ARIB STD-T63-23.038 V11.0.0

Alphabets and language-specific information

(Release 11)

Refer to “Industrial Property Rights (IPR)” in the preface of ARIB STD-T63 for Related Industrial Property Rights. Refer to “Notice” in the preface of ARIB STD-T63 for Copyrights.

3GPP TS 23.038 V11.0.0 (2012-09)

Technical Specification

3rd Generation Partnership Project; Technical Specification Group Core Network and Terminals; Alphabets and language-specific information (Release 11)

The present document has been developed within the 3rd Generation Partnership Project (3GPP TM) and may be further elaborated for the purposes of 3GPP.

The present document has not been subject to any approval process by the 3GPP Organisational Partners and shall not be implemented.
This Specification is provided for future development work within 3GPP only. The Organisational Partners accept no liability for any use of this Specification.
Specifications and reports for implementation of the 3GPP TM system should be obtained via the 3GPP Organisational Partners' Publications Offices.

Keywords
GSM, UMTS, LTE, character

3GPP

Postal address

3GPP support office address
650 Route des Lucioles - Sophia Antipolis
Valbonne - FRANCE
Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Internet
http://www.3gpp.org

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.

© 2012, 3GPP Organizational Partners (ARIB, ATIS, CCSA, ETSI, TTA, TTC).
All rights reserved.

UMTS™ is a Trade Mark 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 registered for the benefit of its Members and of the 3GPP Organizational Partners
GSM® and the GSM logo are registered and owned by the GSM Association

Contents

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 14
6.1 General principles 14
6.1.1 General notes 14
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 16
6.1.2.3.1 Packing of 7 bit characters 16
6.2 Character sets and coding 19
6.2.1 GSM 7 bit Default Alphabet 19
6.2.1.1 GSM 7 bit default alphabet extension table 20
6.2.1.2 National Language Identifier 22
6.2.1.2.1 Introduction 22
6.2.1.2.2 Single shift mechanism 22
6.2.1.2.3 Locking shift mechanism 22
6.2.1.2.4 National Language Identifier 22
6.2.1.2.5 Processing of national language characters 23
6.2.2 8 bit data 24
6.2.3 UCS2 24

Annex A (normative): National Language Tables 25

A.1 Introduction 25
A.2 National Language Single Shift Tables 26
A.2.1 Turkish National Language Single Shift Table 26
A.2.2 Spanish National Language Single Shift Table 27
A.2.3 Portuguese National Language Single Shift Table 28
A.2.4 Bengali National Language Single Shift Table 28
A.2.5 Gujarati National Language Single Shift Table 30
A.2.6 Hindi National Language Single Shift Table 31
A.2.7 Kannada National Language Single Shift Table 32
A.2.8 Malayalam National Language Single Shift Table 33
A.2.9 Oriya National Language Single Shift Table 34
A.2.10 Punjabi National Language Single Shift Table 35
A.2.11 Tamil National Language Single Shift Table 36
A.2.12 Telugu National Language Single Shift Table 37
A.2.13 Urdu National Language Single Shift Table 38
A.3 National Language Locking Shift Tables 39
A.3.1 Turkish National Language Locking Shift Table 39
A.3.2 Void 40
A.3.3 Portuguese National Language Locking Shift Table 40
A.3.4 Bengali National Language Locking Shift Table 40
A.3.5 Gujarati National Language Locking Shift Table 42
A.3.6 Hindi National Language Locking Shift Table 43
A.3.7 Kannada National Language Locking Shift Table 44

A.3.8 Malayalam National Language Locking Shift Table 45
A.3.9 Oriya National Language Locking Shift Table 46
A.3.10 Punjabi National Language Locking Shift Table 47
A.3.11 Tamil National Language Locking Shift Table 48
A.3.12 Telugu National Language Locking Shift Table 49
A.3.13 Urdu National Language Locking Shift Table 50

Annex B (informative): Guidelines for creating language tables 51

B.1 Introduction 51

B.2 Template for Single Shift Language Tables 51

B.3 Template for Locking Shift Language Tables 53

Annex C (Informative): Example for locking shift and single shift mechanisms 54

C.1 Introduction 54

C.2 Example of single shift 54

C.3 Example of locking shift 54

Annex D (informative): Document change history 56

Foreword

This Technical Specification has been produced by the 3rd 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;

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”

[19] 3GPP TS 24.294: "IMS Centralized Services (ICS) Protocol via I1 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].

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	General Data Coding indication Bits 5..0 indicate the following: 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] 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:: 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]) Bits 3 and 2 indicate the character set being used, as follows : Bit 3 Bit2 Character set: 0 0 GSM 7 bit default alphabet 0 1 8 bit data 1 0 UCS2 (16bit) [10] 1 1 Reserved NOTE: The special case of bits 7..0 being 0000 0000 indicates the GSM 7 bit default alphabet with no message class
01xx	Message Marked for Automatic Deletion Group 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. Bit 5..0 are coded exactly the same as Group 00xx
1000..1011	Reserved coding groups
1100	Message Waiting Indication Group: Discard Message The specification for this group is exactly the same as for Group 1101, except that: - after presenting an indication and storing the status, the ME may discard the contents of the message. The ME shall be able to receive, process and acknowledge messages in this group, irrespective of memory availability for other types of short message.
1101	Message Waiting Indication Group: Store Message 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_{MWIS} this information shall be stored within the first record. The contents of the Message Waiting Indication Status should control the

Coding Group Bits 7..4	Use of bits 3..0
	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. 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. Bits 3 indicates Indication Sense: Bit 3 0 Set Indication Inactive 1 Set Indication Active Bit 2 is reserved, and set to 0 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* * Mobile manufacturers may implement the "Other Message Waiting" indication as an additional indication without specifying the meaning.
1110	Message Waiting Indication Group: Store Message 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.
1111	Data coding/message class Bit 3 is reserved, set to 0. Bit 2 Message coding: 0 GSM 7 bit default alphabet 1 8-bit data 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])

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.

5 CBS Data Coding Scheme

The CBS Data Coding Scheme indicates the intended handling of the message at the MS, the character set/coding, and the language (when applicable). Any reserved codings shall be assumed to be the GSM 7 bit default alphabet (the same as codepoint 00001111) 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
0000	Language using the GSM 7 bit default alphabet Bits 3..0 indicate the language: 0000 German 0001 English 0010 Italian 0011 French 0100 Spanish 0101 Dutch 0110 Swedish 0111 Danish 1000 Portuguese 1001 Finnish 1010 Norwegian 1011 Greek 1100 Turkish 1101 Hungarian 1110 Polish 1111 Language unspecified
0001	0000 GSM 7 bit default alphabet; message preceded by language indication. The first 3 characters of the message are a two-character representation of the language encoded according to ISO 639 [12], followed by a CR character. The CR character is then followed by 90 characters of text. 0001 UCS2; message preceded by language indication The message starts with a two GSM 7-bit default alphabet character representation of the language encoded according to ISO 639 [12]. This is padded to the octet boundary with two bits set to 0 and then followed by 40 characters of UCS2-encoded message. An MS not supporting UCS2 coding will present the two character language identifier followed by improperly interpreted user data. 0010..1111 Reserved
0010..	0000 Czech 0001 Hebrew 0010 Arabic 0011 Russian 0100 Icelandic 0101..1111 Reserved for other languages using the GSM 7 bit default alphabet, with unspecified handling at the MS
0011	0000..1111 Reserved for other languages using the GSM 7 bit default alphabet, with unspecified handling at the MS

Coding Group Bits 7..4	Use of bits 3..0
01xx	General Data Coding indication Bits 5..0 indicate the following: 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] 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: 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]) Bits 3 and 2 indicate the character set being used, as follows: Bit 3 Bit 2 Character set: 0 0 GSM 7 bit default alphabet 0 1 8 bit data 1 0 UCS2 (16 bit) [10] 1 1 Reserved
1000	Reserved coding groups
1001	Message with User Data Header (UDH) structure: 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]) Bits 3 and 2 indicate the alphabet being used, as follows: Bit 3 Bit 2 Alphabet: 0 0 GSM 7 bit default alphabet 0 1 8 bit data 1 0 USC2 (16 bit) [10] 1 1 Reserved
1010..1100	Reserved coding groups
1101	I1 protocol message defined in 3GPP TS 24.294 [19]
1110	Defined by the WAP Forum [15]
1111	Data coding / message handling Bit 3 is reserved, set to 0. Bit 2 Message coding: 0 GSM 7 bit default alphabet 1 8 bit data Bit 1 Bit 0 Message Class: 0 0 No message class. 0 1 Class 1 user defined. 1 0 Class 2 user defined. 1 1 Class 3 default meaning: TE specific (see 3GPP TS 27.005 [8])

These codings may also be used for USSD and MMI/display purposes.

