

# ETSI TS 123 038 V8.3.0 (2010-01)

Technical Specification

**Digital cellular telecommunications system (Phase 2+);  
Universal Mobile Telecommunications System (UMTS);  
LTE;  
Alphabets and language-specific information  
(3GPP TS 23.038 version 8.3.0 Release 8)**



**iTeh STANDARD PREVIEW**  
(standards.iteh.ai)

Full standard:  
<https://standards.iteh.ai/catalog/standards/sist/91714f4-51b8-4a0c-aa00-ebf643622e12/etsi-ts-123-038-v8.3.0-2010-01>



## Reference

---

RTS/TSGC-0123038v830

## Keywords

---

GSM, LTE, UMS**ETSI**

650 Route des Lucioles  
F-06921 Sophia Antipolis Cedex - FRANCE

---

Tel.: +33 4 92 94 42 00 - Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C  
Association à but non lucratif enregistrée à la  
Sous-Préfecture de Grasse (06) N° 7803/88

**Important notice**

---

Individual copies of the present document can be downloaded from:

<http://www.etsi.org>

The present document may be made available in more than one electronic version or in print. In any case of existing or perceived difference in contents between such versions, the reference version is the Portable Document Format (PDF). In case of dispute, the reference shall be the printing on ETSI printers of the PDF version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at

<http://portal.etsi.org/tb/status/status.asp>

If you find errors in the present document, please send your comment to one of the following services:

[http://portal.etsi.org/chaicor/ETSI\\_support.asp](http://portal.etsi.org/chaicor/ETSI_support.asp)

**Copyright Notification**

---

No part may be reproduced except as authorized by written permission.  
The copyright and the foregoing restriction extend to reproduction in all media.

© European Telecommunications Standards Institute 2010.  
All rights reserved.

**DECT™**, **PLUGTESTS™**, **UMTS™**, **TIPHON™**, the TIPHON logo and the ETSI logo are Trade Marks of ETSI registered for the benefit of its Members.

**3GPP™** is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners.

**LTE™** is a Trade Mark of ETSI currently being registered

for the benefit of its Members and of the 3GPP Organizational Partners.

**GSM®** and the GSM logo are Trade Marks registered and owned by the GSM Association.

---

## Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: "*Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards*", which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (<http://webapp.etsi.org/IPR/home.asp>).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

---

## Foreword

This Technical Specification (TS) has been produced by ETSI 3rd Generation Partnership Project (3GPP).

The present document may refer to technical specifications or reports using their 3GPP identities, UMTS identities or GSM identities. These should be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between GSM, UMTS, 3GPP and ETSI identities can be found under <http://webapp.etsi.org/key/queryform.asp>.

**PRELIMINARY**  
**STANDARD**  
**PRELIMINARY**  
iTech (standards.itech.ai)  
Full standard:  
<https://standards.itech.ai/catalog/standards/sist/09117444-51b8-4a0c-aa00-ebf643622e12/etsi-ts-123-038-v8.3.0-2010-01>

