chinese Mandarin (PRC)
H. NDVs of the PCL	Taiwanese Mandarin, Hong Kong Mandarin, Singapore Mandarin
I. Standard form:	Mandarin (普通话 - Pǔtōnghuà), Guo Yu, Hua Yum; Standard Cantonese (粵語 - Yuèyǔ); Simplified Chinese (简体字 - Jiǎntǐzì), Used in: Mainland China, Singapore, and Malaysia. Traditional Chinese (繁體字 - Fántǐzì), Used in: Taiwan, Hong Kong, and Macau.
J. Alphabet / WS	Chinese script, based on a system of ideographic characters; Logographs, representing morphemes
K. Standardisation / Codification, regulated by	China: State Language Commission; Taiwan: National Languages Committee; Hakka Affairs Council; Malaysia: Chinese Language Standardisation Council; Singapore: Ministry of Education, Promote Mandarin Council;
L. Mutual intelligibility:	Speakers of any one of the standard varieties of Mandarin are able to understand speakers of another one of the standard Mandarin varieties (the standard varieties being Pu Tong Hua, Guo Yu, Hua Yu and Han Yu). However all standard varieties of Mandarin are not identical, even if they are all mutually intelligible. Chinese regional varieties (Wu, Hokkien (Min), Cantonese, Xiang, Gan, Hakka (Kejia), Jin, Pinghua and Hui) are in their spoken form are as mutually unintelligible as different languages.
M. Ling. fragmen-	Wu, Hokkien (Min), Cantonese, Xiang, Gan, Hakka (Kejia), Jin,

tation:	Pinghua and Hui.
N. Language situation	Large amount of varieties both within Mandarin that differs inside China and Taiwan, and through the many regional languages (Pt. N above).
O. Awareness / Function for identity:	High relevance for identity. Pu Tong Hua Mandarin represents an overarching cultural identity for PRC;
P. Corpora:	China (PRC): Putonghua (Mandarin): Chinese Gigaword (LDC) — newswire texts; Lancaster Corpus of Mandarin Chinese (LCMC) — balanced modern text; Sina Weibo Corpus — social media; CCL Corpus (Center for Chinese Linguistics at Peking University) Taiwan Guoyu (Taiwan Mandarin): Sinica Corpus (Academia Sinica), balanced, annotated; Academia Sinica Balanced Corpus of Modern Chinese — classical source for Taiwan Mandarin. Singapore: Singaporean Mandarin: Singapore Corpus of Spoken English and Mandarin (SCoSEM), collected by NUS; National Speech Corpus (with a Mandarin set) — mainly for speech tech. Hong Kong: Cantonese: HKCanCor (Hong Kong Cantonese Corpus) — spoken Cantonese; Hong Kong Parallel Corpus — translations between English and Chinese (written Chinese here tends to be Standard Written Chinese, close to Mainland style)
Q. Dictionaries and grammars:	Oxford University Press: English-Chinese (Simplified Dictionary) https://dictionary.cambridge.org/dictionary/english-chinese-simplified/ (online); Collins English-Simplified Chinese Dictionary: https://www.collinsdictionary.com/dictionary/english-chinese (online) Ross, Claudia (2014): Modern Mandarin Chinese Grammar (Modern Grammars). London, Routledge.
R. Language Policy:	Policy makers in Beijing and Taipei promote Mandarin as the common “Chinese” national language; In Singapore, Mandarin is promoted as the “mother tongue” of those who are ancestrally from China, other varieties of Chinese are not promoted.
S. Note on the pluricentricity	The pluricentricity of “Chinese” is complex as there is the pluricentricity of Mandarin, the official variety that is used in PRC, Taiwan, Macau, Hong Kong and Singapore. And there is also the pluricentricity of the many regional varieties that in linguistic terms are languages in their own right but for linguistic political reason are considered as dialects of the same language “Chinese”.
T. Literature (selection):	TIEN, Adrian (2016): Perspectives on “Chinese” pluricentricity in China, Greater China and beyond. In: Rudolf Muhr et. al. (eds.) (2016): Pluricentric Languages and non-dominant Varieties worldwide. Volume 1: Pluricentric Languages across continents - Features and usage. Wien et. al., Peter Lang Verlag. p. 33-52. Bradley, David (1992): Chinese as a pluricentric language. In: Pluricentric Languages. Different Norms in Different Countries, edited by Michael Clyne, Berlin/New York. Mouton/de Gruyter 1992, 305-324. Ramsey, Robert S. (1987): The Languages of China. New Jersey. Princeton University Press

11. Coatian

A. Name of the PCL:	Croatian
B. Endonyms - Other name(s) of the PCL	Hrvatski
C. Language family and region:	Slavic, located in the Balkans, Southeast Europe; native to Croatia, Bosnia and Herzegovina, Hungary, Montenegro, Romania, Serbia, Kosovo.
D. Demography and language spread:	ca. 7 mio. in 2007; Croatia 5,1 million L1 speakers, 1,3 million L2 speakers; Bosnia: 545.000 L1 speakers; Austria: Approximately 150.719 people, including Burgenland Croats, speak Croatian as their native language; Germany: ca. 500.000 L1 speakers.
E. Status-OFL in:	Croatia, Bosnia and Herzegovina (co-official), Montenegro (co-official), Serbia (in Vojvodina), European Union
F. Status-RML in:	Austria, Slovakia, Czech Republic, Hungary (in Baranya County), Italy
G. DV:	Croatian Croatian
H. NDVs of the PCL	Bosnia and Herzegovina (co-official), Montenegro (co-official), Serbia
I. Standard form:	Croatian Croatian
J. Alphabet / WS	Latin script
K. Standardisation / Codification, regulated by	Institute of Croatian Language and Linguistics, Council for the Croatian language
L. Mutual intelligibility:	The NVs and regional varieties are mutually intelligible
M. Linguistic fragmentation:	There are three main regional varieties: Shtokavian (on which Standard Croatian is based), Chakavian and Kajkavian
N. Language situation	Croatia: In November 2023, the Croatian Parliament passed the Law on the Croatian Language, which establishes guidelines for its official and public use. The law emphasizes the protection and promotion of the language as a vital component of national identity and culture. In Austria, Serbia, Italy, Romania and Slovenai, Croatian has the status as minority language.
O. Awareness / Function for identity:	Croatian is considered as a separate language that is considered key to national identity
P. Corpora:	RWaC (Croatian Web Corpus): Full Name: Hrvatski web-korpus. Available via Sketch Engine. HNK (Hrvatski Nacionalni Korpus / Croatian National Corpus) Maintained by: Institute of Croatian Language and Linguistics. Size: Varies by subcorpus; total in the hundreds of millions of tokens. srWaC + hrWaC Bilingual Corpus: A combined Serbian and Croatian web corpus
Q. Dictionaries and grammars:	Hrvatski pravopis by the Institute of Croatian Language and Linguistics, available online Hrvatski jezični portal by University Computing Centre (Srce) and Znanje, available online. Rječnik hrvatskoga jezika by Anić Rječnik hrvatskoga jezika by Jure Šonje et al. Hrvatski enciklopedijski rječnik, by a group of authors

	Hrvatska gramatika by Eugenija Barić et al.
R. Language Policy:	Strong endonormative codification through recourse to regional and historical variants.
S. Note on the pluricentricity	Disputed case of pluricentricity. Slavists, mainly from Germany claim that because of the small linguistic difference Croatian is not a language in its own right but a NV of Serbo-Croatian. The position is not viable as there are similar cases like the Scandinavian languages Danish, Norwegian and Swedish that are mutually intelligible but are generally considered as separate languages. The status of the language is enshrined in the constitutions and it is the sole national language.
T. Literature (selection):	Stojanov, Tomislav (2019): Abandoned pluricentricity – the case study of the Serbo-Croatian linguonym. In: Muhr, Rudolf/ Mas Castells, Josep/Rueter, Jack (eds.): European Pluricentric Languages in Contact and Conflict. Berlin et. al.: Peter Lang. p. 163–176. Bugarski, R. (2000). Serbo-Croatian: How many languages? In: Kunzmann-Müller, B. (ed.): Die Sprachen Südosteuropas heute: Umbrüche und Aufbruch. Frankfurt am Main etc., 192-199. Ilic Markovic, Gordana (2016): Creating a name for a pluricentric language: From Serbian to Serbo-Croatian, Bosnian / Croatian / Serbian. In: Muhr, Rudolf et. al. eds. (2016): Vol. 1, p. 409-425. Kapović, Mate (2020): Bosnian/Croatian/Montenegrin/Serbian: Notes on Contact and Conflict. In: Rudolf Muhr, Josep Angel Mas Castells, Jack Rueter (eds.) (2019): European Pluricentric Languages in Contact and Conflict: Wien et. al., Peter Lang Verlag. p. 177-188.

12. Dutch

A. Name of the PCL:	Dutch
B. Endonyms - Other name(s) of the PCL	Nederlands
C. Language family and region	West Germanic, located in the Balkans, Southeast Europe;
D. Demography and language spread	ca. 25 mio. in 2022; native to The Netherlands, Belgium, Suriname, France (Nord) and in the Caribbean
E. Status-OFL in:	Netherlands, Belgium, Suriname, Aruba, Bonaire, Curaçao, Sint Eustatius, Sint Maarten, Saba; European Union, Benelux, Caribbean Community, PROSUR, South American Union
F. Status-RML in:	None
G. DV:	Dutch Dutch / Belgium Dutch
H. NDVs of the PCL	Aruba, Bonaire, Curaçao, Sint Eustatius, Sint Maarten, Suriname, Saba
I. Standard form:	Dutch Dutch / Belgium Dutch
J. Alphabet / WS	Latin script
K. Standardisation / Codification, regulated by	Nederlandse Taalunie (Dutch Language Union) for Belgium and The Netherlands; Instituut voor de Nederlandse Taal (or Dutch Language Institute)

L. Mutual intelligibility:	The NVs are mutually intelligible
M. Linguistic fragmentation:	West Frisian language, Hollandic , Limburgish, Kleverlandish, Westvlaams, Zeelandic Flanders, Gronings dialect, Brabantian
N. Language situation	Dutch is the majority language in the Netherlands (96%) and Belgium (59%) as well as a minority language in Germany and northern France's French Flanders
O. Awareness / Function for identity:	High relevance for identity particularly in Belgium (Flanders)
P. Corpora:	https://ivdnt.org/corpora-lexica/corpora/#corpus-compilation
Q. Dictionaries and grammars:	Van Dale Groot woordenboek van de Nederlandse taal, (Van Dale Great Dictionary of the Dutch Language) Woordenboek der Nederlandsche Taal (Dictionary of the Dutch Language) Dictionary of Contemporary Dutch (Algemeen Nederlands Woordenboek or ANW) Shetter, William Z. & Esther Ham. Dutch: an essential grammar, 10th edn. Abingdon: Routledge, 2016. Spaans, Yolande. A practical Dutch grammar, 3rd unrevised edn. Leiden: Primavera Pers, 2013 Broekhuis, Hans et al. Syntax of Dutch: verbs and verb phrases. 2 vols. Amsterdam: Amsterdam University Press, 2015. Donaldson, Bruce. Dutch: a comprehensive grammar, 3rd edn. Abingdon-on-Thames: Routledge, 2017
R. Language Policy:	No conflicts in The Netherlands but still in Belgium about the use of Dutch and French
S. Note on the pluricentricity	Dutch is the rare case of a symmetrical pluricentricity with the Belgium and Dutch variety being codified together and sharing the same status as standard norm. An intermediate variety called "Tussental" has developed in Belgium that provides a specific Belgium identity.
T. Literature (selection):	Mijts, Eric (2022): Pluricentric languages in the Americas: the case of Dutch in the Dutch Caribbean. Rudolf Muhr, et. al. (eds.) (2022): Pluricentric languages in the Americas. Graz/Berlin, PCL-Press. p. 231-244. Willems, R. (1993): Integration vs. particularism. The undeclared issue at the first "Dutch Congress" in 1849. In: The Earliest Stages of Language Planning, J. Fishman (ed.), 69-83. Berlin/New York, Mouton de Gruyter, De Caluwe, Johan (2012): Dutch as a bicentric language: a lexicographic (r)evolution. In: Rudolf Muhr (ed.) (2012): Non-dominant varieties of pluricentric languages. Getting the picture. In memory of Michael Clyne. Wien et. al., Peter Lang Verlag. p. 143-154. Willems, R. (1996): "Pluricentric Principles in the Standardization of 19th Century Dutch". Word 47: 63-72.

13. English | Global English

A. Name of the PCL:	Global English
B. Endonyms - Other	English

name(s) of the PCL	
C. Language family and region	Western Romance, its origin is located in north-west Europe, now present on all continents
D. Demography and language spread	ca. 400 million native speakers and a total of 1.5 billion speakers, either as a first or second language
E. Status-OFL in:	75 territories where English is used as a first language (L1) or as an unofficial or institutionalised second language (L2) Status-OFL in: 57 independent nations: Antigua and Barbuda, Australia, The Bahamas, Barbados, Belize, Botswana, Burundi, Cameroon, Canada, Dominica, Eswatini, Fiji, The Gambia, Ghana, Grenada, Guyana, India, Ireland, Jamaica, Kenya, Kiribati, Lesotho, Liberia, Malawi, Malta, Marshall Islands, Micronesia, Namibia, Nauru, New Zealand, Nigeria, Pakistan, Palau, Papua New Guinea, Philippines, Rwanda, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Samoa, Seychelles, Sierra Leone, Singapore, Solomon Islands, South Africa, South Sudan, Sudan, Tanzania, Tonga, Trinidad and Tobago, Tuvalu, Uganda, United Kingdom, United States, Vanuatu, Zambia, Zimbabwe De facto OFL (5) in: Barbuda, British Indian Ocean Territory, Guernsey, Montserrat, Saint Helena, Ascension and Tristan da Cunha Working language in government and education, no recognition as OFL (13): Bahrain , Bangladesh , Bhutan , Brunei , Israel , Jordan , Kuwait , Malaysia , Mauritius , Nepal , Oman , Qatar , Sri Lanka , Timor-Leste] , United Arab Emirates
F. Status-RML in:	Non-official minority language in: Lebanon, Louisiana (US), Maine (US), Mauritius
G. DV:	American English, British English, Australian English (in the Pacific)
H. NDVs of the PCL	37 NDVs plus the varieties of the dependent entities (see Pt. E)
I. Standard form:	American English, British English, Australian English
J. Alphabet / WS	Latin script
K. Standardisation / Codification, regulated by	There is no regulatory body. Codification and regulation is done by dictionary publishers: Oxford University Press (UK), Cambridge dictionary, Macquary dictionary (Australia), Merriam Webster dictionary (US), A Dictionary of Canadianisms on Historical Principles (CAN), Dictionary of South African English
L. Mutual intelligibility:	The NVs in Europe are mutually intelligible. The African, Asian and Canada varieties are not always easily intelligible because of the nativisation and influence of local languages.
M. Linguistic fragmentation:	Little linguistic fragmentation in the US and CAN, considerable regional fragmentation in the UK
N. Language situation	English is <i>the</i> global language, this gives the language considerable power but also has some costs by being changed in the many countries it is used; Nobody really owns English any more.
O. Awareness / Function for identity:	Considerable relevance for identity particularly in US and UK

P. Corpora:	Many: https://www.english-corpora.org/ ; Webs site that lists the available corpora for English
Q. Dictionaries and grammars:	Oxford University Press (UK), Cambridge dictionary, Macquary dictionary (Australia), Marriam Webster dictionary (US), A Dictionary of Canadianisms on Historical Principles (CAN), Dictionary of South African English; Oxford Modern English Grammar;
R. Language Policy:	Massive language export via schools and international organisations
S. Note on the pluricentricity	English is a highly decentralised language, present on all continents and <i>the</i> lingua franca in many domains (science, commerce, politics etc.). Many varieties in Africa, Asia and in the Americas have become nativised and are drifting towards becoming languages of its own (India, Pakistan etc.)
T. Literature (selection):	Filppula, Markku / Klemola, Juhani / Sharma, Devyani (2024): The Oxford Handbook of World Englishes. Oxford, OUP.

14. French

A. Name of the PCL:	French
B. Endonyms - Other name(s) of the PCL	Langue Française
C. Language family and region	Western Romance language, located in north-west Europe; native to: France, Belgium, Switzerland, Monaco, Francophone Africa, Canada, and other locations in the Francophonie.
D. Demography and language spread	ca. 80 million native speakers (2020), 321 million people in 2022 are able speak this language;
E. Status-OFL in:	<i>In 26 independent nations:</i> Belgium, Benin, Burundi, Cameroon, Canada, Central African Republic, Chad, Comoros, Congo, Democratic Republic of the Congo, Djibouti, Equatorial Guinea, France, Gabon, Guinea, Haiti, Ivory Coast, Luxembourg, Madagascar, Monaco, Rwanda, Senegal, Seychelles, Switzerland, Togo, Vanuatu, <i>And in dependent entities (overseas departments):</i> New Caledonia , French Polynesia , Saint Martin , Wallis and Futuna , Saint Barthélemy, Saint Pierre and Miquelon , French Southern and Antarctic Lands , Clipperton Island <i>Language of administration and education, de facto Status-Status-OFL in:</i> Algeria, Burkina Faso, Cambodia, Laos, Lebanon, Mali, Mauritania, Mauritius, Morocco, Niger, Tunisia.
F. Status-RML in:	Non-official minority language in: Lebanon, Louisiana (US), Maine (US), Mauritius
G. DV:	French French (le Français standard)
H. NDVs of the PCL	37 NDVs plus the varieties of the dependent entities (see Pt. E)
I. Standard form:	French French,
J. Alphabet / WS	Latin script
K. Standardisation / Codification, regulated by	Académie Française (French Academy, France); Office québécois de la langue française (Quebec Board of the French Language, Quebec); Direction de la langue française [fr] (Belgium)

L. Mutual intelligibility:	The NVs in Europe are mutually intelligible. The African, Asian and Canada varieties are not always mutually intelligible because of the nativisation and influence of local languages.
M. Linguistic fragmentation:	Little linguistic fragmentation
N. Language situation	Language spread is continuing, although the language has lost some ground against English
O. Awareness / Function for identity:	High relevance for identity particularly in Belgium and Canada and for the French nation (language pride)
P. Corpora:	Many: CNRTL: Centre National de Ressources Textuelle et Lexicales https://www.cnrtl.fr/ Corpus de français parlé parisien des années 2000 (http://cfpp2000.univ-paris3.fr/index.html) Corpus de français parlé au Québec (https://applis.flsh.usherbrooke.ca/cfpq); CallFriend (Québécois French Corpus); COCOON (Collection de corpus oraux numériques) (https://cocoon.huma-num.fr/exist/crdo), CLARIN (spoken corpora) (https://www.clarin.eu/resource-families/spoken-corpora) and others.
Q. Dictionaries and grammars:	Petit Larousse; Micro-Robert Plus; Larousse Lexicologos: https://www.lexilogos.com/francais_dictionnaire.htm
R. Language Policy:	Massive language export via schools and international organisations
S. Note on the pluricentricity	French is a highly centralised language and by constitution the only official language of France. There is massive language spread, support of the language via satellite TV and international language organisations (Francophonie, Alliance française)
T. Literature (selection):	Marc Fumaroli (2011). <i>When the World Spoke French</i> . Translated by Richard Howard. ISBN 978-1-59017-375-6. Nadeau, Jean-Benoît, and Julie Barlow (2006). <i>The Story of French</i> . (First U.S. ed.) New York: St. Martin's Press. ISBN 0-312-34183-0. Ursula Reutner (2017). <i>Manuel des francophonies</i> . Berlin/Boston: de Gruyter. ISBN 978-3-11-034670-1

15. Fula/Fulani*

A. Name of the PCL:	Fula
B. Endonyms - Other name(s) of the PCL	Fulani, Fulah, Fulfulde, Peul
C. Language family and region	Niger-Congo, Sengegambian; located in West Africa
D. Demography and language spread	ca. 37 million native speakers, ca. 2,7 million L2 speakers; native to Burkina Faso, Mali and a local language in another 8 countries in West Africa with ca. 95 million speakers
E. Status-OFL in:	Burkina Faso, Mali
F. Status-RML in:	Cameroon, Niger
G. DV:	Burkina Faso
H. NDVs of the PCL	Lingua franca in Mauritania, Sierra Leone, Togo, CAR, Chad, Su-

	dan, Ethiopia and Somalia; Many varieties that are not always mutually intelligible: Maasina Fulfulde, Pular language, Pulaar language, Nigerian Fulfulde, and Adamawa Fulfulde
I. Standard form:	unclear
J. Alphabet / WS	Latin script: African Reference Alphabet (ARA), Adlam script, Arabic script/Ajami; Decree No 2005-990 of 21 October 2005 fixes the official Latin alphabet for Pulaar (Fuuta Tooro) in Senegal, The Gambia and Mauritania; In 1989 Guinea adopted an orthography aligned with the ARA (unified orthography).
K. Standardisation, regulated by	West African Linguistic Society (WALS): Promotes research and development of African languages, including Fulfulde. ACALAN (African Academy of Languages) – under the African Union A pan-African organization promoting the use and development of African languages. Fulfulde is one of the Vehicular Cross-Border Languages (VCBLs) supported by ACALAN. National Language Institutes: Nigeria: National Institute for Nigerian Languages (NINLAN) Senegal: Centre de Linguistique Appliquée de Dakar (CLAD) Mali: Direction Nationale de l’Alphabétisation et des Langues Nationales (DNALN) Cameroon: Ministry of Basic Education and local linguistic bodies
L. Mutual intelligibility:	The varieties are not always intelligible to each other but through strong ethno-linguistic awareness Fulfulde is regarded as a single language.
M. Linguistic fragmentation:	High linguistic fragmentation; Nine different translations are necessary of the Bible to be understood by most Fula speakers.
N. Awareness / Function for identity:	Fulfulde is a core marker of Fulani identity across West and Central Africa. It connects speakers across vast regions — from Senegal to Sudan — giving a sense of unity despite regional differences. Cross-Border Solidarity: Fulfulde is a vehicular cross-border language, used by millions across over 20 countries. It fosters inter-ethnic communication and strengthens pan-Fulani consciousness, especially in trade, religion, and herding communities. In predominantly Muslim Fulani areas, Fulfulde is used to teach and explain Islamic texts and practices. Promotion of Fulfulde is often tied to resisting linguistic marginalisation and asserting ethno-linguistic pride. The language carries indigenous knowledge systems, especially related to cattle herding, agriculture, weather, and medicine. It’s a vehicle for expressing Fulani cosmology and worldview, reinforcing group belonging.
O. Corpora:	Speech Resources Including Fulfulde A collection of unlabeled audio data totaling 671 hours across five languages—French (from Niger), Fulfulde, Hausa, Tamasheq, and Zarma—is available. This dataset supports the development of speech recognition and natural language processing tools for low-resource languages. 1000Langs Super-Parallel Corpus The 1000Langs project includes Fulfulde in its extensive multilin-