The message length specified in this subclause is not applicable for UTRAN and E-UTRAN but only applicable for GSM.

See 3GPP TS 24.090 [11] for specific coding values applicable to USSD for MS originated USSD messages and MS terminated USSD messages. USSD messages using the default alphabet are coded with the GSM 7-bit default alphabet given in clause 6.2.1. The message can then consist of up to 182 user characters.

Cell Broadcast messages using the default alphabet are coded with the GSM 7-bit default alphabet given in clause 6.2.1. The message then consists of 93 user characters.

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 usedCell Broadcast messages using 8-bit data have user-defined coding, and will be 82 octets in length.

UCS2 character set indicates that the message is coded in UCS2 [10]. 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>. Cell Broadcast messages encoded in UCS2 consist of 41 characters.

When a CBS message received by the MS is message class 0 and the MS has the capability of displaying CBS messages, the MS shall display the message immediately. 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 CBS messages or if the immediate display of the message has been disabled through MMI then the ME shall treat the CBS message as though there was no message class, i.e. it will ignore bits 0 and 1 in the TP-DCS but may store the message either on the ME or on the (U)SIM.

Class 1 and Class 2 messages may be routed by the ME to user-defined destinations, but the user may override any default meaning and select their own routing.

Class 3 messages will normally be selected for transfer to a TE, in cases where a ME supports an SMS/CBS interface to a TE, and the TE requests "TE-specific" cell broadcast messages (see 3GPP TS 27.005 [8]). The user may be able to override the default meaning and select their own routing.

Messages with a User Data Header Structure are encoded as described in 3GPP TS 23.040 [4] for SMS, in subclauses 3.10 and 9.2.3.24.

The use of Cell Broadcast DCS values for messages with a User Data Header structure implies that the 82-bytes CB payload has a User Data Header structure.

The CBS message information field will contain the IEs as described in 3GPP TS 23.040. The concatenation IEs will not be used, as CB concatenation will rely in that case on the existing CB mechanism. Note that IEs that cannot be split and that IEs that are too large to fit in one CB segment cannot be transmitted using this mechanism. Also, some IEs as defined for SMS are not applicable for CB:

VALUE (hex)	MEANING
00	Concatenated short messages, 8-bit reference number
01	Special SMS Message Indication
06	SMSC Control Parameters
08	Concatenated short message, 16-bit reference number
20	RFC 822 E-Mail Header
23	Enhanced Voice Mail Information
70-7F	(U)SIM Toolkit Security Headers
80-89	SME to SME specific use

6 Individual parameters

6.1 General principles

6.1.1 General notes

Except where otherwise indicated, the following shall apply to all character sets:

1: The characters marked "1)" are not used but are displayed as a space.

2: The characters of this set, when displayed, should approximate to the appearance of the relevant characters specified in ISO 1073 [16]and the relevant national standards.

3: Control characters:

Code	Meaning
LF	Line feed: Any characters following LF which are to be displayed shall be presented as the next line of the message, commencing with the first character position.
CR	Carriage return: Any characters following CR which are to be displayed shall be presented as the current line of the message, commencing with the first character position.
SP	Space character.

4: The display of characters within a message is achieved by taking each character in turn and placing it in the next available space from left to right and top to bottom.

6.1.2 Character packing

6.1.2.1 SMS Packing

6.1.2.1.1 Packing of 7-bit characters

If a character number α is noted in the following way:

b7 b6 b5 b4 b3 b2 b1
αa αb αc αd αe αf αg

The packing of the 7-bitscharacters in octets is done by completing the octets with zeros on the left.

For examples, packing: α

- one character in one octet:
  - bits number:

    7 6 5 4 3 2 1 0
    0 1a 1b 1c 1d 1e 1f 1g

- two characters in two octets:
  - bits number:

    7 6 5 4 3 2 1 0
    2g 1a 1b 1c 1d 1e 1f 1g
    0 0 2a 2b 2c 2d 2e 2f

- three characters in three octets:
  - bits number:

7	6	5	4	3	2	1	0
2g	1a	1b	1c	1d	1e	1f	1g
3f	3g	2a	2b	2c	2d	2e	2f
0	0	0	3a	3b	3c	3d	3e

- seven characters in seven octets:
  - bits number:

7	6	5	4	3	2	1	0
2g	1a	1b	1c	1d	1e	1f	1g
3f	3g	2a	2b	2c	2d	2e	2f
4e	4f	4g	3a	3b	3c	3d	3e
5d	5e	5f	5g	4a	4b	4c	4d
6c	6d	6e	6f	6g	5a	5b	5c
7b	7c	7d	7e	7f	7g	6a	6b
0	0	0	0	0	0	0	7a

- eight characters in seven octets:
  - bits number:

7	6	5	4	3	2	1	0
2g	1a	1b	1c	1d	1e	1f	1g
3f	3g	2a	2b	2c	2d	2e	2f
4e	4f	4g	3a	3b	3c	3d	3e
5d	5e	5f	5g	4a	4b	4c	4d
6c	6d	6e	6f	6g	5a	5b	5c
7b	7c	7d	7e	7f	7g	6a	6b
8a	8b	8c	8d	8e	8f	8g	7a

The bit number zero is always transmitted first.

Therefore, in 140 octets, it is possible to pack (140x8)/7=160 characters.

6.1.2.2 CBS Packing

6.1.2.2.1 Packing of 7-bit characters

If a character number α is noted in the following way:

b7	b6	b5	b4	b3	b2	b1
αa	αb	αc	αd	αe	αf	αg

the packing of the 7-bits characters in octets is done as follows:

bit number

octet number	7	6	5	4	3	2	1	0
1	2g	1a	1b	1c	1d	1e	1f	1g
2	3f	3g	2a	2b	2c	2d	2e	2f
3	4e	4f	4g	3a	3b	3c	3d	3e
4	5d	5e	5f	5g	4a	4b	4c	4d
5	6c	6d	6e	6f	6g	5a	5b	5c
6	7b	7c	7d	7e	7f	7g	6a	6b
7	8a	8b	8c	8d	8e	8f	8g	7a
8	10g		9a	9b	9c	9d	9e	9f 9g
			.					
			.					
81	93d		93e		93f	93g	92a	92b 92c 92d
82	0	0	0	0	0	93a	93b	93c

The bit number zero is always transmitted first.

Therefore, in 82 octets, it is possible to pack (82x8)/7 = 93.7, that is 93 characters. The 5 remaining bits are set to zero as stated above.

6.1.2.3 USSD packing

6.1.2.3.1 Packing of 7 bit characters

If a character number α is noted in the following way:

b7	b6	b5	b4	b3	b2	b1
αa	αb	αc	αd	αe	αf	αg

The packing of the 7-bit characters in octets is done by completing the octets with zeros on the left.

For example, packing: α

- one character in one octet:
  - bits number:

7	6	5	4	3	2	1	0
0	1a	1b	1c	1d	1e	1f	1g

- two characters in two octets:
  - bits number:

7	6	5	4	3	2	1	0
2g	1a	1b	1c	1d	1e	1f	1g
0	0	2a	2b	2c	2d	2e	2f

- three characters in three octets:
  - bits number:

7	6	5	4	3	2	1	0
2g	1a	1b	1c	1d	1e	1f	1g
3f	3g	2a	2b	2c	2d	2e	2f
0	0	0	3a	3b	3c	3d	3e

- six characters in six octets:
  - bits number:

7	6	5	4	3	2	1	0
2g	1a	1b	1c	1d	1e	1f	1g
3f	3g	2a	2b	2c	2d	2e	2f
4e	4f	4g	3a	3b	3c	3d	3e
5d	5e	5f	5g	4a	4b	4c	4d
6c	6d	6e	6f	6g	5a	5b	5c
0	0	0	0	0	0	6a	6b

- seven characters in seven octets:
  - bits number:

7	6	5	4	3	2	1	0
2g	1a	1b	1c	1d	1e	1f	1g
3f	3g	2a	2b	2c	2d	2e	2f
4e	4f	4g	3a	3b	3c	3d	3e
5d	5e	5f	5g	4a	4b	4c	4d
6c	6d	6e	6f	6g	5a	5b	5c
7b	7c	7d	7e	7f	7g	6a	6b
0	0	0	1	1	0	1	7a

The bit number zero is always transmitted first.

