

## Metadata

<b>Version</b>	1
<b>Created by</b>	Translators Without Borders and University College London
<b>Uploaded on</b>	Feb-20
<b>Notes and caveats</b>	<p>All data is drawn from government census results and is subject to any associated limitations or distortions present in the source data.</p> <p>As a data protection measure, the language dataset provider, IPUMS, has merged various second administrative units in order to provide spatio-temporally harmonized units that meet their 20,000 population threshold required for public dissemination of the data. This has resulted in combined figures for different zones in the dataset. Attempts to disaggregate these locations would likely result in decreased data accuracy.</p> <p>Language data is drawn from the 2001 census, and so will be less accurate due to age and minor area boundary changes.</p> <p>At time of census collection Moldavian was treated as a separate language. It is now commonly recognised as the same language as Romanian and is treated as such here.</p> <p>The Romani language in Ukraine comprises three sub-languages - Balkan Romani in the east, Carpathian Romani in the west, and Vlax Romani in the south and west.</p> <p>The Karaim language was also collected in the census, but fell below 0.1% in all instances and so is not included in this dataset.</p> <p>The methodology by which literacy and ability to speak English were measured during the census is not known.</p> <p>Language shares below 0.1% of an area population have been discarded. As a result, languages may not be present at higher administrative levels.</p> <p>All decimal values have been rounded to a maximum of 3 decimal places. As a result, language shares may not total 100%.</p> <p>Empty values represent non-existent data, and should not be treated as zero values.</p>
<b>Copyright and terms of use</b>	<p>© Translators without Borders 2020</p> <p>Data is available under an Attribution-NonCommercial-ShareAlike 4.0 International license (CC BY-NC-SA 4.0)</p> <p>You are free to share and adapt the data subject to requirements for attribution and non-commercial use. Any derivative work must be distributed under the same license as the original. Full terms at <a href="https://creativecommons.org/licenses/by-nc-sa/4.0/">https://creativecommons.org/licenses/by-nc-sa/4.0/</a>.</p>

<u>Source name</u>	<u>Last accessed</u>	<u>Source URL</u>
2001 population census data, St	25-Mar-22	<a href="http://database.ukrcensus.gov.ua/MULT/Database/Census/databasetree_en.asp">http://database.ukrcensus.gov.ua/MULT/Database/Census/databasetree_en.asp</a>
IPUMS microdata for 2001 popul	25-Mar-22	<a href="https://international.ipums.org/international/">https://international.ipums.org/international/</a>
Open Language Archives Commu	25-Mar-22	<a href="http://www.language-archives.org/index.html">http://www.language-archives.org/index.html</a>

<u>Field heading</u>	<u>HXL code</u>	<u>Description</u>	<u>Sources</u>
admin2_name	#adm2+name	Name of administrative unit	IPUMS microdata for 2001 population census
admin2_pcode	#adm2+code	Unique geographic identification code	IPUMS microdata for 2001 population census
admin1_name	#adm1+name	Name of administrative unit	IPUMS microdata for 2001 population census
admin1_pcode	#adm1+code	Unique geographic identification code	IPUMS microdata for 2001 population census
admin0_name	#country+name	Name of administrative unit	IPUMS microdata for 2001 population census
admin0_pcode	#country+code	Unique geographic identification code	IPUMS microdata for 2001 population census
number_of_named_languages	#indicator+lang+num	Number of named (non-Other) languages reported in area (excluding those v	Calculated field
main_language	#indicator+lang+mair	Name of reported language with highest attributed population in area	Calculated field
main_language_share	#indicator+lang+mair	Percentage share of population in area attributed to main language	Calculated field
..	..	..	..
Language fields	#indicator+lang+pct+	Percentage of population who speak this language as a first or second langu	IPUMS microdata for 2001 population census
Other language	#indicator+lang+pct+	Percentage of population who reportedly speak a non-named language	IPUMS microdata for 2001 population census
Language fields L1	#indicator+first+lang+	Percentage of population who speak this language as a first language	IPUMS microdata for 2001 population census
Language fields L2	#indicator+second+la	Percentage of population who speak this language as a second language	IPUMS microdata for 2001 population census
..	..	..	..

speaks_eng_%	#indicator+lang+pct+ Percentage of population who can speak English	IPUMS microdata for 2001 population census
illiteracy_%	#population+total+pct+ Percentage of population aged 15 and over who are illiterate	IPUMS microdata for 2001 population census
pop_total	#population+total Number of total population	IPUMS microdata for 2001 population census
pop_male	#population+m Number of male population	IPUMS microdata for 2001 population census
pop_female	#population+f Number of female population	IPUMS microdata for 2001 population census
literacy_all	#population+total+pct+ Percentage of population aged 12 and over who are literate	IPUMS microdata for 2001 population census
literacy_male	#population+m+pct+ Percentage of male population aged 12 and over who are literate	IPUMS microdata for 2001 population census
literacy_female	#population+f+pct+ Percentage of female population aged 12 and over who are literate	IPUMS microdata for 2001 population census
data_confidence	#meta+confidence Assessment of data quality (high/medium/low)	-
notes	#meta+text Notes and clarifications on data	-