	<p>gual corpus, primarily sourced from religious texts like the Bible. This resource is beneficial for comparative linguistic studies and machine translation research.</p>
P. Language situation	<p>High fragmentation and many different varieties; The African Academy of Languages under the African Union (ACALAN) promotes cross-border “vehicular” languages like Fulfulde, endorses unified orthographic standards (ARA).</p>
Q. Dictionaries and grammars:	<p>Wolf, P. P. (1994): English-Fula Dictionary. A Multidialectal Approach. Reimer Verlag GmbH ISBN 9783496025498 Gambian Fula-english Dictionary: https://www.smcm.edu/gambia/wp-content/uploads/sites/31/2015/03/gamble-12.pdf https://lughayangu.com/fula (activist web site *** Stennes (2008): Fula: A Reference Grammar of Adamawa Fulani Publisher: SIM Burkina Faso, Edition: Second Edition, January 2008. 236 pages Harrison, Annette (2003): Fulfulde Language Family Report. Publisher: SIL International Richard W. Smith (2008): Fulfulde Grammar – Burkina Faso Publisher: SIM Burkina Faso. Edition: Second Edition.</p>
R. Language Policy:	<p>Niger: Fulfulde is one of eight recognized national languages. Used in adult literacy programs, local radio, and some early education contexts. Fulfulde is promoted for cultural and practical communication. Nigeria: Recognized under the National Policy on Education as a language of wider communication in the north. May be used in early primary education and local broadcasting. Senegal: Known locally as Pulaar. It is a national language along with Wolof, Serer, and others. Actively promoted by government-funded programs and used in literacy, education, and public broadcasting. Senegal has strong institutional support for Pulaar. Mali: Fulfulde is a national language, used in non-formal education, radio, and increasingly in media and politics. The Malian government promotes mother-tongue education, though implementation remains uneven. Cameroon: Fulfulde is widely spoken in the north but lacks official recognition. Used in radio and religious education. Challenges: Implementation often lags behind policy; Limited resources for teacher training and material development; Some policies are symbolic or underfunded.</p>
S. Note on the pluricentricity	<p>Fula is a PCL by occurrence in several countries. Its status as a PCL is complicated due to high fragmentation and lack of institutional support. It is a highly fragmented language, the varieties are not always intelligible to each other but through strong ethno-linguistic awareness is regarded as a single language. There is little research in this language, additional research is need.</p>
T. Literature (selection):	<p>Ka, Fary (1991). "Problématique de la standardisation linguistique: Le cas du pulaar/fulfulde". In Cyffer, N. (ed.). Language Standardization in Africa. Hamburg: Helmut Buske Verlag. pp. 35–</p>

	<p>38. Arnott, D. W. (2003). "Fula". In Frawley, W. (ed.). International Encyclopedia of Linguistics. Vol. 2. Oxford University Press SIL: Fulfulde Language Family Report https://web.archive.org/web/20180111154945/http://www-01.sil.org/silesr/2003/silesr2003-009.html Fula varieties websites: Nigeria: www.mafindi.com/words www.mafindi.com/words, Nigeria: fulfulde24.com; Nigeria: www.rfi.fr/ff/, Mauritania: pulaar.org, Mauritania: pulaagu.com, Guinea: jowlol.org, Guinea: tabaldefouta.org, Guinea: misiide.net, Guinea: webpulaaku.net, Sierra Leone: peeral.com, Fuuta Tooro: www.pulaaronline.com</p>
--	--

16. German

A. Name of the PCL:	German
B. Endonyms - Other name(s) of the PCL	Deutsch
C. Language family and region	West Germanic, located in Central Europe
D. Demography and language spread	ca. 95 million native speakers, about 60 million L2 speakers; native to Germany, Austria, Switzerland, Liechtenstein, Luxembourg, Belgium, South Tyrol (Italy)
E. Status-OFL in:	8 independent nations or regions: Germany, Austria, Switzerland, Liechtenstein, Luxembourg, Belgium, South Tyrol (Italy), Namibia; 12 International organisations: European Patent Organisation , Unified Patent Court , European Space Agency , EU, OECD , International Union of Railways , European Centre for Medium-Range Weather Forecasts , International Trade Union Confederation , World Association of Newspapers and News Publishers , Danube Commission , European Investment Bank , European Bank for Reconstruction and Development
F. Status-RML in:	8 countries: Brazil (6 municipalities), Czech Republic, Denmark, Hungary, Poland, Romania, Russia, Slovakia
G. DV:	German German
H. NDVs of the PCL	Austria, Switzerland, Liechtenstein, Luxembourg, Belgium, South Tyrol (Italy), Namibia
I. Standard form:	German German, Austrian German, Swiss German
J. Note on the pluricentricity	German is a highly researched language that has no official regulatory body. The dominance of German German is strong and enforced by language export via language teaching and the media. Efforts to improve the status of Austrian and Swiss German are made.
K. Alphabet / WS	Latin script
L. Standardisation / Codification, regulated by	No official regulatory body except for the orthography: Deutscher Rechtschreibrat (German Council for Orthography); the language is regulated; Indirectly regulated by dictionary office of Duden dictionary and other institutions; Indirectly regulated by the dictionary office of the Österreichische Wörterbuch; Switzerland

	has no German dictionary of its own
M. Mutual intelligibility:	The NVs are mutually intelligible. The regional varieties differ substantially from each other and may not be mutually intelligible
N. Linguistic fragmentation:	Substantial linguistic fragmentation, especially in the south of the language area
O. Language situation	German is an important language in Europe
P. Awareness / Function for identity:	High relevance for identity particularly in Germany and (partly) in Switzerland
Q. Corpora:	Many: Datenbank für Gesprochenes Deutsch - DGD (Database for Spoken German); Deutscher Referenzkorpus - DeReKo over 50 billion written words. Korpora in DWDS (Corpora in German Vocabulary 1600 till the Present); Deutsches Textarchiv (DTA); Kernkorpus (German Text Archive - Key Corpus);
R. Dictionaries and grammars:	DWDS: Digitales Wörterbuch der deutschen Sprache Duden Grammatik der deutschen Sprache Gisela Zifonun, Ludger Hoffmann, Bruno Strecker et al.: Grammatik der deutschen Sprache (= Schriften des Instituts für Deutsche Sprache. Band 7, Nr. 1–3). 3 Bände. De Gruyter, Berlin/New York 1997 https://grammis.ids-mannheim.de/ Information system about the Grammar of German
S. Language Policy:	Germany practises massive language export via schools and international organisations (Goethe institute); The One Standard German Axiom is now again maintained as a core assumption of German linguistics by some German linguists that deny the pluricentric character of German.
T. Literature (selection):	Ammon, Ulrich; Bickel, Hans; Ebner, Jakob; Gasser, Markus; Esterhammer, Ruth (2004). Variantenwörterbuch des Deutschen: Die Standardsprache in Österreich, der Schweiz und Deutschland sowie in Liechtenstein, Luxemburg, Ostbelgien und Südtirol Berlin: W. de Gruyter. ISBN 978-3-11-016575-3. Barbour, Stephen; Stevenson, Patrick (1990). Variation in German. Cambridge: Cambridge University Press. ISBN 978-0-521-35704-3. Muhr, Rudolf (2020): Eine kurze politische Geschichte des Österreichischen Deutsch. In: Köhler, Thomas Walter/ Mertens, Christian / Pelinka, Anton (Hg.): Ein Hauch von Welt. Österreich vor und nach Saint Germain. Wien: Braumüller Verlag. S. 499-524. Dollinger, Stefan (2021). Österreichisches Deutsch oder Deutsch in Österreich? Identitäten im 21. Jahrhundert [Austrian German or German in Austria: Identities in the 21st Century] (3rd ed.). Vienna: New Academic Press. ISBN 978-3-99036-023-1.

17. Greek

A. Name of the PCL:	Greek
B. Endonyms - Other name(s) of the PCL	Ελληνικά, Hellēniké
C. Language family and region	Indo-European language, Hellenic branch, located in south east Europe
D. Demography and language spread	ca. 13,5 million native speakers; native to Greece and Cyprus and Akrotiri and Dhekelia (British military base in Cyprus), Large minorities in Albania, southern Italy, and by the diaspora in the US, Australia
E. Status-OFL in:	8 independent nations or regions: Greece, Cyprus, European Union, Akrotiri and Dhekelia; International organisations: EU
F. Status-RML in:	Armenia, Hungary, Romania, Ukraine
G. DV:	Greek Greek
H. NDVs of the PCL	Cyprus Greek, Dhekelia Greek
I. Standard form:	Greek Greek
J. Note on the pluricentricity	Greek is a highly centralised language where there is the believe that there is a single norm from ancient times until today
K. Alphabet / WS	Greek alphabet
L. Standardisation / Codification, regulated by	Centre for the Greek Language (Κέντρον Ελληνικής Γλώσσας): Established in 1994 by the Government of Greece, it operates under the supervision of the Ministry of Education (and co-supervision of the Ministries of Foreign Affairs and Culture). Academy of Athens (Ακαδημία Αθηνών): Founded in 1926 as Greece's National Academy and highest research institution, under the Ministry of Education. Through its "Research Centre for Modern Greek Dialects" and its Committee on Modern Greek Language, it conducts fundamental research on the language's history, dialectology and lexicography.
M. Mutual intelligibility:	The NVs are mutually intelligible.
N. Ling. fragmentation:	Little linguistic fragmentation
O. Language situation	Greek was a culturally important language in ancient times and is still relevant in the historical subjects of science. There is diglossia between the Dimotiki variety and Katharevousa, the conservative variety that is used by the Orthodox church and the diaspora.
P. Awareness / Function for identity:	High relevance for identity
Q. Corpora:	Corpus of modern Greek: http://web-corpora.net/ GreekCorpus/search/; eITenTen: Corpus of the Greek Web: https://www.sketchengine.eu/eltenten-greek-corpus/ ; Hellenic National Corpus (HNC) TM ; Hellenic National Corpus (HNC) TM : https://hnc.ilsp.gr/index.php?current_page=main&lang=en
R. Dictionaries and grammars:	The European Dictionary Portal – Greek dictionaries: http://www.dictionaryportal.eu/en/ctlg/?objLang=el Holton, David; Mackridge, Peter; Philippaki-Warbuton, Irene

	(1997). <i>Greek: A Comprehensive Grammar of the Modern Language</i> . London and New York: Routledge.
S. Language Policy:	There is no explicit language policy known
T. Literature (selection):	Karyolemou, Marilena (2012): <i>Cypriot Greek as a non dominant variety of Greek</i> . In: Rudolf Muhr (ed.) (2012): <i>Non-dominant Varieties of pluricentric Languages. Getting the Picture</i> . Lang Verlag. p.167-186 <i>Greek language</i> ". Encyclopædia Britannica. Encyclopædia Britannica Zacharia, Katerina, ed. (2008). <i>Hellenisms: culture, identity, and ethnicity from antiquity to modernity</i> . Aldershot, England ; Burlington, VT: Ashgate; Horrocks, Geoffrey (1997). <i>Greek: A History of the Language and Its Speakers</i> . London and New York: Longman Linguistics Library

18. Guarani

A. Name of the PCL:	Guarani
B. Endonyms - Other name(s) of the PCL:	avañe'ẽ (the language of the people" or "language of man)
C. Language family and region:	South American, Tupian language, Tupi–Guarani branch; native to Paraguay, southeastern Bolivia and southwestern Brazil, Argentina.
D. Demography and language spread:	Paraguay: L1 3.915.089 (85,9% of population), L2 30.232 (0.7 %) (2002); Bolivia L1/L2 ~79.000 (2012); Argentina L1/L2 115.000; Brazil: ~33.500.
E. Status-OFL in:	Paraguay (co-official with Spanish), Bolivia
F. Status-RML in:	None
G. DV:	Paraguay Guarani
H. NDVs of the PCL:	Bolivian Guarani
I. Standardisation, regulated by:	Paraguay: Academia de la Lengua Guaraní (Guaraní Ñe'ẽ Rerekupavẽ), created by Paraguay's Language Law 4251/10 (2010),
J. Alphabet / WS:	Latin, Guarani alphabet with 33 letters and digraphs
K. Standard form:	Paraguayan Guaraní (Avañe'ẽ) is the de facto standard variety of Guaraní used in education, media and official documents in Paraguay.
L. Ling. fragmentation:	Regional varieties: Jopará, Taragui
M. Mutual intelligibility:	High Intelligibility between Paraguayan, Argentinian, and most Bolivian varieties. Speakers can usually understand each other easily, although vocabulary and pronunciation differ slightly. Lower Intelligibility between Paraguayan Guaraní and some Brazilian varieties like Kaiowá. Jopará speakers (a blend of Spanish and Guaraní): People who speak "mixed" Spanish-Guaraní might find it a bit harder to understand the more traditional or "pure" varieties like Bolivian Guaraní.
N. Dictionaries and grammars:	Máximo Ricardo, Dacunda Díaz (2006): <i>Gran Diccionario de la Lengua Guaraní: Vocabulario Bilingüe, Gramática, Sintaxis, Cultura Guaraní</i> . Publisher: Moglia Ediciones. ISBN: 978-9871035922. 472 p.

	<p>***</p> <p>Estigarribia, Bruno (2020): A Grammar of Paraguayan Guarani Publisher: UCL Press. ISBNs: Paperback: 978-1-78735-292-6, 386p.</p>
O. Corpora:	<p>GitHub: Jojajovai Corpus. Reference: Chiruzzo et al., LREC 2022 Guarani News and Social Media Corpus: Comprises a parallel Guarani–Spanish news corpus and a monolingual set of Guarani tweets. Reference: Góngora et al., AmericasNLP 2021. Corpus of Paraguayan Spanish and Guarani (CEGPA): Data: Contains 14 hours of audio interviews with residents of Asunción, Paraguay, aged 18 to 91. Hosted by the University of California, Santa Cruz. Guarani Wikipedia: An active platform with over 3,000 articles in Guarani, offering a rich source of contemporary written language.</p>
P. Language situation:	<p>Paraguay: Law No. 4.251/2010 mandates the State to promote Guarani’s recognition in supranational organizations and to preserve and use native American languages nationally and internationally. In addition, the Secretaría de Política Lingüística (2011) and the Academia de la Lengua Guaraní (2012) were established to coordinate language policy and standardization efforts.</p> <p>Bolivia: Although Guaraní has constitutional status, Spanish remains dominant in government, mass media, commerce, and most secondary and higher education contexts. Guaraní is still used in rural communities for daily communication, traditional ceremonies, and local markets, but is largely absent from national broadcasting and formal business. Since 2013, all Bolivian schools are required to teach at least one Indigenous language, leading many institutions in Guaraní-speaking areas to incorporate Guaraní into their curricula. Bolivian Guaraní is classified as vulnerable by UNESCO, with younger generations increasingly shifting to Spanish in urban settings and among mestizo populations.</p> <p>Argentina: Spanish predominates in formal domains—schools, churches, government offices—while Guaraní thrives in home, folk culture, humor, and intimate social contexts, reflecting a diglossic pattern common in minority-language communities. Bilingual education in Corrientes has integrated Guaraní into primary school curricula since 2004, though implementation quality and resources remain uneven. Urban migration and social prejudice have led to uneven language maintenance, with younger generations often favoring Spanish. Nonetheless, local cultural movements, social media content, and academic programs are fostering renewed interest in Guaraní as a marker of regional identity.</p> <p>Brazil: Guaraní in Brazil is primarily a community language used in rituals, inter-village communication, cultural ceremonies, and land-rights mobilizations. Portuguese dominates in all official, commercial, and most educational contexts. While national policy allows for Indigenous-language instruction, implementation depends on local demand and resources. A handful of Indigenous schools incorporate Guaraní in early grades, often supported by</p>

	<p>FUNAI and NGOs. Guaraní faces intense pressure from Portuguese, territorial conflicts, and socio-economic marginalization. Community activists continue to press for expanded bilingual programs, media content in Guaraní, and legal recognition of municipal co-official status in Guaraní-speaking areas.</p>
<p>Q. Awareness / Function for identity:</p>	<p>Paraguay: Approximately 70 % of Paraguayans over age five speak Guaraní habitually, and the country commemorates August 25 as “Día del Idioma Guaraní” in recognition of its deep social penetration. The Yvy Marãe’ỹ Foundation (est. 2007) advocates for Guaraní diffusion, teaching, and research, and partners with the Senate and other bodies to provide interpreter services and develop terminology resources.</p> <p>Bolivia: There is a resurgence of Guaraní ethnic identity and collective self-assertion over the past two decades. Indigenous communication initiatives such as “YASAREKOMO” and bilingual education programs underscore Guaraní’s role as a carrier of history, memory, and culture, demanding formal recognition after centuries of marginalization.</p> <p>Argentina: Regional Co-Officialisation and Cultural Symbols. In Corrientes Province, where Guaraní is co-official, the language features in public signage, school curricula, and cultural festivals, signaling local identity and pride.</p> <p>Brazil: The “Guaraní Way of Being”: Guaraní life in Brazil is ideologically structured around <i>ava ñe’ë</i> (language as soul), <i>tamöi</i> (mythic ancestors), and <i>ava reko</i> (social customs), which together define a distinctive ethnic identity. The Guaraní Yvyrupa Commission (CGY) unites leaders from Kaiowá, Mbya, and Ñandeva villages to advocate for territorial recovery and cultural sovereignty at the national level.</p>
<p>R. Language Policy:</p>	<p>Bolivia: The 2009 Political Constitution of the plurinational State recognizes “Castilian” and 36 indigenous languages (including Guaraní) as official throughout the territory.</p> <p>Bolivia: Bolivia’s General Law of Linguistic Rights and Policies requires all government institutions to operate in at least two languages—Spanish and one indigenous language—thereby institutionalizing bilingual public services. Education Law 070 “Avelino Siñani–Elizardo Pérez” (2010) mandates intracultural, intercultural, and plurilingual education, ensuring that Guaraní and other indigenous languages are used as mediums of instruction in communities where they predominate.</p> <p>Argentina: Corrientes Province enacted Provincial Law No. 5.598 on October 22, 2004, declaring Guaraní co-official alongside Spanish and obligating its use in government and education. In 2010, Chaco Province followed suit, also granting Guaraní co-official status and integrating it into local administrative and educational contexts.</p> <p>Brazil: Although Brazil lacks a national policy co-officializing indigenous languages, several municipalities in Mato Grosso do Sul have enacted local laws recognizing Guaraní varieties as co-official. These include Amambai (Guaraní Kaiowá), Coronel Sapucaia (Guaraní), Tacuru (Guaraní since 2010), and Paranhos (ap-</p>

	proved pending regulation), each requiring the use of Guarani in municipal administration and public education.
S. Note on the pluricentricity:	The pluricentricity of Guarani is not well developed as there is a strong asymmetry in the number of speakers and in its status as national versus regional language. The language only partly fulfils all criteria. There is a strong asymmetry between the better equipped and promoted Paraguayan Guarani and the other varieties.
T. Literature (selection):	Estigarribia & Pinta (Eds.), Guarani Linguistics in the 21st Century. Publisher: Brill. DOI: 10.1163/9789004322578_009 Estigarribia, Bruno (2015). Guarani-Spanish code-switching in Asunción: Socio-linguistic dimensions. Journal: International Journal of Bilingualism, 19(5), 539–558. DOI: 10.1177/1367006913516034. Gynan, Shaw N. (2001). Language policy and planning in Paraguay. Journal: Current Issues in Language Planning, 2(1), 53–118. DOI: 10.1080/14664200108668016 Michael, Lev (2008). "Language ideologies and the production of linguistic differentiation: the case of Amazonian languages." Journal: Language in Society, 37(5), 517–549. DOI: 10.1017/S0047404508080745

19. Hausa

A. Name of the PCL:	Hausa
B. Endonyms - Other name(s) of the PCL	Harshen/Halshen, Hausa; Ajami: هَرْشَن هَوْسَا
C. Language family and region	Afro-Asiatic language, West Chadic, located in Central Africa
D. Demography and language spread	ca. 94 million native and L2 speakers; native to a large number of African countries: Nigeria, Niger, Cameroon, Ghana, Benin, Togo, Chad, Sudan, Ivory Coast.
E. Status-OFL in:	In three independent nations: Nigeria, Niger
F. Status-RML in:	Hausa is lingua franca in many countries of Central Africa but does not have an official status as ML there (Cameroon, Ghana, Benin, Togo, Chad, Sudan, Ivory Coast).
G. DV:	Nigerian Hausa
H. NDVs of the PCL	The African Academy of Languages (ACALAN) harmonizes orthographies and standards for Hausa across borders. Within Nigeria, the National Institute for Nigerian Languages (NINLAN) oversees research, teaching, documentation, and coordination of Hausa, operating under the National Universities Commission. In Niger, the Ministry of Primary Education, Literacy, Promotion of National Languages, and Civic Education implements constitutional and statutory provisions for Hausa as a "national language." Supplementary efforts by civil-society and educational programs (e.g., HLTA, ELAN) further support technological development and classroom integration of Hausa.

I. Standard form:	Daura (Dauranchi) and Kano (Kananci) variety
J. Alphabet / WS	Boko (Latin), Ajami (Arabic)
K. Standardisation, regulated by	No official regulatory body, but norm setting is done in Nigeria via text books for schools and media use
L. Mutual intelligibility:	The NVs are mutually intelligible.
M. Ling. fragmentation:	High uniformity and low linguistic fragmentation
N. Language situation	Expanding in use throughout Central Africa
O. Awareness / Function for identity:	High relevance for identity
P. Corpora:	Hausa corpus (haWaC) collected from the internet: https://www.sketchengine.eu/hawac-hausa-corpus/ Ibrahim, Umar A. (2021): Hausa Speech Corpus. https://data.mendeley.com/datasets/j6kjmfrbby/2
Q. Dictionaries and grammars:	https://www.lexilogos.com/english/hausa_dictionary.htm Robinson, Charles Henry (2011): Dictionary of the Hausa Language. 2 Volume Paperback Set. Cambridge: OUP Ma, Roxana/Paul Newman (2020): Hausa Dictionary for Everyday Use: Hausa-English/English-Hausa. Baltimore, Maryland, USA, Bayero University Press Newman Paul (2000): The Hausa Language: An Encyclopedic Reference Grammar. Yale University Press. 800 pages
R. Language Policy:	There is no explicit language policy known, except that it is an indigenous language that is promoted in primary school
S. Note on the pluricentricity	Hausa is an important and expanding language in Central Africa. Mostly found in southern Niger and northern Nigeria. It serves as an important lingua franca across Central Africa. The pluricentricity is weak and fulfils only the criteria 1-3.
T. Literature (selection):	Ziegelmayr, Georg (2016): Hausa – A pluricentric language of West Africa? In: Rudolf Muhr, et. al. (eds.) (2016): Pluricentric Languages and non-dominant Varieties worldwide. Volume 1: Pluricentric Languages across continents - Features and usage. Wien et. al., Peter Lang Verlag. p. 127-138. Philips, John Edward. "Hausa in the Twentieth Century: An Overview." in Sudanic Africa, vol. 15, 2004, pp. 55–84 Cyffer, Norbert (1977): Sprachplanung in Nigeria. Africa Spectrum 12/3. 239-262.; Jaggard, Philip J. (2001): Hausa. Amsterdam/Philadelphia. John Benjamins. Furniss, Graham (1991): Standards in speech, spelling and style - The Hausa case. In: Norbert Cyffer, Klaus Schubert, Hans-Ingolf Weier, and Ekkehard Wolff (eds.), Language Standardization in Africa – Sprachstandardisierung in Afrika – Standardisation des Langues en Afrique. Hamburg. Helmut Buske. 97-111.