- eight characters in seven octets:
  - bits number:

7	6	5	4	3	2	1	0
2g	1a	1b	1c	1d	1e	1f	1g
3f	3g	2a	2b	2c	2d	2e	2f
4e	4f	4g	3a	3b	3c	3d	3e
5d	5e	5f	5g	4a	4b	4c	4d
6c	6d	6e	6f	6g	5a	5b	5c
7b	7c	7d	7e	7f	7g	6a	6b
8a	8b	8c	8d	8e	8f	8g	7a

- nine characters in eight octets:
  - bits number:

7	6	5	4	3	2	1	0
2g	1a	1b	1c	1d	1e	1f	1g
3f	3g	2a	2b	2c	2d	2e	2f
4e	4f	4g	3a	3b	3c	3d	3e
5d	5e	5f	5g	4a	4b	4c	4d
6c	6d	6e	6f	6g	5a	5b	5c
7b	7c	7d	7e	7f	7g	6a	6b
8a	8b	8c	8d	8e	8f	8g	7a
0	9a	9b	9c	9d	9e	9f	9g

- fifteen characters in fourteen octets:
  - bits number:

7	6	5	4	3	2	1	0
2g	1a	1b	1c	1d	1e	1f	1g
3f	3g	2a	2b	2c	2d	2e	2f
4e	4f	4g	3a	3b	3c	3d	3e
5d	5e	5f	5g	4a	4b	4c	4d
6c	6d	6e	6f	6g	5a	5b	5c
7b	7c	7d	7e	7f	7g	6a	6b
8a	8b	8c	8d	8e	8f	8g	7a
10g	9a	9b	9c	9d	9e	9f	9g
11f	11g	10a	10b	10c	10d	10e	10f
12e	12f	12g	11a	11b	11c	11d	11e
13d	13e	13f	13g	12a	12b	12c	12d
14c	14d	14e	14f	14g	13a	13b	13c
15b	15c	15d	15e	15f	15g	14a	14b
0	0	0	1	1	0	1	15a

- sixteen characters in fourteen octets:
  - bits number:

7	6	5	4	3	2	1	0
2g	1a	1b	1c	1d	1e	1f	1g
3f	3g	2a	2b	2c	2d	2e	2f
4e	4f	4g	3a	3b	3c	3d	3e
5d	5e	5f	5g	4a	4b	4c	4d
6c	6d	6e	6f	6g	5a	5b	5c
7b	7c	7d	7e	7f	7g	6a	6b
8a	8b	8c	8d	8e	8f	8g	7a
10g	9a	9b	9c	9d	9e	9f	9g
11f	11g	10a	10b	10c	10d	10e	10f
12e	12f	12g	11a	11b	11c	11d	11e
13d	13e	13f	13g	12a	12b	12c	12d
14c	14d	14e	14f	14g	13a	13b	13c
15b	15c	15d	15e	15f	15g	14a	14b
16a	16b	16c	16d	16e	16f	16g	15a

The bit number zero is always transmitted first.

Therefore, in 160 octets, is it possible to pack (160*8)/7 = 182.8, that is 182 characters. The remaining 6 bits are set to zero as stated above.

Packing of 7 bit characters in USSD strings is done in the same way as for SMS (clause 6.1.2.1). The character stream is bit padded to octet boundary with binary zeroes as shown above.

If the total number of characters to be sent equals (8n-1) where n=1,2,3 etc. then there are 7 spare bits at the end of the message. To avoid the situation where the receiving entity confuses 7 binary zero pad bits as the @ character, the carriage return or <CR> character (defined in clause 6.1.1) shall be used for padding in this situation, just as for Cell Broadcast.

If <CR> is intended to be the last character and the message (including the wanted <CR>) ends on an octet boundary, then another <CR> must be added together with a padding bit 0. The receiving entity will perform the carriage return function twice, but this will not result in misoperation as the definition of <CR> in clause 6.1.1 is identical to the definition of <CR><CR>.

The receiving entity shall remove the final <CR> character where the message ends on an octet boundary with <CR> as the last character.

6.2 Character sets and coding

This section provides list of character sets and codings to be supported by SMS, CBS and USSD. Implementation of the GSM 7 bit default alphabet is mandatory. Support of other character sets is optional.

It should be noted that support of Latin and non-Latin languages by GSM 7 bit default alphabet is limited. It is therefore essential to introduce UCS 2 character set in mobile stations, SCs and systems handling SMSs, CBSs and USSDs.

6.2.1 GSM 7 bit Default Alphabet

Bits per character: 7

CBS/USSD pad character: CR

Character table:

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	@	Δ	SP	0	¡	P	¿	p
0	0	0	1	1	£	_	!	1	A	Q	a	q
0	0	1	0	2	$	Φ	"	2	B	R	b	r
0	0	1	1	3	¥	Γ	#	3	C	S	c	s
0	1	0	0	4	è	Λ	¤	4	D	T	d	t
0	1	0	1	5	é	Ω	%	5	E	U	e	u
0	1	1	0	6	ù	Π	&	6	F	V	f	v
0	1	1	1	7	ì	Ψ	'	7	G	W	g	w
1	0	0	0	8	ò	Σ	(	8	H	X	h	x
1	0	0	1	9	Ç	Θ	)	9	I	Y	i	y
1	0	1	0	10	LF	Ξ	*	:	J	Z	j	z
1	0	1	1	11	Ø	1)	+	;	K	Ä	k	ä
1	1	0	0	12	ø	Æ	,	<	L	Ö	l	ö
1	1	0	1	13	CR	æ	-	=	M	Ñ	m	ñ
1	1	1	0	14	Å	ß	.	>	N	Ü	n	ü
1	1	1	1	15	å	É	/	?	O	§	o	à

NOTE 1):	This code is an escape to an extension of this table (either to the GSM 7 bit default alphabet extension table, see subclause 6.2.1.1, or a National Language Single Shift Table, see subclause 6.2.1.2.2). A receiving entity which does not understand the meaning of this escape mechanism shall display it as a space character.

6.2.1.1 GSM 7 bit default alphabet extension table

The table below is reserved for symbols of international significance (e.g currency symbols). It also contains a mechanism to permit escape (Note 1) to additional tables for symbols of international significance in the event that the table below becomes fully populated.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0					|			
0	0	0	1	1								
0	0	1	0	2								
0	0	1	1	3								
0	1	0	0	4		^						
0	1	0	1	5							€	
0	1	1	0	6								
0	1	1	1	7								
1	0	0	0	8			{					
1	0	0	1	9			}					
1	0	1	0	10	3)							
1	0	1	1	11		1)						
1	1	0	0	12				[				
1	1	0	1	13				~				
1	1	1	0	14				]				
1	1	1	1	15			\					

In the event that an MS receives a code where a symbol is not represented in the above table then the MS shall display either the character shown in the main GSM 7 bit default alphabet table in subclause 6.2.1., or the character from the National Language Locking Shift Table in the case where the locking shift mechanism as defined in subclause 6.2.1.2.3 is used.

NOTE 1): This code is reserved for the extension to another extension table. On receipt of this code, a receiving entity shall display a space until another extension table is defined. It is not intended that this extension mechanism should be used as an alternative to UCS2 to enhance the 7bit default alphabet character repertoire for national specific character sets.

NOTE 2): Void

NOTE 3): This code is defined as a Page Break character and may be used for example in compressed CBS messages. Any mobile station which does not understand the GSM 7 bit default alphabet table extension mechanism will treat this character as Line Feed.

6.2.1.2 National Language Identifier

6.2.1.2.1 Introduction

The national language tables are used for adding the special characters of certain languages that cannot be expressed using the GSM default 7 bit alphabet.

The principle is to use the National Language Identifier to indicate to a receiving entity that the message has been encoded using a national language table. Both single shift and locking shift mechanisms are defined.

The single shift mechanism, as defined in subclause 6.2.1.2.2, applies to a single character and it replaces the GSM 7 bit default alphabet extension table defined in subclause 6.2.1.1 with a National Language Single Shift Table (see subclause A.2).

The locking shift mechanism, as defined in subclause 6.2.1.2.3, applies throughout the message, or the current segment in case of a concatenated message, and it replaces the GSM 7 bit default alphabet defined in subclause 6.2.1 with a National Language Locking Shift Table (see subclause A.3) that defines the whole character set needed for the language.

In case that several languages are used, which require different national language tables, it is recommended to encode the message in UCS-2, however it is possible to use both single shift and locking shift with the corresponding tables in a single message.

Implementations based on older reference versions (so-called "legacy implementations") will use the fallback mechanisms that are defined in the earlier versions of the specification for handling of unknown characters.

6.2.1.2.2 Single shift mechanism

In the case where single shift is not combined with locking shift, single shift means that the receiving entity shall decode all characters in the message (or the current segment in case of a concatenated message) using the GSM 7 bit default alphabet unless the escape mechanism is used, i.e <escape><character>, as defined in subclause 6.2.1.

The case where single shift and locking shift (which may be for the same or different languages) are combined is described in subclause 6.2.1.2.3.

