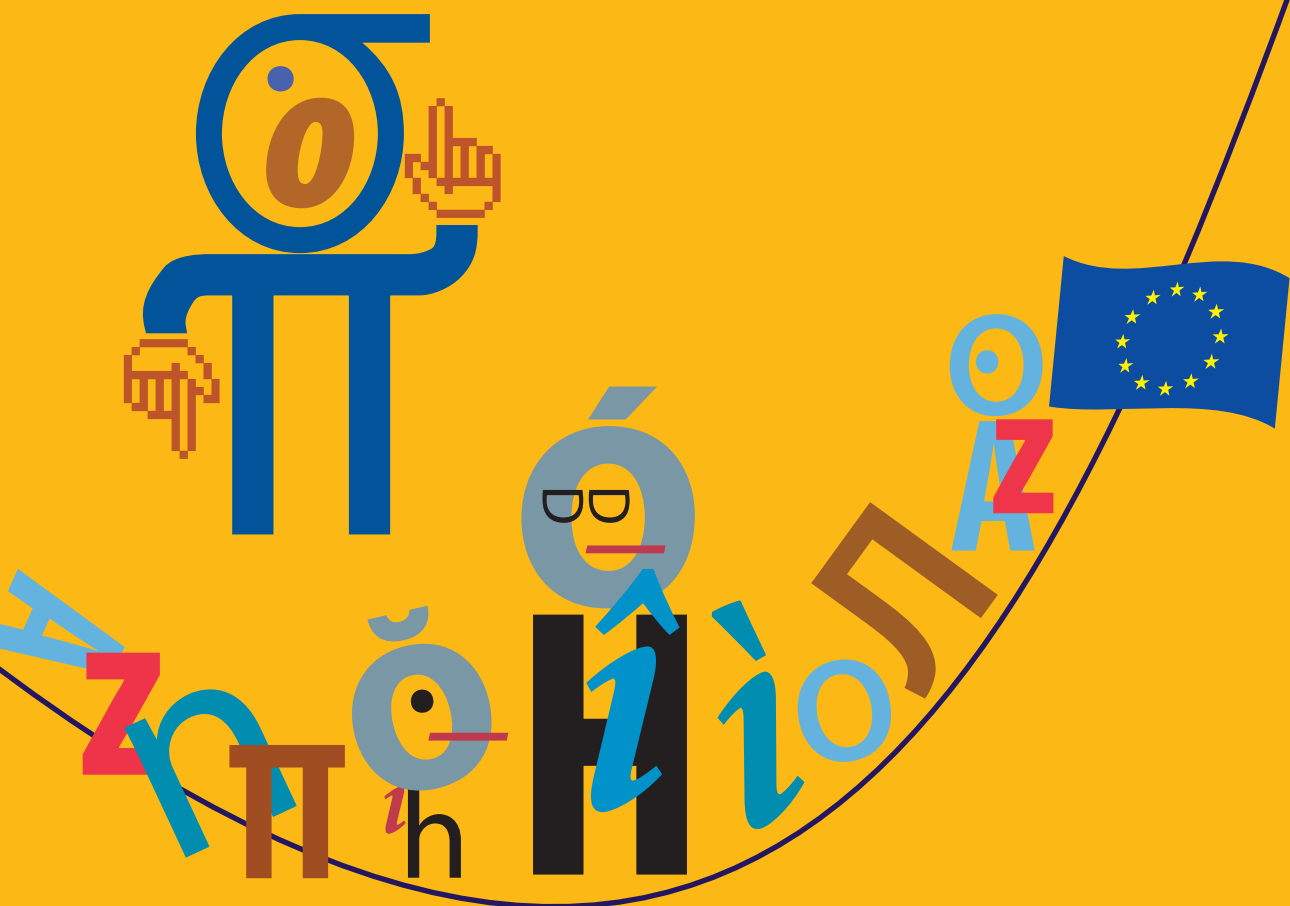


# A FIELD GUIDE TO THE MAIN LANGUAGES OF EUROPE

— Spot that language —  
and tell it apart from others



Editor and compiler: **Bill Fraser (DGT/B/2)**  
Contributors: **DGT language departments and units**  
Graphics and distribution: **Communication and Information Unit (DGT.02)**

**Third Edition, July 2008**  
**© European Communities, 2008**  
**Reproduction is authorised if the source is acknowledged**

## Languages covered in this Guide

Albanian, Armenian, Azeri,  
Basque, Belarusian, Bosnian, Breton,  
Bulgarian,  
Castilian Spanish, Catalan, Croatian, Czech,  
Danish, Dutch,  
English, Estonian,  
Faeroese, Finnish, French,  
Galician, Georgian, German, Greek,  
Hungarian,  
Icelandic, Irish, Italian,  
Latvian, Lëtzebuergesch, Lithuanian,  
Macedonian, Maltese, Moldovan,  
Norwegian,  
Polish, Portuguese,  
Romani, Romanian, Russian,  
Sámi, Scots Gaelic, Serbian, Slovakian, Slovenian, Swedish,  
Turkish,  
Ukrainian,  
Welsh,  
Yiddish



## Introduction

This guide has three purposes: it is a summary by language family of the main languages of the European Union and its European neighbours, it gives practical hints on how to recognise these languages and distinguish between them, and it is meant to encourage you to explore these languages further.