20. Hindi

A. Name of the PCL:	Hindi
B. Endonyms - Other name(s) of the PCL	Modern Standard Hindi, आधुनिक मानक हिन्दी, Ādhunik Mānak Hindī
C. Language family and region	Asiatic language, Central Indo-Ayran, Hindustani, located in northern India and eastern Pakistan.
D. Demography and language spread	ca. 909.470 million speakers; L1 ca. 422.048 mio. and 487 mio. L2 speakers; 41% of the population of India and 7,4% of Pakistan speak Hindi; native to India and Pakistan (Official language in the Indian states: Bihar, Chhattisgarh, Haryana, Himachal Pradesh, Jharkhand, Madhya Pradesh, Rajasthan, Uttar Pradesh, Uttarakhand)
E. Status-OFL in:	India
F. Status-RML in:	South Africa, UAE
G. DV:	Indian Hindi
H. NDVs of the PCL	Pakistani Hindi
I. Standard form:	Khari Boli
J. Note on the pluricentricity	Hindi is a cover term – a kind of dachsprache –with a number of languages that are subsumed in the Indian constitution as "mother tongues": 50 languages that often strongly differ linguistically from each other are subsumed under the term "Hindi". Everyday Hindi and everyday Urdu are mutually intelligible and were considered as single language under the name "Hindustani". Hindi and Urdu can be considered as NVs even though each of them are only an Status-OFL in: a single nation.
K. Alphabet / WS	Devanagari script
L. Standardisation / Codification, regulated by	Regulated by Central Hindi Directorate
M. Mutual intelligibility:	The NVs are mutually intelligible.
N. Ling. fragmentation:	High linguistic fragmentation through the linguistically strongly different "mother tongues" that are subsumed under the term Hindu.
O. Language situation	Hindi is the de facto national language of India and linked to Hinduism as religion. It is in some varieties intelligible with Urdu which linked to Islam
P. Awareness / Function for identity:	High relevance for religious identity (Hindu) and Urdu (Muslim)
Q. Corpora:	Linguistic Data consortium for Indian Languages: https://data.ldcil.org/a-gold-standard-hindi-raw-text-corpus Hindi corpora at Sketch Engine (web corpora, parallel corpora etc.): https://www.sketchengine.eu/corpora-and-languages/hindi-text-corpora/ ; Twine Hindi Language Datasets of 2022: https://www.twine.net/blog/hindi-language-datasets/
R. Dictionaries and grammars:	Agnihotri, Rama Kant (2006): Hindi. An Essential Grammar-London: Routledge; McGregor, R.S. (1995). Outline of Hindi

	<p>grammar: With exercises (3. ed.). Oxford: Clarendon Pr. ISBN 978-0-19-870008-1.</p> <p>McGregor, R.S. (1993) Oxford Hindi–English Dictionary (2004 ed.), Oxford University Press, USA.</p> <p>Hindi Dictionaries: https://www.lexilogos.com/english/hindi_dictionary.htm;</p> <p>The Oxford Hindi-English Dictionary online (Part of Digital Dictionaries of South Asia): https://dsal.uchicago.edu/dictionaries/mcgregor/ ***</p> <p>Koul, Omkar N. (2008): Modern Hindi Grammar. Publisher: Dunwoody Press, ISBN: 978-1-931546-06-5.</p> <p>McGregor, R. S. (1995): Outline of Hindi Grammar: With Exercises by Publisher: Oxford University Press. 3rd Edition: 978-0-19-870008-1</p>
S. Language Policy:	Policy of Sanskritisation in India that makes Urdu and Hindi differ from each other;
T. Literature (selection):	<p>Rahman, Tariq (2021): From Hindu to Urdu. A social and political history. Graz, PCL-Press. 476p.</p> <p>Ghosh, Aditi (2012): Bhojpuri as a non-dominant variety of Hindi. In: Rudolf Muhr (ed.) (2012): Non-dominant Varieties of pluricentric Languages. Getting the Picture. In memory of Michael Clyne. Wien et. al., Peter Lang Verlag. p. 435-452.</p> <p>Dua, Hans R. (1992): Hindi-Urdu as a pluricentric language. In: Pluricentric languages: differing norms in different nations. Michael Clyne (ed.). Berlin/New York. Mouton de Gruyter, 381-400.</p> <p>Rai, Alok (2001): Hindi nationalism. Delhi. Orient Longman.</p>

21. Hungarian

A. Name of the PCL:	Hungarian
B. Endonyms - Other name(s) of the PCL	Magyar (magyar nyelv)
C. Language family and region	Uralic, Ugric language, located in Central Europe
D. Demography and language spread	ca. 13 million speakers; native to Austria, Bosnia and Herzegovina, Croatia, Romania, Serbia, Slovakia, Slovenia, and Ukraine
E. Status-OFL in:	Hungary, large regional language in Romania and Slovakia
F. Status-RML in:	Austria, Bosnia and Herzegovina, Croatia, Romania, Serbia, Slovakia, Slovenia, Ukraine
G. DV:	Hungarian Hungarian
H. NDVs of the PCL	Slovakian Hungarian, Romanian Hungarian, Croatian Hungarian, Ukrainian Hungarian, Slovenian Hungarian, Serbian Hungarian
I. Standard form:	Hungarian Hungarian, Slovakian Hungarian, Romanian Hungarian
J. Note on the pluricentricity	Hungarian is an Status-OFL in: Hungary only. But it is also a large regional language especially in Romania, Slovakia, Serbia, and

	Ukraine, which makes it to a PCL, even though it is only recognised as a RML.
K. Alphabet / WS	Latin script, Hungarian alphabet (44 letters)
L. Standardisation / Codification, regulated by	Hungarian Research Centre for Linguistics at the Hungarian Academy of Sciences, (https://nytud.hu/en)
M. Mutual intelligibility:	The NVs and the regional varieties in the different countries are mutually intelligible.
N. Ling. fragmentation:	High linguistic fragmentation
O. Language situation	The pluricentric character of Hungarian is no recognised and the other national varieties are codified.
P. Awareness / Function for identity:	Strong connection between nation and national language.
Q. Corpora:	Hungarian national corpus: http://corpus.nytud.hu/mnsz/index_eng.html
R. Dictionaries and grammars:	https://www.lexilogos.com/english/hungarian_dictionary.htm https://szotar.sztaki.hu/en/dictionaries Hungarian-English/English-Hungarian Concise Dictionary (1995): (Hippocrene Concise Dictionary) *** Keresztes, László (1999). A practical Hungarian grammar (3rd, rev. ed.). Debrecen: Debreceni Nyári Egyetem. Törkenczy, Miklós (2002). Practical Hungarian grammar: [a compact guide to the basics of Hungarian Grammar. Budapest, Corvina. http://www.hungarianreference.com/
S. Language Policy:	No specific language policy at the moment.
T. Literature (selection):	Muhr Rudolf et. al. (eds.) (2020): Hungarian as a Pluricentric Language in Language and Literature. Frankfurt et. al., Lang Verlag. 290 p. Huber, Máté Imre (2016): Problems and advantages of treating Hungarian as a pluricentric language. In: Rudolf Muhr, et. al. (eds.) (2016): Pluricentric Languages and non-dominant Varieties worldwide. Volume 1: Pluricentric Languages across continents - Features and usage. Wien et. al., Peter Lang, p. 303-314. Bíró, Anna (2016): An insight into Serbian Hungarian. In: Rudolf Muhr, et. al. (eds.) (2016): Pluricentric Languages and non-dominant Varieties worldwide. Volume 1: Pluricentric Languages across continents - Features and usage. Wien et. al., Peter Lang Verlag. p. 351-362.

22. Italian

A. Name of the PCL:	Italian
B. Endonyms - Other name(s) of the PCL	Italiano, lingua italiana
C. Language family and region	Romance, Italo-Romance language, located in south Europe; native to Italy, San Marino, Switzerland (Ticino and the Grisons), and Vatican City;
D. Demography and language spread	ca. 67 million native speakers; 85 million including the L2 speakers in other nations; Language spread via the Italian Società Dante Alighieri (400 offices abroad).
E. Status-OFL in:	Italy, Switzerland (Ticino and the Grisons), San Marino
F. Status-RML in:	Bosnia and Herzegovina, Croatia, Romania, Slovenia
G. DV:	Italian Italian
H. NDVs of the PCL	Swiss Italian, San Marino Italian
I. Standard form:	Italian Italian, Swiss Italian, San Marino Italian
J. Alphabet / WS	Latin script
K. Standardisation / Codification, regulated by	Accademia della Crusca (Italy), Research institute: Osservatorio linguiscodella Svizzera italiana in Bellinzona, Switzerland
L. Mutual intelligibility:	The NVs are mutually intelligible.
M. Ling. fragmentation:	Large number of regional languages and regional varieties and high degree of linguistic fragmentation: Friulian, Ligurian, Lombard, Neapolitan, Piedmont, Sardinia, Sicilian and Venetian, Trentino, Tuscany etc.
N. Language situation	The pluricentric character of Italian is now acknowledged and the Swiss variety is codified.
O. Awareness / Function for identity:	Language pride is considerable.
P. Corpora:	PAISA: Italian corpus: https://www.corpusitaliano.it/en/contents/paisa.html Corpora of Italian language: https://biblio.sns.it/en/corpora-italian-language CORDIS/CODIS: https://corpora.ficlit.unibo.it/coris_eng.html
Q. Dictionaries and grammars:	https://www.lexilogos.com/english/italian_dictionary.htm Grande dizionario della lingua italiana: https://www.gdli.it/Ricerca/Pagina?q= University of Michigan Library: https://guides.lib.umich.edu/c.php?g=282798&p=1884354 Langenscheidt Komplett-Grammatik Italienisch, ISBN: 978-3-12-563468-8
R. Language Policy:	In 2023, the Italian government proposed legislation to penalize the use of foreign words, particularly English, in official communications. In March 2025, the Ministry of Education banned the use of gender-neutral symbols, such as the asterisk (*) and the schwa (ə), in schools. The government has introduced curriculum changes emphasizing classical studies, including Latin and

	Italian history, to instill patriotism and reinforce national identity. These measures are linked to the post-fascist governments under prime minister Meloni.
S. Note on the pluricentricity	Due to the small number of speakers in San Marino, Italian only has a one real linguistic centre outside Italy, namely in Switzerland, where it is a recognised national language in the canton of Ticino and the Grisons. Codification work began after 1991 and the codification centre "Osservatorio linguistico della Svizzera italiana" was set up in Bellinzona, Switzerland
T. Literature (selection):	Baranzini, Laura , Ricci, Claudia (2023): TV news in Italian-speaking Switzerland: Linguistic authorities and standard-setting. In: Reglindis De Ridder (ed.) (2023): "One size fits all"? Linguistic standards in the media of pluricentric language areas: PCL-PRESS/epubli. Graz/Berlin p. 197-219. Hajek, John (2012): (Non-) dominant varieties of a (non-) pluricentric language? Italian in Italy and Switzerland. In: Rudolf Muhr (ed.) (2012): Non-dominant Varieties of pluricentric Languages. Getting the Picture. In memory of Michael Clyne. Wien et. al., Peter Lang Verlag. p. 155-166.

23. Irish

A. Name of the PCL:	Irish
B. Endonyms - Other name(s) of the PCL:	Gaeilge, Gaelic, Irish Gaelic Gaeilg/Gaeilic/Gaeilig/Gaelainn/Gaoluin (names in the local varieties)
C. Language family and region:	Celtic, Goidelic language, located in north-west Europe on the British Isles
D. Demography and language spread:	ca. 1,8 million native speakers (2022); 40% can of the population can speak Irish, but only 10% speak it very well; Irish; Irish is the national and first language of the Republic of Ireland and Ulster (Northern Ireland) alongside English, Ulster Scots
E. Status-Status-OFL in:	Republic of Ireland and Ulster (Northern Ireland), Scotland, EU
F. Status-RML in:	None
G. DV:	Irish Irish
H. NDVs of the PCL:	Ulster Irish
I. Standard form:	An Caighdeán Oifigiúil ("The Official [Written] Standard")
J. Alphabet / WS:	Latin (Irish alphabet)
K. Standardisation / Codification, regulated by:	Foras na Gaeilge (Irish Institute) for both varieties
L. Mutual intelligibility:	The NVs are largely mutually intelligible.
M. Ling. fragmentation:	Three regional varieties: Connacht, Munster and Ulster Irish.
N. Language situation:	Irish was an endangered language until the late 1990ies. The number of speakers is now rising both in Ireland and in Ulster.
O. Awareness / Function for identity:	Language pride is considerable.
P. Corpora:	National Corpus of Irish: https://www.corpas.ie/en/cng/

	gaTenTen: Corpus of the Irish Web: https://www.sketchengine.eu/gatenten-irish-corpus/
Q. Dictionaries and grammars:	Concise English-Irish Dictionary (2020) from Foras na Gaeilge eDIL - Electronic Dictionary of the Irish Language: https://dil.ie/ Grammar Database: https://www.teanglann.ie/en/gram/
R. Language Policy:	No specific language policy at the moment.
S. Note on the pluricentricity:	The pluricentricity of Irish fulfils all 7 criteria and is now fully acknowledged with the codification jointly done by the Irish language institute.
T. Literature (selection):	Hickey, Raymond (2019): Pluricentricity and Irish English. In: Rudolf Muhr, et. al. (eds.) (2019): European Pluricentric Languages in Contact and Conflict: Wien et. al., Peter Lang Verlag. p. 141-154.

24. Swahili/Kishwahili

A. Name of the PCL:	Swahili
B. Endonyms - Other name(s) of the PCL:	Kishwahili, Standard Swahili (Kiswahili Sanifu)
C. Language family and region:	Niger–Congo, Bantu language, Sabaki branch; located in south-east Africa; native to Tanzania, Kenya, Mozambique, Democratic Republic of the Congo, Bajuni Islands and Brava, southern Somalia.
D. Demography and language spread:	ca. 200 million L1 and L2 speakers (UNESCO 2021); 14 countries: Tanzania, Kenya, Uganda, Rwanda, Burundi, the Democratic Republic of the Congo (DRC), South Sudan, Somalia, Mozambique, Malawi, Zambia, Comoros, and as far as Oman and Yemen in the Middle East. Southern African countries such as South Africa and Botswana have introduced it in schools, while Namibia and others are considering doing so.
E. Status-Status-OFL in:	Tanzania, Rwanda, Kenya, Uganda; International: African Union, East African Community
F. Status-RML in:	Democratic Republic of the Congo (DRC): recognized minority language with official functions. South Africa: recognized in education as a minority language. In Rwanda, Burundi, Mozambique, and Somalia, Swahili is spoken by some communities but does not have formal recognition as a minority language.
G. DV:	Tanzanian Swahili
H. NDVs of the PCL:	Rwanda Swahili, Kenya Swahili, Uganda Swahili;
I. Standard form:	Standard Swahili (Kiswahili Sanifu), based on the Kiunguja variety in Zanzibar.
J. Alphabet / WS:	Latin (Irish alphabet)
K. Standardisation, regulated by:	Taasisi ya Uchunguzi wa Kiswahili (Institute of Kiswahili Research) Baraza la Kiswahili la Taifa (Tanzania); Chama cha Kiswahili cha Taifa (Kenya)
L. Mutual intelligibility:	The NVs are largely partly intelligible.

M. Ling. fragmentation:	Many different varieties in the single countries where Swahili is spoken.
N. Language situation:	Swahili is expanding by the number of its speakers.
O. Awareness / Function for identity:	Swahili as “our language.” Many speakers—especially in Tanzania and Kenya—are taught from early school that Kiswahili is the glue of the nation. Awareness of dialectal diversity. Swahili speakers recognize regional varieties (e.g. Kiunguja, Kimvita, Kingwana) but often rally around Kiswahili Sanifu as the prestige standard. Nation-building: Tanzania’s post-independence government deliberately elevated Swahili to foster cross-ethnic unity (versus English, which was seen as elitist). Pan-East African identity: Beyond national borders, Swahili serves as a shared cultural and communicative code across Kenya, Uganda, DR Congo, Rwanda/Burundi borderlands, and even parts of southern Somalia and Mozambique. Swahili pop music (Bongo Flava), television soap operas (e.g. Mambo Leo), and literature reinforce collective memory and shared values
P. Corpora:	Language Bank of Finland: Helsinki Corpus of Swahili 2.0 https://www.kielipankki.fi/corpora/hcs2-group/ Scetch Engine swWaC: Swahili corpus from the web https://www.sketchengine.eu/swwac-swahili-corpus/
Q. Dictionaries and grammars:	New Standard Swahili Dictionary (KAMUSI), published by the University of Dar es Salaam; Kapinga’s ‘Sarufi Maumbo ya Kiswahili Sanifu’ was an accompanying publication. Mpiranya, Fidèle (2014): Swahili Grammar and Workbook. Taylor & Francis, Bad Hersfeld 978-1-138-80826-3
R. Language Policy:	No specific language policy at the moment.
S. Note on the pluricentricity:	The pluricentricity of Swahili fulfils all 7 criteria. However there is no general agreement on a unified standard variety.
T. Literature (selection):	Waldburger, Daniela (2016): Swahili in Eastern Congo: Status, role and attitudes. In: Rudolf Muhr, et. al. (eds.) (2016): Pluricentric Languages and non-dominant Varieties worldwide. Volume 1: Pluricentric Languages across continents - Features and usage. Wien et. al., Peter Lang Verlag. p. 139-154. Robinson, Morgan J.: A Language for the World. The Standardization of Swahili. Athens 2022 , ISBN 978-0-8214-2495-7 Blommaert, Jan (2013): State Ideology and Language in Tanzania. Second and revised edition. Tilburg Papers in Culture Studies, Vol. 80.

25. Kongo/Kikongo

A. Name of the PCL:	Kongo
B. Endonyms - Other name(s) of the PCL:	<i>Kikongo ya Leta</i> (Democratic Republic of the Congo), <i>Kituba</i> (Republic of the Congo), Kileta, Kibula-matari
C. Language family and region:	Niger–Congo, Bantu language, Sabaki branch; located at the West coast of Central Africa; native to DR Congo, Republic of the Congo, Angola, Gabon
D. Demography and	ca. 6.0 million L1 and 5.0 million L2 speakers

language spread:	
E. Status-Status-OFL in:	<i>Kikongo ya Leta</i> : recognised as "national language" in the constitution of the Democratic Republic of the Congo and under the name <i>Kituba</i> is recognised as a "national language" in the Republic of the Congo and in Angola
F. Status-RML in:	unclear
G. DV:	Kikongo ya Leta Kituba
H. NDVs of the PCL:	Angolan Kikongo
I. Standard form:	Kikongo ya Leta, Munukutuba (Republic of the Congo)
J. Alphabet / WS:	Latin
K. Standardisation / Codification, regulated by:	The Centre de Linguistique Appliquée de Kinshasa (CELTA) and other language institutes in universities (like the University of Kinshasa) engage in linguistic research and curriculum development for national languages, including Kikongo and Kituba. Republic of the Congo: Kituba (also known as Munukutuba there) is promoted in schools and government, with some support from the government's education and culture departments. In Angola: Kikongo is recognized as one of the national languages, and the Instituto Nacional das Línguas Nacionais (National Institute of National Languages) under the Ministry of Culture works on documentation and promotion of Kikongo and other local languages.
L. Ling. fragmentation:	Kiyombe, Kibeni, Kizombo, Kikongo-Kimanyanga
M. Mutual intelligibility:	Speakers of neighboring dialects (like Kiyombe and Kizombo) can usually understand each other to some extent, especially with context. With more distant varieties the mutual intelligibility drops considerably.
N. Language situation:	Kituba has become a lingua franca in many urban centres. The language is strongly under resourced and lacks research
O. Awareness / Function for identity:	Unclear
P. Corpora:	Very few: Bible translations (e.g., <i>Biblia ya Kikongo</i> or <i>Biblia ya Kituba</i>) are often used as de facto corpora. Available from Bible.com, Bible Gateway, or through SIL International. SIL International has conducted extensive fieldwork on Kikongo dialects and Kituba. There are some publicly available glossed texts, dictionaries, and grammars: Look at SIL's Language & Culture Archives for resources tagged with "kon" (the ISO 639-3 code for Kikongo).
Q. Dictionaries and grammars:	Kongo language course : a course in the dialect of Zoombo, northern Angola = <i>Maloongi makikoongo</i> (1987) by Carter, Hazel and João Makoondekwa. Madison, WI : African Studies Program, University of Wisconsin—Madison. <i>KIKONGO, Notions grammaticales, Vocabulaire Français – Kikongo – Néerlandais - Latin</i> [usurped] (1960) par A. Coene, Imprimerie Mission Catholique Tumba. 102 pages.
R. Language Policy:	unclear

S. Note on the pluricentricity:	The pluricentricity of Kongo/Kikongo fulfils 2 out of 7 criteria as there is little reliable information about this language. It is named to be a "national language" in three countries while French and Portuguese (Angola) are the official languages.
T. Literature (selection):	Nkaya, Kimbouala, Didace Mouhouelo, and Merveille Ngoulou (2022): Congolese Languages Vis-à-vis foreign Languages: Status, Forms and Functions. In: International Journal of Language, Linguistics, Literature and Culture 01, No. 01 (2022): 28–39. Müri, Sabine. Guide orthographique de la langue Kituba. Brazzaville, République du Congo: SIL-Congo, 2012. Tsoumou, Jean Mathieu. "Language Policy in Congo-Brazzaville." In The Palgrave Handbook of Language Policies in Africa, 179–94. Cham: Springer International Publishing, 2024.

26. Korean

A. Name of the PCL:	Korean
B. Endonyms - Other name(s) of the PCL:	South Korea: 한국어 (Hanguk-eo) (South Korea language); North Korea: 조선 (Chosŏn), (Historical Korean kingdom + language)
C. Language family and region:	Koreanic, located at the Korean peninsula, in southern China (Yanbian Korean Autonomous Prefecture), and eastern Russia
D. Demography and language spread:	Korean is spoken natively by approximately 80–81 million people. L1 + L2 ca. 80 million, China, recognised mority language, ca. 2.3 million; Diaspora: Ca. 7.184.872 speakers were living outside Korea in some 180 countries
E. Status-Status-OFL in:	South Korea (ca. 51 million), North Korea (ca. 26 million), China (Official at a regional level (Yanbian Korean Autonomous Prefecture), ca. 2 million;
F. Status-RML in:	China, Yanbian Korean Autonomous Prefecture
G. DV:	South Korea Korean
H. NDVs of the PCL:	North Korea Korean
I. Standard form:	South Korea: Pyojun-eo (표준어), meaning "Standard Language". North Korea: Munhwaŏ (문화어) meaning "cultured language".
J. Alphabet / WS:	Hangul (한글) with 4 basic consonants and 10 basic vowels, with additional combined forms (like double consonants and complex vowels)
K. Standardisation / regulated by:	The National Institute of the Korean Language (국립국어원) Creating and updating orthography rules, spelling reforms, standard dictionaries, etc. Managing language planning (for example, guidelines for loanwords, technical terms, etc.). North Korea (Democratic People's Republic of Korea) Developed their own form called "Munhwaŏ" (문화어; "Cultured Language"). North Korea removed many loanwords, especially English and Japanese ones, and made their own spelling reforms, creating a different standard from South Korea's.

	<p>Joseon Language Society (조선어학회): Founded in 1921. Goal: To preserve Korean language and standardise it. South Korea (Republic of Korea)</p>
L. Ling. fragmentation:	<p>Seoul (Gyeonggi) variety: Basis of standard South Korean. Pyongan variety: Basis of standard North Korean. Jeolla and Gyeongsang varieties: Southern dialects with strong phonological and lexical differences. Jeju variety (Jeju-eo): Sometimes considered a separate language because of low mutual intelligibility with standard Korean. Hamgyong and Yukjin varieties: Northern varieties, some preserved by diaspora in China and Russia.</p>
M. Mutual intelligibility:	<p>Basic mutual intelligibility persists between most varieties. Political division and cultural isolation have been the biggest accelerators of divergence.</p>
N. Language situation:	<p>South Korea promotes Pyojun-eo (표준어), "standard language", based on the Seoul dialect; North Korea promotes Munhwaŏ (문화어, "cultured language") based on the Pyongyang dialect.</p>
O. Awareness / Function for identity:	<p>The Korean language is deeply tied to Korean national identity — in both South and North Korea. Language symbolizes independence, culture, and unity, especially after periods of colonization (like during Japanese rule) and division. Hangeul (한글, the Korean alphabet) is seen as a uniquely scientific, accessible writing system, invented by King Sejong in the 15th century. Many Koreans feel proud that their language was created with the goal of widespread literacy. Unification Tool: Despite political division between North and South Korea, the Korean language remains a common cultural bond. Marker of Authenticity: Fluency in Korean (especially traditional expressions, proverbs, honorifics) is seen as a marker of being authentically "Korean." Cultural Transmission: Through Korean, traditional values (like Confucian hierarchy, respect for elders) are taught and reinforced.</p>
P. Corpora:	<p>Sejong Corpus (세종말뭉치) Managed by: National Institute of Korean Language (국립국어원) Size: Over 200 million words Korean National Corpus (한국어기초사전말뭉치) Managed by: National Institute of Korean Language Size: About 12 million words (more specialized) AI Hub Korean Corpora Managed by: NIA (National Information Society Agency) Size: Billions of words across multiple datasets. Korean Web Corpus (한국어웹말뭉치) Managed by: Open research projects or collected individually.</p>
Q. Dictionaries and	<p>Standard Korean Language Dictionary"</p>

grammars:	<p>(표준국어대사전) (1999ff): Author(s): National Institute of the Korean Language (한국어학회) Publisher: The National Institute of the Korean Language. The Korean-English Dictionary" by John J. Schenck and In K. Oh Author(s): John J. Schenck, In K. Oh. Publisher: Hippocrene Books Publication Year: 1997. ISBN: 978-0781801676 Naver Korean Dictionary" (네이버한국어사전) Author(s): Naver (Online platform) Publisher: Naver Corporation Publication Year: Ongoing updates (online resource) *** Samuel E. Martin (1992): A Reference Grammar of Korean". Publisher: Routledge. ISBN: 978-0415004861 Sohn, Ho-min (1999): Korean Grammar for International Learners" (한국어문법). University of Hawaii Press. ISBN: 978-0824818705 Amen IV, Henry J. and Park, Kyubyong (2014): Kyubyong Korean for Beginners: Mastering Conversational Korean". Publisher: Tuttle Publishing. ISBN: 978-0804845172</p>
R. Language Policy:	<p>North Korea: Policy Goals: Purify Korean by eliminating foreign (especially English and Chinese) loanwords. Promote a "socialist" and "nationalistic" Korean identity. Language used as a tool for political ideology (e.g., glorifying the Kim family, nationalism). Standardization of grammar, spelling, and vocabulary along "revolutionary" lines. Heavy emphasis on the idea that the Korean language should reflect the Juche (self-reliance) philosophy. South Korea : Promote Korean nationally and internationally (including through Hallyu or the Korean Wave). Modernize Korean while balancing foreign influences. English loanwords and foreign terms are widely accepted, though efforts exist to replace excessive foreign usage with native Korean terms. Policies exist to preserve regional dialects and promote the correct use of Korean (e.g., language purity campaigns, standard speech contests). Active promotion of Korean language education both domestically and abroad (e.g., King Sejong Institutes worldwide). Hanja (Chinese characters) were historically used but are now mostly phased out in daily life; Hangeul is dominant.</p>
S. Note on the pluricentricity:	<p>The pluricentricity of Korean is a special case as the two countries are still at war and do not co-operate on their common language, which is drifting apart due to widely differing language policies. There is still mutual intelligibility, but it is getting weaker.</p>
T. Literature (selection):	<p>Sohn, Ho-min (1999). The Korean Language. Publisher: Cambridge, CCUP. ISBN: 9780521661898</p>

Janet E. Kim (2020). *Language Contact and Change in Korean*. Publisher: Cambridge University Press. ISBN: 9781108498345
 Lee, Sang-Oak (1997). *Language, Society, and Culture in Korea*.