If the escape mechanism is used then instead of the GSM 7 bit default alphabet extension table in subclause 6.2.1.1 the receiving entity shall decode the subsequent character using the National Language Single Shift Table for the indicated language in table 6.2.1.2.4.1. Each time a sending entity requires to send a character from the National Language Single Shift Table the sending entity shall encode this as <escape><character>, where the <character> is encoded using the indicated National Language Single Shift Table.

6.2.1.2.3 Locking shift mechanism

Locking Shift means that the receiving entity shall decode all characters in the message (or the current segment in case of a concatenated message) using the National Language Locking Shift Table unless the escape mechanism is used. i.e. <escape><character>, as defined in subclause 6.2.1.

If the escape mechanism is used and no National Language Single Shift Table is indicated (see subclause 6.2.1.2.4), the receiving entity shall decode the message (or the current segment in case of a concatenated message) using the GSM 7 bit default alphabet extension table as defined in subclause 6.2.1.1.

If the escape mechanism is used and a National Language Single Shift Table is indicated (see subclause 6.2.1.2.4), the receiving entity shall decode the message (or the current segment in case of a concatenated message) using the National Language Single Shift Table as defined in subclause 6.2.1.2.2.

6.2.1.2.4 National Language Identifier

A National Language Single Shift IE and a National Language Locking Shift IE can be included in the TP User Data Header, as defined in 3GPP TS 23.040 [4]. The receiving entity shall decode using single shift or locking shift as applicable for the language indicated in the National Language Identifier within these IEs.

The National Language Identifier octet is encoded as shown in table 6.2.1.2.4.1.

Table 6.2.1.2.4.1

Language code b7……b0	Language	National Language Single Shift Table	National Language Locking Shift Table
00000000	Reserved	n/a	n/a
00000001	Turkish	Subclause A.2.1	Subclause A.3.1
00000010	Spanish	Subclause A.2.2	Not defined – fallback to GSM 7 bit default alphabet (see subclause 6.2.1)
00000011	Portuguese	Subclause A.2.3	Subclause A.3.3
00000100	Bengali	Subclause A.2.4	Subclause A.3.4
00000101	Gujarati	Subclause A.2.5	Subclause A.3.5
00000110	Hindi	Subclause A.2.6	Subclause A.3.6
00000111	Kannada	Subclause A.2.7	Subclause A.3.7
00001000	Malayalam	Subclause A.2.8	Subclause A.3.8
00001001	Oriya	Subclause A.2.9	Subclause A.3.9
00001010	Punjabi	Subclause A.2.10	Subclause A.3.10
00001011	Tamil	Subclause A.2.11	Subclause A.3.11
00001100	Telugu	Subclause A.2.12	Subclause A.3.12
00001101	Urdu	Subclause A.2.13	Subclause A.3.13
00001110 to 11111111	Reserved	n/a	n/a

6.2.1.2.5 Processing of national language characters

When supporting a specific national language, the sending entity shall support the encoding of messages using the corresponding National Language Identifier defined in subclause 6.2.1.2.4.

The receiving entity should be able to decode messages usingthe National Language Identifiers defined in subclause 6.2.1.2.4 for the languages that are supported by that entity.

If a message is received, containing a National Language Identifier indicating a reserved value or a value that is not supported by the receiving entity, the receiving entity shall ignore the IE (see 3GPP TS 23.040 [4]) in which the National Language Identifier was indicated.

The receiving entity shall be capable of processing both single shift and locking shift within the same message.

It is an implementation option for the sending entity whether to use the single shift mechanism, the locking shift mechanism or both.

NOTE 1: A message using the locking shift mechanism cannot make use of characters from the GSM 7 bit Default Alphabet table unless such characters are replicated in the National Language Locking Shift Table or (in the case of locking shift and single shift), the National Language Single Shift table.

NOTE 2: Encoding of a message using the national locking shift mechanism is not intended to be implemented until a formal request is issued by the relevant national regulatory body. This is because a receiving entity not supporting the relevant locking-shift decoding will present different characters from the ones intended by the sending entity.

NOTE 3: An SMS message using a locking shift table for a language may not be properly displayed when the terminal does not support the locking shift table for that language. When the network is aware of the list of the locking shift tables supported by the UE, the network can deliver the SMS messages using an appropriate encoding.

6.2.2 8 bit data

8 bit data is user defined

Padding:	CR in the case of an 8 bit character set
	Otherwise - user defined
Character table:	User Specific

6.2.3 UCS2

Bits per character:	16
CBS/USSD pad character:	CR
Character table:	ISO/IEC 10646 [10]

Annex A (normative): National Language Tables

A.1 Introduction

This annex contains character tables for specific languages whose characters are not wholly or partially present within the GSM 7 bit default alphabet.

A.2 National Language Single Shift Tables

A.2.1 Turkish National Language Single Shift Table

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	B1		0	1	2	3	4	5	6	7
0	0	0	0	0								
0	0	0	1	1								
0	0	1	0	2								
0	0	1	1	3						Ş	Ç	ş
0	1	0	0	4		^						
0	1	0	1	5							€	
0	1	1	0	6								
0	1	1	1	7					Ğ		ğ	
1	0	0	0	8			{					
1	0	0	1	9			}		İ		ı	
1	0	1	0	10	3)							
1	0	1	1	11		1)						
1	1	0	0	12				[				
1	1	0	1	13	4)			~				
1	1	1	0	14				]				
1	1	1	1	15			\					

NOTE 1): This code is reserved for the extension to another extension table. On receipt of this code, a receiving entity shall display a space until another extension table is defined. NOTE 2): Void

NOTE 3): This code is defined as a Page Break character and may be used for example in compressed CBS messages. Any mobile station which does not understand the GSM 7 bit default alphabet table extension mechanism will treat this character as Line Feed.

NOTE 4): This code represents a control character and therefore must not be used for language specific characters.

A.2.2 Spanish National Language Single Shift Table

NOTE: This table also includes the character "ç" used in Catalan.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0					|			
0	0	0	1	1					Á		á	
0	0	1	0	2								
0	0	1	1	3								
0	1	0	0	4		^						
0	1	0	1	5						Ú	€	ú
0	1	1	0	6								
0	1	1	1	7								
1	0	0	0	8			{					
1	0	0	1	9	ç		}		Í		í	
1	0	1	0	10	3)							
1	0	1	1	11		1)						
1	1	0	0	12				[				
1	1	0	1	13	4)			~				
1	1	1	0	14				]				
1	1	1	1	15			\		Ó		ó	

NOTE 1): This code is reserved for the extension to another extension table. On receipt of this code, a receiving entity shall display a space until another extension table is defined.
NOTE 2): Void
NOTE 3): This code is defined as a Page Break character and may be used for example in compressed CBS messages. Any mobile station which does not understand the GSM 7 bit default alphabet table extension mechanism will treat this character as Line Feed.
NOTE 4): This code represents a control character and therefore must not be used for language specific characters.

A.2.3 Portuguese National Language Single Shift Table

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0					|			
0	0	0	1	1					À		Â	
0	0	1	0	2		Φ						
0	0	1	1	3		Γ						
0	1	0	0	4		^						
0	1	0	1	5	ê	Ω				Ú	€	ú
0	1	1	0	6		Π						
0	1	1	1	7		Ψ						
1	0	0	0	8		Σ	{					
1	0	0	1	9	ç	Θ	}		Í		í	
1	0	1	0	10	3)							
1	0	1	1	11	Ô	1)				Ã		ã
1	1	0	0	12	ô			[		Õ		õ
1	1	0	1	13	4)			~				
1	1	1	0	14	Á			]				
1	1	1	1	15	á	Ê	\		Ó		ó	â

NOTE 1): This code is reserved for the extension to another extension table. On receipt of this code, a receiving entity shall display a space until another extension table is defined.
NOTE 2): Void.
NOTE 3): This code is defined as a Page Break character and may be used for example in compressed CBS messages. Any mobile station which does not understand the GSM 7 bit default alphabet table extension mechanism will treat this character as Line Feed.
NOTE 4): This code represents a control character and therefore must not be used for language specific characters.

A.2.4 Bengali National Language Single Shift Table

NOTE: In the table below, the Bengali characters are represented using Unicode.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	@	<	09EC	09F6	|	P		
0	0	0	1	1	£	=	09ED	09F7	A	Q		
0	0	1	0	2	$	>	09EE	09F8	B	R		
0	0	1	1	3	¥	¡	09EF	09F9	C	S		
0	1	0	0	4	¿	^	09DF	09FA	D	T		
0	1	0	1	5	"	¡	09E0		E	U	€	
0	1	1	0	6	¤	_	09E1		F	V		
0	1	1	1	7	%	#	09E2		G	W		
1	0	0	0	8	&	*	{		H	X		
1	0	0	1	9	'	09E6	}		I	Y		
1	0	1	0	10	3)	09E7	09E3		J	Z		
1	0	1	1	11	*	1)	09F2		K			
1	1	0	0	12	+	09E8	09F3	[	L			
1	1	0	1	13	4)	09E9	09F4	~	M			
1	1	1	0	14	-	09EA	09F5	]	N			
1	1	1	1	15	/	09EB	\		O			

NOTE 1): This code is reserved for the extension to another extension table. On receipt of this code, a receiving entity shall display a space until another extension table is defined.
NOTE 2): Void
NOTE 3): This code is defined as a Page Break character and may be used for example in compressed CBS messages. Any mobile station which does not understand the GSM 7 bit default alphabet table extension mechanism will treat this character as Line Feed.
NOTE 4): This code represents a control character and therefore must not be used for language specific characters.