# Contents

Intellectual Property Rights .....	2
Foreword.....	2
Foreword.....	5
1 Scope .....	6
2 References .....	6
3 Abbreviations and definitions .....	7
4 SMS Data Coding Scheme .....	8
5 CBS Data Coding Scheme .....	11
6 Individual parameters .....	13
6.1 General principles.....	13
6.1.1 General notes .....	13
6.1.2 Character packing .....	14
6.1.2.1 SMS Packing.....	14
6.1.2.1.1 Packing of 7-bit characters .....	14
6.1.2.2 CBS Packing.....	15
6.1.2.2.1 Packing of 7-bit characters .....	15
6.1.2.3 USSD packing.....	15
6.1.2.3.1 Packing of 7 bit characters .....	15
6.2 Character sets and coding.....	18
6.2.1 GSM 7 bit Default Alphabet .....	18
6.2.1.1 GSM 7 bit default alphabet extension table .....	19
6.2.1.2 National Language Identifier .....	21
6.2.1.2.1 Introduction .....	21
6.2.1.2.2 Single shift mechanism.....	21
6.2.1.2.3 Locking shift mechanism.....	21
6.2.1.2.4 National Language Identifier .....	21
6.2.1.2.5 Processing of national language characters .....	22
6.2.2 8 bit data .....	22
6.2.3 UCS2 .....	22
<b>Annex A (normative): National Language Tables.....</b>	<b>24</b>
A.1 Introduction .....	24
A.2 National Language Single Shift Tables.....	25
A.2.1 Turkish National Language Single Shift Table .....	25
A.2.2 Spanish National Language Single Shift Table .....	26
A.2.3 Portuguese National Language Single Shift Table.....	27
A.3 National Language Locking Shift Tables.....	28
A.3.1 Turkish National Language Locking Shift Table .....	28
A.3.2 Void.....	29
A.3.3 Portuguese National Language Locking Shift Table.....	29
<b>Annex B (informative): Guidelines for creating language tables.....</b>	<b>30</b>
B.1 Introduction .....	30
B.2 Template for Single Shift Language Tables.....	30
B.3 Template for Locking Shift Language Tables.....	32
<b>Annex C (Informative): Example for locking shift and single shift mechanisms.....</b>	<b>33</b>
C.1 Introduction .....	33

C.2 Example of single shift.....33

C.3 Example of locking shift .....33

**Annex D (informative): Document change history .....35**

History .....36

**iTeh STANDARD PREVIEW**  
**(standards.iteh.ai)**  
Full standard:  
<https://standards.iteh.ai/catalog/standards/sist/091714f4-51b8-4a0c-aa00-ebf643622e12/etsi-ts-123-038-v8.3.0-2010-01>

---

## Foreword

This Technical Specification has been produced by the 3<sup>rd</sup> Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

- x the first digit:
  - 1 presented to TSG for information;
  - 2 presented to TSG for approval;
  - 3 Indicates TSG approved document under change control.
- y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.
- z the third digit is incremented when editorial only changes have been incorporated in the specification;

**ITeH STANDARD PREVIEW**  
(standards.iteh.ai)  
Full standard:  
<https://standards.iteh.ai/catalog/standards/sist/123-038-v8.3.0-2010-01>  
51b8-4a0c-aa00-ebf643622e12/etsi-ts-123-038-v8.3.0-

---

# 1 Scope

The present document defines the character sets, languages and message handling requirements for SMS, CBS and USSD and may additionally be used for Man Machine Interface (MMI) (3GPP TS 22.030 [2]).

The specification for the Data Circuit terminating Equipment/Data Terminal Equipment (DCE/DTE) interface (3GPP TS 27.005 [8]) will also use the codes specified herein for the transfer of SMS data to an external terminal.

---

# 2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

- [1] void
- [2] 3GPP TS 22.030: "Man-Machine Interface (MMI) of the User Equipment (UE)".
- [3] 3GPP TS 23.090: "Unstructured Supplementary Service Data (USSD) - Stage 2".
- [4] 3GPP TS 23.040: "Technical realization of the Short Message Service (SMS)".
- [5] 3GPP TS 23.041: "Technical realization of Cell Broadcast Service (CBS)".
- [6] 3GPP TS 24.011: "Point-to-Point (PP) Short Message Service (SMS) support on mobile radio interface".
- [7] Void.
- [8] 3GPP TS 27.005: "Use of Data Terminal Equipment - Data Circuit terminating Equipment (DTE - DCE) interface for Short Message Service (SMS) and Cell Broadcast Service (CBS)".
- [10] ISO/IEC 10646: "Information technology; Universal Multiple-Octet Coded Character Set (UCS)".
- [11] 3GPP TS 24.090: "Unstructured Supplementary Service Data (USSD); Stage 3".
- [12] ISO 639: "Code for the representation of names of languages".
- [13] 3GPP TS 23.042: "Compression algorithm for text messaging services".
- [14] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".
- [15] "Wireless Datagram Protocol Specification", Wireless Application Protocol Forum Ltd.
- [16] ISO 1073-1 and ISO 1073-2 Alphanumeric character sets for optical recognition – Parts 1 and 2: Character sets OCR-A and OCR-B, respectively - Shapes and dimensions of the printed image.
- [17] 3GPP TS 31.102: 'Characteristics of the USIM application'
- [18] 3GPP TS 51.011 Release 4 (version 4.x.x): 'Specification of the Subscriber Identity Module - Mobile Equipment (SIM - ME) interface'