We have extended "Europe" as far as the Caucasus in order to include mentions of *Armenian*, *Georgian* and *Azeri*. However, the local languages of EU overseas territories are not covered\* nor are the so-called community languages (with a small 'c'), of which the main ones are Arabic, Bengali, various forms of Chinese, Gujarati, Indonesian, Vietnamese. We have also decided to leave out regional languages that have little chance of arriving at the European Commission for translation, such as Frisian or Corsican, while keeping in those that do come to us occasionally, like Galician and Welsh. This is a purely practical decision and no discrimination is intended.

**\* There are about 50 languages in the French overseas departments and territories alone.**

### How to use this guide

The guide is designed for those who do not know **any** of the languages in a given family group and it therefore concentrates on just the one or two essential telltales which such a person will need to tell the different languages apart. It does not attempt to list exhaustively all the distinguishing marks of each language.

*For example*, if you can already read **Danish** you are likely to spot immediately the 'unusual-looking' vocabulary, spellings and syntax which give away a text as **Norwegian**, even though you do not 'know' Norwegian. However, someone who knows none of the Scandinavian languages at all will have extreme trouble separating Danish from Norwegian at sight.

Apart from the difficulty of distinguishing e.g. Danish and Norwegian in language terms, another problem which can arise in the Directorate-General for Translation is the misidentification of a text because of the Member State from which it comes. For instance, documents in Swedish concerning the Åland Islands are often marked erroneously as Finnish, while texts in Galician can be marked as Spanish.

If you have your text in electronic form, you can try the following website, which provides links to a number of automatic language identifiers:  
<http://www.yourdictionary.com/morph.html#guessers>

## Introduction

### What next?

General websites worth investigating are –

- for a comprehensive survey of world languages: <http://www.omniglot.com/writing/index.htm>
- for the languages of Europe and where they are used: [http://www.bbc.co.uk/languages/european\\_languages/index.shtml](http://www.bbc.co.uk/languages/european_languages/index.shtml)
- for characters needed to write specific languages: <http://www.eki.ee/letter/>

For more websites, to show what individual languages look like, see the Appendix.

If you want to explore some of these languages further, you could investigate the phrasebooks for travellers produced by Lonely Planet (under the slogan "Don't just stand there, say something!"), which cover a large number of the languages dealt with here. Details at: [http://shop.lonelyplanet.com/Primary/Product/Language\\_Guides/Europe.2](http://shop.lonelyplanet.com/Primary/Product/Language_Guides/Europe.2)

The pace of political change is continually affecting the status of Europe's languages. Montenegro has recently become independent but the designation of its official language(s) remains controversial, while Kosovo's declaration of independence itself is still controversial along with its new constitution naming Albanian and Serbian as official languages. Irish was already an EU Treaty language from 1973 and is now being used for limited publication of EU law since 2007. Spain is so far the only Member State to have given enhanced status at EU level to three regional languages (Basque, Catalan and Galician) under a decision taken by the Council of the European Union at the same time.

*Bill Fraser, European Commission, June 2008*

## Table of contents

OVERVIEW	Language-Family Finder	7
SECTION I	Slavic Languages (with Latin script)	9
SECTION II	Romance Languages	11
SECTION III	Germanic Languages	13
SECTION IV	Baltic Languages	15
SECTION V	Celtic Languages	15
SECTION VI	Albanian	16
SECTION VII	Finno-Ugrian Languages	16
SECTION VIII	Semitic: Maltese	17
SECTION IX	Turkic Languages	18
SECTION X	Slavic Languages (with Cyrillic script)	19
SECTION XI	Greek	20
SECTION XII	Armenian, Georgian, Yiddish	20
SECTION XIII	Indic, Basque	21
APPENDIX	Text samples online	23





# Overview

## Language-Family Finder

Use telltale characters in your text to identify the group of languages and refer to the relevant section

Latin script? ↓	or another script? ⇒	⇒	Cyrillic ↓	or ⇒	Other ↓
↓			Ѡ, Ѧ, Ѣ, Ѥ, Ѧ		α, á, β, δ, η, φ, ρ
đ, č, š, ž ⇒	<b>South Slavic</b> Bosnian Croatian Slovenian (Serbian) <i>See section I</i>		<b>South Slavic</b> Bulgarian Macedonian Serbian <i>See section X</i>		<b>Greek</b> (Indo-European) <i>See section XI</i>
ą, ł, ń, Ź ⇒	<b>West Slavic</b> Czech Polish Slovakian <i>See section I</i>		Я, Д, Э, Ы, Ф Б, Г, Ы, І ↓ <b>East Slavic</b> Belarusian Russian Ukrainian <i>See section X</i>		Ամբողջ երկիրը ↓ <b>Armenian</b> or Hayeren (Indo-European) <i>See section XII</i>
â, ê, ç, ã, õ, ñ, nh ⇒	<b>Ibero-Romance</b> Spanish Catalan Galician Portuguese <i>See section II</i>	î, ç, ş, è, à, ò ⇒	<b>Other Romance</b> French Italian Romanian Moldovan  <i>See section II</i>		ყოველი საიდუმლოა ↓ <b>Georgian</b> (non-Indo-European) <i>See section XII</i>
æ, å, ä, ø, ð, þ ⇒	<b>North Germanic</b> Danish Faeroese Icelandic Norwegian Swedish <i>See section III</i>	ä, ö, ü, è, ee, éé ⇒	<b>West Germanic</b> Dutch English German Lëtze- buergesch <i>See section III</i>		מצפוניות ורבת השראה ↓ <b>Hebrew</b> script, often used for writing <b>Yiddish</b> (Indo-European) <i>See section XII</i>
ā, ū, ā, è, ķ, ņ ⇒	<b>Baltic</b> Latvian Lithuanian <i>See section IV</i>				