A.2.5 Gujarati National Language Single Shift Table

NOTE: In the table below, the Gujarati characters are represented using Unicode.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	@	<	0AEA		|	P		
0	0	0	1	1	£	=	0AEB		A	Q		
0	0	1	0	2	$	>	0AEC		B	R		
0	0	1	1	3	¥	¡	0AED		C	S		
0	1	0	0	4	¿	^	0AEE		D	T		
0	1	0	1	5	"	¡	0AEF		E	U	€	
0	1	1	0	6	¤	–			F	V		
0	1	1	1	7	%	#			G	W		
1	0	0	0	8	&	*	{		H	X		
1	0	0	1	9	'	0964	}		I	Y		
1	0	1	0	10	3)	0965			J	Z		
1	0	1	1	11	*	1)			K			
1	1	0	0	12	+	0AE6		[	L			
1	1	0	1	13	4)	0AE7		~	M			
1	1	1	0	14	-	0AE8		]	N			
1	1	1	1	15	/	0AE9	\		O			

NOTE 1): This code is reserved for the extension to another extension table. On receipt of this code, a receiving entity shall display a space until another extension table is defined.
NOTE 2): Void
NOTE 3): This code is defined as a Page Break character and may be used for example in compressed CBS messages. Any mobile station which does not understand the GSM 7 bit default alphabet table extension mechanism will treat this character as Line Feed.
NOTE 4): This code represents a control character and therefore must not be used for language specific characters.

A.2.6 Hindi National Language Single Shift Table

NOTE: In the table below, the Hindi characters are represented using Unicode.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	@	<	096A	095B	|	P		
0	0	0	1	1	£	=	096B	095C	A	Q		
0	0	1	0	2	$	>	096C	095D	B	R		
0	0	1	1	3	¥	¡	096D	095E	C	S		
0	1	0	0	4	¿	^	096E	095F	D	T		
0	1	0	1	5	"	¡	096F	0960	E	U	€	
0	1	1	0	6	¤	_	0951	0961	F	V		
0	1	1	1	7	%	#	0952	0962	G	W		
1	0	0	0	8	&	*	{	0963	H	X		
1	0	0	1	9	'	0964	}	0970	I	Y		
1	0	1	0	10	3)	0965	0953	0971	J	Z		
1	0	1	1	11	*	1)	0954		K			
1	1	0	0	12	+	0966	0958	[	L			
1	1	0	1	13	4)	0967	0959	~	M			
1	1	1	0	14	-	0968	095A	]	N			
1	1	1	1	15	/	0969	\		O			

NOTE 1): This code is reserved for the extension to another extension table. On receipt of this code, a receiving entity shall display a space until another extension table is defined.
NOTE 2): Void
NOTE 3): This code is defined as a Page Break character and may be used for example in compressed CBS messages. Any mobile station which does not understand the GSM 7 bit default alphabet table extension mechanism will treat this character as Line Feed.
NOTE 4): This code represents a control character and therefore must not be used for language specific characters.

A.2.7 Kannada National Language Single Shift Table

NOTE: In the table below, the Kannada characters are represented using Unicode.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	@	<	0CEA		|	P		
0	0	0	1	1	£	=	0CEB		A	Q		
0	0	1	0	2	$	>	0CEC		B	R		
0	0	1	1	3	¥	¡	0CED		C	S		
0	1	0	0	4	¿	^	0CEE		D	T		
0	1	0	1	5	"	¡	0CEF		E	U	€	
0	1	1	0	6	¤	_	0CDE		F	V		
0	1	1	1	7	%	#	0CF1		G	W		
1	0	0	0	8	&	*	{		H	X		
1	0	0	1	9	'	0964	}		I	Y		
1	0	1	0	10	3)	0965	0CF2		J	Z		
1	0	1	1	11	*	1)			K			
1	1	0	0	12	+	0CE6		[	L			
1	1	0	1	13	4)	0CE7		~	M			
1	1	1	0	14	-	0CE8		]	N			
1	1	1	1	15	/	0CE9	\		O			

NOTE 1): This code is reserved for the extension to another extension table. On receipt of this code, a receiving entity shall display a space until another extension table is defined.
NOTE 2): Void
NOTE 3): This code is defined as a Page Break character and may be used for example in compressed CBS messages. Any mobile station which does not understand the GSM 7 bit default alphabet table extension mechanism will treat this character as Line Feed.
NOTE 4): This code represents a control character and therefore must not be used for language specific characters.

A.2.8 Malayalam National Language Single Shift Table

NOTE: In the table below, the Malayalam characters are represented using Unicode.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	@	<	0D6A	0D7B	|	P		
0	0	0	1	1	£	=	0D6B	0D7C	A	Q		
0	0	1	0	2	$	>	0D6C	0D7D	B	R		
0	0	1	1	3	¥	¡	0D6D	0D7E	C	S		
0	1	0	0	4	¿	^	0D6E	0D7F	D	T		
0	1	0	1	5	"	¡	0D6F		E	U	€	
0	1	1	0	6	¤	_	0D70		F	V		
0	1	1	1	7	%	#	0D71		G	W		
1	0	0	0	8	&	*	{		H	X		
1	0	0	1	9	'	0964	}		I	Y		
1	0	1	0	10	3)	0965	0D72		J	Z		
1	0	1	1	11	*	1)	0D73		K			
1	1	0	0	12	+	0D66	0D74	[	L			
1	1	0	1	13	4)	0D67	0D75	~	M			
1	1	1	0	14	-	0D68	0D7A	]	N			
1	1	1	1	15	/	0D69	\		O			

NOTE 1): This code is reserved for the extension to another extension table. On receipt of this code, a receiving entity shall display a space until another extension table is defined.
NOTE 2): Void
NOTE 3): This code is defined as a Page Break character and may be used for example in compressed CBS messages. Any mobile station which does not understand the GSM 7 bit default alphabet table extension mechanism will treat this character as Line Feed.
NOTE 4): This code represents a control character and therefore must not be used for language specific characters.

A.2.9 Oriya National Language Single Shift Table

NOTE: In the table below, the Oriya characters are represented using Unicode.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	@	<	0B6A		|	P		
0	0	0	1	1	£	=	0B6B		A	Q		
0	0	1	0	2	$	>	0B6C		B	R		
0	0	1	1	3	¥	¡	0B6D		C	S		
0	1	0	0	4	¿	^	0B6E		D	T		
0	1	0	1	5	"	¡	0B6F		E	U	€	
0	1	1	0	6	¤	_	0B5C		F	V		
0	1	1	1	7	%	#	0B5D		G	W		
1	0	0	0	8	&	*	{		H	X		
1	0	0	1	9	'	0964	}		I	Y		
1	0	1	0	10	3)	0965	0B5F		J	Z		
1	0	1	1	11	*	1)	0B70		K			
1	1	0	0	12	+	0B66	0B71	[	L			
1	1	0	1	13	4)	0B67		~	M			
1	1	1	0	14	-	0B68		]	N			
1	1	1	1	15	/	0B69	\		O			

NOTE 1): This code is reserved for the extension to another extension table. On receipt of this code, a receiving entity shall display a space until another extension table is defined.
NOTE 2): Void
NOTE 3): This code is defined as a Page Break character and may be used for example in compressed CBS messages. Any mobile station which does not understand the GSM 7 bit default alphabet table extension mechanism will treat this character as Line Feed.
NOTE 4): This code represents a control character and therefore must not be used for language specific characters.