27. Lingala

A. Name of the PCL:	Lingala
B. Endonyms - Other name(s) of the PCL:	Ngala, liNgála, Bangala
C. Language family and region:	Niger–Congo, Bantu language; located at the West coast of Central Africa; native to DR Congo, Republic of the Congo, Angola
D. Demography and language spread:	ca. 70 million L1 and 20 million L2 speakers;
E. Status-Status-OFL in:	Democratic Republic of the Congo, Republic of the Congo
F. Status-RML in:	none
G. DV:	Literary Lingala
H. NDVs of the PCL:	Angolan Kongo
I. Standard form:	Lingala littéraire or Lingala classique (primarily used in educational settings, media broadcasts, and religious services, particularly within the Catholic Church.) <i>Lingala parlé</i> or <i>Lingala populaire</i> is the variant most commonly used in daily conversations across Lingala-speaking region
J. Alphabet / WS:	Latin
K. Standardisation / Codification, regulated by:	unclear: Some activities: Jomo Kenyatta University of Agriculture and Technology: https://github.com/neulab/AfricanVoices The language is insufficiently researched and is under-resourced.
L. Mutual intelligibility:	yes
M. Ling. fragmentation:	Northwestern Lingala, Kinshasa Lingala and Brazzaville Lingala.
N. Language situation:	It is widely used in the media, the army, the police, official speeches and music, which has also contributed to its widespread use since decolonisation. It also enjoys a great deal of local prestige among the general public. It functions as the primary vernacular of both Kinshasa and Brazzaville and it is associated with modernity and urban culture.
O. Awareness / Function for identity:	Has a strong identity function.
P. Corpora:	Lingala Read Speech Corpus LRSC Congoese Speech Radio Corpus CSRC
Q. Dictionaries and grammars:	Malcolm Guthrie: <i>Lingala Grammar and Dictionary: English-Lingala, Lingala-English</i> . Baptist Missionary Society 1988. Budibunene Ngandu: <i>Glossary Of Business Terms: English-lingala, Lingala-English</i> . Aglob Pub 2004, ISBN 1-59427-034-1. Ashem Tem Kawata: <i>Dictionnaire Français-Lingala et Lingala-Français</i> . Dictionnaire et encyclopédie. Karthala René Van Everbroeck: <i>Maloba Ma Lokota, Lingala-français / français-lingala</i> . Éditions L'Épiphanie, Limete (Kinshasa) 1985.
R. Language Policy:	unclear

S. Note on the pluricentricity:	The pluricentricity of Lingala fulfils 3 out of 7 criteria. It is named to be a "national language" in three countries with French and Portuguese (Angola) the official languages .
T. Literature (selection):	Atibakwa Baboya Edema: L'orthographe des langues de la République démocratique du Congo: entre usages et norme. In: Cahiers du Rifal: Le traitement informatique des langues africaines. Band 23, 2003, ISSN 1726-5363, S. 76–83 (französisch, ulb.ac.be)

28. Malay

A. Name of the PCL:	Malay
B. Endonyms - Other name(s) of the PCL:	Bahasa Melayu
C. Language family and region:	Austronesian language; native to Brunei, Christmas Island, Cocos (Keeling) Islands, East Timor, Indonesia, Malaysia, Singapore, South Thailand; Located on islands in maritime Southeast Asia and on the Malay Peninsula on the Asian mainland
D. Demography and language spread:	L1: ca. 77-85 million, L2: 200 million;
E. Status-Status-OFL in:	Indonesia (Bahasa Indonesia), Malaysia (Bahasa Malaysia), Brunei (Bahasa Melayu Brunei), Singapore (Bahasa Melayu)
F. Status-RML in:	none
G. DV:	Bahasa Indonesia
H. NDVs of the PCL:	Bahasa Malaysia, Brunei Malay, Singapore Malay;
I. Standard form:	Indonesia (Bahasa Indonesia), Malaysia (Bahasa Malaysia), Brunei, Singapore
J. Alphabet / WS:	Two forms of the Latin alphabet, one used in Indonesia and one in Malaysia; and also a form of the Arabic alphabet called Jawi;
K. Standardisation / Codification, regulated by:	Language Council of Malaysia-Indonesia-Brunei (MABBIM) Codification centres: Pusat Bahasa (Language Centre) in Indonesia; Dewan Bahasa dan Pustaka (Institute of Language and Literature) in Malaysia and Brunei
L. Mutual intelligibility:	yes
M. Ling. fragmentation:	Iban (Sea Dayak), Brunei Malay, Sambas Malay, Kutai Malay, and Banjarese.
N. Language situation:	Malay exists in several regional forms or dialects, with notable variations between the standard language and the colloquial forms spoken in different countries. It is an official in 4 countries: Indonesia (Bahasa Indonesia), Malaysia (Bahasa Malaysia), Brunei (Bahasa Melayu Brunei), Singapore (Bahasa Melayu) and spoken by a significant number of people across the borders of these nations.
O. Awareness / Function for identity:	Has a strong identity function.
P. Corpora:	Malay Corpus (MalayWeb); MALINDO (Malay-Indonesian) Corpus) ASEAN Multilingual Corpus; BNC (British National Corpus) Malay Subset)

Q. Dictionaries and grammars:	Kamus Dewan (The Dewan Dictionary) for Bahasa Malaysia Kamuss.org; Online Malay Dictionary for both Bahasa Malaysia and Bahasa Indonesia; Pusat Rujukan Persuratan Melayu (PRPM) (Malay Language Reference Center Dictionary): official online dictionary for the Malay language published by the Dewan Bahasa dan Pustaka (DBP), Malaysia's national language agency.
R. Language Policy:	There is an explicit language policy to facilitate mutual intelligibility and to let the varieties of Indonesia and Malaysia not to become too different
S. Note on the pluricentricity:	The pluricentricity of Lingala fulfils 3 out of 7 criteria. It is named to be a "national language" in three countries with French and Portuguese (Angola) the official languages .
T. Literature (selection):	Omar, Asmah Haji (2018) Pluricentricity & Identity in the Malay World. In: Rudolf Muhr / Benjamin Meisnitzer (eds.) (2018): Pluricentric Languages and non-dominant Varieties worldwide: New pluricentric languages-old problems. Wien et. al., Peter Lang Verlag. p. 45-60. Omar, Asmah Haji (ed.) (2016): Languages in the Malaysian Education System: Monolingual Strands in Multilingual Settings. Kuala Lumpur: Dewan Bahasa dan Pustaka. Simpson, Andrew (2007): Language & National Identity: Indonesia. In: Andrew Simpson (ed.) (2007). pp. 312-336. Simpson, Andrew (ed.): 2007. Language & National Identity in Asia. Oxford: Oxford University Press.

29. Maninka

A. Name of the PCL:	Maninka
B. Endonyms - Other name(s) of the PCL:	Maninkakan, Malinke, Maninka
C. Language family and region:	Niger-Congo, Western Mande language; native in West-Africa to Guinea, Mali, Senegal, Côte d'Ivoire, The Gambia, and Guinea-Bissau.
D. Demography and language spread:	L1: ca. 15 million;
E. Status-Status-OFL in:	Guinea: Recognised national language (used in media, adult literacy, and education alongside French); Mali: also recognized alongside another national languages; Senegal: recognised national language
F. Status-RML in:	none
G. DV:	Guinea Maninka
H. NDVs of the PCL:	unclear
I. Standard form:	Guinea Maninka (unclear)
J. Alphabet / WS:	Latin (Guinea and Mali); N'Ko script, Grassroots standard, used across borders by educated Manding speakers. It has a growing base of literacy materials and digital usage.
K. Standardisation / Codification, regu-	There is no centralized, international regulatory body for Maninkan but several national language commissions:

lated by:	Guinea: Direction Nationale de l'Alphabétisation et des Langues Nationales Mali: Institut des Langues et des Cultures (ILC) N'Ko Institute of Research and Literacy: A transnational body (informal) promoting the N'Ko script and literacy National orthographic norms in Guinea and Mali exist
L. Mutual intelligibility:	yes
M. Ling. fragmentation:	Kankan Maninka (Guinea), Bamako Maninka (Mali), Eastern Maninka (Ivory Coast and Guinea), Mandinka (Senegal, The Gambia)
N. Language situation:	Mutually intelligible with Bambara, Dyula, and Mandinka. In Guinea (actively used in education and media). N'Ko has created a grassroots movement for literacy, literature, and education, especially in Mali and Guinea
O. Awareness / Function for identity:	Deeply tied to the history of the Mali Empire, and it carries strong cultural and musical traditions, especially through griots (oral historians and musicians)..
P. Corpora:	Kirill Maslinsky and Andrij Rovenchak: Maninka Reference Corpus Maninkakan Dictionary Project : aims to build electronic dictionaries for Mandekan languages, including Maninkakan.
Q. Dictionaries and grammars:	Valentin Vydrin (1999): <i>Manding-English Dictionary (Maninka, Bamana)</i> ; Bamba, Moussa: Maninkakan Lexicon (LDC2013L01): Provides Maninkakan–English and Maninkakan–French translations. Malidaba Multilingual Dictionary (online resource): Provides Maninka translations in English, French, Russian, and N'Ko script Creissels, Denis (2009): <i>Malinké de Kita</i> by (2009). Köln, Rüdiger Köppe Verlag.
R. Language Policy:	Guinea promotes mother-tongue education in Mandinka In Mali government support of Maninka in adult literacy programs and community education In The Gambia Mandinka is one of the most widely spoken languages used in local radio, public information campaigns, and informal education. In Senegal not widely promoted by the state but may appear in community radio, cultural programs, and local literacy efforts.
S. Note on the pluricentricity:	The pluricentricity of Maninka fulfils 3 out of 7 criteria and very weakly developed. It is named to be a "national language" in two countries.
T. Literature (selection):	Vydrin, Valentin (2000): Mandé language family of West Africa: Location and genetic classification. SIL Electronic Survey Report. Dallas, SIL International. (with T.G. Bergman and Matthew Benjamin), 2000

30. Mongol/Mongolian

A. Name of the PCL:	Mongolian
B. Endonyms - Other name(s) of the PCL:	Mongolia: Монгол хэл (Mongol khel) [Mongol language]; Inner Mongolia, China: 在中国, 蒙古语的自称 (Mongyol kele); Kalmykia (Russia): хальмг келн (xalmg keln); Mongol khelen (Buryat);
C. Language family and region:	Altaic language; native to Mongolia and in the autonomous regions of Inner Mongolia and Xinjiang and the provinces of Qinghai and Gansu in China.
D. Demography and language spread:	L1: ca. 7 million;
E. Status-Status-OFL in:	Mongolia, China: Inner Mongolia (co-official with Mandarin), Xinjiang, Russia (recognised regional language): Kalmykia, as Kalmyk;
F. Status-RML in:	none
G. DV:	Khalkha Mongolian (Mongolia)
H. NDVs of the PCL:	Inner Mongolia Mongolian, Kalmyk Mongolian
I. Standard form:	Khalkha (Mongolia), Chakhar (China), Oirat
J. Note on the pluricentricity:	The pluricentricity of Mongolian fulfils 7 out of 7 criteria. It is the official language in Mongolia and co-official in Inner Mongolia (China). It is taught in schools and present in the media on all levels in Mongolia; In Inner Mongolia it is the language of instruction in "ethnic Mongolian schools", especially in rural and Mongolian-majority areas, subjects are taught in Mongolian, but since 2020 there's been a shift toward Mandarin, with key subjects (like politics, language, history) now obligatory in Mandarin.
K. Alphabet / WS:	Mongolian alphabet 26 letters, written vertically and left to right (in China and Mongolia), Mongolian Cyrillic (in Mongolia and Russia)
L. Standardisation / Codification, regulated by:	Mongolian Academy of Sciences (MAS), Institute of Language and Literature; China: Council for Language and Literature Work
M. Mutual intelligibility:	yes
N. Ling. fragmentation:	Mongolian proper, Oirat, Peripheral
O. Language situation:	Mongolia: Khalkha Mongolian is the official language (uses Cyrillic script). Widely used in government, education, media, and everyday life. Youth-driven interest in Mongolian hip-hop, poetry, and calligraphy. Inner Mongolia: Inner Mongolia (China): Mongolian (mainly the Chakhar dialect) is co-official alongside Mandarin. Uses the traditional vertical Mongolian script.
P. Awareness / Function for identity:	In Mongolia, the language is a cornerstone of post-socialist national identity. It symbolizes independence, tradition, and unity. Cyrillic is official; traditional script is being reintroduced gradually. In Inner Mongolia, the language (in its traditional script) is a key part of ethnic Mongol identity, often in tension with Sinicisation

	<p>pressures.</p> <p>There's a growing language revitalization movement in both Mongolia and China: Teaching the traditional script. Promoting Mongolian in digital media and tech.</p>
Q. Corpora:	<p>Mongolian National Corpus projects (Mongolia).</p> <p>Online dictionaries and grammar tools supported by MAS.</p> <p>Unicode and font development for traditional script use.</p>
R. Dictionaries and grammars:	<p>Mongolia: "Modern Mongolian Dictionary": Орчин цагийн монгол хэлний толь бичиг, published by the Institute of Language and Literature. Phrase dictionaries and idiom compilation: "Богино өгүүлбэрийн толь бичиг" often used in education.</p> <p>Inner Mongolia: Bilingual dictionaries (Mongolian-Chinese and vice versa); "Монголоор хятад, хятад-монгол толь бичгүүд".</p> <p>Traditional script explanatory dictionaries: "Монгол бичгийн товч тайлбар толь"</p> <p>Mongolia: Modern Mongolian Grammar: "Сүүлийн үеийн монгол хэлний зүй" from the Mongolian Academy of Sciences</p> <p>Nicholas Poppe: "The Mongolian Language": A classic grammar in English</p>
S. Language Policy:	<p>Mongolia: Constitution (1992): Declares Mongolian the official state language. Language law: Promotes proper use of Mongolian in public domains and protects it from excessive foreign-language influence.</p> <p>Inner Mongolia: Bilingual education policy: Previously allowed Mongolian-medium instruction, but since 2020, shifts toward Mandarin-only curriculum.</p>
T. Literature (selection):	<p>Nicholas Poppe – Grammar of Written Mongolian; foundational.</p> <p>Juha Janhunen – The Mongolic Languages; typological overview.</p> <p>John Street – Introduction to Classical Mongolian.</p> <p>Linda T. Waugh, Tserenpil D. – studies on Mongolian identity and language revival.</p> <p>Gerelchimeg B. – works on youth language, media, and diglossia in Mongolia.</p> <p>Zhou, Minglang; Sun, Hongkai (2006-04-11). Language Policy in the People's Republic of China: Theory and Practice Since 1949. Springer Science & Business Media. ISBN 978-1-4020-8039-5</p>

31. Norwegian

A. Name of the PCL:	Norwegian
B. Endonyms - Other name(s) of the PCL:	Norsk
C. Language family and region:	North Germanic language; native to Norway.
D. Demography and language spread:	L1: ca. 5,6 million; No language spread, Norwegian is limited to Norway
E. Status-Status-OFL in:	Norway

F. Status-RML in:	none
G. DV:	Bokmål (used by ~85–90% of the population) Nynorsk (used by ~10–15%)
H. NDVs of the PCL:	None
I. Standard form:	Bokmål and Nynorsk
J. Note on the pluricentricity:	The pluricentricity of Norwegian fulfils 7 out of 7 criteria. It is a special case of pluricentricity as there are two standard varieties within the same country. Both are taught in schools and are present in the media on all levels
K. Alphabet / WS:	Dano-Norwegian alphabet, it has 29 letters
L. Standardisation / Codification, regulated by:	Regulated by: Language Council of Norway (Språkrådet) (Bokmål and Nynorsk); Norwegian Academy (Riksmål)
M. Mutual intelligibility:	yes
N. Ling. fragmentation:	Nordnorsk: Northern Norway; Trøndersk: Trøndelag; Vestnorsk: Western Norway; Østnorsk: Eastern Norway; Sørlandske: Southern Norway
O. Language situation:	State policy attempted to merge Nynorsk and Bokmål into a single language, to be called Samnorsk. However, the attempt failed. The Samnorsk policy had little influence after 1960, and was officially abandoned in 2002. Norway has two linguistic standards that are equally by law.
P. Awareness / Function for identity:	Norwegians are generally very aware of the two official written standards. Bokmål: Often associated with urban, modern, and sometimes elite or standardized identity; Perceived as more "neutral" or even "mainstream". Strongly tied to regional and cultural identity, especially in rural and western parts of Norway. Seen as a symbol of national pride, independence from Danish influence, and cultural authenticity. Used intentionally as a marker of identity in politics, education, and literature.
Q. Corpora:	NoWaC (Norwegian Web as Corpus): ~700 million words of web text. Useful for modern vocabulary and frequency studies. Oslo Corpus / Bokmålskorpuset: Academic corpus hosted by UiO; literature, newspapers, and web. Bokmål: Norsk aviskorpus: Corpus of Norwegian newspapers; great for journalistic language.
R. Dictionaries and grammars:	Bokmålsordboka & Nynorskordboka (ordbokene.no) Official dictionaries maintained by Språkrådet and UiB NAOB – Det Norske Akademis Ordbok (naob.no) Extensive historical and contemporary Bokmål Lexin (lexin.udir.no) For learners & multilingual users Cappelen Damm Norsk–Engelsk Ordbok Collins Norwegian Dictionary Norsk referansegrammatikk Faarlund, Lie, Vannebo Most comprehensive grammar of Norwegian (both varieties) Jan Olav Fretland: Norsk grammatikk – Nynorsk og Bokmål: Parallel explanation of both standards

	Moen & Hove: Grammatikk for lærere: Education-focused, includes exercises
S. Language Policy:	Both standard varieties are legally equal under the Language Usage Act (Språklova), in force since 2022. Key principles of language policy: Linguistic Equality: Bokmål and Nynorsk are given equal legal status in public institutions. Protection of Dialects: Spoken varieties are encouraged and respected across all settings, even in media and schools. Language Diversity: Minority languages (Sami, Kven, Romani) and immigrant languages are also recognized and protected. Clear Language Emphasis on using klarspråk (clear, accessible language) in government and public communication. Technological Support for both standard varieties
T. Literature (selection):	De Smedt, Koenraad; Lyse, Gunn Inger; Gjesdal, Anje Müller; Losnegaard, Gyri S. (2012). The Norwegian Language in the Digital Age. White Paper Series. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 45. doi:10.1007/978-3-642-31389-9. ISBN 9783642313882. Jahr, Ernst Håkon (ed.) (1993): <i>Language Conflict and Language Planning in Scandinavia</i> Lane, Pia (2010): <i>Multilingualism and Language Policy in Norway</i> ; Vikør, Lars S. (1993): <i>The Nordic Languages: Their Status and Interrelations</i> ; Ryen, Else (2000): <i>Språkvalg og språkholdninger i Noreg</i> .

