A FIELD GUIDE
TO THE MAIN LANGUAGES
OF EUROPE

Spot that Language
- and how to tell them apart

by Bill Fraser and friends
** Second Edition 11/12/2006 **

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

A FIELD GUIDE TO THE MAIN LANGUAGES OF EUROPE
Spot that language and how to tell them apart

<u>A few words about the second edition</u>
Our thanks go to the colleagues who pointed out shortcomings in the first edition. All their comments have been incorporated. A note has been added about the language of Montenegro, which has become independent since the first edition. All web links have been reviewed. A few extra notes have been added. We welcome any and all feedback, so don't hesitate to add your knowledge to the store contained in this Field Guide.

<u>Intro</u>
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, Gujerati, Indonesian, Vietnamese. We have also decided to leave out regional languages that have little chance of arriving in DGT for translation, such as Frisian or Corsican, while keeping in those that do come to us occasionally, like Galician and Welsh.

<u>How to use this guide</u>
The guide is designed for those who do not know <u>any</u> 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.

> *<u>For example</u>, 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 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. This guide notes such possible confusions, introduced by the sign ►.

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

---

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

What next?

*If you want to explore some of these languages further, we can recommend the phrasebooks for travellers produced by Lonely Planet (under the slogan "Don't just stand there, say something!") as useful aids at an affordable price. This series includes: "**Europe Phrasebook**" (ISBN 11-86450-224-X), which covers Basque, Irish, Maltese and Turkish along with the more obvious languages, "**Scandinavian Phrasebook**" (ISBN 1-86450-225-8), which includes Faeroese and Icelandic, "**Central Europe Phrasebook**" (ISBN 1-86450-226-6), which includes Hungarian, Slovakian and Slovenian, "**Eastern Europe Phrasebook**" (ISBN 1-86450-227-4), which includes Albanian and the other Balkan languages, and "**Baltic Phrasebook**" (ISBN 1-86450-369-6), which covers Estonian along with the two 'Baltic-family' languages. All these phrasebooks provide basic elements of grammar and aids to pronunciation together with background information about the culture of the country and useful word lists such as months of the year. They cost around EUR 8 or EUR 9 each, depending on the number of pages. Some of the languages in these regional phrasebooks are also given a whole phrasebook on their own, e.g. Czech, Greek, Hungarian, Turkish.*
More at:
[*http://shop.lonelyplanet.com/series_products.cfm?seriesID=4&groupID=13&affil=lp*](http://shop.lonelyplanet.com/series_products.cfm?seriesID=4&groupID=13&affil=lpfp-msp)
[*fp-msp*](http://shop.lonelyplanet.com/series_products.cfm?seriesID=4&groupID=13&affil=lpfp-msp)

General websites worth investigating are –

for a comprehensive survey of world languages:
[*http://www.omniglot.com/writing/index.htm*](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/*](http://www.eki.ee/letter/)

For more websites, see **Annex I**.

Table of Contents

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 ⇓
⇓			ъ, ќ, љ, њ, џ ⇓		α, ά, β, δ, η, φ, ή ⇓
đ, č, š, ž ⇒	**South Slavic** Bosnian Croatian Slovenian (Serbian) *See section I*		**South Slavic** Bulgarian Macedonian Serbian *See section X*		**Greek** (Indo-European) *See section XI*
ą, ł , ř, ů, ž ⇒	**West Slavic** Czech Polish Slovakian *See section I*		я, д, э ,й, ф ы, г, ў, ї ⇓ **East Slavic** Belarusian Russian Ukrainian *See section X*		Ամբողջ երկիրը ⇓ **Armenian** *or* Hayeren (Indo-European) *See section XII*
â, ê, ç, ã, õ, ñ, nh ⇒	**Ibero-Romance** Spanish Catalan Galician Portuguese *See section II*	î, ç, ş, è, à, ò ⇒	**Other Romance** French Italian Romanian Moldovan *See section II*		ქოველი საიდუმლოe ⇓ **Georgian** (non-Indo-European) *See section XII*
æ, å, ä, ø, ð, þ ⇒	**North Germanic** Danish Faeroese Icelandic Norwegian Swedish *See section III*	ä, ö, ü, ë, ee, éé ⇒	**West Germanic** Dutch English German Lëtze-buergesch *See section III*		מצפונית ורבת השראה ⇓ **Hebrew** script, often used for writing **Yiddish** (Indo-European) *See section XII*
ā, ū, ą, ė, ķ, ņ ⇒	**Baltic** Latvian Lithuanian *See section IV*				