A.2.10 Punjabi National Language Single Shift Table

NOTE: In the table below, the Punjabi characters are represented using Unicode.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	@	<	0A6A		|	P		
0	0	0	1	1	£	=	0A6B		A	Q		
0	0	1	0	2	$	>	0A6C		B	R		
0	0	1	1	3	¥	¡	OA6D		C	S		
0	1	0	0	4	¿	^	0A6E		D	T		
0	1	0	1	5	"	¡	0A6F		E	U	€	
0	1	1	0	6	¤	_	0A59		F	V		
0	1	1	1	7	%	#	0A5A		G	W		
1	0	0	0	8	&	*	{		H	X		
1	0	0	1	9	'	0964	}		I	Y		
1	0	1	0	10	3)	0965	0A5B		J	Z		
1	0	1	1	11	*	1)	0A5C		K			
1	1	0	0	12	+	0A66	0A5E	[	L			
1	1	0	1	13	4)	0A67	0A75	~	M			
1	1	1	0	14	-	0A68		]	N			
1	1	1	1	15	/	0A69	\		O			

NOTE 1): This code is reserved for the extension to another extension table. On receipt of this code, a receiving entity shall display a space until another extension table is defined.
NOTE 2): Void
NOTE 3): This code is defined as a Page Break character and may be used for example in compressed CBS messages. Any mobile station which does not understand the GSM 7 bit default alphabet table extension mechanism will treat this character as Line Feed.
NOTE 4): This code represents a control character and therefore must not be used for language specific characters.

A.2.11 Tamil National Language Single Shift Table

NOTE: In the table below, the Tamil characters are represented using Unicode.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	@	<	0BEA		|	P		
0	0	0	1	1	£	=	0BEB		A	Q		
0	0	1	0	2	$	>	0BEC		B	R		
0	0	1	1	3	¥	¡	0BED		C	S		
0	1	0	0	4	¿	^	0BEF		D	T		
0	1	0	1	5	"	¡	0BEF		E	U	€	
0	1	1	0	6	¤	_	0BF3		F	V		
0	1	1	1	7	%	#	0BF4		G	W		
1	0	0	0	8	&	*	{		H	X		
1	0	0	1	9	'	0964	}		I	Y		
1	0	1	0	10	3)	0965	0BF5		J	Z		
1	0	1	1	11	*	1)	0BF6		K			
1	1	0	0	12	+	0BE6	0BF7	[	L			
1	1	0	1	13	4)	0BE7	0BF8	~	M			
1	1	1	0	14	-	0BE8	0BFA	]	N			
1	1	1	1	15	/	0BE9	\		O			

NOTE 1): This code is reserved for the extension to another extension table. On receipt of this code, a receiving entity shall display a space until another extension table is defined.
NOTE 2): Void
NOTE 3): This code is defined as a Page Break character and may be used for example in compressed CBS messages. Any mobile station which does not understand the GSM 7 bit default alphabet table extension mechanism will treat this character as Line Feed.
NOTE 4): This code represents a control character and therefore must not be used for language specific characters.

A.2.12 Telugu National Language Single Shift Table

NOTE: In the table below, the Telugu characters are represented using Unicode.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	@	<	0C6A	0C7D	|	P		
0	0	0	1	1	£	=	0C6B	0C7E	A	Q		
0	0	1	0	2	$	>	06CC	0C7F	B	R		
0	0	1	1	3	¥	¡	06CD		C	S		
0	1	0	0	4	¿	^	0C6E		D	T		
0	1	0	1	5	"	¡	0C6F		E	U		
0	1	1	0	6	¤	_	0C58		F	V		
0	1	1	1	7	%	#	0C59		G	W		
1	0	0	0	8	&	*	{		H	X		
1	0	0	1	9	'		}		I	Y		
1	0	1	0	10	3)		0C78		J	Z		
1	0	1	1	11	*	1)	0C79		K			
1	1	0	0	12	+	0C66	0C7A	[	L			
1	1	0	1	13	4)	0C67	0C7B	~	M			
1	1	1	0	14	-	0C68	0C7C	]	N			
1	1	1	1	15	/	0C69	\		O			

NOTE 1): This code is reserved for the extension to another extension table. On receipt of this code, a receiving entity shall display a space until another extension table is defined.
NOTE 2): Void
NOTE 3): This code is defined as a Page Break character and may be used for example in compressed CBS messages. Any mobile station which does not understand the GSM 7 bit default alphabet table extension mechanism will treat this character as Line Feed.
NOTE 4): This code represents a control character and therefore must not be used for language specific characters.

A.2.13 Urdu National Language Single Shift Table

NOTE: In the table below, the Urdu characters are represented using Unicode.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	@	<	06F4	0613	|	P		
0	0	0	1	1	£	=	06F5	0614	A	Q		
0	0	1	0	2	$	>	06F6	061B	B	R		
0	0	1	1	3	¥	¡	06F7	061F	C	S		
0	1	0	0	4	¿	^	06F8	0640	D	T		
0	1	0	1	5	"	¡	06F9	0652	E	U	€	
0	1	1	0	6	¤	_	060C	0658	F	V		
0	1	1	1	7	%	#	060D	066B	G	W		
1	0	0	0	8	&	*	{	066C	H	X		
1	0	0	1	9	'	0600	}	0672	I	Y		
1	0	1	0	10	3)	0601	060E	0673	J	Z		
1	0	1	1	11	*	1)	060F	06CD	K			
1	1	0	0	12	+	06F0	0610	[	L			
1	1	0	1	13	4)	06F1	0611	~	M			
1	1	1	0	14	-	06F2	0612	]	N			
1	1	1	1	15	/	06F3	\	06D4	O			

NOTE 1): This code is reserved for the extension to another extension table. On receipt of this code, a receiving entity shall display a space until another extension table is defined.
NOTE 2): Void
NOTE 3): This code is defined as a Page Break character and may be used for example in compressed CBS messages. Any mobile station which does not understand the GSM 7 bit default alphabet table extension mechanism will treat this character as Line Feed.
NOTE 4): This code represents a control character and therefore must not be used for language specific characters.

A.3 National Language Locking Shift Tables

A.3.1 Turkish National Language Locking Shift Table

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	@	Δ	SP	0	İ	P	ç	p
0	0	0	1	1	£	_	!	1	A	Q	a	q
0	0	1	0	2	$	Φ	"	2	B	R	b	r
0	0	1	1	3	¥	Γ	#	3	C	S	c	s
0	1	0	0	4	€	Λ	¤	4	D	T	d	t
0	1	0	1	5	é	Ω	%	5	E	U	e	u
0	1	1	0	6	ù	Π	&	6	F	V	f	v
0	1	1	1	7	ı	Ψ	'	7	G	W	g	w
1	0	0	0	8	ò	Σ	(	8	H	X	h	x
1	0	0	1	9	Ç	Θ	)	9	I	Y	i	y
1	0	1	0	10	LF	Ξ	*	:	J	Z	j	z
1	0	1	1	11	Ğ	1)	+	;	K	Ä	k	ä
1	1	0	0	12	ğ	Ş	,	<	L	Ö	l	ö
1	1	0	1	13	CR	ş	-	=	M	Ñ	m	ñ
1	1	1	0	14	Å	ß	.	>	N	Ü	n	ü
1	1	1	1	15	å	É	/	?	O	§	o	à

NOTE 1): This code is an escape to an extension of this table (either to the GSM 7 bit default alphabet extension table, see subclause 6.2.1.1, or a National Language Single Shift Table, see subclause 6.2.1.2.2). A receiving entity which does not understand the meaning of this escape mechanism shall display it as a space character.

A.3.2 Void

A.3.3 Portuguese National Language Locking Shift Table

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	@	Δ	SP	0	Í	P	~	p
0	0	0	1	1	£	_	!	1	A	Q	a	q
0	0	1	0	2	$	ª	"	2	B	R	b	r
0	0	1	1	3	¥	Ç	#	3	C	S	c	s
0	1	0	0	4	ê	À	º	4	D	T	d	t
0	1	0	1	5	é	∞	%	5	E	U	e	u
0	1	1	0	6	ú	^	&	6	F	V	f	v
0	1	1	1	7	í	\	'	7	G	W	g	w
1	0	0	0	8	ó	€	(	8	H	X	h	x
1	0	0	1	9	ç	Ó	)	9	I	Y	i	y
1	0	1	0	10	LF	|	*	:	J	Z	j	z
1	0	1	1	11	Ô	1)	+	;	K	Ã	k	ã
1	1	0	0	12	ô	Â	,	<	L	Õ	l	õ
1	1	0	1	13	CR	â	-	=	M	Ú	m	`
1	1	1	0	14	Á	Ê	.	>	N	Ü	n	ü
1	1	1	1	15	á	É	/	?	O	§	o	à

NOTE 1): This code is an escape to an extension of this table (either to the GSM 7 bit default alphabet extension table, see subclause 6.2.1.1, or a National Language Single Shift Table, see subclause 6.2.1.2.2). A receiving entity which does not understand the meaning of this escape mechanism shall display it as a space character.