.../

## Overview

... Language-Family Finder

ò, í, bh, mh, c'h, ÿ ⇒	<b>Celtic</b> Breton Irish Scots Gaelic Welsh <i>See section V</i>	ë, xh, shp, shq ⇒	<b>Other Indo-European</b>  Albanian <i>See section VI</i>		
ä, õ, ö, ú, đ, á ⇒	<b>Finno-Ugric</b> Estonian Finnish Hungarian Sámi <i>See section VII</i>		ǵ, ħ, ź ↓ <b>Semitic</b> Maltese <i>See section VIII</i>		
ç, ş, ğ, ü, ə ⇒	<b>Turkic</b> Azeri Turkish <i>See section IX</i>				
â, î⇒	<b>Indic</b> Romani <i>See section XIII</i>		il, in, ts, tx, tz ↓ Basque <b>Family not clarified</b> <i>See section XIII</i>		

## Section I – Slavic Languages (with Latin script)

### South Slavic

#### *Bosnian/Croatian /Serbian/ Montenegrin/Slovenian*

Bosnian, Croatian and Serbian used to be known collectively as Serbo-Croat but are differentiating all the time since the break-up of former Yugoslavia.

While **Croatian** is written almost exclusively in the Latin alphabet, **Serbian** is just as likely to be written in Cyrillic script (see section X) as in Latin characters. When written in Latin script, Serbian uses the same accented characters as Croatian, but is distinguishable from Croatian by certain grammatical and lexical features.

**NOTE:** Possible confusion of *Serbian* with *Croatian* if the former is written in Latin script.

**Bosnian** used at one time to be written in Arabic script but now appears in either Latin or Cyrillic script.

A dialect of Serbian may be set to become the national language of Montenegro (to be called "Montenegrin"?) following its independence in June 2006.

For more details on the differences between four of these five languages, see [http://en.wikipedia.org/wiki/Differences\\_in\\_official\\_languages\\_in\\_Serbia%2C\\_Croatia\\_and\\_Bosnia](http://en.wikipedia.org/wiki/Differences_in_official_languages_in_Serbia%2C_Croatia_and_Bosnia).

In general, the South Slavic languages have fewer accents and diacritics than the West Slavic languages (Czech, Polish, Slovakian).

**Croatian** has a characteristic barred d: **đ / Đ**, as well as **ć, č, š, ž**.

Another notable feature is **h** in combination with other consonants, as in **hl, hr, hv** (e.g. the name of the country: Hrvatska).

**NOTE:** Documents in *Slovenian* and *Slovakian* are sometimes confused because the Slovak name for Slovakia is 'Slovensko', reminiscent of 'Slovenia'.

**Bosnian** also uses the barred **đ** and **ć, č, š, ž**.

#### **Slovenian**

Slovenian has a *háček* (upside-down circumflex) on **č, š** and **ž** only, with absence of **ć** and **đ**.

## Section I – Slavic Languages (with Latin script)

### West Slavic

*Czech/Slovakian/Polish*

**Czech and Slovakian:** Both make use of many accents and diacritics (as compared e.g. to Slovene or Croatian). The upside-down circumflex, called *háček* in Czech and *mäkčeň* in Slovakian, is perhaps the most noticeable feature of Czech and Slovakian. On *t* and *d* the *háček* is printed as an apostrophe very close to the letter in lower case, but as an upside-down circumflex when in upper case.

Because of different spelling rules in the two languages, a Czech text will typically have more accents than a Slovakian one. In both languages it is common to find words containing many accents (*váčší, příště*) and unusual-looking words with no apparent vowels (*prst, dlh, stíp, trň, tíct, krb, štvrt*).

**Czech:** the main distinguishing character is **ů**, unique to Czech.

**Slovakian** also has **Ľ** with a *mäkčeň*, printed as an apostrophe in upper and lower case. In **Slovakian**, but not **Czech**, syllabic **r** and **l** can also be long and have an acute accent.

Characters common to Czech and Slovakian: **č, d' / Ď, ň, š, ě / Ě, ý, ž**

Characters used in Czech, but not in Slovakian: **ě, ř, ů**

Characters used in Slovakian, but not in Czech: **ä, ĺ, ľ / Ľ, ô, í**

**Polish** has :

- no *háček* at all, but characteristic digraphs and consonant clusters instead: for example, *cz, dz, rz* (common ending for masculine nouns), *sz, szcz*
- neither 'ů' nor 'ä'; *but* note that
- the crossed or barred 'ł' (capital or small) is unique to Polish;
- Polish uses **w** where other Slavic languages with Latin script use **v**.