ò, í, bh, mh, c'h, ŷ ⇒	**Celtic** Breton Irish Scots Gaelic Welsh *See section V*	ë, xh, shp shq ⇒	**Other Indo-European** Albanian *See section VI*
ä, õ, ő, ű, đ, č, **á** ⇒	**Finno-Ugric** Estonian Finnish Hungarian Sámi *See section VII*	ġ, ħ, ż ⇓ **Semitic** Maltese *See section VIII*	
ç, ş, ğ, ü, ə ⇒	**Turkic** Azeri Turkish *See section IX*		
â, î ⇒	**Indic** Romani *See section XIII*	ll, ñ ⇓ Basque ***Family not clarified*** *See section XIII*	

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.

Bosnian used at one time to be written in Arabic script but now appears in either latin or cyrillic script.

> ► "Montenegrin" may be destined to become the name of the national language of Montenegro (independence in June 2006). The 1992 constitution of Montenegro defines the official language as the Ijekavian dialect of **Serbian**.

For more details on the differences between the four languages, see 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).

> ► Possible confusion of Serbian with Croatian if the former is written in latin script.

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**.

> ► Slovenian and Slovakian, a West Slavic language, are often confused because the Slovak name for Slovakia is 'Slovensko', reminiscent of 'Slovenia', and because the two-letter ISO language codes are easily mixed up (SL - SK). Quick way of distinguishing SL from SK: Slovenian does not use **y** at all, whereas Slovakian does fairly frequently.

West Slavic
Czech/Slovakian/Polish

Czech and Slovakian: dense use of accents and diacritics (as compared 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 (väčší, příště) and unusual-looking words with no apparent vowels (prst, dlh, stĺp, třň, tĺcť, krb, štvrť).

Czech: the main distinguishing character is **ů**, unique to Czech.
Slovakian also has **l** 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, ď/Ď, ň, š, ť/Ť, 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, ŕ**

> ► Slovakian and Slovenian, a South Slavic language, are often confused because the Slovak name for Slovakia is 'Slovensko', reminiscent of 'Slovenia', and because the language codes are easily mixed up (SK - SL). Quick way of distinguishing SK from SL: Slovenian does not use **y** at all, whereas Slovakian does fairly frequently.

Polish has
(a) no *háček* at all, but characteristic digraphs and consonant clusters instead: for example, cz, dz, rz (common ending for masculine nouns), sz, szcz
(b) neither '**ů**' nor '**ä**'; <u>but</u> note that
(c) the crossed or barred '**Ł**' (capital or small) is unique to Polish;
(d) 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

> ► Polish is also an important regional language in Lithuania, so a Lithuanian text in PL could get wrongly assigned to Poland.

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 <u>not</u> on vowels, in contrast to **Portuguese**. **Catalan** has **ny** where Castilian has **ñ**

Both Castilian and Catalan make use of double ¡...! 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: *Catalan is the official language of Andorra, which is not a member of the European Union.*]

[Note: *While Latin American Spanish <u>usage</u> is not uniform across the Americas, there are no striking differences from Castilian in <u>spelling</u> or <u>punctuation</u>.*]

Portuguese: common use of final **m** (unlike Castilian) - circumflex accents **ê, â, ô** - no double **ll** - tilde on the <u>vowels</u> **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).

> ► Galego is sometimes misidentified as Spanish because of the Member State of origin.