A.3.4 Bengali National Language Locking Shift Table

NOTE: In the table below, the Bengali characters are represented using Unicode.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	0981	0990	SP	0	09AC	09BE	09CE	p
0	0	0	1	1	0982		!	1	09AD	09BF	a	q
0	0	1	0	2	0983		099F	2	09AE	09C0	b	r
0	0	1	1	3	0985	0993	09A0	3	09AF	09C1	c	s
0	1	0	0	4	0986	0994	09A1	4	09B0	09C2	d	t
0	1	0	1	5	0987	0995	09A2	5		09C3	e	u
0	1	1	0	6	0988	0996	09A3	6	09B2	09C4	f	v
0	1	1	1	7	0989	0997	09A4	7			g	w
1	0	0	0	8	098A	0998	)	8			h	x
1	0	0	1	9	098B	0999	(	9		09C7	i	y
1	0	1	0	10	LF	099A	09A5	:	09B6	09C8	j	z
1	0	1	1	11	098C	1)	09A6	;	09B7		k	09D7
1	1	0	0	12		099B	,		09B8		l	09DC
1	1	0	1	13	CR	099C	09A7	09AA	09B9	09CB	m	09DD
1	1	1	0	14		099D	.	09AB	09BC	09CC	n	09F0
1	1	1	1	15	098F	099E	09A8	?	09BD	09CD	o	09F1

NOTE 1): This code is an escape to an extension of this table (either to the GSM 7 bit default alphabet extension table, see subclause 6.2.1.1, or a National Language Single Shift Table, see subclause 6.2.1.2.2). A receiving entity which does not understand the meaning of this escape mechanism shall display it as a space character.

A.3.5 Gujarati National Language Locking Shift Table

NOTE: In the table below, the Gujarati characters are represented using Unicode.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	0A81	0A90	SP	0	0AAC	0ABE	0AD0	p
0	0	0	1	1	0A82	0A91	!	1	0AAD	0ABF	a	q
0	0	1	0	2	0A83		0A9F	2	0AAE	0AC0	b	r
0	0	1	1	3	0A85	0A93	0AA0	3	0AAF	0AC1	c	s
0	1	0	0	4	0A86	0A94	0AA1	4	0AB0	0AC2	d	t
0	1	0	1	5	0A87	0A95	0AA2	5		0AC3	e	u
0	1	1	0	6	0A88	0A96	0AA3	6	0AB2	0AC4	f	v
0	1	1	1	7	0A89	0A97	0AA4	7	0AB3	0AC5	g	w
1	0	0	0	8	0A8A	0A98	)	8			h	x
1	0	0	1	9	0A8B	0A99	(	9	0AB5	0AC7	i	y
1	0	1	0	10	LF	0A9A	0AA5	:	0AB6	0AC8	j	z
1	0	1	1	11	0A8C	1)	0AA6	;	0AB7	0AC9	k	0AE0
1	1	0	0	12	0A8D	0A9B	,		0AB8		l	0AE1
1	1	0	1	13	CR	0A9C	0AA7	0AAA	0AB9	0ACB	m	0AE2
1	1	1	0	14		0A9D	.	0AAB	0ABC	0ACC	n	0AE3
1	1	1	1	15	0A8F	0A9E	0AA8	?	0ABD	0ACD	o	0AF1

NOTE 1): This code is an escape to an extension of this table (either to the GSM 7 bit default alphabet extension table, see subclause 6.2.1.1, or a National Language Single Shift Table, see subclause 6.2.1.2.2). A receiving entity which does not understand the meaning of this escape mechanism shall display it as a space character.

A.3.6 Hindi National Language Locking Shift Table

NOTE: In the table below, the Hindi characters are represented using Unicode.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	0901	0910	SP	0	092C	093E	0950	p
0	0	0	1	1	0902	0911	!	1	092D	093F	a	q
0	0	1	0	2	0903	0912	091F	2	092E	0940	b	r
0	0	1	1	3	0905	0913	0920	3	092F	0941	c	s
0	1	0	0	4	0906	0914	0921	4	0930	0942	d	t
0	1	0	1	5	0907	0915	0922	5	0931	0943	e	u
0	1	1	0	6	0908	0916	0923	6	0932	0944	f	v
0	1	1	1	7	0909	0917	0924	7	0933	0945	g	w
1	0	0	0	8	090A	0918	)	8	0934	0946	h	x
1	0	0	1	9	090B	0919	(	9	0935	0947	i	y
1	0	1	0	10	LF	091A	0925	:	0936	0948	j	z
1	0	1	1	11	090C	1)	0926	;	0937	0949	k	0972
1	1	0	0	12	090D	091B	,	0929	0938	094A	l	097B
1	1	0	1	13	CR	091C	0927	092A	0939	094B	m	097C
1	1	1	0	14	090E	091D	.	092B	093C	094C	n	097E
1	1	1	1	15	090F	091E	0928	?	093D	094D	o	097F

NOTE 1): This code is an escape to an extension of this table (either to the GSM 7 bit default alphabet extension table, see subclause 6.2.1.1, or a National Language Single Shift Table, see subclause 6.2.1.2.2). A receiving entity which does not understand the meaning of this escape mechanism shall display it as a space character.

A.3.7 Kannada National Language Locking Shift Table

NOTE: In the table below, the Kannada characters are represented using Unicode.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0		0C90	SP	0	0CAC	0CBE	0CD5	p
0	0	0	1	1	0C82		!	1	0CAD	0CBF	a	q
0	0	1	0	2	0C83	0C92	0C9F	2	0CAE	0CC0	b	r
0	0	1	1	3	0C85	0C93	0CA0	3	0CAF	0CC1	c	s
0	1	0	0	4	0C86	0C94	0CAA	4	0CB0	0CC2	d	t
0	1	0	1	5	0C87	0C95	0CA2	5	0CB1	0CC3	e	u
0	1	1	0	6	0C88	0C96	0CA3	6	0CB2	0CC4	f	v
0	1	1	1	7	0C89	0C97	0CA4	7	0CB3		g	w
1	0	0	0	8	0C8A	0C98	)	8		0CC6	h	x
1	0	0	1	9	0C8B	0C99	(	9	0CB5	0CC7	i	y
1	0	1	0	10	LF	0C9A	0CA5	:	0CB6	0CC8	j	z
1	0	1	1	11	0C8C	1)	0CA6	;	0CB7		k	0CD6
1	1	0	0	12		0C9B	,		0CB8	0CCA	l	0CE0
1	1	0	1	13	CR	0C9C	0CA7	0CAA	0CB9	0CCB	m	0CE1
1	1	1	0	14	0C8E	0C9D	.	0CAB	0CBC	0CCC	n	0CE2
1	1	1	1	15	0C8F	0C9E	0CA8	?	0CBD	0CCD	o	0CE3

NOTE 1): This code is an escape to an extension of this table (either to the GSM 7 bit default alphabet extension table, see subclause 6.2.1.1, or a National Language Single Shift Table, see subclause 6.2.1.2.2). A receiving entity which does not understand the meaning of this escape mechanism shall display it as a space character.

A.3.8 Malayalam National Language Locking Shift Table

NOTE: In the table below, the Malayalam characters are represented using Unicode.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0		0D10	SP	0	0D2C	0D3E	0D57	p
0	0	0	1	1	0D02		!	1	0D2D	0D3F	a	q
0	0	1	0	2	0D03	0D12	0D1F	2	0D2E	0D40	b	r
0	0	1	1	3	0D05	0D13	0D20	3	0D2F	0D41	c	s
0	1	0	0	4	0D06	0D14	0D21	4	0D30	0D42	d	t
0	1	0	1	5	0D07	0D15	0D22	5	0D31	0D43	e	u
0	1	1	0	6	0D08	0D16	0D23	6	0D32	0D44	f	v
0	1	1	1	7	0D09	0D17	0D24	7	0D33		g	w
1	0	0	0	8	0D0A	0D18	)	8	0D34	0D46	h	x
1	0	0	1	9	0D0B	0D19	(	9	0D35	0D47	i	y
1	0	1	0	10	LF	0D1A	0D25	:	0D36	0D48	j	z
1	0	1	1	11	0D0C	1)	0D26	;	0D37		k	0D60
1	1	0	0	12		0D1B	,		0D38	0D4A	l	0D61
1	1	0	1	13	CR	0D1C	0D27	0D2A	0D39	0D4B	m	0D62
1	1	1	0	14	0D0E	0D1D	.	0D2B		0D4C	n	0D63
1	1	1	1	15	0D0F	0D1E	0D28	?	0D3D	0D4D	o	0D79

NOTE 1): This code is an escape to an extension of this table (either to the GSM 7 bit default alphabet extension table, see subclause 6.2.1.1, or a National Language Single Shift Table, see subclause 6.2.1.2.2). A receiving entity which does not understand the meaning of this escape mechanism shall display it as a space character.