32. Occitan*

A. Name of the PCL:	Occitan
B. Endonyms - Other name(s) of the PCL:	Occitan, Lengua d'òc, Provençal / Provençau, langue d'oc
C. Language family and region:	Western Romance language; native to Southern France, Spain (Val d'Aran), Italy (Alpine valleys of Piedmont and Liguria), Monaco
D. Demography and language spread:	ca. 1-2 million; active fluent speakers at around 0,5 million.
E. Status-Status-OFL in:	Regional language status in Catalonia, Spain: (Val d'Aran – co-official status along with Catalan and Spanish); Unofficially recognised as regional language in France.
F. Status-RML in:	none
G. DV:	Occitan larg
H. NDVs of the PCL:	Gascon / Aranese, Provençal, Limousin; Auvergnat; Vivaro-Alpine
I. Standard form:	Occitan larg (Occitan standard), sometimes referred to as Occitan commun. Primarily based on the Languedocien varieties, with elements from other major varieties (e.g., Provençal, Gascon, Limousin) to ensure broader acceptance.
J. Note on the pluricentricity:	The pluricentricity of Occitan fulfils 3 out of 7 criteria. It is a borderline case of pluricentricity as there is high fragmentation, only official status as regional language in Spain and unofficially in France, taught in Spain, only taught in some French and Italian schools, and it counts as endangered language in France

K. Alphabet / WS:	Latin, Occitan alphabet, 23 letters
L. Standardisation / Codification, regulated by:	Regulated by: Institut d'Estudis Occitans (IEO), (Institute of Occitan Studies); Congrès Permanent de la Lengua Occitana (CPLO); Academia Occitana
M. Mutual intelligibility:	High Mutual Intelligibility between Languedocien, Provençal, Vivaro-Alpine. Moderate mutual intelligibility between Limousin, Auvergnat, and southern varieties Languedocien, and Provençal
N. Ling. fragmentation:	Gascon – includes Aranese (spoken in Spain). Languedocien – often used as a basis for standard Occitan. Provençal – includes the maritime variety and Niçard (Nice area). Limousin; Auvergnat; Vivaro-Alpine (sometimes called Alpine Provençal).
O. Language situation:	Occitan is considered a seriously endangered language by UNESCO. Diglossia is common. Most Occitan speakers are bilingual, often shifting entirely to the dominant national language, especially among younger generations.
P. Awareness / Function for identity:	Low Awareness in France, people living in the Occitan speaking area are not aware that their local language is part of a broader Occitan language; In France, Occitan has been adopted as a symbol of resistance against centralisation and linguistic homogenization. Aranese Occitan is a marker of local identity and pride in Val d'Ara: Children grow up learning it in school; it's used in media and official documents; Seen as part of a trilingual identity: Aranese, Catalan, and Spanish. Occitan represents a link to regional roots, rural life, oral tradition, and historical continuity; Occitan Valleys (Italy): Medium awareness, Cultural revival, local pride, linked to minority laws and activism
Q. Corpora:	BaTelOc (Base de Textes Occitans Languedociens) A linguistic corpus developed by Université Paul-Valéry Montpellier 3. Focused on Languedocien variety. Ortòc / Corpus Ortografic Occitan: A corpus for the study and normalization of Occitan orthography. Managed by Lo Congrès permanent de la lenga occitana. Aranese Corpus: A smaller corpus specifically for Aranese, the Gascon dialect in Spain. Often used in NLP projects in Catalonia and Spain
R. Dictionaries and grammars:	Carrera, Aitor (2007a): Gramatica aranesa. Lleida. Pagès Editors. Mistral Frédéric: Tresor dóu Felibrige Diccionari General Occitan – Lo Congrès Permanent de la Lengua Occitana *** Alibert, Loui: Grammaire Occitane Sauzet, Patric: Nouvelle Grammaire Occitane Bec, Pierre: Initiation à l'occitan moderne Mistral, Frédéric: Grammaire Provençale
S. Language Policy:	France: weak and symbolic support, Occitan is taught in some schools (Calandretas); Protected under Spanish law and Catalan language policies; Taught in schools as a main language. Used in public administration, media, and signage;

	Italy: Can be taught in schools and used in cultural events, Local municipalities may use bilingual signage and fund Occitan initiatives; Lack of consistent policy implementation.
T. Literature (selection):	<p>Carrera, Aitor (2019): Occitan as a pluricentric language – Problems standardizing a fragmented language. In: Rudolf Muhr / Josep Àngel Mas Castells / Jack Rueter (eds.): European Pluricentric Languages in Contact and Conflict. Berlin / Wien u.a., Peter Lang Verlag. p. 121-140.</p> <p>Taupiac, Jacme (2001): L'Occitan modèrne. Montalban. Institut d'Estudis Occitans.</p> <p>Sumien, Domergue (2017): La recèrca fàcia a un novèl frenèsi occitan de planificacion lingüística (2004-2014). In: Occitània en Catalonha: de tempses novèls, de novèlas perspectives. Actes de l'XIè Congrès de l'Associacion Internacionala d'Estudis Occitans. Ed. of Aitor Carrera and Isabel Grifoll, Lleida/Barcelona, 365-376.</p> <p>Taupiac, Jacme (2001): L'Occitan modèrne. Montalban. Institut d'Estudis Occitans.</p>

33. Pashto*

A. Name of the PCL:	Pashto
B. Endonyms - Other name(s) of the PCL:	Afghanistan: Pəxtó / Pašto, پښتو Pakistan: Pakhto
C. Language family and region:	Eastern Iranian language; native to Afghanistan, and Pakistan
D. Demography and language spread:	L1 ca. 51 million; L2 ca. 4,9 million
E. Status-Status-OFL in:	Afghanistan (alongside Dari); Pakistan: Province Khyber Pakhtunkhwa: official provincial status; Balochistan: status as minority language;
F. Status-RML in:	Balochistan (Pakistan)
G. DV:	Afghanistan: Central Pashto
H. NDVs of the PCL:	Pakistan: North-Western Pashto
I. Standard form:	Afghanistan: Central Pashto Pakistan: North-Western Pashto
J. Note on the pluricentricity:	The pluricentricity of Pashto fulfils 4 out of 7 criteria. It is a borderline case of pluricentricity as there is high fragmentation, with reduced official status outside Afghanistan, however with a high value socially and for identity. There is strong ethno-linguistic awareness, which compensates low support in Pakistan.
K. Alphabet / WS:	Modified Arabic alphabet.
L. Standardisation / Codification, regulated by:	Academy of Sciences of Afghanistan Pashto Academy of Pakistan Pashto Academy Quetta
M. Ling. fragmentation:	Northern Pashto (spoken in parts of Afghanistan) Southern Pashto (spoken in southern Afghanistan and southwestern Pakistan)
N. Mutual intelligibility:	Mutually intelligible, especially with effort, but differences in

	accent, vocabulary, and usage
O. Language situation:	<p>The Pashto language is culturally and regionally strong, however, it faces challenges in formalization, education, and technology.</p> <p>Pakistan: Pashto is taught in schools in Pashtun-majority areas, but implementation is uneven. Only regional language status; Lack of standardisation;</p> <p>In both countries, Pashto-medium education often suffers from underfunding and low-quality materials.</p>
P. Awareness / Function for identity:	<p>Pashto is a core marker of Pashtun identity, social cohesion, and cultural continuity. Pashto speakers typically possess a strong, conscious awareness of their native language as a distinct code, fuels pride and motivates efforts to preserve linguistic purity and local varieties.</p> <p>Mastery and use of Pashto contribute to individual self-esteem, reinforcing the notion "I am Pashtun," particularly in multilingual setting. Shared language fosters trust and mutual support among Pashtuns—both within Afghanistan and Pakistan.</p>
Q. Corpora:	<p>Pashto Language Corpus by Uppsala University (Sweden) Accessible through Uppsala's Corpus Portal.</p> <p>Leipzig Corpora Collection (Pashto): A publicly available Pashto web corpus</p> <p>BBC Pashto Corpus: Archive of news stories—useful for learners and media researchers.</p>
R. Dictionaries and grammars:	<p>Pashto–English Dictionary by D. M. Collins</p> <p>Online Pashto Dictionary by Academy of Sciences of Afghanistan: Contains standardized vocabulary, mostly in the Kandahari variety.</p> <p>Online Pashto Dictionary by Academy of Sciences of Afghanistan</p> <p>University of Arizona's Pashto–English Dictionary: A more modern, searchable online database</p> <p>***</p> <p>Boyle David, Anne; Brugman, Claudia, eds. (2014). <i>Descriptive Grammar of Pashto and its Dialects</i>. Berlin, Boston: De Gruyter Mouton. doi:10.1515/9781614512318. ISBN 978-1-61451-303-2.</p> <p>Tegey & Robson (2002): <i>A Reference Grammar of Pashto</i>" by (1996, updated 2002)</p> <p>David, John: <i>Pashto: A Descriptive Grammar</i>", newer academic grammar.</p> <p>Penzl, Herbert (1955): <i>A Grammar of Pashto A Descriptive Study of the Dialect of Kandahar, Afghanistan</i>. Washington, DC: American Council of Learned Societies, ISBN 0-923891-72-2</p>
S. Language Policy:	<p>Afghanistan: Pashto is used in government, media, and education, especially in Pashtun-majority areas.</p> <p>Both Pashto and Dari are taught in schools, and educational materials are developed in both languages;</p> <p>Significant presence of Pashto in radio, television, and print media.</p> <p>Cultural preservation and promotion of Pashto literature, poetry, and folklore are encouraged by the state.</p>

	Pakistan: Pashto-language newspapers, TV channels (like AVT Khyber), and radio stations are active. Universities offer Pashto literature and language degrees, Lack of consistent policy enforcement. Urdu and English dominate in elite education.
T. Literature (selection):	Rahman, Tariq (2022): Language and Politics in Pakistan. Graz, PCL-Press. Government of Pakistan: Population by Mother Tongue" (PDF). statpak.gov.pk. Pakistan Bureau of Statistics.

34. Punjabi*

A. Name of the PCL:	Punjabi
B. Endonyms - Other name(s) of the PCL:	Pañjābī Pakistan: پنجابی India: Pañjābī; ਪੰਜਾਬੀ, Lahnda
C. Language family and region:	Eastern Iranian language; native to Afghanistan, and Pakistan
D. Demography and language spread:	Pakistan: L1 88,915,544 (2023) India: 33,124,726 (2021)
E. Status-Status-OFL in:	Afghanistan (alongside Dari); Pakistan: Province Khyber Pakhtunkhwa: official provincial status; Balochistan: status as minority language;
F. Status-RML in:	Balochistan (Pakistan)
G. DV:	Afghanistan: Central Pashto
H. NDVs of the PCL:	Pakistan: North-Western Pashto
I. Standard form:	Standard Punjabi (sometimes referred to as Majhi)
J. Note on the pluricentricity:	The pluricentricity of Punjab fulfils most criteria. It is clearly pluricentric even though it is only recognised as a regional language in some areas. It has a high value socially and for identity. There is strong ethno-linguistic awareness.
K. Alphabet / WS:	Pakistan: Shahmukhi alphabet, based on the Perso-Arabic script India: Gurmukhi alphabet, based on Indic scripts Romanized Punjabi alphabet
L. Standardisation, regulated by:	Pakistan: Punjab Institute of Language, Art and Culture India: Department of Languages
M. Ling. fragmentation:	Eastern Punjabi (refers to the varieties of Punjabi spoken in Pakistani Punjab (specifically Northern Punjabi), and most of Indian Punjab) Western Punjabi (Lahnda) (spoken in the majority of Pakistani Punjab, the Hazara region, most of Azad Kashmir and small parts of Indian Punjab)
N. Mutual intelligibility:	Mutual intelligibility is partial, esp. southern Pashto and Central/Ghilzai Pashto (spoken in central Afghanistan) can be quite distinct phonetically and lexically. Educated speakers, especially those exposed to media in other dialects, tend to understand others better.
O. Language situation:	In Pakistan, despite widespread home use, Punjabi is often stigmatized as a "rustic" tongue and omitted from higher-level instruction; activists press for its elevation. Even in Indian Punjab,

	<p>pressures to prioritize English–medium education can erode Punjabi-medium schools; rural areas face teacher shortages. Grassroots movements in Pakistan’s Punjab province call for a formal “Punjabi Language Act” to guarantee its use in courts and higher education.</p>
P. Awareness / Function for identity:	<p>There is community mobilization: Grassroots organizations, literary societies (e.g. Punjabi Sahit Akademi branches), and social-media movements (hashtags like #LearnPunjabi or #PunjabiPride) actively promote visibility of the language. Punjabi-language television channels (PTC Punjabi, Doordarshan Punjabi), radio stations, YouTube channels, and podcasts increase both passive and active engagement. Punjabi remains the strongest boundary-marker distinguishing “Punjabi” from adjacent groups.</p> <p>In Indian Punjab, Punjabi is central to state-level politics: parties’ campaign on “Punjabi suba” (Punjabi Province) slogans, legislative debates occur in Punjabi, and language policy is a vote-winning asset.</p> <p>For Punjabis abroad, speaking or learning the language becomes both a bridge to “home” and a signal to host-society multiculturalism policies. Punjabi weekend schools, cultural festivals (Vaisakhi parades in Canada), and inter-generational storytelling embed language in identity formation:</p>
Q. Corpora:	<p>IndicCorp: Web–crawled collection of ca. 9 billion tokens covering 12 Indian languages (including Punjabi in Gurmukhi). OSCAR (CommonCrawl–derived): The Punjabi subset (in Gurmukhi script) provides hundreds of millions of words of contemporary web text.</p> <p>Wikipedia Dumps: Full Punjabi Wikipedia XML dumps (Gurmukhi), updated regularly via Wikimedia</p> <p>Indian Parallel Corpus (Samaantar et al.): includes Punjabi–English sentence pairs drawn from government sites and literature, totalling several hundred thousand aligned segments.</p> <p>CVIT–IIITH Multilingual Parallel: Sentence–aligned Punjabi–English data.</p> <p>OPUS Collection: Open parallel corpora such as JW300 (Bible translations), GNOME, KDE4, and TED talks—all include Punjabi↔English, with sizes from tens of thousands to millions of sentences.</p>
R. Dictionaries and grammars:	<p>Singh, Prem Prakash (2005): Sanskrit–Punjabi Kosh: A 1084–page heritage dictionary linking modern Punjabi to its Sanskrit roots; invaluable for historical linguistics.</p> <p>Publication Bureau, Punjabi University, Patiala (1971): English–Punjabi Dictionary. 703 pp.;</p> <p>Punjab State Univ. Textbook Board (1982): PSUTB English–Punjabi Dictionary: 1407 pp.; completion of the earlier Teja Singh project, with bold–face Gurmukhi pronunciation.</p> <p>SriGranth (online) — aggregates Mahan Kosh and other Gurmukhi–script dictionaries in one searchable interface</p> <p>Akhar software (2016; Akhar–2021) by Punjabi University — desktop/web tool offering Punjabi–English lookup plus Gur-</p>

	<p>mukhi↔Shahmukhi transliteration. Wikipedia U-Dictionary (mobile app) — cross-platform Punjabi↔English dictionary with 40 000+ headwords; includes everyday jargon *** Bhardwaj, Mangat Rai (2006): Panjabi: A Comprehensive Grammar. Routledge. Singh Mann, Gurinder et. al. (2011): Introduction to Punjabi: Grammar, Conversation and Literature. Publication Bureau, Punjabi University. Pedagogical grammar with exercises and readings.</p>
S. Language Policy:	<p>Pakistan: Punjabi is the majority mother tongue in Punjab province but has no status at the federal level (where Urdu and English are official). Provincial government of Punjab — increasingly promoting the teaching of Punjabi in primary schools, but Urdu/English dominate secondary and tertiary education. India: Punjabi is one of India's 22 "Scheduled Languages" (Eighth Schedule, Constitution of India), entitling it to use in Parliament and courts and to have its own grammar, dictionary and literature supported by the state. Official-language status - Punjab state (and its capital Chandigarh) ; recognised as a "second official" in Delhi and Haryana; Medium-of-instruction up to secondary school in most Punjab-state government schools; "Three-language formula" – Punjabi + Hindi + English across many districts. General: Digraphia & mutual intelligibility: Lack of a unified script policy; materials in Gurmukhi and Shahmukhi are seldom interoperable, which hampers cross-border literary exchange.</p>
T. Literature (selection):	<p>Rahman, Tariq (2022): Language and Politics in Pakistan. Graz, PCL-Press. Singh, Joga. "Linguistic Publications in Punjabi (2000–2008): A Brief Survey." In: Annual Review of South Asian Languages and Linguistics: 2008, pp. 243–264. "Phonemic Inventory of Punjabi." International Journal of English Linguistics, Vol. 6 No. X (2018). Sani, Bushra (2022): Explicit and Implicit Attitudes of Pakistani Rural Punjabis towards Punjabi.... PhD thesis, University of East Anglia. Language Shift and Maintenance: A Sociolinguistic Study of the Punjabi Language. In: Journal of Education and Social Sciences Research (2022). Gill, Harjeet S. (1973): Linguistic Atlas of the Punjab. Punjabi University Publication Bureau,</p>

35. Romanian

A. Name of the PCL:	Romanian
B. Endonyms - Other name(s) of the PCL:	limba română, românește, ромынэ (Transistrien Cyrillic)
C. Language family and region:	Eastern Romance; native to Romania, Moldova, Ukraine (Districts in Chernivtsi, Odesa and Zakarpattia oblasts), Serbia (20 municipalities) and Hungary; located in south-eastern Europe
D. Demography and language spread:	Romania: L1 17.263.561 (2011); Moldova 2.184.000, Transnistria 156,600, Vojvodina (Serbia) 18,038 (2023) plus 21,000 “Vlach” Romanians in the eastern Timok Valley, Hungary 13,886 (2011), Ukraine (150.00); large diaspora in Italy, Spain, Germany, UK, France, USA, Canada, Israel, Australia
E. Status-Status-OFL in:	Romania, Moldova, Transnistria (as “Moldovan”), EU
F. Status-RML in:	Bosnia and Herzegovina, Croatia, Hungary, Serbia, Ukraine
G. DV:	Romanian Romanian
H. NDVs of the PCL:	Moldova Romanian, Ukraine Romanian
I. Standard form:	Limba română literară or română standard—is a single, codified variety used in education, the media, administration and literature across both Romania and Moldova.
J. Note on the pluricentricity:	The pluricentricity of Punjab fulfils all criteria. Reluctance of the Romanian Romanian language community to accept its pluricentricity. The monocentric, mononormative and highly prescriptive attitude of Romanian Romanian (RR) towards its non-dominant varieties is even prescribed in a law (299/2007 law, modified by the 176/2013 law) concerning the rights of Romanians from everywhere". Official Name of the Language in Moldova: A long-standing debate was resolved in 2023 when the Moldovan authorities officially affirmed “Romanian” as the name of the state language. On 16 March 2023, when the Parliament of Moldova voted to replace “Moldovan language” with “Romanian language” in all laws and the constitution; President Maia Sandu promulgated this law shortly after.
K. Alphabet / WS:	Latin (Romanian alphabet), Moldovan Cyrillic alphabet (Transnistria only)
L. Standardisation, regulated by:	Romanian Academy (Academia Română), Institute of Linguistics “Iorgu Iordan – Al. Rosetti” Academy of Sciences of Moldova (Academia de Științe a Moldovei)
M. Ling. fragmentation:	Transylvanian, Crișana, Moldavian, Banat, Wallachian, Maramureș, Bukovinian, Oltenian
N. Mutual intelligibility:	All the main regional varieties of Romanian (the “Romanian proper” spoken in Romania and Moldova) - notably Wallachian, Moldavian, Banat and the various Transylvanian subvarieties - are fully mutually intelligible. Aromanian and Megleno-Romanian (vlăhește) considered as languages of their own and not mutually intelligible to Romanian proper.

O. Language situation:	<p>There is an ongoing effort to produce a unified Romanian language dictionary and academic publications that include contributions from both Romanian and Moldovan scholars, further integrating the linguistic research.</p> <p>There is an ongoing debate on preserving linguistic purity versus openness.</p>
P. Awareness / Function for identity:	<p>High language awareness: There is a strong movement in both Romania and Moldova to celebrate the Romanian language. Since 2011, August 31 is officially “The Day of the Romanian Language” (Ziua Limbii Române) in Romania, and it coincides with “Limba Noastră” (Our Language) Day in Moldova (a holiday that has existed since 1990).</p>
Q. Corpora:	<p>CoRoLa – Reference Corpus of Contemporary Romanian: Over 1 billion words and 300 hours of transcribed spoken Romanian;</p> <p>FuLG – 150B Token Romanian Corpus: 150 billion tokens extracted from Common Crawl, designed for pretraining Romanian language models.</p> <p>ROMBAC – Romanian Balanced Annotated Corpus: 41 million words with morphosyntactic annotations.</p> <p>Sketch Engine: Offers access to various Romanian corpora</p>
R. Dictionaries and grammars:	<p>Dicționarul explicativ al limbii române, Dicționarul ortografic, ortoepic și morfologic al limbii române) from the Romanian Academy.</p> <p>Dicționarul Limbii Române (DLR): Thesaurus Dictionary of the Romanian Language, 175.000 words.</p> <p>WordReference English–Romanian Dictionary, 59,000 terms and 165,000 translations.</p> <p>***</p> <p>Pană Dindelegan, Gabriela: The Grammar of Romanian</p> <p>Gönczöl, Ramona: Romanian: An Essential Grammar</p> <p>The Elements of Romanian: A Complete Romanian Grammar</p>
S. Language Policy:	<p>Romanian is the only official language of Romania and Moldova;</p> <p>In Romanian localities where a national minority constitutes at least 20% of the population, that minority’s language can be used in public administration, the justice system, and education</p> <p>In Moldova Russian holds the status of a language of interethnic communication, semi-official in practice. Gagauzia (autonomy), Transnistria (de facto independence).</p>
T. Literature (selection):	<p>Huțanu, Monica / Sorescu-Marinković, Annemarie (2018): Non-dominant varieties of Romanian in Serbia: between pluricentricity and division. Muhr, Rudolf / Meisnitzer, Benjamin (eds.) (2018): Pluricentric Languages and non-dominant Varieties worldwide: New pluricentric languages-old problems. Wien et. al.: Peter Lang Verlag. p. 205-218. Wilson Center: The Politics of Language in Romania and Moldova.</p> <p>Lozovanu, Dorin: Romanian-Speaking Communities Outside Romania: Linguistic Identities.</p>

36. Russian

A. Name of the PCL:	Russian
B. Endonyms - Other name(s) of the PCL:	русский язык (russskiy yazyk) русский = Russian (adjective) язык = language or tongue
C. Language family and region:	East Slavic language; native to Russia; located in eastern Europe and Asia;
D. Demography and language spread:	Russia: L1 118,6 Millionen (85,7 % der Bevölkerung); Ukraine 12,4 Millionen (32,9 %), Belarus 5 million, Kasachstan 3,9 million, Germany 3,5 million, Israel 1 million
E. Status-Status-OFL in:	Russia (state official), Belarus (co-official), Kazakhstan (co-official), Kyrgyzstan (co-official), Tajikistan (as inter-ethnic language designated by the constitution); International organisations: IAEA; ICAO; UNESCO; WHO; CIS; EAEU; CSTO; SCO; OSCE; ATS; ISO
F. Status-RML in:	Armenia, Finland, Poland, Romania, Slovakia, Ukraine
G. DV:	Russian Russian
H. NDVs of the PCL:	Belarus Russian, Kazakhstan Russian, Ukraine Russian, Kyrgyzstan Russian
I. Alphabet / WS:	Cyrillic (Russian alphabet consists of 33 letters)
J. Standardisation, regulated by:	V.V. Vinogradov Russian Language Institute of the Russian Academy of Sciences
K. Standard form:	Modern Russian Literary Language (<i>современный русский литературный язык</i>),
L. Ling. fragmentation:	Northern Russian varieties: Predominantly spoken in the northern and north-eastern parts of European Russia, from Veliky Novgorod to the Perm and northern Ural regions; Central (Middle) Russian Dialects: Found in regions such as Pskov, Tver, Moscow, Nizhny Novgorod, and the Lower Volga area; Southern Russian Dialects: Spoken in the western and southern parts of European Russia, including regions like Belgorod, Bryansk, Kaluga, Kursk, Lipetsk, Oryol, Ryazan, Smolensk, Tambov, Tula, and Voronez.
M. Mutual intelligibility:	The mutual intelligibility between the regional varieties of Russian remains high. This is largely due to the dominance of Standard Russian.
N. Dictionaries and grammars:	С. И. Ожегов (Sergey Ozhegov): Толковый словарь русского языка (Explanatory Dictionary of the Russian Language), most influential modern dictionary for native speakers. Russian Academy of Sciences: Большой академический словарь русского языка (Great Academic Dictionary of Russian) 33 volumes Lopatina, V. V.: Орфографический словарь русского языка (Orthographic Dictionary) *** Institute of Russian Language, Academy of Sciences: Русская грамматика (Russian Grammar) – 1980 edition. Gold-standard academic grammar.