[Note*: Brazilian Portuguese differs in spelling and grammar from European Portuguese, e.g. not showing the consonant groups cc, cç:* **acção = ação.**]

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: *French is an official language in 41 countries including France, Belgium, Luxembourg and Switzerland (for certain purposes also in Jersey/Channel Islands), as well as being used in Italy. French is spoken extensively in the Americas (in particular Canada), Africa, Asia and the Pacific, with many dialect variations. In addition, it is an official language of international organisations which are too numerous to mention.]*

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.

[Note*: Italian is also an official language of Switzerland and was an official language of Malta between 1921 and 1934.*]

Romanian/Moldovan

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

Accented characters used in Romanian: **â, ă, î, ş, ţ**.

> ► Romanian, at least in short texts like titles and headings, might conceivably be confused with Turkish, to which it is not linguistically related, see Section IX for details.

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 Transdniestria, 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 <u>and</u> 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**.

(Contrast with **German**, which has capital letters for nouns and three umlaut accents (**ä**, **ö** <u>and</u> **ü**), although the umlauts can be replaced by **ae** (<u>not</u> the same as the Scandinavian ligature **æ**), **oe** and **ue.)**

> ► Under the Åland Autonomy Act, Swedish is the sole official language in Finland's Åland Islands. Swedish documents sometimes get misclassified as Finnish as a result. Likewise, documents in Swedish from the Åland Islands can get assigned to Sweden by mistake.

[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. BM is used for EEA texts and was used for the EU Treaties prior to the accession that did not happen.*]

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

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

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 <u>not</u> 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, representing a sibilant 's' sound.

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

[Note 1: *in the new, reformed spelling since 1998, German has done away with 'ß' after short vowels and uses 'ss' instead. So expect ß still after long vowels only - i.e. that much rarer. Swiss German had already abandoned ß altogether before the reform.*]

[Note 2: *High German has several variants as an official national language, in terms of both grammar and vocabulary (in Germany, Austria, Switzerland, Luxembourg and Liechtenstein, with regional status in Belgium, Denmark and Italy), as well as a host of major dialects, while Low German also has several dialects but is not an official language in any country.*]

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** - use of **ë** in some plurals – frequent use of double vowels **aa**, **ee**, **oo** – also **éé** and **óó** with accents for emphasis - nouns are <u>not</u> capitalised, unlike in German - <u>no</u> umlaut accents on a, o, u.

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

[Note: *Dutch as used in the Netherlands and Belgium is officially called Algemeen Beschaafd Nederlands (ABN). Flemish (Vlaams) is a dialect of Dutch spoken in Belgium and northern France, while Afrikaans is an official national language in South Africa. In Afrikaans, **ij** is written **y**.*]

Lëtzebuergesch (also Luxemburgish**)** 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

Lithuanian: all the <u>vowels</u> except 'o' can also take a <u>right-bending</u> cedilla below: **ą ę į ų** .
Only Lithuanian has **ė** .
There is a high frequency of endings in '-os'.

Latvian: the <u>consonants</u> 'k', 'l' and 'n' can take a <u>left-bending</u> cedilla: **ķ ļ ņ** but never the vowels. When **g** takes the cedilla, it is placed above the letter: **ġ**.
There is a high frequency of endings in '-as'.

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

> ►Confusion between Lithaunian and Latvian is quite common at the Commission: this may be because superficially they look similar (but the diacritics below consonants in one language and below vowels in the other are immediately apparent), or because the ISO codes LV (Latvian) and LT (Lithuanian) are easily mixed up.

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 spoken widely in Kosovo*.]

Section VII – Finno-Ugric Languages

As the complicated 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**).

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**.

[Note: *Hungarian is also used by 15 million people in Slovakia, Romania, Serbia and Ukraine.*]

Sámi
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 20th century. Sámi has given us the word "tundra".

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.

Section VIII – Semitic Maltese

Maltese
The crossed 'h' is the most obvious distinguishing mark: **ħ, Ħ, għ.**
Maltese also has **ċ**, **ġ** and **ż**, and uses a lot of hyphens.