#### **Data confidence descriptions** **Examples**

High	Data from well regarded government census or other specialist organisation.
Medium	Data from quality source but without clear methodology, or inverse.
Low	Data known to be erroneous, or based on non-statistical or imprecise data.
Factors increasing confidence	Data is corroborated. Data can be directly applied to administrative areas without aggregation or manipulation.
Factors decreasing confidence	Data is disputed. Data requires transformation that may affect accuracy, for example combining admin 2 data to form admin 1 data.

<b><u>Language</u></b>	<b><u>ISO 639-3 code</u></b>	<b><u>Alternative names and spellings</u></b>
Armenian	hyw	
Azeri	azj	Azerbaijani Belarusan, Belorussian, Bielorussian, Byelorussian, Central Belarusan, Grodnen-Baranovich, Northeast Belarusan, Polots, Slutska-Mazyrski, Slutsko-Mozyr, Southwest Belarusan, Viteb-Mogilev, White Russian, White Ruthenian
Belarusian	bel	
Bulgarian	bul	Balgarski, Bogomil, Palitiani, Palityan, Pomak, Pomakci, Pomakika Central Crimean, Crimean, Crimean Nogai, Crimean Turkish, Northern Crimean, Southern Crimean, Steppe Crimean
Crimean Tatar	crh	
English	eng	
Gagauz	gag	Bulgar Gagauz, Bulgar Gagauzi, Gagauzi, Maritime Gagauz, Maritime Gagauzi
Georgian	kat	
German	deu	Deutsch, Standard German, Kärntnerisch, Tedesco, Transylvanian Aspromonte, Cargese, Crimeo-Rumeic, Cypriot Greek, Dimotiki, Ellinika, Graecae, Grec, Greco, Greek, Griko, Katharevousa, Mariupol Greek, Neo-Hellenic, Romaic, Salento, Saracatsan, Tavro-Rumeic
Greek	ell	Arabized Hebrew, Europeanized Hebrew, General Israeli, Israeli, Ivrit, Oriental Hebrew, Standard Hebrew, Yemenite Hebrew
Hebrew	heb	Central Transdanubian, Csángó, Magyar, Mezőség, Northeastern Hungarian, Oberwart, Palóc, Southern Great Plains, Southern Transdanubian, Székely, Tisza-Körös, Western Transdanubian
Hungarian	hun	
Karaim	kdr	Galits, Halych, Karaite, Trakai, Trakay, Turkic Karaite

Polish	pol	Polnisch, Polski, Upper Silesian
Romani	rom	
		Arlija, Beás, Cigány, Dzambazi, East Bulgarian Romani, Erli, Greek Romani, Gypsy, Ironworker Romani, Paspatian, Roma, Balkan Romani, Tinnars
Romani, Balkan	rmn	Romani, Ursári, Usari
		Bashaldo, East Slovakian Romani, Galician, Hungarian-Slovak Romani, Moravian Romani, Carpathian Romani, Romungro, Selice Romani,
Romani, Carpathian	rmc	Transylvanian, West Slovakian Romani
		Arli, Bisa, Central Vlax Romani, Churari, Churarícko, Coppersmith, Danubian, Eastern Vlax Romani, Ghagar, Greco, Grekurja, Gurbet, Gypsy, Kaldarári, Kalderash, Kelderashícko, Lovari, Lovarícko, Machvanmcko, Machvano, Machwaya, North Albanian, Rom, Romanés, Romanese, Vlax
		Romani, Romenes, Sedentary Bulgarian, Sedentary Romanian, Serbo-Bosnian, Sievemakers, Sinto, South Albanian, Southern Vlax, Southern Vlax
Romani, Vlax	rmy	Romani, Tsigane, Tsigene, Tsingani, Ukraine-Moldavia, Vlax, Vlax Romany, Zagundzi, Zigenare
		Banat, Bayash, Boyash Romanian, Chrishana, Daco-Romanian, Daco-Rumanian, Maramuresh, Moldavan, Moldavian, Moldovan, Muntean, Muntenian, Oltean, Oltenia-Lesser Wallachia, Roumanian, Rumanian,
Romanian	ron	Transylvanian, Walachian
Russian	rus	Eluosi, North Russian, Olossu, Russ, Russit, Russki, South Russian
Slovak	slk	Slovakian, Slovencina
Tatar	tat	Tatarian
Ukrainian	ukr	East Ukrainian, Northwest Ukrainian, Southwest Ukrainian