	<p>Shanskii, N. M., Rozental, D. E. etc. Практическая грамматика русского языка. (Practical Grammar of the Russian Language) Rau, Gerlad B.: Russian Grammar in Use (for English speakers) Title: Russian Grammar in Use. For bilingual learners and linguists. Online access: Gramota.ru – government-supported portal with dictionaries and grammar tools. Runivers.ru – historical dictionaries like Dahl Corpora-linked tools – Dictionaries often integrated into ruscorpora.ru or sketchengine.eu.</p>
O. Corpora:	<p>Russian National Corpus (RNC) / Национальный корпус русского языка: https://ruscorpora.ru OpenCorpora https://opencorpora.org annotated Sketch Engine – Russian Web 2011 (ruTenTen11): https://www.sketchengine.eu; Corpus of Contemporary Russian Language (CoRuS): Higher School of Economics (HSE), Moscow, https://ruscorpora.ru/new/corus.html</p>
P. Language situation:	<p>Russian is an important international language with around 250 million speakers. It is an official language in several post-Soviet countries (Belarus, Kazakhstan, Kyrgyzstan and some de facto states). In countries such as Ukraine, Latvia, Estonia and Georgia, the use and reputation of the language is declining, while in Kazakhstan and Kyrgyzstan it is making a comeback, especially in urban areas and in government and business circles. Has a strong online presence (top 10 most used languages on the internet) with major platforms such as VK, Telegram, Yandex and Runet.</p>
Q. Awareness / Function for identity:	<p>Russian language is not just a communication tool – it is a symbol of Russianness. Speaking Russian = belonging to the Russian cultural sphere. Language policy reinforces this (e.g., education in Russian, state media). Preserving “correct” Russian is tied to preserving national values.</p>
R. Language Policy:	<p>Constitution of the Russian Federation (1993) Article 68: Declares Russian as the state language throughout the Russian Federation. Federal Law “On the State Language of the Russian Federation” (2005). Specifies the official status, standard usage, and regulations of Russian. Goals: Preserve and promote standard Russian, Guard against slang, borrowings, and regional “errors”. Support Russian as a unifying national identity; Maintain Russian as a global language Programs to promote Russian abroad (Russiky Mir Foundation, Pushkin Institute). Russian is mandatory in all federal institutions and most schools. Primary and secondary education: Russian is the default language of instruction, even in non-Russian regions. Further Goals: Reinforce a pan-Russian identity, assert influence in “near abroad” (former Soviet republics); Protect Russian speakers outside of Russia (claimed justification in some geopolitical actions).</p>

	<p>The Russian state heavily promotes Russian as a symbol of national identity and cultural heritage.</p> <p>There is increasing pressure within Russia to prioritize Russian over minority languages (e.g., Tatar, Bashkir), sometimes sparking controversy.</p>
S. Note on the pluricentricity:	<p>The pluricentricity of Russian only fulfils the first three criteria. There is strong reluctance on the side of the Russian elites to accept the pluricentricity of Russian. The relationship between Russian and the other national varieties is marked by a strong asymmetry. Using and preserving “correct” Russian is tied to preserving national values. Using a different national variety is therefore seen like a betrayal of the language and culture and is therefore rejected.</p>
T. Literature (selection):	<p>Del Gaudio, Salvatore (2013): Russian as a non-dominant variety in post-Soviet states: a comparison. In: C. Amoros Negra, R. Muhr et al. (eds.). Exploring linguistic standards in non-dominant varieties of pluricentric languages. / Explorando estándares lingüísticos en variedades no dominantes de lenguas pluricéntricas. Wien et al., Peter Lang Verlag, 343-362.</p> <p>Del Gaudio, Salvatore / Dorofeev, Jurij (2016): The Russian Language in Crimea: From Pluricentricity to Monocentricity. In: Rudolf Muhr, et. al (eds.) (2016): Pluricentric Languages and non-dominant Varieties worldwide. Volume 1: Pluricentric Languages across continents - Features and usage. Wien et. al., Peter Lang Verlag. p. 423-442.</p> <p>Del Gaudio, Salvatore & Ivanova, Olga (2015): A variety in formation? Morphosyntactic variation in Ukraine-Russian speech and press. In: Rudolf Muhr, Dawn Marley et. al. (eds.) (2015): Pluricentric Languages. New Perspectives in Theory and Description. Wien et. al., Peter Lang Verlag. p. 157-180.</p> <p>Kulyk, V. (2010): Ideologies of language use in post-Soviet Ukrainian media. International Journal of Sociology of Language. Languages and Language Ideologies in Ukraine, 201.79-104.</p> <p>Miller, Alexei (2003): The Ukrainian Question. The Russian Empire and Nationalism in the Nineteenth Century. Budapest-New York: Central European University Press.</p>

37. Scots

A. Name of the PCL:	Scots
B. Endonyms - Other name(s) of the PCL:	Scots Leid (where leid means "language") Braid Scots (or "broad Scots" in English), Lallans. Ulster-Scots (in Northern Ireland), Ullans", a recent neologism merging Ulster and Lallans
C. Language family and region:	West Germanic language; native to United Kingdom, and Republic of Ireland
D. Demography and language spread:	Around 1,508,540 (2022) million people (30% of the population aged 3+) reported some ability in Scots. About 1.1 million said they could speak Scots.
E. Status-Status-OFL in:	Scots is an official language of Scotland. The other official languages are English, Gaelic, and British Sign Language (2022).

F. Status-RML in:	Northern Ireland (as Ulster Scots); Republic of Ireland (County Donegal; also as Ulster Scots); Officially recognized under the European Charter for Regional or Minority Languages.
G. DV:	Scottish Scots
H. NDVs of the PCL:	Ulster Scots of Northern Ireland and the Republic of Ireland
I. Alphabet / WS:	Latin, 26 letters, which are the same as the English alphabet
J. Standardisation, regulated by:	No official language regulator in Scotland, but organisations that promote and guide Scots: Scots Language Centre (SLC): Offers resources, guides, and educational support. Promotes consistent spelling and usage. Scots Language Society: Publishes the journal <i>Lallans</i> and supports literary Scots. Ane Braw Scots Wiki: A Scots version of Wikipedia—helps establish written norms informally Regulator in Ulster/Republic of Ireland: Ulster-Scots Agency (Tha Boord o Ulstèr-Scotch)
K. Standard form:	Scottish Scots
L. Ling. fragmentation:	Mostly spoken in the Lowlands and Northern Isles of Scotland: Northeast (Doric dialect) – e.g., Aberdeenshire, Moray Central Belt – including Glasgow, Edinburgh, and surrounding areas Southern Scotland – e.g., Borders and Dumfries and Galloway Orkney and Shetland – Norn-influenced dialects
M. Mutual intelligibility:	yes
N. Dictionaries and grammars:	DSL – Dictionary of the Scots Language: https://dsl.ac.uk SND (Scottish National Dictionary) – covers Scots from 1700 onward Concise Scots Dictionary: Published by Edinburgh University Press; useful for students, teachers, and writers. Scots School Dictionary: Produced by Scots Language Centre / Scottish Language Dictionaries. Designed for use in schools. User-friendly, with example sentences and comparisons to English. *** Purves, David: <i>The Grammar of Modern Scots</i> Murison, David: <i>A Scots Grammar</i> Scots Wikipedia (https://sco.wikipedia.org) Scots Language Centre (scotslanguage.com) offers grammar guides, spelling help.
O. Corpora:	The Scottish Corpus of Texts & Speech (SCOTS) https://www.scottishcorpus.ac.uk Corpus of Modern Scots Writing (CMSW) https://www.scottishcorpus.ac.uk/cmsw/
P. Language situation:	Scots is often spoken informally at home or with friends, rather than in formal or professional settings. Younger generations may understand Scots but use it less fluently, especially in urban areas where Scottish Standard English is dominant in education and media. There's a growing push in schools and literature to preserve and promote Scots, especially in poetry, folk traditions, and community projects. In Scotland, Scots has some support from

	<p>the Scottish Government, which acknowledges it as a part of the nation's cultural heritage.</p> <p>The status of the language was raised in Scottish schools, with Scots being included in the new national school curriculum.</p> <p>Since 2016, the newspaper <i>The National</i> has regularly published articles in the language. The 2010s also saw an increasing number of English books translated in Scots and becoming widely available, particularly those in popular children's fiction series.</p>
Q. Awareness / Function for identity:	<p>Education Scotland: Body that encourages Scots language awareness in schools.</p> <p>National Library of Scotland: Archives and promotes Scots literature.</p>
R. Language Policy:	<p>In Scotland: Scottish Government Scots Language Policy (2015) – outlines intentions to promote Scots in education, the arts, and public life.</p> <p>Curriculum for Excellence – allows for Scots language learning and creative expression in schools.</p> <p>Ulster: Good Friday Agreement (1998) – commits to promoting Ulster Scots culture and linguistic heritage.</p> <p>St Andrews Agreement (2006) – expanded commitments to support Ulster Scots.</p> <p>Identity and Language Act (2022) – commits the UK Government to actions promoting Ulster Scots as a recognized part of the region's cultural identity.</p>
S. Note on the pluricentricity:	<p>The pluricentricity of Scots fulfils six out of seven criteria. However there is a lack of codification and status. Language promotion and schooling in Scots have been introduced in Scotland but not in Ulster and in the Republic of Ireland. The Ulster-Scots Agency (Tha Boord o Ulstèr-Scotch), part of the North/South Language Body set up under the Good Friday Agreement, is responsible for promoting the language and culture across the Irish island. Limited presence in schools, mostly extracurricular or cultural programs.</p> <p>BBC Radio Ulster and BBC Northern Ireland occasionally feature Ulster Scots.</p>
T. Literature (selection):	<p>McClure, J. Derrick (1995). <i>Why Scots Matters</i>.</p> <p>Aitken, A.J. (1984). <i>Scottish Speech: A Historical View with Special Reference to the Standard English of Scotland</i>.</p> <p>Murison, David (1977). <i>The Guid Scots Tongue</i>.</p> <p>Corbett, John, McClure, J. Derrick, and Stuart-Smith, Jane (2003). <i>The Edinburgh Companion to Scots</i>.</p> <p>Millar, Robert McColl (2016). <i>Modern Scots: An Analytical Survey</i>.</p> <p>Douglas, Fiona (2009). <i>Language, Identity and the Scots Language</i>.</p> <p>Nic Craith, Máiréad (2003). <i>Culture and Identity Politics in Northern Ireland</i>.</p> <p>McColl Millar, Robert (2021). <i>Scots and Scottish English</i>.</p>

38. Serbian

A. Name of the PCL:	Serbian
B. Endonyms - Other name(s) of the PCL:	српски језик (in Cyrillic) (Serbian language) srpski jezik (in Latin script) (Serbian language)
C. Language family and region:	South Slavic language; native to Serbia, Bosnia and Herzegovina, Montenegro and Kosovo.
D. Demography and language spread:	Around 1,508,540 (2022) million people (30% of the population aged 3+) reported some ability in Scots. About 1.1 million said they could speak Scots.
E. Status-Status-OFL in:	OFL in Serbia, one of the three official languages of Bosnia and Herzegovina and co-official in Montenegro and Kosovo.
F. Status-RML in:	Croatia, Hungary, Slovakia, Slovenia, Romania, Czech Republic
G. DV:	Serbian Serbian
H. NDVs of the PCL:	Bosnian Serbian, Montenegrin Serbian, Kosovo Serbian
I. Alphabet / WS:	Serbian Cyrillic, Serbian Latin
J. Standardisation, regulated by:	Serbian Academy of Sciences and Arts (SANU) Serbian: Srpska akademija nauka i umetnosti (Српска академија наука и уметности) Board for Standardization of the Serbian Language (Odbor za standardizaciju srpskog jezika)
K. Standard form:	Ekavian (Екавски изговор) (Dominant standard form in Serbia) Ijekavian (Ијекавски изговор) (Used in: Eastern Bosnia and Herzegovina, Montenegro, and parts of Western Serbia)
L. Ling. fragmentation:	Ekavian and Ijekavian, are 100% mutually intelligible. Serbian spoken in Montenegro or Bosnia may feature localisms or slightly different intonation, but it's still very much mutually intelligible.
M. Mutual intelligibility:	The two main varieties of Serbian within Serbia are 100% mutually intelligible. The varieties outside Serbia are also mutually intelligible.
N. Dictionaries and grammars:	Rečnik srpskohrvatskog književnog i narodnog jezika Published by: Serbian Academy of Sciences and Arts (SANU) Matica Srpska Dictionary (Rečnik Matice srpske) Published by: Matica srpska (Novi Sad), Completed: 2005, 6 volumes, around 85,000 words. Šipka, Milan, KlajnPravopis: Ivan srpskoga jezika (Orthography of the Serbian Language). Matica Srpska editions. *** Ivić, Pavle et. al.: Gramatika srpskog jezika – Morfologija i sintaksa Stanojević, Ljubomir: Gramatika srpskoga jezika za gimnazije i srednje škole Klajn Ivan: Savremena srpska jezička norma Klajn, Ivan: Tvorba reči u savremenom srpskom jeziku (I & II). The definitive work on word formation and morphology Online Resources: Moj jezik (https://www.mojjezik.rs/) – Online orthography and grammar rules Rečnici.com – Hosts several Serbian dictionaries for free search
O. Corpora:	Corpus of Contemporary Serbian Language (Користни корпус савременог српског језика)

	<p>Maintained by: Faculty of Philology, University of Belgrad, 100 million words; (http://www.korpus.matf.bg.ac.rs/) SrWaC – Serbian Web Corpus: Built by: Faculty of Mathematics, University of Belgrade & Sketch Engine Over 555 million words (https://www.sketchengine.eu/) PAROLE Corpus (Serbian), Part of: PAROLE/SIMPLE project by EU Manually annotated corpus of 250,000 words ReLDI Corpus (Regional Linguistic Data Initiative) Languages: Serbian, Croatian, Slovene (https://reldi.spur.uzh.ch/)</p>
P. Language situation:	<p>Serbian is the official language of Serbia, co-official in Bosnia and Kosovo, widely used in Montenegro; Widely used in Serbia in education, media, administration, and literature. In Bosnia and Herzegovina: Serbian is common among ethnic Serbs. Kosovo: Usage is geographically limited to Serb enclaves. The choice of script (Cyrillic vs Latin) can carry political connotations—Cyrillic is more associated with nationalism and tradition, while Latin is seen as more modern and Western-facing.</p>
Q. Awareness / Function for identity:	<p>Serbian speakers generally have a high degree of awareness about their language, being digraphic (Cyrillic and Latin). Speakers are often aware of the distinction between standard Serbian and the languages of Croatia, Bosnia, and Montenegro. Speaking Serbian (especially using Cyrillic script or Ekavian pronunciation) can be seen as a signifier of Serbian identity.</p>
R. Language Policy:	<p>The Constitution of Serbia (2006) recognizes Serbian as the official language of the country. It is used in all state institutions, including the judiciary, legislature, and executive branches. Serbian is standardized and it is the primary language of instruction in schools. The Serbian language is predominantly used in print, radio, and television media. ☐ Serbia also guarantees the rights of ethnic minorities to use their languages in certain regions. For example, in areas with significant Hungarian, Croatian, or Bosniak populations, minority languages can be used in local government and education.</p>
S. Note on the pluricentricity:	<p>The pluricentricity of Serbian fulfils seven out of seven criteria. There is a lack of codification in the non-dominant varieties, which is due to the split of Yugoslavia into several independent nations. The language situation is also still marked by split of former Serbo-Croatian into several languages due to the Yugoslav war. The newly founded languages are still mutually intelligible, which makes Serbian language activists refuse the split into independent languages-</p>
T. Literature (selection):	<p>Ilić Marković, Gordana (2016): Creating a name for a pluricentric language: From Serbian to Serbo-Croatian, Bosnian / Croatian / Serbian. In: Muhr, Rudolf et. al. eds. (2016): Vol. 1, p. 409-425. Ilić Marković, Gordana (2009): Bosnisch/Kroatisch/Serbisch: Die Auswirkung der Koexistenz mehrerer Standardsprachen auf den Sprachunterricht In: Hanns-Jürgen Krumm/ Paul R. Portmann-Tselikas: Theorie und Praxis. Nr.13/2009, Studien Verlag, Innsbruck/Wien/Bozen, p. 162-179. Ranko Bugarski – Language and Identity in the Balkans. English-language book that explores how language relates to national</p>

	<p>identity in Serbia and the former Yugoslavia SANU Digital Library: https://dais.sanu.ac.rs/ Greenberg, Robert D. (2004). <i>Language and Identity in the Balkans: Serbo-Croatian and its Disintegration</i>. New York: Oxford University Press. ISBN 9780191514555. Sorescu-Marinković, Annemarie (2010). "Serbian Language Acquisition in Communist Romania" (PDF). <i>Balcanica</i> (41): 7–31. doi:10.2298/BALC1041007S.</p>
--	--

39. Setswana

A. Name of the PCL:	Setswana
B. Endonyms - Other name(s) of the PCL:	Tswana
C. Language family and region:	Bantu language; native to Botswana, South Africa, Zimbabwe, Namibia
D. Demography and language spread:	4.1 million in South Africa (2011), 1.1 million in Botswana (1993), unknown numbers in Namibia and Zimbabwe, 7,7 million L2 speakers in South Africa (2002) located in the North West Province, with significant numbers in Gauteng, Limpopo, and Northern Cape.
E. Status-Status-OFL in:	One of the official languages of Botswana and South Africa.
F. Status-RML in:	Namibia (a national language)
G. DV:	Botswana Setswana
H. NDVs of the PCL:	South African Setswana
I. Alphabet / WS:	Latin with tones marked
J. Standardisation, regulated by:	<p>Botswana: Setswana Language Council (Informal Name); No formal, single-body regulator, but Setswana is heavily supported by the government. University of Botswana plays a key role through: The Department of African Languages and Literature; The Botswana Examinations Council (BEC) oversees the use of standardized Setswana in schools.</p> <p>South Africa: Pan South African Language Board (PanSALB); National Language Bodies (NLBs) under PanSALB focus on specific languages—there is a dedicated Setswana NLB.</p>
K. Standard form:	Standard Setswana is based on the Sehurutshe variety. Used in schools across Botswana and South Africa (especially in the North West Province) and in print media, radio, TV, government, and legal systems.
L. Ling. fragmentation:	There is a large number of regional varieties: Rolong, Hurutshe, Kwena, Lete, Melete, Ngwaketse, Ngwatu, Kgatla, Tawan, Tlharo, Tlhaping, Thlahaping, Thlaro.
M. Mutual intelligibility:	The regional varieties are largely mutually intelligible.
N. Dictionaries and grammars:	Molefe, Z.I. et al. (1985): <i>Dintlha tsa Setswana: A monolingual Setswana dictionary</i> widely used in Botswana and South African schools. Botswana Ministry of Education: <i>Thanodi ya Setswana</i> (1995, updated editions);

	<p>Oxford University Press South Africa: Oxford Bilingual School Dictionary: Setswana and English. ***</p> <p>T. Cole, Desmond 1955, revised 1980s): Setswana Grammar Pheko Mpho: Modern Setswana Grammar by (various editions) Essential Grammar of Setswana (various authors, educational publishers)</p>
O. Corpora:	South African Centre for Digital Language Resources (SADiLaR): Supports Setswana corpora, computational linguistics, and digital preservation
P. Language situation:	<p>Official Recognition: One of the few African languages with official language status in more than one country. Has a standardized grammar and orthography. There are dictionaries, grammars, school books, radio and TV programming, and news publications in Setswana. Setswana is a strong marker of Tswana identity. Used in traditional institutions, such as the kgotla (public court), and in proverbs, songs, and ceremonies. However, there is lack of widespread use in science, technology, law, and academia. Shortage of technical vocabulary and insufficient investment in modern language development.</p> <p>In Botswana, Setswana's dominance may undermine smaller indigenous languages, leading to language shift and loss for groups like the Kalanga, Wayeyi, or †Hoan speakers. In cities and among younger speakers, there's growing use of English and code-switching.</p>
Q. Awareness / Function for identity:	Setswana is a carrier of Tswana identity, culture, and traditions.
R. Language Policy:	<p>Botswana: National language of Botswana. Medium of instruction in early primary education (Grades 1–3), after which English takes over. Used in government, broadcasting, and community life. The 1994 Revised National Policy on Education promotes mother tongue instruction (primarily Setswana) in early schooling. English remains the official language, especially in administration, law, and higher education.</p> <p>Promotes standardization of Setswana. Encourages literature development, broadcasting in Setswana (e.g., Radio Botswana), and teacher training Lacks comprehensive support for multilingualism beyond Setswana and English.</p> <p>South Africa: One of 11 official languages of South Africa. Recognized in the Constitution (1996) as deserving equal treatment and development.</p> <p>Pan South African Language Board (PanSALB): Mandated by law to develop and protect all official languages, including Setswana.</p> <p>Language-in-Education Policy (LiEP): Supports mother-tongue instruction. Setswana is used as a language of learning and teaching (LoLT) in early grades (Grades R–3) in areas where it is dominant, especially in the North West Province.</p> <p>Setswana is used in broadcasting (e.g., SABC radio and TV), basic education, and cultural programming.</p>
S. Note on the pluri-	The pluricentricity of Setswana fulfils seven out of seven criteria.

centricity:	It is an national language in Botswana and South Africa and has strong support by the respective governments. Both varieties (Botswana and South Africa) are largely symmetrical in their power relation.
T. Literature (selection):	Cole, Desmond (1955), An Introduction to Tswana Grammar, Cape Town: Longmans, Green and Co. Janson, Tore, Tsonope, Joseph (1991), Birth of a National Language: The History of Setswana, Heinemann Botswana, ISBN 0-435-91620-3 University of Botswana (2001), The Sound System of Setswana, Lightbooks, ISBN 99912-71-21-X

40. Somali

A. Name of the PCL:	Somali
B. Endonyms - Other name(s) of the PCL:	Af Soomaali; Wadaad: <i>أف صومالي</i> ; Osmanya:
C. Language family and region:	Afro-asiatic language, Cushitic branch; native to Somalia, Somaliland, Ethiopia, Kenya, Djibuti.
D. Demography and language spread:	Somalia: 16 million, Ethiopia: ca. 5 million, Kenya: ca. 2.5 million (especially in the north-eastern part), Djibouti: ca. 500.000 speakers.
E. Status-OFL in:	Somalia (co-official with Arabic), Somaliland (OFL), Djibouti (national language), Ethiopia (working language in the Somali Region);
F. Status-RML in:	Kenya
G. DV:	Somalia Somali
H. NDVs of the PCL:	Kenyan Somali, Ethiopian Somali, Somaliland Somali
I. Alphabet / WS:	Latin alphabet
J. Standardisation, regulated by:	Regional Somali Language Academy, established in 2013 by the governments of Djibouti, Somalia and Ethiopia;
K. Standard form:	Standard Somali (Af Soomaali Heerka ah), primarily based on the variety spoken by the Isaaq, Darod, and Dir clans in northern Somalia;
L. Ling. fragmentation:	Three main varieties: Northern Somali, Benadir and Maay.
M. Mutual intelligibility:	The main varieties are fairly mutually intelligible.
N. Dictionaries and grammars:	Cabdulle (Xashi), Maxamud C.: Af Soomaali – English Dictionary. One of the most commonly used Somali-English dictionaries; widely referenced in Somali schools and academic circles. Somali Academy of Sciences and Arts (1976): Somali Dictionary (Qaamuuska Af-Soomaaliga). Awil Jama: English-Somali Dictionary. Helpful for learners of both languages. *** Saeed, John (1999): Somali Reference Grammar. Dunwoody Press; Andrzejewski, B. W., Lewis, M. S. (1965): A Grammar of Somali.