[Note*: 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 generally 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** also puts a cedilla under **ş** but <u>not</u> **t** (cf Romanian) - most characteristic letter: **ğ**. Turkish uses the letter **i** both with and without a dot in upper and lower case (representing two separate vowel sounds), with occasional circumflex on **a**, **i** and **u**.

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

[Note 1: *Turkish is the 'other language' of our current Member State Cyprus. The website of the Commission Office in Cyprus has some content in Turkish.*]

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.

[Note 2: *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/definition.htm* *for the details.*

Section X – Slavic languages (with cyrillic script)

[Note: *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.

► Possible confusion of Bulgarian with Russian (and Belarusian/Ukrainian).

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).

[Note: *The ISO code for Bulgarian is BG, not BU as occasionally encountered.*]

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

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

► Possible confusion of Serbian with Croatian when written in latin script.

East Slavic

Russian/Belarusian/Ukrainian

Distinguishing letters that do <u>not</u> occur in **Russian**:

Belarusian: r, i, ў
Ukrainian: r, є, i, ï

Unique distinguishing characters: Only Belarusian has **ў** - **o**nly Ukrainian has **ï.**

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

> ► Possible confusion of Russian/Belarusian/Ukrainian with Bulgarian.

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

Section XI – Greek

The **Greek script** has a rounded and cursive look (e.g. α, β, δ, η, φ) compared to **cyrillic**, which is more straight-lined, up-and-down (e.g. я, в, д, й, ф). Hand-written Cyrillic is a bit closer to Greek but not all that much. Cyrillic was originally adapted from the Greek script in 10th century Macedonia.

Greek can have accents on some letters, e.g. ά, ή.

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

> ► Documents written in Greek can be assigned to the wrong Member State (Greece/Cyprus).

[Note 1: *The military dictatorship in Greece (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.*]

[Note 2: *Many Greek characters are of course used internationally in mathematics and in science generally.*]

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 may also be 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 საიდუმლოდ

Yiddish is essentially a Germanic language with much Slavic and Hebrew vocabulary. It classicaly 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 uses **ñ** and **ll** as in Spanish, otherwise typical digraphs are: **dd, rr, ts, tt, tx, tz.**

Annex I - 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 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.

Albanian
http://www.bbc.co.uk/albanian/

Armenian
http://hy.wikipedia.org/wiki/%D5%80%D5%A1%D5%B5%D5%A1%D5%BD%D5%BF%D5%A1%D5%B6

http://www.omniglot.com/writing/armenian.htm

Azeri
http://www.bbc.co.uk/azeri/

Basque
http://www.egaztea.com/web/

Belarusian
http://www.svaboda.org/default_text.aspx

Bosnian
http://www.fbihvlada.gov.ba/

Breton
http://www.breman.org/

Bulgarian
http://www.europarl.europa.eu/hearings/commission/2006_enlarg/default_bg.htm