---

## 3 Abbreviations and definitions

For the purposes of the present document, the following terms and definitions apply:

**National Language Identifier:** A code representing a specific language and thereby selecting a specific National Language Table.

**National Language Locking Shift Table:** A national language table which replaces the GSM 7 bit default alphabet table in the case where the locking shift mechanism as defined in subclause 6.2.1.2.3 is used.

**National Language Single Shift Table:** A national language table which replaces the GSM 7 bit default alphabet extension table in the case where the single shift mechanism as defined in subclause 6.2.1.2.2 is used.

**National Language Table:** A table containing the characters of a specific national language.

For the purposes of the present document, the abbreviations used in the present document are listed in 3GPP TR 21.905 [14].

**ITeH STANDARD PREVIEW**  
(standards.iteh.ai)  
Full standard:  
<https://standards.iteh.ai/catalog/standards/sist/091714f4-51b8-4a0c-aa00-ebf643622e12/etsi-ts-123-038-v8.3.0-2010-01>

## 4 SMS Data Coding Scheme

The TP-Data-Coding-Scheme field, defined in 3GPP TS 23.040 [4], indicates the data coding scheme of the TP-UD field, and may indicate a message class. Any reserved codings shall be assumed to be the GSM 7 bit default alphabet (the same as codepoint 00000000) by a receiving entity. The octet is used according to a coding group which is indicated in bits 7..4. The octet is then coded as follows:

Coding Group Bits 7..4	Use of bits 3..0																														
00xx	<p>General Data Coding indication Bits 5..0 indicate the following:</p> <p>Bit 5, if set to 0, indicates the text is uncompressed Bit 5, if set to 1, indicates the text is compressed using the compression algorithm defined in 3GPP TS 23.042 [13]</p> <p>Bit 4, if set to 0, indicates that bits 1 to 0 are reserved and have no message class meaning Bit 4, if set to 1, indicates that bits 1 to 0 have a message class meaning::</p> <table border="1"> <thead> <tr> <th>Bit 1</th> <th>Bit 0</th> <th>Message Class</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0</td> <td>Class 0</td> </tr> <tr> <td>0</td> <td>1</td> <td>Class 1 Default meaning: ME-specific.</td> </tr> <tr> <td>1</td> <td>0</td> <td>Class 2 (U)SIM specific message</td> </tr> <tr> <td>1</td> <td>1</td> <td>Class 3 Default meaning: TE specific (see 3GPP TS 27.005 [8])</td> </tr> </tbody> </table> <p>Bits 3 and 2 indicate the character set being used, as follows :</p> <table border="1"> <thead> <tr> <th>Bit 3</th> <th>Bit 2</th> <th>Character set:</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>0</td> <td>GSM 7 bit default alphabet</td> </tr> <tr> <td>0</td> <td>1</td> <td>8-bit data</td> </tr> <tr> <td>1</td> <td>0</td> <td>UCS2 (16bit) [10]</td> </tr> <tr> <td>1</td> <td>1</td> <td>Reserved</td> </tr> </tbody> </table> <p>NOTE: The special case of bits 7..0 being 0000 0000 indicates the GSM 7 bit default alphabet with no message class</p>	Bit 1	Bit 0	Message Class	0	0	Class 0	0	1	Class 1 Default meaning: ME-specific.	1	0	Class 2 (U)SIM specific message	1	1	Class 3 Default meaning: TE specific (see 3GPP TS 27.005 [8])	Bit 3	Bit 2	Character set:	0	0	GSM 7 bit default alphabet	0	1	8-bit data	1	0	UCS2 (16bit) [10]	1	1	Reserved
Bit 1	Bit 0	Message Class																													
0	0	Class 0																													
0	1	Class 1 Default meaning: ME-specific.																													
1	0	Class 2 (U)SIM specific message																													
1	1	Class 3 Default meaning: TE specific (see 3GPP TS 27.005 [8])																													
Bit 3	Bit 2	Character set:																													
0	0	GSM 7 bit default alphabet																													
0	1	8-bit data																													
1	0	UCS2 (16bit) [10]																													
1	1	Reserved																													
01xx	<p>Message Marked for Automatic Deletion Group</p> <p>This group can be used by the SM originator to mark the message ( stored in the ME or (U)SIM ) for deletion after reading irrespective of the message class. The way the ME will process this deletion should be manufacturer specific but shall be done without the intervention of the End User or the targeted application. The mobile manufacturer may optionally provide a means for the user to prevent this automatic deletion.</p> <p>Bit 5..0 are coded exactly the same as Group 00xx</p>																														
1000..1011	Reserved coding groups																														
1100	<p>Message Waiting Indication Group: Discard Message</p> <p>The specification for this group is exactly the same as for Group 1101, except that:</p> <ul style="list-style-type: none"> <li>- after presenting an indication and storing the status, the ME may discard the contents of the message.</li> </ul> <p>The ME shall be able to receive, process and acknowledge messages in this group, irrespective of memory availability for other types of short message.</p>																														
1101	<p>Message Waiting Indication Group: Store Message</p> <p>This Group defines an indication to be provided to the user about the status of types of message waiting on systems connected to the GSM/UMTS PLMN. The ME should present this indication as an icon on the screen, or other MMI indication. The ME shall update the contents of the Message Waiting Indication Status on the SIM (see 3GPP TS 51.011 [18]) or USIM (see 3GPP TS 31.102 [17]) when present or otherwise should store the status in the ME. In case there are multiple records of EF<sub>MWIS</sub> this information shall be stored within the first record. The contents of the Message Waiting Indication Status should control the</p>																														

Coding Group Bits 7..4	Use of bits 3..0
	<p>ME indicator. For each indication supported, the mobile may provide storage for the Origination Address. The ME may take note of the Origination Address for messages in this group and group 1100.</p> <p>Text included in the user data is coded in the GSM 7 bit default alphabet. Where a message is received with bits 7..4 set to 1101, the mobile shall store the text of the SMS message in addition to setting the indication. The indication setting should take place irrespective of memory availability to store the short message.</p> <p>Bits 3 indicates Indication Sense:</p> <p>Bit 3  0 Set Indication Inactive  1 Set Indication Active</p> <p>Bit 2 is reserved, and set to 0</p> <p>Bit 1 Bit 0 Indication Type:  0 0 Voicemail Message Waiting  0 1 Fax Message Waiting  1 0 Electronic Mail Message Waiting  1 1 Other Message Waiting*</p> <p>* Mobile manufacturers may implement the "Other Message Waiting" indication as an additional indication without specifying the meaning.</p>
1110	<p>Message Waiting Indication Group: Store Message</p> <p>The coding of bits 3..0 and functionality of this feature are the same as for the Message Waiting Indication Group above, (bits 7..4 set to 1101) with the exception that the text included in the user data is coded in the uncompressed UCS2 character set.</p>
1111	<p>Data coding/message class</p> <p>Bit 3 is reserved, set to 0.</p> <p>Bit 2 Message coding:  0 GSM 7 bit default alphabet  1 8-bit data</p> <p>Bit 1 Bit 0 Message Class:  0 0 Class 0  0 1 Class 1 default meaning: ME-specific.  1 0 Class 2 (U)SIM-specific message.  1 1 Class 3 default meaning: TE specific (see 3GPP TS 27.005 [8])</p>