Accented letters and common combinations used in Polish:

**ą, ć, ę, ł, ń, ó, ś, ść, ź, ż, źdź**

**NOTE:** *Polish* is also an important regional language in Lithuania.

## Section II – Romance Languages

### Ibero-romance

*Castilian Spanish/Catalan/Galician/Portuguese*

**Castilian:** has **y** as a single word - frequent use of final **-d**

**Catalan:** has **i** as a single word - frequent use of final **-t**

**Castilian** and **Catalan:** both use double **ll** to represent the sound [λ].

However, in **Catalan** a very distinctive mid-height stop is used to identify a double 'l' pronounced as two 'l's (e.g. *excel·lents*).

**Castilian** has a tilde on **ñ** but *not* on vowels, in contrast to **Portuguese**.

**Catalan** has **ny** where Castilian has **ñ**.

Both Castilian and Catalan make use of double i...! and ¿....? for emphasis and questions, though this is optional in Catalan.

**Catalan** has grave accents, and makes use of dieresis on **i** (e.g. *reïna = resin*) and on **u** following **g** and **q** (e.g. *eloqüència*) to show that the vowel is pronounced separately.

**NOTE:** *Brazilian Portuguese* differs slightly in spelling and grammar from European Portuguese.

**Portuguese:** common use of final **m** (unlike Castilian) - circumflex accents **ê, â, ô** - no double **ll** - tilde on the vowels **ã** and **õ** - the equivalent sound to **ñ** in Spanish is written **nh** in Portuguese while **Galician** has both **ñ** and **nh**.

**Galician (or Galego):** looks at first sight like Portuguese, but is a Spanish regional language - frequent use of **'x'** for a soft 'g' sound - use of **ll** (unlike Portuguese) - use of **ñ** alongside **nh** (unlike Portuguese).

**NOTE:** Documents in *Galego* are sometimes misidentified as Spanish because of the Member State of origin.

## Section II – Romance Languages

### Other Romance languages

#### *French/Italian*

**French :** The most obvious feature of French texts is the large number of apostrophes, marking the elision of vowels. The commonest are **l'**, **d'** and **s'** at the start of a word.

Accents and diacritics used are: **é, è, à, ç, ë** - and note that **à** can be a single word.

**NOTE:** In Europe, *French* is an official language in France, Belgium, Luxembourg, Monaco and Switzerland, and for certain purposes also in Jersey/Channel Islands.

**NOTE:** *Italian* is also an official language of Switzerland.

**Italian:** Almost all Italian words end in a vowel. Most typical characters: **ò** at the end of words (not found in French at all) and **è** as a single word (also not found in French). Another noteworthy feature is initial capitals for the forms of the polite 'you' even when in the middle of a sentence, e.g. Lei, Loro.

#### *Romanian/Moldovan*

**Romanian** has no **ç** (unlike French) but does put a cedilla under **ș** and **ț**. Also frequent circumflex on **î**.

Accented characters used in Romanian: **â, ă, î, ș, ț**.

**Moldovan:** Romanian is also the official language of Moldova, although its official name there is "Moldovan". It uses Romanian spelling.

Russian is also a major language in Moldova. In the self-declared republic of Transdnistria, Moldovan is still written in the Cyrillic script, as in Soviet times and before that in the Austro-Hungarian Empire, and has official status alongside Russian and Ukrainian.

## Section III – Germanic Languages

### North Germanic (Scandinavian)

*Danish/Norwegian/Swedish*

**Danish, Norwegian** and **Swedish** all use **å**.

**Danish** and **Norwegian**: both have **æ** and **ø**, where **Swedish** uses **ä** and **ö** instead.

**Danish** and **Norwegian** use **og** (English 'and'), where **Swedish** has **och**.

**Danish** uses **af** (=English 'from', 'of'), where **Norwegian** and **Swedish** have **av**.

**Danish** uses **at** (=English infinitive 'to'), where **Norwegian** has **å** and **Swedish** **att**.

Under the Åland Autonomy Act, **Swedish** is the sole official language in Finland's Åland Islands.

*Icelandic/Faeroese*

Both languages use **ð**, and are the only modern languages that do.

Both languages have an acute accent on vowels, that is **á, é, í, ó, ú**, and **ý**.

**Icelandic** has **þ** and **æ**, Faeroese does not use either of these.

**Faeroese** has **ø**, Icelandic uses **ö** instead. No other Icelandic vowel takes an umlaut.

### West Germanic

*English/German/Dutch/Lëtzebuergesch*

**NOTE:** *English* in Europe has official status in the Isle of Man, the Channel Islands and Gibraltar, Ireland (alongside Irish) and Malta (alongside Maltese).

**English:** The most noteworthy feature of English is the many digraph vowels such as **ee, ea, ie, ei, ay, eu, ey**.

English spelling has two main variants, U.S. and British, but given the status of English as a first, second or third language for large numbers of users around the globe, spelling is becoming very fluid. British English spelling is preferred for EU texts.

More generally, it is becoming essential to preface 'English' with a qualifier like 'American', 'Australian', 'British', 'Indian', 'International', etc. and to speak of 'Englishes' in the plural.

**NOTE:** *Norwegian* exists in two variants: **bokmål** and **nynorsk**, but BM is used by at least 80% of the population and to almost 100% in the media.

## Section III – Germanic Languages

### West Germanic

(continued)

**German:** The most striking feature of German is initial capital letters within a sentence to mark all nouns.

German has umlaut accents on a, o and u only (**ä, ö, ü**). However, the umlauts may be replaced by 'ae' (NB *not* the same as Scandinavian ligature æ), 'oe' and 'ue', e.g. in proper names, and in e-mails and other cases where the character set is restricted – but most noteworthy is the character **ß**, unique to German (though no longer used in Swiss German), representing a sibilant 's' sound.