Castilian Spanish
http://www.rtve.es/rne/web/index.php

Catalan
http://www.catradio.com/pcatradio/crHome.jsp

Corsican
http://adecec.net/vocenustrale/

Croatian
http://www.hrt.hr/index.xml

Czech
http://www.radio.cz/cz/

Danish
http://www.dr.dk/

Dutch
http://www.wereldomroep.nl/

English
http://www.bbc.co.uk/worldservice/

Estonian
http://www.er.ee/index.php

Faeroese
http://www.framtak.com/fo/index.html

Finnish
http://www.yle.fi/radiosuomi/

French
http://www.radiofrance.fr/

Galician
http://www.crtvg.es/cgi-bin/iniciocrtvg.asp?idioma=galego

Georgian
http://www.tavisupleba.org/

German
http://www.dw-world.de/dw/0,2142,265,00.html

Greek
http://www.hri.org/news/latest/greek.latest.html

Greenlandic
http://knr.gl/nutaarsiassat/index.html

Hungarian
http://www.hullamvadasz.hu/

Icelandic
http://www.utvarpsaga.is/

Irish
http://www.bnag.ie/default1.asp?lang=ga

Italian

http://www.radio.rai.it/

Latvian
http://www.radio.org.lv/

Lithuanian
http://www.lrt.lt/

Lëtzebuergesch
http://lb.wikipedia.org/wiki/Haapts%C3%A4it

Macedonian
http://www.bbc.co.uk/macedonian/

Maltese
http://www.rtk.org.mt/

Moldovan
http://www.protv.md/

Norwegian
http://www.nrk.no/

Polish
http://www.gazeta.pl/0,0.html

Portuguese
http://www.rdp.pt/

Romani
http://www.romnet.hu/hirek/romaindex.html

Romanian
http://www.radior.ro/

Russian
http://www.rambler.ru/

Sámi
http://www.saamicouncil.net

Scots Gaelic
http://www.bbc.co.uk/scotland/alba/

Serbian (latin script)
http://www.bbc.co.uk/serbian/

Serbian (cyrillic script)

http://www.srpska.com/cir/index.php

Slovakian
http://www.skonline.sk/

Slovenian
http://www.rtvslo.si/

Spanish *see* **Castilian Spanish**

Swedish
http://www.sr.se/

Turkish
http://www.bbc.co.uk/turkish/

Ukrainian
http://www.bbc.co.uk/ukrainian/domestic/

Welsh
http://www.bbc.co.uk/cymru

Yiddish
http://yiddish.forward.com/

Annex II
Useful words for recognition purposes

This list is restricted to current and likely future EU languages and one or two others that are regularly received by DGT Planning.

Language and code	Native name of country/countries	Native name of capital(s)	"European Commission"	"Confidential"
Bulgarian BG	България	София	Европейската комисия	
Castilian ES	España	Madrid	Comisión europea	
Croatian HR	Hrvatska	Zagreb	Europske komisije	
Czech CS	Česká republika	Praha	Evropská komise	důvěrný
Danish DA	Dansk	København	Europa-kommissionen	
Dutch NL	België, Nederland	Brussel, Amsterdam (Den Haag seat of govt)	Europese Commissie	vertrouwelijk
English EN	United Kingdom, Ireland	London, Dublin	European Commission	confidential
Estonian ET	Eesti	Tallinn	Euroopa Komisjon	
Finnish FI	Suomi	Helsinki	Euroopan komissio	
French FR	Belgique, France	Bruxelles, Paris	Commission européenne	confidentiel
German DE	Deutschland, Österreich	Berlin, Wien	Europäische Kommission	vertraulich
Greek EL	Ελλάδα, Κύπρος	Αθήνα, Λευκωσία	Ευρωπαϊκή Επιτροπή	
Hungarian HU	Magyarország	Budapest	Európai Bizottság	bizalmas
Irish GA	Éire	Baile Átha Cliath	Coimisiún Eorpach	
Italian IT	Italia	Roma	Commissione europea	
Latvian LV	Latvija	Rīga	Eiropas Komisija	
Lithuanian LT	Lietuva	Vilnius	Europos Komisija	slaptai
Maltese MT	Malta	Valetta	Kummissjoni Ewropea	
Norwegian NO	Norge	Oslo	Europa-kommisjonen	

Polish PL	Polska	Warszawa	Komisja Europejska	poufne
Portuguese PT	Portugal	Lisbõa	Comissão Europeia	
Romanian RO	România, Moldova (Молдова)	Bucureşti, Chişinău (Кишинёв)	Comisia Europeană	confidenţial
Russian RU	Россия	Москва	Европейская Комиссия	
Slovakian SK	Slovensko	Bratislava	Europska komisia	dôverné
Slovenian SL	Slovenija	Ljubljana	Evropska komisija	
Swedish SV	Sverige	Stockholm	Europeiska kommissionen	
Turkish TR	Türkiye, (Kıbrıs)	Ankara, (Lefkoşa)	Avrupa Komisyonu	