GSM 7 bit default alphabet indicates that the TP-UD is coded from the GSM 7 bit default alphabet given in clause 6.2.1. When this character set is used, the characters of the message are packed in octets as shown in clause 6.1.2.1.1, and the message can consist of up to 160 characters. The GSM 7 bit default alphabet shall be supported by all MSs and SCs offering the service. If the GSM 7 bit default alphabet extension mechanism is used then the number of displayable characters will reduce by one for every instance where the GSM 7 bit default alphabet extension table is used. 8-bit data indicates that the TP-UD has user-defined coding, and the message can consist of up to 140 octets.

UCS2 character set indicates that the TP-UD has a UCS2 [10] coded message, and the message can consist of up to 140 octets, i.e. up to 70 UCS2 characters. The General notes specified in clause 6.1.1 override any contrary specification in UCS2, so for example even in UCS2 a <CR> character will cause the MS to return to the beginning of the current line and overwrite any existing text with the characters which follow the <CR>.

When a message is compressed, the TP-UD consists of the GSM 7 bit default alphabet or UCS2 character set compressed message, and the compressed message itself can consist of up to 140 octets in total.

When a mobile terminated message is class 0 and the MS has the capability of displaying short messages, the MS shall display the message immediately and send an acknowledgement to the SC when the message has successfully reached the MS irrespective of whether there is memory available in the (U)SIM or ME. The message shall not be automatically stored in the (U)SIM or ME.

The ME may make provision through MMI for the user to selectively prevent the message from being displayed immediately.

If the ME is incapable of displaying short messages or if the immediate display of the message has been disabled through MMI then the ME shall treat the short message as though there was no message class, i.e. it will ignore bits 0 and 1 in the TP-DCS and normal rules for memory capacity exceeded shall apply.

When a mobile terminated message is Class 1, the MS shall send an acknowledgement to the SC when the message has successfully reached the MS and can be stored. The MS shall normally store the message in the ME by default, if that is possible, but otherwise the message may be stored elsewhere, e.g. in the (U)SIM. The user may be able to override the default meaning and select their own routing.

When a mobile terminated message is Class 2 ((U)SIM-specific), an MS shall ensure that the message has been transferred to the SMS data field in the (U)SIM before sending an acknowledgement to the SC. The MS shall return a "protocol error, unspecified" error message (see 3GPP TS 24.011 [6]) if the short message cannot be stored in the (U)SIM and there is other short message storage available at the MS. If all the short message storage at the MS is already in use, the MS shall return "memory capacity exceeded". This behaviour applies in all cases except for an MS supporting (U)SIM Application Toolkit when the Protocol Identifier (TP-PID) of the mobile terminated message is set to "(U)SIM Data download" (see 3GPP TS 23.040 [4]).

When a mobile terminated message is Class 3, the MS shall send an acknowledgement to the SC when the message has successfully reached the MS and can be stored, irrespectively of whether the MS supports an SMS interface to a TE, and without waiting for the message to be transferred to the TE. Thus the acknowledgement to the SC of a TE-specific message does not imply that the message has reached the TE. Class 3 messages shall normally be transferred to the TE when the TE requests "TE-specific" messages (see 3GPP TS 27.005 [8]). The user may be able to override the default meaning and select their own routing.

The message class codes may also be used for mobile originated messages, to provide an indication to the destination SME of how the message was handled at the MS.

The MS will not interpret reserved or unsupported values but shall store them as received. The SC may reject messages with a Data Coding Scheme containing a reserved value or one which is not supported.