German makes frequent use of abbreviations in running text: common examples are *bzw.* and *usw.*

**NOTE:** *German* is an official language in Germany, Austria, Switzerland, Luxembourg and Liechtenstein, with regional status also in Belgium, Denmark and Italy.

**NOTE:** In Europe, Dutch is an official language in the Netherlands and Belgium. Flemish (Vlaams) is a dialect of Dutch spoken in Belgium and northern France.

**Dutch:** A notable feature is the combination **ij**, including at the start of words. At the start of a sentence the double character is uppercased **IJ**. Note the use of **ë** in some plurals. Dutch makes frequent use of double vowels **aa, ee, oo** – also **éé** and **óó** with accents for emphasis. Nouns are *not* capitalised, unlike in German and there are no umlaut accents on a, o, u.

Dutch, like German, also makes use of many abbreviations in running text: examples are *dmv*, *dwz*, *enz*, *tgw*.

**Lëtzebuergesch** (*also* Luxembourgish) is recognised as a national language of Luxembourg alongside French and German for use in the administration and judiciary – though all legislation is published in French only. The most noteworthy features of Lëtzebuergesch are very frequent use of **ë** (almost non-existent in German and quite rare in Dutch) and also of single **é** (quite rare in German and Dutch, where it is used almost exclusively in foreign loan words).



## Section IV – Baltic Languages

### *Lithuanian/Latvian*

These languages are the sole survivors of a separate family of Indo-European origin. They are not related to Estonian (see Section VII).

**Lithuanian:** all the vowels except 'o' can also take a right-bending cedilla below: **ą ę į ų**.

Only Lithuanian has **ė**. There is a high frequency of endings in '-os'.

**Latvian:** the consonants 'k', 'l' and 'n' can take a left-bending cedilla: **ķ ļ ņ** but never the vowels. When **g** takes the cedilla, it is placed above the letter: **ģ**.

There is a high frequency of endings in '-as'.

Both languages have **ū**, but **ā**, **ē** and **ī** are found only in Latvian.

## Section V – Celtic Languages

The most striking feature of the Celtic languages is the mutation of the initial sounds of words following certain prepositions and in oblique cases.

### **Q-Celtic: Irish and Scots Gaelic**

In Irish and Scots Gaelic, the mutation is marked by adding a letter, so that characteristic initial consonant clusters in **Irish** are: gc-, bhf-, mb-, ng-, bp-, and in **Scots Gaelic** ch-, fh-, mh-, bh-.

**Irish** has acute accents only: **mé, tú, sé, sí** etc., whereas **Scots Gaelic** traditionally had a 'grave' accent as well: **òg** (Irish **óg**), **sgìth** (Irish **scíth**). However, a spelling reform for **Scots Gaelic** in 1981 recommended acute accents only and practice is no longer uniform. Any 'grave' accent betrays **Scots Gaelic**, in other words.

The most striking feature of both languages, but also enabling a distinction to be readily made between them, is the appearance of capital letters within words, which in **Irish** gives e.g. *Gaeilge na hÉireann* ("Irish Gaelic") with no separator, while **Scots Gaelic** inserts a hyphen: *Gàidhlig na h-Alba* ("Scots Gaelic").

### **P-Celtic: Welsh and Breton**

In **Welsh** and **Breton**, the initial mutation is less evident because often one letter replaces another rather than a letter being added. In **Welsh**, for example, **c** becomes **g**, **ci** becomes **gi**, **m** becomes **f**.

**Welsh** has a distinctive double **ll**, including at the start of words, and makes wide use of **y** and **ŷ**, **w** and **ŵ** for vowel sounds.

**Breton** represents the same sound as the Welsh hard "c" with "k". It also has a distinctive tilde on **ñ** to indicate nasalisation. One very characteristic final cluster is -**añv**. There is also a typical **c'h**, when **ki** (cf **ci** in Welsh) mutates to e.g. **ur c'hi**. Another frequent pattern is final **-ou** or **-où**. The cluster **"zh"** is also common.

## Section VI – Albanian

**Albanian** is an Indo-European language which forms its own branch in the Indo-European family and has no close living relatives. It has been written in a standardised Latin script since 1909. There are very few distinguishing characters, the main one being **ë**. However, a striking feature are the consonant clusters such as **shq** (e.g. the name of the language itself: Shqip), **shp**, **sht**.

Standard Albanian is a conflation of the two major dialects, Tosk and Gheg.

**NOTE:**  
Albanian is also used widely in Kosovo.

## Section VII – Finno-Ugric Languages

As the complex name suggests, this is a very disparate group of languages with only distant ties between most of them. Several minority languages in Russia are also members of this family.

### *Estonian/Finnish*

Estonian and Finnish can look very similar at first sight to the uninitiated.

**Estonian:** **ü** and **üü**, corresponding to **Finnish y** and **yy**.

Both languages have **ä** and **ö** and frequent double vowels, but **õ** and **ü** occur only in Estonian.

Generally speaking, in contrast to Finnish, Estonian has more words ending in consonants (**-d**, the plural ending, and **-b**, the third person singular verb ending, are common) and consonant clusters (such as **-st**, **-lt** and **-ks**), whereas **Finnish** typically has no word-final consonant clusters and the only common single final consonants are **n** and **t** (and less frequently **s**, **r** and **l**).