	<p>Longman; Ministry of Education, Somali (1970ff): Af Soomaali Grammar. Standard grammar texts used in Somali schools; Online Resources: Websites: Somali Language Resource Center and Qurbejoog.com offer modern resources and digital dictionaries</p>
O. Corpora:	<p>The Somali Language Corpus (SoNaR-Somali) Developed by: Tilburg University NL; Parallel Corpora (Somali-English) UN & NGO documents: Many translated documents from organizations like the UN, World Bank, and WHO have Somali-English versions. Tatoeba Project: Contains Somali-English sentence pairs. OPUS Corpus (from University of Helsinki): (https://opus.nlpl.eu/) The Somali Language and Culture Archive (Indiana University) https://archive.iu.edu/ (search for "Somali")</p>
P. Language situation:	<p>Somalia: Somali has been the primary language of administration and education since 1973. Somali serves as the primary language of instruction in schools, especially in the early grades. Arabic is taught for religious studies, while English is increasingly used in higher education and international communication. Kenya: No formal support or curriculum in Somali from the central government, though there are local efforts for mother-tongue education.</p>
Q. Awareness / Function for identity:	<p>Despite linguistic differences between the regional varieties, Somali speakers collectively view themselves as speaking a common language.</p>
R. Language Policy:	<p>Somali is the main official language in government, media, and education. The Latin script is the standard writing system (adopted in 1972). Education policy mandates Somali as the medium of instruction in primary and secondary school. Challenges: Lack of centralized enforcement due to federalism and instability.</p>
S. Note on the pluricentricity:	<p>The pluricentricity of Somali fulfils seven out of seven criteria, even though the status and recognition apart from Somalia are weak. Lack of centralized enforcement due to federalism and instability.</p>
T. Literature (selection):	<p>Nilsson, Morgan (2018): Somali as a Pluricentric Language: corpus based evidence from schoolbooks. In: Rudolf Muhr / Benjamin Meisnitzer (eds.) (2018): Pluricentric Languages and non-dominant Varieties worldwide: New pluricentric languages-old problems. Wien et. al., Peter Lang Verlag. p. 61-76. Lamberti, M. (1984): The linguistic situation in the Somali Democratic Republic. In: Labahn (ed.), Proceedings of the Second International Congress of Somali Studies, Vol. I, Hamburg, pp. 155-200. Lamberti, M. (1986): Die Somali-Dialekte. Hamburg. Appleyard, D. / Orwin, M. (2008): The Horn of Africa : Eritrea, Djibouti, and Somalia. In: Simpson, A. (ed.), Language and national identity, pp. 267-290.</p>

41. Soninke

A. Name of the PCL:	Soninke
B. Endonyms - Other name(s) of the PCL:	Sooninkanxanne, (سونكنخانس), also known as Serakhulle or Azer of Maraka
C. Language family and region:	Niger-Cong language, Mande branch; native to Mali, Senegal, Ivory Coast, Gambia, Mauritania, Guinea-Bissau, Guinea, Ghana, Burkina Faso
D. Demography and language spread:	Mali, ca. 1.8 million; Sénégal, ca. 1 million; Mauritanie, ca. 180.000; Gambie, ca. 200 000; Guinée-Conakry, ca. 10.000; Guinée-Bissau, ca. 5000), et à l'ouest du Burkina Faso
E. Status-OFL in:	Mali, Senegal, Mauritania (national language)
F. Status-RML in:	Burkina Faso, Gambia, Guinea-Bissau, Guinea, Ghana, Ivory Coast,
G. DV:	Malian Soninke
H. NDVs of the PCL:	Senegalese Soninke, Mauritan Soninke
I. Alphabet / WS:	Latin Script (officially promoted in schools and literacy programs); Ajami Script (Arabic-based), traditionally used in religious and cultural contexts
J. Standardisation, regulated by:	Mali: Institut des Langues et des Sciences Appliquées (ILSA) and the Direction Nationale de l'Alphabétisation Fonctionnelle et des Langues Nationales (DNAFLN) are involved in language planning and standardization for national languages, including Soninke. Senegal: Direction de l'Alphabétisation et des Langues Nationales (DALN) works on orthographic standardization and literacy materials. Mauritania: Government literacy programs support Soninke, with involvement from the Institut Pédagogique National (IPN).
K. Standard form:	Malian Soninke
L. Ling. fragmentation:	Kayes, Nioros, Nara, Wangara, Kaarta, Kingui, Kagorota, Bakel, Kaniaga, Garawol, Guidimakan, Guidimé, Gadiaga, Diombokoh, Diafounou, Bahkonou, Soroma, Banamba, Diangounté, Diawara, Diongaga, Koussané, Marena, Yélimané, Asowana, Barouéli, Trougoumbé, Kremis, Alahina, Banzan, and Dafort.
M. Mutual intelligibility:	The local variants are mostly mutually intelligible
N. Dictionaries and grammars:	https://www.asawan.org/fr/Soninke-accueil http://www.soninkara.org/langue-soninke https://de.glosbe.com/ Lexicon Soninke-French-English (https://www.asawan.org/fr/lexique_soninke) Moussa Diagana, Ousmane (2013). Dictionnaire soninké-français (Mauritanie). Karthala. Gràcia; Contreras, Lluïsa; Joan Miquel (2005). El Soninké i el Mandinga. Universitat de Girona.
O. Corpora:	There are no real corpora of Soninke except for the Leipzig collection of sentences. Leipzig Corpora Collection – Soninke Corpus: up to 1 million sentences

	Asawan.org – Soninke Language and Literature: Asawan.org is dedicated to the Soninke language and literature, offering a collection of Soninke books available for free download
P. Language situation:	Mali, Senegal, and Mauritania: National language departments support literacy programs in Soninke. Promotion of Soninke oral traditions, especially epic poetry (e.g., Epic of Wagadu). Soninke-language festivals and storytelling events in both West Africa and diaspora communities. In France and other European countries, Soninke associations organize language classes for children. Social media platforms and YouTube are increasingly used to teach Soninke basics.
Q. Awareness / Function for identity:	In Mali, Senegal, and Mauritania — often have a clear sense of their linguistic identity. Many recognize Soninke (or Sooninkanxanne) as a distinct, historical, and prestigious language. In traditional settings, it's often associated with cultural pride, oral history, and Islamic scholarship (due to Ajami usage). Language is tied to lineage, caste, and regional origins in many Soninke-speaking societies. There is growing awareness of Soninke's grammatical structure, dialectal diversity, and its relation to other Mande languages. Soninke is increasingly being seen as part of a broader movement to recognize and promote African languages alongside French, Arabic, or English. Schooling in colonial languages: French or Arabic often dominate the classroom, limiting Soninke's visibility. Lack of institutional standardization in some countries (e.g., Guinea, Côte d'Ivoire) means people may speak Soninke but not read/write it.
R. Language Policy:	In Mali, the government promotes the use of national languages in education, literacy, and administration; In Senegal Soninke is used in non-formal education, community radio, and occasionally in public health campaigns. Materials like literacy manuals, orthographies, and teacher guides have been published. Soninke has less political and economic visibility compared to Wolof. In Mauritania limited use in formal education due to prioritization of Arabic and French. More visible in community radio, cultural programming, and religious contexts (e.g., Ajami texts)
S. Note on the pluricentricity:	Soninke's pluricentricity is weak. It fulfils criteria 1-5, but lacks a sufficient number of speakers in the countries where Soninke is a minority language (Burkina Faso, Gambia, Guinea-Bissau, Guinea, Ghana, Ivory Coast). It is a national language in three countries (Mali, Senegal, Mauritania) alongside other national languages and the official languages (French and Arabic). This lowers its status. It is a completely underresourced language. Although it is officially recognised and used in the media, it is not implemented in the school curriculum, where French and Arabic dominate.
T. Literature (selection):	Bandiougou Dramé et Marianne Hagg, Guide d'écriture de la langue soninké, SIL Mali, 2020 (lire en ligne [archive]) Direction de la promotion des langues nationales du Sénégal, Li-

vret d'auto-formation en Soninké, éditions Kalaama-Edicef, 2001
 Christian Girier, Parlons soninké, Paris, l'Harmattan, 1996, 311 p.
 (ISBN 2-7384-3769-9,

42. Sotho

A. Name of the PCL:	Sotho
B. Endonyms - Other name(s) of the PCL:	Sesotho, Southern Sotho, or Sesotho sa Borwa
C. Language family and region:	Niger-Cong language, Southern Bantu branch; native to Lesotho, South Africa, Zimbabwe
D. Demography and language spread:	Lesotho: ca. 1.965.000 L1 speakers (85% of the population); South Africa, ca. 804.000 L1 speakers (7,6%); Zimbabwe, L1 ca. 49.000)
E. Status-OFL in:	Lesotho: National language (co-official with English); South Africa: Sesotho is one of South Africa's twelve official languages Zimbabwe: Sesotho is one of sixteen official languages of Zimbabwe. Sesotho holds de jure status and may be used in education and local government where speaker communities exist.
F. Status-RML in:	None
G. DV:	Lesotho Sotho
H. NDVs of the PCL:	South African Sesotho, Zimbabwean Sesotho
I. Alphabet / WS:	Latin with additional letters and digraphs
J. Standardisation, regulated by:	Pan South African Language Board
K. Standard form:	Lesotho Sotho
L. Ling. fragmentation:	Minimal Internal Dialect Variation: Despite its geographic spread across Lesotho and South Africa, Sesotho proper exhibits remarkably little internal dialect stratification
M. Mutual intelligibility:	High among Sesotho, Tswana, Northern Sotho
N. Dictionaries and grammars:	Phakeng, D. M.: Southern Sotho-English Dictionary: One of the most well-known bilingual dictionaries Sesotho National Lexicography Unit: Thanodi ya Sesotho: A comprehensive monolingual Sesotho dictionary Milubi, M.A.: English-Southern Sotho Dictionary". A reverse companion to the Sotho-English dictionary. Focuses on practical vocabulary and usage. Pukuntsu (Online Sesotho Dictionary) https://pukuntšu.co.za Sesotho Online A gentle introduction to the Sesotho language and culture. *** Mkhize, Z.L. (2003): A Practical Guide to the Grammar of Sesotho Milubi, M.A. 1988: English-Southern Sotho Dictionary Doke, Clement Martyn (1931): Outline of Sotho Grammar van Wyk, J.A. & Snyman, Jean (1994): A Grammar of Southern Sotho
O. Corpora:	The Sesotho National Corpus (SNC): Developed by: Sesotho Na-

	<p>tional Lexicography Unit (NLU), The South African National Centre for Digital Language Resources (SADiLaR) – Sesotho Corpora. Hosted by: https://sadilar.org The Leipzig Corpora Collection – Sesotho Corpus. Maintained by: Universität Leipzig, Germany. Autshumato Project – Sesotho Parallel Corpus. Funded by: Department of Arts and Culture, South Africa. Official government documents in multiple languages. https://autshumato.sourceforge.net</p>
P. Language situation:	<p>South Africa: It's widely used in the Free State and parts of Gauteng and North West provinces—in education, provincial government communications, and public broadcasting—alongside the other official tongues of the post-1994 constitutional order</p>
Q. Awareness / Function for identity:	<p>For the Basotho people, Sesotho is the primary emblem of their shared heritage—from oral praise poems extolling clan lineages to ceremonies like the Morija Arts & Cultural Festival In Lesotho, Sesotho is enshrined in the constitution as an official language alongside English; official speeches, legal proceedings, and parliamentary debates routinely employ Sesotho to signal national unity. Code-switching between Sesotho and English (or Afrikaans) allows speakers to negotiate social contexts—switching into Sesotho to signal closeness or solidarity, or into English to index formality or wider reach. Young urban Sesotho speakers blend “tsotsi-taal” slang and digital borrowings into everyday speech, creating a distinct youth identity that still roots itself in Sesotho grammar and sound patterns.</p>
R. Language Policy:	<p>Mother-tongue instruction in early grades has been promoted in Lesotho and South Africa to strengthen literacy and pride in Sesotho. Radio Sesotho stations (e.g., Lesotho's Radio Lesotho, South Africa's Lesedi FM) broadcast news, drama, and discussion programs entirely in Sesotho, normalizing its use in public life. Annual book fairs in Maseru and Bloemfontein often feature Sesotho authors, translators, and poets. Sesotho language days in schools celebrate traditional storytelling, praise poetry (lit. “diboko”), and folk music, embedding language pride from childhood.</p>
S. Note on the pluricentricity:	<p>The pluricentricity of Sesotho is well developed. It fulfils criteria all seven criteria, codification is still underdeveloped, but under way.</p>
T. Literature (selection):	<p>Demuth, Katherine (1995). Sesotho Verbal Morphology: A Prosodic-Morphological Account. <i>Journal of African Languages and Linguistics</i> 17(2): 161–184. <i>Journal of African Languages and Linguistics</i> 17(2): 161–184. Marten, Lisa & Demuth, Katherine (2001). “Sesotho Prosody: Foot Structure and Reduplication.” <i>Phonology</i> 18(1): 77–117. Mokitimi, Molefi & Sebe, Andrew (2012). “Code-Switching and Identity among Urban Sesotho Speakers.”</p>

	<p>Journal of Multilingual and Multicultural Development 33(5): 429–442.</p> <p>Mabatle, Ratiba (2015). “Discourse Markers in Sesotho: A Pragmatic Analysis.” <i>Pragmatics</i> 25(1): 1–27.</p> <p>Mahlatsi, Bulelani (2018). “Language Attitudes and the Lesotho Education Act.” <i>International Journal of the Sociology of Language</i> 255: 121–140.</p>
--	--

43. Swazi

A. Name of the PCL:	Swazi
B. Endonyms - Other name(s) of the PCL:	siSwati
C. Language family and region:	Bantu language, Nguni group, Southern Bantu; native to Eswatini (formerly Swaziland) and South Africa by the Swati people;
D. Demography and language spread:	L1: Eswatini: ca. 1.3 million, South Africa, ca. 2.3 million, Mozambique & Lesotho: a few thousand;
E. Status-OFL in:	Eswatini (formerly Swaziland) and South Africa
F. Status-RML in:	none
G. DV:	Eswatini Swazi
H. NDVs of the PCL:	none
I. Alphabet / WS:	Latin (Swazi alphabet); Ditema tsa Dinoko
J. Standardisation, regulated by:	<p>Swaziland (Eswatini) National Curriculum Centre (NCC): plays a key role in developing and standardising siSwati for education, produces school syllabi, textbooks, and literary materials in siSwati.</p> <p>Pan South African Language Board (PanSALB) – South Africa, has a dedicated siSwati Language Body, responsible for: Standardisation of grammar, spelling, and terminology. Promoting the development and use of siSwati in media, education, and government. Grammar and Lexicon: Standard grammatical rules are taught in schools; new terms are developed for modern concepts (especially in technology and science).</p>
K. Standard form:	Eswatini Swazi, South African Swazi
L. Ling. fragmentation:	Central Swazi: Manzini region (Eswatini); Considered the standard variety; Northern Swazi: Hhohho and northern Eswatini; (minor phonetic differences); Southern Swazi: Eswatini; (influenced by Zulu in border regions); South African Swazi: (more lexical and phonetic borrowing from Zulu and Afrikaans)
M. Mutual intelligibility:	There isn't significant linguistic fragmentation of siSwati yet. siSwati remains fairly cohesive across Eswatini and South Africa. The dialects are mutually intelligible, and the language remains cohesive.
N. Dictionaries and grammars:	<p>Shongwe, J.S. (1986): <i>English-Swati Dictionary</i>. Publisher: Macmillan Boleswa Publishers</p> <p>Guma, S. M. (1971): <i>An Outline Structure of the Swati Language</i> University of South Africa Press.</p> <p>Gwebu, T.S. (2006). <i>A Concise SiSwati-English Dictionary</i>. Longman.</p>

	<p>***</p> <p>Rycroft, D.K., & Ngcobo, A.B. (1979). <i>The Lexicon of the Swazi Language</i>. Department of African Languages, University of Natal. Pietermaritzburg.</p> <p>Dlamini, M.P. (1993). <i>An Introduction to the Grammar of siSwati</i>. Swaziland National Curriculum Centre. Mbabane, Swaziland</p> <p>Ziervogel, D., & Mokgokong, P.C. (1975). <i>A Handbook of the Swazi Language</i>. J.L. van Schaik. Pretoria</p>
O. Corpora:	<p>The South African Centre for Digital Language Resources (SADi-LaR) https://www.sadilar.org; SADiLaR works with official South African languages, including siSwati.</p> <p>The Leipzig Corpora Collection: https://corpora.uni-leipzig.de; web based corpus.</p>
P. Language situation:	<p>siSwati is used in newspapers, radio, TV, and increasingly in digital content. siSwati is both a subject and medium of instruction in lower primary levels in Eswatini, and taught as a subject in South Africa.</p> <p>Terminology Development: Language bodies actively create new vocabulary for education, science, and governance.</p>
Q. Awareness / Function for identity:	<p>SiSwati it's a symbol of national pride and unity in Eswatini. It's deeply tied to Swazi culture (umhlanga, traditional ceremonies, kinship structures) and oral storytelling traditions. The monarchy and cultural institutions actively promote siSwati through ceremonies and schools. Many Swazis see siSwati as a marker of authenticity and belonging, especially in rural areas.</p> <p>siSwati in South Africa: A Minority Voice</p> <p>In Mpumalanga province, siSwati is one of South Africa's 11 official languages. However, awareness and use depend on regional demographics — it's not as prominent as Zulu, Xhosa, or Afrikaans nationally. Urban migration and education in English sometimes dilute siSwati's everyday use among younger generations. Some siSwati speakers feel their language gets marginalized, prompting cultural initiatives to revive pride and usage.</p> <p>Radio stations like Ligwalagwala FM promote siSwati music, news, and dialogue.</p> <p>There's increasing visibility in education, translation efforts, and even social media activism. There's often a code-switching culture — English for business/school, siSwati for home/community.</p>
R. Language Policy:	<p>Eswatini (Swaziland): siSwati and English are the two official languages. This is enshrined in the Constitution of the Kingdom of Eswatini (2005), which protects siSwati's status and usage in public life.</p> <p>siSwati is the dominant language in daily life, traditional structures, and the home. English dominates in official documents, education beyond early grades, and business. There is no fully developed, detailed national language policy document with implementation strategies like some multilingual countries have. siSwati is taught in schools, especially in early grades. It is used as a medium of instruction in lower primary, then shifts to English. There's curriculum development in siSwati literature and grammar. There is a lack of advanced resources for siSwati (e.g., text-</p>

	books, scientific materials). siSwati in one of South Africa's 11 official languages. The policy recognizes language rights and equal status for all official languages. The Pan South African Language Board (PanSALB) works to promote and protect language rights, including siSwati. The department of Basic Education and Higher Education encourage mother-tongue instruction and local language development. siSwati is offered as a home language and additional language subject in schools in Mpumalanga. Radio stations like Ligwalagwala FM promote cultural and linguistic content. Despite constitutional protection, practical implementation is inconsistent. Resources, funding, and teacher training in siSwati are often lacking.
S. Note on the pluricentricity:	The pluricentricity of siSwati fulfils 7 out of seven criteria. The language has its centre in Eswatini, with South Africa as a non-dominant variety that suffers from being less cared for than in Eswatini.
T. Literature (selection):	Dlamini, Nkosinathi Benson (2013): An Investigation of the Teaching of siSwati as a First Language in Swaziland Secondary Schools. PhD Dissertation. University of South Africa (UNISA) Msimang, C.T. (1991): An Introduction to siSwati Phonetics and Phonology. University of South Africa Press. Nkabinde, Z. P. (2008): An Analysis of the Language Policy in Eswatini. Working Paper. University of Eswatini (UNESWA) PanSALB Reports: Status of the Use of Official Languages in South Africa (Annual Reports). Vary (e.g., 2017, 2019) Pan South African Language Board

44. Swedish

A. Name of the PCL:	Swedish
B. Endonyms - Other name(s) of the PCL:	Svenska
C. Language family and region:	North Germanic, East Scandinavian language ; native to Sweden and Finland (including Åland)
D. Demography and language spread:	L1: Sweden: ca. 10,45 million, Finland ca. 263.000, International: EU, Nordic Council
E. Status-OFL in:	Sweden, Finland (co-official alongside Finish)
F. Status-RML in:	None
G. DV:	Swedish Swedish
H. NDVs of the PCL:	Finland Swedish
I. Alphabet / WS:	Latin (Swedish alphabet with 29 letters)
J. Standardisation, regulated by:	Swedish Language Council (in Sweden), Swedish Academy (in Sweden) Institute for the Languages of Finland (in Finland)
K. Standard form:	Rikssvenska / Standardsvenska (Standard Swedish) In Sweden; Högsvenska (High Swedish) in Finland
L. Ling. fragmentation:	Norrland, Finland Swedish varieties, Svealand, Gotland, Götaland, South Swedish; Rinkeby Swedish (immigrant sociolect)
M. Mutual intelligibility:	High intelligibility within Standard + Svealand/Götaland, moderate with Scanian and Norrland dialects and lower with Gotlandic

	and older rural dialects. Finland Swedish is generally mutually intelligible with Standard Swedish.
N. Dictionaries and grammars:	<p>Svenska Akademiens ordlista (SAOL): https://www.saob.se/Svenska Akademiens Ordbok Språkrådet (2008), Svenska skrivregler (in Swedish) (3rd ed.), Stockholm: Liber, ISBN 978-91-47-08460-9 (Swedish orthography).</p> <p>Swedish-English/Swedish-Arabic/Swedish-Russian/Swedish-Spanish Dictionaries from Språkrådet – Institute for Language and Folklore: http://lexin2.nada.kth.se/</p> <p>***</p> <p>Bolander, Maria (2002), Funktionell svensk grammatik (in Swedish), Stockholm: Liber, ISBN 978-91-47-05054-3, OCLC 67138445</p> <p>Swedish: A Comprehensive Grammar Second Edition. Holmes, Philip; Hinchliffe, Ian; (2003). London; New York: Routledge. ISBN 0-415-27884-8.</p> <p>Svenska utifrån. Schematic grammar-Swedish structures and everyday phrases Byrman, Gunilla; Holm, Britta; (1998) ISBN 91-520-0519-4.</p>
O. Corpora:	<p>Korp (Språkbanken) Host: Språkbanken (University of Gothenburg)</p> <p>https://spraakbanken.gu.se/korp (annotated)</p> <p>SweCorp Host: Various universities (often in cooperation with KTH or Stockholm University), General-purpose text corpus</p> <p>Swedish Language Bank Parallel Corpora</p> <p>Languages: Swedish + English, Finnish, German, etc.</p> <p>Swell (Swedish Learner Language Corpus)</p> <p>Focus: Learner texts (L2)</p>
P. Language situation:	<p>Swedish is a well cared for language both in Sweden and Finland. In Finland, there is occasional political debate about the requirement to learn Swedish in Finnish-speaking schools. Some advocate for more flexibility, while others stress the importance of preserving bilingualism as part of Finnish identity</p>
Q. Awareness / Function for identity:	<p>In Sweden, Swedish is a unifying symbol and a marker of Swedishness.</p> <p>In Finland, Swedish functions as a minority identity marker—important for maintaining cultural distinctiveness, especially in areas like Ostrobothnia and the Åland Islands (where Swedish is the only official language).</p>
R. Language Policy:	<p>Official Language Status in Sweden (since 2009) and Finland (since 1920)</p> <ol style="list-style-type: none"> Swedish should be the main language in Sweden. Everyone has the right to develop and use their mother tongue. Everyone should have the opportunity to learn and use Swedish. The public sector should use clear, understandable language (also known as klarspråk or "plain language"). <p>Sweden officially recognizes five national minority languages: Sami (all varieties), Finnish, Meänkieli (Tornedalen Finnish), Ro-</p>

	<p>mani Chib, Yiddish (Education in the minority language) In Finland: Government and Administration: Finnish citizens have the right to use either Finnish or Swedish when interacting with state and municipal authorities. Municipalities are either unilingual or bilingual, depending on the proportion of Swedish/Finnish speakers. Education: Both Finnish and Swedish are mediums of instruction in schools. Students in Finnish-speaking schools are required to study Swedish as a second national language (pakkoruotsi, or "mandatory Swedish"). Conversely, Swedish-speaking students must study Finnish.</p>
S. Note on the pluricentricity:	<p>The pluricentricity of Swedish is well developed and fulfils 7 out of 7 criteria. There is an implicit policy ot to let the linguistic differences between Swedish Swedish and Finald Swedish getting to large. Swedish in Finland has high prestige because of its status as national language – a rare case among NDVs.</p>
T. Literature (selection):	<p>Andersson, Erik (2002), "Swedish", in König, Ekkehard; van der Auwera, Johan (eds.), <i>The Germanic Languages</i>, Routledge language family descriptions, Routledge, pp. 271–312, ISBN 978-0-415-28079-2</p> <p>Haugen, Einar (2009). "Danish, Norwegian and Swedish". In Bernard Comrie (ed.). <i>The World's Major Languages</i>. New York: Routledge. pp. 125–144. ISBN 978-0-415-35339-7.</p> <p>Norrby, Catrin; Wide, Camilla; Lindström, Jan; Nilsson, Jenny (2012): <i>Finland-Swedish as a non-dominant variety of Swedish – extending the scope to pragmatic and interactional aspects</i>. In: Rudolf Muhr (ed.) (2012): <i>Non-dominant Varieties of pluricentric Languages. Getting the Picture</i>. In memory of Prof. Michael Clyne. Wien et. al., Peter Lang Verlag. p. 49-62.</p> <p>Henricson, Sofie; Nelson, Marie; Wide, Camilla; Catrin Norrby; Nilsson, Jenny; Lindström, Jan (2015): <i>You and I in Sweden-Swedish and Finland-Swedish supervision meetings</i>. In: Rudolf Muhr, Dawn Marley et. al. (eds.) (2015): <i>Pluricentric Languages. New Perspectives in Theory and Description</i>. Wien et. al., Peter Lang Verlag. p. 117-130.</p> <p>Bernsand, Niklas and Nilsson, Jenny (eds.): <i>Urban Voices: Sociolinguistic Variation in Swedish</i>.</p>

Tamazight/Berber

A. Name of the PCL:	Tamazight, Amazight
B. Endonyms - Other name(s) of the PCL:	Berber, Tamazight, Amazigh تَمَزِيْغَت, Tamaziɣt ⵜⴰⴳⴷⵓⴷⴰⵢⵜ, Tashelhit (Shilia) and Tarifit (Riffian) Regional names: Taqbaylit: Kabyle (northeastern Algeria), Tarifit or Tamaragt: Riffian Berber (northern Morocco), Tamasheq or Tamaheq: Tuareg varieties (spoken in parts of Mali, Niger, Algeria, Libya), Tachenwit: Shawiya (Aurès Mountains, eastern Algeria), Tamahaq: Tuareg of the Hoggar region (southern Algeria).
C. Language family and region:	Afroasiatic; native to Morocco, Algeria, Mali, Niger, Burkina Faso, Tunisia, Egypt, Lybia.
D. Demography and language spread:	L1: Morocco: ca. 9,1 million, Algeria ca. 4,5million, As of 1998, there were an estimated 450,000 Tawellemmet speakers, 250,000 Air Tamajeq speakers, and 20,000 Tamahaq speakers in Niger, as of 2018 and 2014 respectively, there were an estimated 420,000 speakers of Tawellemmet and 378,000 of Tamasheq in Mali.
E. Status-OFL in:	Morocco and Algeria (co-official); Mali and Niger (recognised as a national language)
F. Status-RML in:	None
G. DV:	Moroccan Tamazight, Mali Tamazight, Niger Tamazight
H. NDVs of the PCL:	Algerian Tamazight,
I. Alphabet / WS:	Tifinagh alphabet, the Arabic script, and the Berber Latin alphabet
J. Standardisation, regulated by:	Morocco: Royal Institute of Amazigh Culture (IRCAM); No regulation body in Algeria;
K. Standard form:	Standard Moroccan Amazigh (Tamaziɣt) ⵜⴰⴳⴷⵓⴷⴰⵢⵜ ⵜⴰⴷⵓⴷⴰⵢⵜ ⵜⴰⴳⴷⵓⴷⴰⵢⵜ (Tanawassit), (Tamaziɣt Tanawadt)
L. Ling. fragmentation:	High regional fragmentation across Western Africa (Niger, Mali, Mauritania, Burkina Faso) and North Africa (Tunisia, Lybia, Egypt).
M. Mutual intelligibility:	The mutual intelligibility of Berber (Tamazight) varieties is low to moderate and varies significantly across dialect. Varieties spoken far apart (e.g., Tamasheq in Mali vs. Kabyle in Algeria) are not mutually intelligible. Neighboring varieties (e.g., Tashelhit and Central Atlas Tamazight in Morocco) may have partial intelligibility.
N. Dictionaries and grammars:	Chafik, Mohamed: Dictionnaire amazigh-français Variety: Standardized Tamazight. Publisher: IRCAM (Royal Institute of Amazigh Culture), Morocco) Roux, Arsène (2002): Dictionnaire Tamazight-Français. Variety: Tachelhit (Southern Morocco), Publisher: Peeters. ISBN: 9789042912276 Mouloud Mammeri (1980): Dictionnaire général français-tamazight" variety: Kabyle. Editions La Découverte / L'Harmattan. Foucauld, Charles de (1951): Dictionnaire Touareg-Français" (2 vols.) Variety: Ahaggar Tuareg. Publisher: Imprimerie Nationale.