**NOTE:** Hungarian has about 10 million speakers in Hungary itself and more than 3 million in Slovakia, Romania, Serbia and Ukraine.

**Hungarian:** The most striking feature of **Hungarian** is the lengthened double stroke above **ő** and **ű**. Hungarian also has both **ö** and **ü** as in German (by contrast, **German** has no lengthened strokes, and has capitalised nouns and umlaut **ä** as well). Hungarian can take an acute accent on all vowels: **á, é, í, ó, ú**. Typical consonant clusters are **sz, gy** and **ty**.

## Section VII – Finno-Ugric Languages

**Sámi** (also known as Lapp) is the language of a still largely nomadic people in Northern Europe, whose biggest population is found in Norway, followed by Sweden, Finland and Russia in declining order. The status of the Sámi language(s) has slowly become more important in the late 20<sup>th</sup> century.

Typical characters in Sámi: **č**, **š** and **á**, also a barred **đ**, which can be doubled: **đđ**.

Confusion is possible between Sámi and Scandinavian languages, especially Icelandic and Faeroese, because of the similarity of **đ** to **ð** and the fairly large number of Scandinavian loan words in Sámi.

**NOTE:** Sámi has given us the word "tundra".

## Section VIII – Semitic: *Maltese*

### *Maltese*

The crossed 'h' is the most obvious distinguishing mark: **ħ, Ħ, għ**.

Maltese also has **č**, **ġ** and **ż**, and uses many hyphens.

Maltese developed out of Arabic, overlaid with Romance family elements, in the Middle Ages. In 1921, Italian and English were declared the official languages of Malta. Maltese then replaced Italian in 1934. Today Maltese is used in most sectors of public life, including parliament, the church, the press and other media, and in general conversation. English is the preferred medium of instruction in schools, especially at the higher levels of the educational system.

## Section IX – Turkic Languages

**Turkish:** in addition to **ç**, Turkish also puts a cedilla under **ş** but not **t** (cf Romanian) - most characteristic letter: **ğ**. Turkish uses the letter **ı** both with and without a dot in upper and lower case (representing two separate vowel sounds), with occasional circumflex on **â, î** and **û**.

Characteristic letters of Turkish: **â, ç, ğ, ı/I, î/İ, î, ö, ş, û, ü**.

**NOTE:** Turkish is the 'other language' of our current Member State Cyprus, but not an EU official language.

**Azeri** (Azerbaijani), which is very similar to Turkish, is written with the same accented characters as Turkish, except without the circumflex and with the addition of a very characteristic schwa = ə (a back-to-front, upside-down e), which, for typographical reasons, may sometimes be replaced with **ä**. Azeri also uses **q** and **x**, which Turkish does not.

Turkish was written in Arabic script until the 1920s. Azeri was also written in Arabic script up to the 1920s, then Cyrillic was imposed from the 1930s until 1991. Azeri is also a minority language in Georgia and Iran.

Although the languages of the other Turkic-speaking former Soviet Republics (Turkmen, Kazakh, Uzbek and Kyrgyz) are currently in the process of converting from the Cyrillic to the Latin alphabet, when written in Cyrillic they could potentially be mistaken for Russian, even if there are some distinctive non-Russian characters.

See <http://www.omniglot.com/writing/cyrillic.htm> for details.

## Section X – Slavic languages (with Cyrillic script)

The Cyrillic script exists in printed and hand-written versions which differ considerably. A normalised version of hand-written Cyrillic is often used for printing 'italics', e.g. in Russian год = *год*. We consider here only the non-cursive printed versions.

### South Slavic

*Bulgarian/Macedonian/Serbian*

**Bulgarian** is not easy to distinguish from **Russian**, an East Slavic language.

**Bulgarian** does not have the letters **ё** or **э** (which are found in Russian).

The "hard sign" **ъ**, rather rare in Russian, is frequently used as a full vowel in Bulgarian, e.g. the second letter in the name of the country: България.

Frequent endings in Bulgarian are **-ът**, **-та**, **-то** and **-те** (post-positive articles).

**Macedonian** has several special characters distinguishing it from Bulgarian: **ѐ**, **ѕ**, **ј**, **љ**, **њ**, **ќ**, **џ**.

**Serbian**: has its characteristic **ђ** and **ћ**, but shares **ј**, **љ**, **њ** and **џ** with Macedonian. Note Serbian is not always written in Cyrillic, but can be found in Latin script too.

### East Slavic

*Russian/Belarusian/Ukrainian*

Distinguishing letters that do *not* occur in **Russian**:

**Belarusian**: **р**, **і**, **ў**

**Ukrainian**: **ґ**, **є**, **і**, **ї**

**NOTE:** The EU has substantial Russian-speaking minorities in Latvia - especially in Rīga - and Estonia.

Unique distinguishing characters: Only Belarusian has **ў** - only Ukrainian has **ї**.

Both also make use of an apostrophe, which Russian does not.

## Section XI – Greek

The **Greek script** has a rounded and cursive look (e.g.  $\alpha$ ,  $\beta$ ,  $\delta$ ,  $\eta$ ,  $\phi$ ) compared to **Cyrillic**, which is more straight-lined, up-and-down (e.g.  $\mathbf{\text{Я, В, Д, Ы, Ф}}$ ). Hand-written Cyrillic is a bit closer to Greek but not all that much.

Greek can have accents on some letters, e.g.  $\acute{\alpha}$ ,  $\grave{\eta}$ .

The Greek used in Cyprus is virtually identical to that used in Greece.

**NOTE:** Cyrillic was originally adapted from the Greek script in 10<sup>th</sup> century Macedonia.

In Greece, the military dictatorship (1967-74) tried to impose a purist form of literary Greek, *katharevousa*, over the form of Greek which up until then had come to be predominant, *dhimotiki*. The latter was re-established after the regime of the colonels and further simplified in 1981. However, *katharevousa* is still theoretically in use, especially in the Orthodox Church and some professional circles (e.g. some legal documents). In general, *dhimotiki* has fewer accents and diacritics than *katharevousa*, and while *dhimotiki* has only an acute accent, *katharevousa* has a grave and tilde as well.

## Section XII – Armenian, Georgian, Yiddish

These languages are grouped here solely because of their non-Latin scripts.

**Armenian** is an Indo-European language with remote links to the Balto-Slavic and Indo-Iranian language groups. It is also referred to as **Hayeren**.

*Sample:* Ամբողջ երկիրը

**Georgian** is a Caucasian language, i.e. belongs to a group or family not otherwise represented in this Field Guide.

*Sample:* ყოველი საიდუმლოა

**Yiddish** is essentially a Germanic language with much Slavic and Hebrew vocabulary. It classically uses Hebrew script but is also regularly written in Latin script.

*Sample:* מצפונית ורבת השראה

## Section XIII – Romani, Basque

**Romani** is Indic by origin, sometimes described as “Balkanised” Indo-Aryan, that is, consisting of mainly Aryan grammar with heavily Balkan vocabulary, especially Greek of the Byzantine period. It is the language of the **Roma** (we no longer say ‘gypsies’), who represent important minorities in Bulgaria, the Czech Republic, Slovakia, Hungary, Poland, Romania, Moldova and Russia. In most cases it is written with the script and spelling conventions of the host community, and has enjoyed a considerable boost in recognition since 1990.

**Basque (or Euskara)** is the language of the Basque Country (Euskadi), straddling the French-Spanish frontier. The origins of Basque are very obscure, with no apparent relationship to other European languages. It has full regional status alongside Castilian Spanish in Spain’s *País Vasco*, and also regional status in Navarre, but in France it is recognised only as a teaching medium in schools.

Basque can use **ñ** and **ll** as in Spanish but these sounds are more commonly written as **in** and **il**, otherwise typical digraphs are: **rr, ts, tx, tz**.

However, Spanish also has **rr** and Catalan has **tx**.





## Appendix - Text samples online

There is not enough space here to give textual examples of all the languages covered.

For the EU official languages, you can go to our own Language Portal [http://europa.eu.int/abc/european\\_countries/languages/index\\_en.htm](http://europa.eu.int/abc/european_countries/languages/index_en.htm) to see a text in each one and hear it being read.

We also offer you a wider range of typical websites where you can see “real life” samples of the chosen language at a glance. (Some of the sites are radio homepages, where with the right software you can listen to the language as well!)

In alphabetical order of the name of the language in English.

<b>Albanian</b>	<a href="http://www.bbc.co.uk/albanian/">http://www.bbc.co.uk/albanian/</a>
<b>Armenian</b>	<a href="http://hy.wikipedia.org/wiki/%D5%80%D5%A1%D5%B5%D5%A1%D5%BD%D5%BF%D5%A1%D5%B6">http://hy.wikipedia.org/wiki/%D5%80%D5%A1%D5%B5%D5%A1%D5%BD%D5%BF%D5%A1%D5%B6</a> <a href="http://www.omniglot.com/writing/armenian.htm">http://www.omniglot.com/writing/armenian.htm</a>
<b>Azeri</b>	<a href="http://www.bbc.co.uk/azeri/">http://www.bbc.co.uk/azeri/</a>
<b>Basque</b>	<a href="http://www.egaztea.com/web/">http://www.egaztea.com/web/</a>
<b>Belarusian</b>	<a href="http://www.svaboda.org/default_text.aspx">http://www.svaboda.org/default_text.aspx</a>
<b>Bosnian</b>	<a href="http://www.fbihvlada.gov.ba/bosanski/federacija/index.html">http://www.fbihvlada.gov.ba/bosanski/federacija/index.html</a>
<b>Breton</b>	<a href="http://www.breman.org/">http://www.breman.org/</a>
<b>Bulgarian</b>	<a href="http://www.bnr.bg/RadioBulgaria/Emission_Bulgarian/">http://www.bnr.bg/RadioBulgaria/Emission_Bulgarian/</a>
<b>Castilian Spanish</b>	<a href="http://www.rtve.es/rne/web/index.php">http://www.rtve.es/rne/web/index.php</a>
<b>Catalan</b>	<a href="http://www.catradio.com/pcatradio/crHome.jsp">http://www.catradio.com/pcatradio/crHome.jsp</a>
<b>Corsican</b>	<a href="http://adecec.net/vocenustrale/">http://adecec.net/vocenustrale/</a>
<b>Croatian</b>	<a href="http://www.hrt.hr/index.xml">http://www.hrt.hr/index.xml</a>
<b>Czech</b>	<a href="http://www.radio.cz/cz/">http://www.radio.cz/cz/</a>
<b>Danish</b>	<a href="http://www.dr.dk/">http://www.dr.dk/</a>
<b>Dutch</b>	<a href="http://www.wereldomroep.nl/">http://www.wereldomroep.nl/</a>
<b>English</b>	<a href="http://www.bbc.co.uk/worldservice/">http://www.bbc.co.uk/worldservice/</a>
<b>Estonian</b>	<a href="http://www.er.ee/index.php">http://www.er.ee/index.php</a>
<b>Faeroese</b>	<a href="http://www.framtak.com/fo/index.html">http://www.framtak.com/fo/index.html</a>
<b>Finnish</b>	<a href="http://www.yle.fi/radiosuomi/">http://www.yle.fi/radiosuomi/</a>
<b>French</b>	<a href="http://www.radiofrance.fr/">http://www.radiofrance.fr/</a>
<b>Galician</b>	<a href="http://www.crtvg.es/cqi-bin/iniciocrtvg.asp?idioma=galego">http://www.crtvg.es/cqi-bin/iniciocrtvg.asp?idioma=galego</a>
<b>Georgian</b>	<a href="http://www.tavisupleba.org/">http://www.tavisupleba.org/</a>