	<p>Acherchou, Amar (2012): Lexique français-tamazight (berbère). Variety: Central Atlas Tamazight. Publisher: Editions La Croisée des Chemins, 2012 ***</p> <p>Maarten Kossmann (2011): Taqbaylit: Grammaire du berbère kabyle" Variety: Kabyle. Publisher: Peeters Publishers. ISBN: 9789042924569</p> <p>Jeffrey Heath (2005): Grammar of Tamashek (Tuareg of Mali)" Variety: Kidal Tuareg (Mali), Publisher: Mouton de Gruyter.</p> <p>Mouloud, Mammeri (1976): Grammaire berbère (kabyle). Variety: Kabyle Publisher: Editions La Découverte.</p>
O. Corpora:	<p>Corpus of Amazigh Language (Morocco): Managed by: Institut Royal de la Culture Amazighe (IRCAM): www.ircam.ma TALAA Corpus: Stands for: Traitement Automatique de la Langue Amazighe et Arabe. OPUS Open Corpus: https://opus.nlpl.eu/ Multilingual corpora with some Tamazight content. University Projects in Morocco, Algeria, and France often include linguistic data collection efforts. Github Repos: Some researchers and NLP developers have uploaded tokenizers, POS taggers, and basic corpora for Tamazight (especially under tags like "tamazight", "berber", or "amazigh"). ISO codes for web searches: ber = Berber languages (macrolanguage), zgh = Standard Moroccan Tamazight, tzm = Central Atlas Tamazight, shi = Tachelhit, rif = Tarifit</p>
P. Language situation:	<p>After independence, all the Maghreb countries pursued a policy of Arabisation, aimed partly at displacing French from its colonial position as the dominant language of education and literacy. The use of the Berber languages was suppressed or even banned. Berbers in Morocco and Algeria—especially Kabylie – contested this policy. This was addressed in both countries by affording the language official status and introducing it in some schools. In 2011, Tamazight was added as an official language to the Moroccan constitution. In Algeria, Tamazight was added to the constitution as a national and official language in 2016. In Mali and Niger, some Tuareg languages have been recognized as national languages and have been part of school curriculums since the 1960s. It is taught in Moroccan schools since 2003. Used in official signage, some media broadcasts, and public communication. Still undergoing standardization, especially for technical vocabulary and grammar rules.</p>
Q. Awareness / Function for identity:	<p>Many Tamazight speakers are acutely aware of their language's endangered status, especially in regions where it has historically been marginalized. Tamazight distinguishes Amazigh speakers from the dominant Arabophone and Francophone cultures of North Africa. Symbol of Resistance: Use of Tamazight became a rallying point against Arabization policies and for linguistic rights—speaking it publicly is itself a political act. Transnational Networks: Shared language fosters connections</p>

	<p>among Amazigh communities from Morocco to Libya and across the diaspora, enabling cross-border media, conferences, and online activism.</p> <p>Digital Identity: Hashtags (#Tamazight), YouTube channels and TikTok creators in Tamazight build a virtual “homeland” where Amazigh youth affirm a global Amazigh identity</p>
R. Language Policy:	<p>The endangered status has led to community-driven revitalization efforts; increased activism for official recognition; awareness of language standardization processes, particularly through initiatives by organizations like IRCAM (Royal Institute of Amazigh Culture in Morocco). Even among its dialects (e.g. Tarifit, Tachelhit, Tamahaq), each variety carries local prestige and fosters pride in one’s regional roots.</p>
S. Note on the pluricentricity:	<p>The pluricentricity of Tamazight is fulfils 7 out of 7 criteria, even though the the implementation of the promotion of the language is problematic in Algeria and in all other countries except Morocco. "Berber" is a metalinguistic expression that refers to a large number of regional varieties that vary from each other. The language itself is highly fragmented and pluricentricity can be claimed by being the official language in Morocco and Algeria.</p>
T. Literature (selection):	<p>Lafkioui, Mena B. (2024): Pluricentricity, iconisation, and instrumentalisation of the Tamazigh language in North Africa and its diaspora. In: Máté Huber / Benjamin Meisnitzer (eds.) (2023): Pluricentric languages in Africa and in other Regions of the World Graz. PCL-Press. p. 11-34.</p> <p>Lafkioui, Mena B. (2018): Berber languages and linguistics. Oxford Bibliographies. DOI: 10.1093/OBO/9780199772810-0219.</p> <p>Arezki, Abdenour (2016): Between official recognition and social reality: The case of Tamazight/Berber in Algeria. In: Rudolf Muhr, Kelen Ernesta Fonyuy, Zeinab Ibrahim, Corey Miller (eds.) (2016): Pluricentric Languages and non-dominant Varieties worldwide. Volume 1: Pluricentric Languages across conti-nents - Features and usage. Wien et. al., Peter Lang Verlag. p. 155-162.</p> <p>Lafkioui, Mena B. (2024): Tamazight. A multi-millennial journey of a language family and its cultural practices. In: García-Ehrenfeld, C. (ed.) <i>Amazige, Griego, Latín. Una aproximación multidisciplinaria</i>. México, UNAM, p. 368–395.</p> <p>Chaker, Salem (1991): <i>Manuel de linguistique berbère I</i>, Edition Bouchène, Alger</p> <p>Chaker, Salem (1995): <i>Linguistique berbère: Études de syntaxe et de diachronie</i>. M. S.—Ussun amaziɣ 8, ser. ed. Salem Chaker. Paris and Leuven: Uitgeverij Peeters.</p>

A. Name of the PCL:	Urdu
B. Endonyms - Other name(s) of the PCL:	اُردُو (Urdu) [ʊrdu:] Older endonyms: زَبان اُردُو (Zabān-e-Urdu) – “Language of Urdu”; زَبان اُردُوئے مُعَلّٰی (Zabān-e-Urdu-e-Mu'allā) – “The exalted language of the camp” [court]
C. Language family and region:	Indo-Aryan, Hindustani language; native to Pakistan (widely used as lingua franca), India: Hindi-Urdu Belt and Deccan, Afghanistan, Terai, Nepal, Old Dhaka, Bangladesh
D. Demography and language spread:	Pakistan: L1: Ca. 14 million (about 8% of the population), L2: ca. 94 million, Total: ca. 108 million; India: L1: ca. 85 million (6.7% of the population, per 2011 Census), L2: ca. 6 million, Total: ca. 91 million Bangladesh: L1: ca. 250.000, L2: ca. 140.000, total: Ca. 390.000 Nepal: L1: ca. 692.000, L2: ca. 23.000, total: ca. 715.000
E. Status-OFL in:	Pakistan (national); India (scheduled language) in: Jammu and Kashmir (co-official); National Capital Territory of Delhi (additional); Bihar (additional); Uttar Pradesh (additional); Jharkhand (additional); Andhra Pradesh (additional), Telangana (additional), West Bengal.
F. Status-RML in:	South Africa (protected language)
G. DV:	Pakistani Urdu
H. NDVs of the PCL:	Indian Urdu, Afghanistan Urdu, Bangladesh Urdu
I. Standardisation, regulated by:	National Language Promotion Department (Pakistan) National Council for Promotion of Urdu Language (India)
J. Alphabet / WS:	Perso-Arabic (Urdu alphabet), Latin (Roman Urdu), Bengali-Assamese script (in Bangladesh)
K. Standard form:	Modern Standard Urdu (ادبی اُردُو or معیاری اُردُو), Perso-Arabic script, written in Nasta'liq style. Closely aligned with Standard Hindi, but with key lexical and phonetic differences.
L. Ling. fragmentation:	Punjabi-influenced Urdu: Punjab (both urban and rural areas): Mutual Intelligibility: Very high, as the main differences are related to accent and some vocabulary; Pashto-influenced Urdu: Region: Khyber Pakhtunkhwa, Balochistan, and areas near the Afghanistan border: Mutual Intelligibility: High; Sindh-influenced Urdu, Region: Sindh: Mutual Intelligibility: Very high. Balochi-influenced Urdu: Region: Balochistan: Mutual Intelligibility: Very high. Hyderabad Urdu, Region: Hyderabad (Sindh): Mutual Intelligibility: Very high. Indian regional varieties: Deccani (Dakhini) Urdu (Andhra Pradesh, Karnataka, Maharashtra); Hyderabad Urdu: Hyderabad and surrounding areas; North Indian Urdu: Uttar Pradesh, Bihar, Delhi, parts of Madhya Pradesh; Kolkata Urdu: West Bengal
M. Mutual intelligibility:	Pakistan: Despite linguistic differences, they are mutually intelligible for speakers of standard Urdu. India: Most regional varieties maintain a high degree of mutual intelligibility with Standard Urdu, particularly in spoken communication.
N. Corpora:	Linguistic Data Consortium for Indian Languages (LDC-IL): LDC-IL Gold urWaC Sketch Engine: Urdu Web Corpus - Standard Urdu Raw Text Corpus. 53 million words:

	<p>https://www.sketchengine.eu/urwac-urdu-corpus/ LDC-IL Urdu Sentence Aligned Speech Corpus: Linguistic Data Consortium for Indian Languages (LDC-IL). Size: 32,384 audio segments; 50 hours of speech; 434 speakers</p>
O. Dictionaries and grammars:	<p>John T. Platts (1884): A Dictionary of Urdu, Classical Hindi, and English Publisher: Sampson Low, Marston & Co., London. ISBN-13: 978-1881338604. Reprinted in 2011 by Nataraj Book. Available online through the Digital South Asia Library.</p> <p>Haqquee, Shanul Haq (2002): Oxford English–Urdu Dictionary. Oxford University Press, Pakistan. ISBN: 9780195793406, 2000p.</p> <p>Qureshi, Bashir / Ahmad Haq, Abdul (1991): Standard Twenty First Century Urdu–English Dictionary. Educational Publishing House, India</p> <p>Maulvi, Ferozuddin (2005): Feroz-ul-Lughat (Urdu to Urdu). Ferozsons (Private) Limited, Lahore, Pakistan. ISBN: 9789690019424. Pages: 1544</p> <p>Haq Haqquee, Shanul (2014): Little Oxford English–Urdu Dictionary Oxford University Press, Pakistan. ISBN: 9780195978995 ***</p> <p>Schmidt, Ruth Laila (1999): Urdu: An Essential Grammar. London, Routledge. ISBN: 978-0415163811, 300p.</p> <p>Fatihi, Ali R. (2022): A Reference Grammar of Urdu. Publisher: LAP LAMBERT Academic Publishing. ISBN: 978-6203198027. 320p.</p>
P. Language situation:	<p>Pakistan: Urdu is the national language, but not the mother tongue of the majority. Used widely as a second language and in media, education, and government (though English often dominates in elite domains). Unifying Language: Intended to unify a multi-ethnic, multilingual country. Medium of Instruction: Used in public schools and madrassas. Tension with Regional Languages. Stronger presence in urban centers; rural populations often prefer their native regional languages. Elite schools favor English, while public schools use Urdu, reinforcing socio-economic divisions.</p> <p>India: Official language in some Indian states. Considered the mother tongue of Indian Muslims. Cultural and Literary Language: Rich tradition of poetry, fiction. Medium of Education: Exists, but shrinking. Religious Language: Plays a role in Islamic education. Declining Institutional Support: Urdu-medium schools are underfunded and closing. Political Marginalization: Script Barrier: The Perso-Arabic script is seen as a barrier to learning for non-native users; some suggest promoting a Roman or Devanagari script alternative. Declining use as a first language in younger generations, especially in urban India. Increased symbolic and cultural value, particularly among Indian Muslims and cultural enthusiasts.</p>
Q. Awareness / Function for identity:	<p>Pakistan: Urdu is the symbol of national unity in Pakistan. It functions as a neutral lingua franca, not affiliated with any major ethnic group in power at the time. Urdu is tied to Islamic identity and heritage, drawing from Persian and Arabic vocabulary. Its literary</p>

	<p>tradition — especially poetry (like Ghalib, Iqbal) — is considered central to Pakistan's cultural and national ethos. Despite its national status, Urdu is seen by some as an "imposed language", especially by Punjabi nationalists and Sindhi speakers.</p> <p>India: Urdu in India is closely associated with Muslim identity and Minority Identity. It retains cultural prestige among Indian Muslims and is often seen as a symbol of cultural resistance and pride. Urdu is linked to the Ganga-Jamuni tehzeeb (composite Hindu-Muslim culture) of North India. Urdu has often been marginalized politically.</p> <p>It's used as a symbol of minority rights and is frequently at the center of debates over language and secularism.</p>
R. Language Policy:	<p>Pakistan: Official Language (in practice): Although English has long functioned as the de facto official language; Urdu is the medium of instruction in public schools in many regions. Provinces may prescribe measures for promoting provincial languages alongside Urdu. Higher education and technical disciplines are often taught in English, creating a diglossic environment. However, there is a lack of unified policy implementation across provinces. Tension between national and regional languages. Dominance of English in elite and bureaucratic circles.</p> <p>India: Is one of the 22 scheduled languages of India (Eighth Schedule of the Constitution) and the official language in six Indian states.</p> <p>Urdu is recognized as a medium of instruction in various states.</p> <p>Three-language formula: Allows Urdu to be taught as a first, second, or third language depending on the state policy. Disadvantages include a decline in Urdu-medium schools. Political marginalization and reduction of institutional support in some regions. Competition with Hindi and regional languages in public and educational domains.</p>
S. Note on the pluricentricity:	<p>The pluricentricity of Urdu fulfils 7 out of 7 criteria. Urdu is a well-established language in both Pakistan and India. The linguistic differences between the two NVs are present, but not too great, so that mutual understanding is achieved.</p>
T. Literature (selection):	<p>Rahman, Tariq (2021): <i>From Hindu to Urdu. A social and political history.</i> Graz, PCL-Press. 476p.</p> <p>Rahman, Tariq (2022): <i>Language and Politics in Pakistan.</i> Graz, PCL-Press.</p> <p>Dua, Hans R. (1992): <i>Hindi-Urdu as a pluricentric language.</i> In: <i>Pluricentric languages: differing norms in different nations.</i> Michael Clyne (ed.). Berlin/New York. Mouton de Gruyter, 381-400.</p> <p>Rauf, Maqbool (2011). <i>Sociolinguistics of Urdu: Language, Society, and Identity.</i> Publisher: Cambridge Scholars Publishing. ISBN: 978-1443827881</p> <p>Zia, Amina (2009). <i>Language Contact and Change: Urdu in the World</i> Publisher: Multilingual Matters. ISBN: 978-1847691553</p> <p>Rahman, Tariq (2022). <i>Language and Politics in Pakistan</i> Publisher: PCL-Press. Graz.</p> <p>Mishra, R. & Rani, R. (2017). <i>Corpus-Based Studies on Urdu Lin-</i></p>

guistics
 Publisher: Springer. ISBN: 978-9811064361

46. Venda

A. Name of the PCL:	Venda
B. Endonyms - Other name(s) of the PCL:	Tshivenda or Tshivenda, also Luvenda
C. Language family and region:	Niger-Congo, Bantu language; native to South Africa, Zimbabwe and Mozambique
D. Demography and language spread:	South Africa: L1: ca. 1,2 million, L2: ca. 1,7 million, Total: ca. 2,9 million; Zimbabwe: L1: ca. 71.600 but estimated to about one million; L2: unclear, but higher. No data available for the number of speakers in Mozambique.
E. Status-OFL in:	South Africa, Zimbabwe
F. Status-RML in:	None
G. DV:	South African Venda
H. NDVs of the PCL:	Zimbabwe Venda
I. Standardisation, regulated by:	In South Africa: The Pan South African Language Board (PanSALB) Zimbabwe: The National Language Institute at the University of Zimbabwe, plays a role in developing educational materials and promoting the usage of all official languages; The Zimbabwean Languages Association (ZLA) works towards the preservation and development of Zimbabwe's indigenous languages. While there are similarities in how Venda is written in both countries, some regional differences in vocabulary, pronunciation, and usage may exist, requiring adaptation to ensure consistent communication. The standardization of Venda in both South Africa and Zimbabwe is focused on creating a uniform and consistent language form for formal, educational, and literary purposes. In South Africa, the standard form of Venda is more fully developed, supported by government policy, educational systems, and media. In Zimbabwe, the language is less prominent but still standardized for use in education and by communities where it is spoken.
J. Alphabet / WS:	Latin (Venda alphabet with 58 letters and digraphs in all), Ditema tsa Dinoko
K. Standard form:	South African Venda, Zimbabwe Venda: The NVs differ slightly in orthography, vocabulary and pronunciation due to the different linguistic environments in which the language is spoken.
L. Ling. fragmentation:	Regional varieties: Guvhu, Ilafuri, Lembetu, Manda, Mbedzi, Phani, Tavha-Tsindi
M. Mutual intelligibility:	The mutual intelligibility of Venda across South Africa and Zimbabwe is relatively high. Most speakers of Venda from either country will understand each other without much difficulty.
N. Dictionaries and grammars:	Munyai, Tshifhiwa Isaac (2005): Tshivenda-English Dictionary The University of Venda Press. ISBN: 978-1-86822-459-4 Stedman, Rudolf S. R. (1975): Venda Dictionary: Tshivenda-English, English-Tshivenda. Publisher: Oxford University Press. ISBN: 978-0195702460

	<p>The Venda-English Dictionary (1981): Thuma, M. G. H. T. /Stassen, A. H. J. The Venda-English Dictionary. Publisher: Human & Rousseau. ISBN: 978-0798102450 ***</p> <p>Munyai, T.E.D. (2002): A Grammar of Tshivenda. Publisher: The University of Venda Press. ISBN: 978-1-86822-219-4</p> <p>Thuma, M.G.H.T. (1996): A Practical Introduction to Tshivenda Grammar. Publisher: Maskew Miller Longman. ISBN: 978-0639004410</p> <p>McCullough, Lorna W.S. (1983): A Reference Grammar of Tshivenda. Publisher: University of Cape Town Press. ISBN: 978-0797205360</p>
O. Corpora:	<p>Tshivenda NCHLT Text Corpora; Tshivenda NCHLT Speech Corpora; Tshivenda Word Lists and Frequency Lists. Hosted by: Tshivenda National Centre for Human Language Technology (NCHLT) Text and Speech Corpora: SADiLaR Language Resources Portal</p> <p>Pretoria Lexical Database for Tshivenda: Compiled by: Department of African Languages, University of Pretoria Language Resource Management Agency (LRMA), Zimbabwe: Venda Oral Literature Collection (Pilot Project). Managed by: University of Zimbabwe (African Languages Department)</p>
P. Language situation:	<p>South Africa has a more developed system for teaching Venda, with greater resources, textbooks, and a structured curriculum for the language. In Zimbabwe, the language is taught in fewer schools, and the focus is more on preserving the language rather than developing a full-fledged curriculum.</p>
Q. Awareness / Function for identity:	<p>Tshivenda carries the collective memory of the Venda people—proverbs, folktales, rituals, songs, and traditional knowledge. Oral traditions passed down in Tshivenda preserve indigenous wisdom, moral values, and worldviews. Used in greetings, rituals, weddings, and funerals, the language strengthens community cohesion. Is the medium for sacred practices: ancestor veneration, rainmaking, initiation ceremonies, and traditional healing. During colonial and apartheid eras, speaking Tshivenda was a form of cultural resistance. Today, it is a symbol of pride, especially among youth reclaiming African identity through indigenous languages. Tshivenda unites Venda communities across South Africa and Zimbabwe, reinforcing cross-border ethnic identity in a post-colonial context.</p>
R. Language Policy:	<p>South Africa: National Language Policy Framework (2003), developed by the Department of Arts and Culture. Promotes the use of indigenous African languages like Tshivenda in: Government communication, Education, Broadcasting and media, Translation and interpreting services; Emphasizes the intellectualization and modernization of languages like Tshivenda to be used in science, technology, and higher education.</p> <p>Language-in-Education Policy (1997): Issued by the Department of Basic Education: Encourages mother-tongue instruction in the early years of schooling. Supports the development of Tshivenda textbooks, teaching materials, and teacher training. Learners</p>

	<p>have the right to be educated in Tshivenda where there is demand and sufficient resources.</p> <p>Zimbabwe: Language Policy Draft (Ministry of Primary and Secondary Education, 2010ff): Proposes using local languages in early childhood education (ECD). Encourages curriculum development in all officially recognized languages, including Venda, though implementation varies widely. Limited use (some pilot projects)</p>
S. Note on the pluricentricity:	<p>The pluricentricity of Venda is well developed and fulfils all criteria. However there is a strong asymmetrie between the better equipped and promoted Venda in South Africa versus Zambia, where the situation is rather dire.</p>
T. Literature (selection):	<p>Poulos, George (1990). A Linguistic Analysis of Venda. Via Afrika. Tshivenda Grammar Guide by Zach Gershkoff, US Peace Corps (2012).</p> <p>Madadzhe R.N., Mashamba, T. (2004): The Role of Language in the Construction of Identity: The Case of Tshivenda. Journal: Southern African Linguistics and Applied Language Studies, Vol. 22, No. 1–2, pp. 101–107. Publisher: Taylor & Francis. DOI: 10.2989/16073610409486376</p> <p>Naudé, Piet & Mulaudzi, M. P.: Language Policy, Language Planning and Attitudes in South Africa (2004): Tshivenda as a Case Study. Journal for Language Teaching, Vol. 38, No. 2. Publisher: South African Association for Language Teaching (SAALT)</p> <p>Marashe, J., Pfukwa, C. & Mangena, T. (2009). Language Endangerment and Language Revitalisation: The Case of Tshivenda in Zimbabwe. Journal: NAWA Journal of Language and Communication, Vol. 3, No. 1, pp. 37–46</p>