## Appendix - Text samples online

<b>German</b>	<a href="http://www.dw-world.de/dw/0,2142,265,00.html">http://www.dw-world.de/dw/0,2142,265,00.html</a>
<b>Greek</b>	<a href="http://www.hri.org/news/latest/greek.latest.html">http://www.hri.org/news/latest/greek.latest.html</a>
<b>Greenlandic</b>	<a href="http://knr.gl/nutaarsiassat/index.html">http://knr.gl/nutaarsiassat/index.html</a>
<b>Hungarian</b>	<a href="http://www.hullamvadasz.hu/">http://www.hullamvadasz.hu/</a>
<b>Icelandic</b>	<a href="http://www.utvarpsaga.is/">http://www.utvarpsaga.is/</a>
<b>Irish</b>	<a href="http://www.bnag.ie/default1.asp?lang=ga">http://www.bnag.ie/default1.asp?lang=ga</a>
<b>Italian</b>	<a href="http://www.radio.rai.it/">http://www.radio.rai.it/</a>
<b>Latvian</b>	<a href="http://www.radio.org.lv/">http://www.radio.org.lv/</a>
<b>Lithuanian</b>	<a href="http://www.lrt.lt/">http://www.lrt.lt/</a>
<b>Lëtzebuergesch</b>	<a href="http://lb.wikipedia.org/wiki/Haapts%C3%A4it">http://lb.wikipedia.org/wiki/Haapts%C3%A4it</a>
<b>Macedonian</b>	<a href="http://www.bbc.co.uk/macedonian/">http://www.bbc.co.uk/macedonian/</a>
<b>Maltese</b>	<a href="http://www.rtk.org.mt/">http://www.rtk.org.mt/</a>
<b>Moldovan</b>	<a href="http://www.protv.md/">http://www.protv.md/</a>
<b>Norwegian</b>	<a href="http://www.nrk.no/">http://www.nrk.no/</a>
<b>Polish</b>	<a href="http://www.gazeta.pl/0,0.html">http://www.gazeta.pl/0,0.html</a>
<b>Portuguese</b>	<a href="http://www.rdp.pt/">http://www.rdp.pt/</a>
<b>Romani</b>	<a href="http://www.romnet.hu/hirek/romaindex.html">http://www.romnet.hu/hirek/romaindex.html</a>
<b>Romanian</b>	<a href="http://www.rador.ro/">http://www.rador.ro/</a>
<b>Russian</b>	<a href="http://www.rambler.ru/">http://www.rambler.ru/</a>
<b>Sámi</b>	<a href="http://www.saamicouncil.net">http://www.saamicouncil.net</a>
<b>Scots Gaelic</b>	<a href="http://www.bbc.co.uk/scotland/alba/">http://www.bbc.co.uk/scotland/alba/</a>
<b>Serbian (Latin script)</b>	<a href="http://www.bbc.co.uk/serbian/">http://www.bbc.co.uk/serbian/</a>
<b>Serbian (Cyrillic script)</b>	<a href="http://www.srpska.com/cir/index.php">http://www.srpska.com/cir/index.php</a>
<b>Slovakian</b>	<a href="http://www.skonline.sk/">http://www.skonline.sk/</a>
<b>Slovenian</b>	<a href="http://www.rtvlo.si/">http://www.rtvlo.si/</a>
<b>Spanish</b> see <b>Castilian Spanish</b>	
<b>Swedish</b>	<a href="http://www.sr.se/">http://www.sr.se/</a>
<b>Turkish</b>	<a href="http://www.bbc.co.uk/turkish/">http://www.bbc.co.uk/turkish/</a>
<b>Ukrainian</b>	<a href="http://www.bbc.co.uk/ukrainian/domestic/">http://www.bbc.co.uk/ukrainian/domestic/</a>
<b>Welsh</b>	<a href="http://www.bbc.co.uk/cymru">http://www.bbc.co.uk/cymru</a>
<b>Yiddish</b>	<a href="http://yiddish.forward.com/">http://yiddish.forward.com/</a>