A.3.9 Oriya National Language Locking Shift Table

NOTE: In the table below, the Oriya characters are represented using Unicode.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	0B01	0B10	SP	0	0B2C	0B3E	0B56	p
0	0	0	1	1	0B02		!	1	0B2D	0B3F	a	q
0	0	1	0	2	0B03		0B1F	2	0B2E	0B40	b	r
0	0	1	1	3	0B05	0B13	0B20	3	0B2F	0B41	c	s
0	1	0	0	4	0B06	0B14	0B21	4	0B30	0B42	d	t
0	1	0	1	5	0B07	0B15	0B22	5		0B43	e	u
0	1	1	0	6	0B08	0B16	0B23	6	0B32	0B44	f	v
0	1	1	1	7	0B09	0B17	0B24	7	0B33		g	w
1	0	0	0	8	0B0A	0B18	)	8			h	x
1	0	0	1	9	0B0B	0B19	(	9	0B35	0B47	i	y
1	0	1	0	10	LF	0B1A	0B25	:	0B36	0B48	j	z
1	0	1	1	11	0B0C	1)	0B26	;	0B37		k	0B57
1	1	0	0	12		0B1B	,		0B38		l	0B60
1	1	0	1	13	CR	0B1C	0B27	0B2A	0B39	0B4B	m	0B61
1	1	1	0	14		0B1D	.	0B2B	0B3C	0B4C	n	0B62
1	1	1	1	15	0B0F	0B1E	0B28	?	0B3D	0B4D	o	0B63

NOTE 1): This code is an escape to an extension of this table (either to the GSM 7 bit default alphabet extension table, see subclause 6.2.1.1, or a National Language Single Shift Table, see subclause 6.2.1.2.2). A receiving entity which does not understand the meaning of this escape mechanism shall display it as a space character.

A.3.10 Punjabi National Language Locking Shift Table

NOTE: In the table below, the Punjabi characters are represented using Unicode.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	0A01	0A10	SP	0	0A2C	0A3E	0A51	p
0	0	0	1	1	0A02		!	1	0A2D	0A3F	a	q
0	0	1	0	2	0A03		0A1F	2	0A2E	0A40	b	r
0	0	1	1	3	0A05	0A13	0A20	3	0A2F	0A41	c	s
0	1	0	0	4	0A06	0A14	0A21	4	0A30	0A42	d	t
0	1	0	1	5	0A07	0A15	0A22	5			e	u
0	1	1	0	6	0A08	0A16	0A23	6	0A32		f	v
0	1	1	1	7	0A09	0A17	0A24	7	0A33		g	w
1	0	0	0	8	0A0A	0A18	)	8			h	x
1	0	0	1	9		0A19	(	9	0A35	0A47	i	y
1	0	1	0	10	LF	0A1A	0A25	:	0A36	0A48	j	z
1	0	1	1	11		1)	0A26	;			k	0A70
1	1	0	0	12		0A1B	,		0A38		l	0A71
1	1	0	1	13	CR	0A1C	0A27	0A2A	0A39	0A4B	m	0A72
1	1	1	0	14		0A1D	.	0A2B	0A3C	0A4C	n	0A73
1	1	1	1	15	0A0F	0A1E	0A28	?		0A4D	o	0A74

NOTE 1): This code is an escape to an extension of this table (either to the GSM 7 bit default alphabet extension table, see subclause 6.2.1.1, or a National Language Single Shift Table, see subclause 6.2.1.2.2). A receiving entity which does not understand the meaning of this escape mechanism shall display it as a space character.

A.3.11 Tamil National Language Locking Shift Table

NOTE: In the table below, the Tamil characters are represented using Unicode.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0		0B90	SP	0		0BBE	0BD0	p
0	0	0	1	1	0B82		!	1		0BBF	a	q
0	0	1	0	2	0B83	0B92	0B9F	2	0BAE	0BC0	b	r
0	0	1	1	3	0B85	0B93		3	0BAF	0BC1	c	s
0	1	0	0	4	0B86	0B94		4	0BB0	0BC2	d	t
0	1	0	1	5	0B87	0B95		5	0BB1		e	u
0	1	1	0	6	0B88		0BA3	6	0BB2		f	v
0	1	1	1	7	0B89		0BA4	7	0BB3		g	w
1	0	0	0	8	0B8A		)	8	0BB4	0BC6	h	x
1	0	0	1	9		0B99	(	9	0BB5	0BC7	i	y
1	0	1	0	10	LF	0B9A		:	0BB6	0BC8	j	z
1	0	1	1	11		1)		;	0BB7		k	0BD7
1	1	0	0	12			,	0BA9	0BB8	0BCA	l	0BF0
1	1	0	1	13	CR	0B9C		0BAA	0BB9	0BCB	m	0BF1
1	1	1	0	14	0B8E		.			0BCC	n	0BF2
1	1	1	1	15	0B8F	0B9E	0BA8	?		0BCD	o	0BF9

NOTE 1): This code is an escape to an extension of this table (either to the GSM 7 bit default alphabet extension table, see subclause 6.2.1.1, or a National Language Single Shift Table, see subclause 6.2.1.2.2). A receiving entity which does not understand the meaning of this escape mechanism shall display it as a space character.

A.3.12 Telugu National Language Locking Shift Table

NOTE: In the table below, the Telugu characters are represented using Unicode.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	0C01	0C10	SP	0	0C2C	0C3E	0C55	p
0	0	0	1	1	0C02		!	1	0C2D	0C3F	a	q
0	0	1	0	2	0C03	0C12	0C1F	2	0C2E	0C40	b	r
0	0	1	1	3	0C05	0C13	0C20	3	0C2F	0C41	c	s
0	1	0	0	4	0C06	0C14	0C21	4	0C30	0C42	d	t
0	1	0	1	5	0C07	0C15	0C22	5	0C31	0C43	e	u
0	1	1	0	6	0C08	0C16	0C23	6	0C32	0C44	f	v
0	1	1	1	7	0C09	0C17	0C24	7	0C33		g	w
1	0	0	0	8	0C0A	0C18	)	8		0C46	h	x
1	0	0	1	9	0C0B	0C19	(	9	0C35	0C47	i	y
1	0	1	0	10	LF	0C1A	0C25	:	0C36	0C48	j	z
1	0	1	1	11	0C0C	1)	0C26	;	0C37		k	0C56
1	1	0	0	12		0C1B	,		0C38	0C4A	l	0C60
1	1	0	1	13	CR	0C1C	0C27	0C2A	0C39	0C4B	m	0C61
1	1	1	0	14	0C0E	0C1D	.	0C2B		0C4C	n	0C62
1	1	1	1	15	0C0F	0C1E	0C28	?	0C3D	0C4D	o	0C63

NOTE 1): This code is an escape to an extension of this table (either to the GSM 7 bit default alphabet extension table, see subclause 6.2.1.1, or a National Language Single Shift Table, see subclause 6.2.1.2.2). A receiving entity which does not understand the meaning of this escape mechanism shall display it as a space character.

A.3.13 Urdu National Language Locking Shift Table

NOTE: In the table below, the Urdu characters are represented using Unicode.

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0	0627	062B	SP	0	0635	06BA	0654	p
0	0	0	1	1	0622	062C	!	1	0636	06BB	a	q
0	0	1	0	2	0628	0681	068F	2	0637	06BC	b	r
0	0	1	1	3	067B	0684	068D	3	0638	0648	c	s
0	1	0	0	4	0680	0683	0630	4	0639	06C4	d	t
0	1	0	1	5	067E	0685	0631	5	0641	06D5	e	u
0	1	1	0	6	06A6	0686	0691	6	0642	06C1	f	v
0	1	1	1	7	062A	0687	0693	7	06A9	06BE	g	w
1	0	0	0	8	06C2	062D	)	8	06AA	0621	h	x
1	0	0	1	9	067F	062E	(	9	06AB	06CC	i	y
1	0	1	0	10	LF	062F	0699	:	06AF	06D0	j	z
1	0	1	1	11	0679	1)	0632	;	06B3	06D2	k	0655
1	1	0	0	12	067D	068C	,	069A	06B1	064D	l	0651
1	1	0	1	13	CR	0688	0696	0633	0644	0650	m	0653
1	1	1	0	14	067A	0689	.	0634	0645	064F	n	0656
1	1	1	1	15	067C	068A	0698	?	0646	0657	o	0670

NOTE 1): This code is an escape to an extension of this table (either to the GSM 7 bit default alphabet extension table, see subclause 6.2.1.1, or a National Language Single Shift Table, see subclause 6.2.1.2.2). A receiving entity which does not understand the meaning of this escape mechanism shall display it as a space character.

Annex B (informative): Guidelines for creating language tables

B.1 Introduction

This annex provides guidelines for creating language tables.

It is recommended that the characters and their positions in the table are checked by people fluent in the appropriate language, and preferably endorsed by an appropriate responsible body.

It is recommended that character positions are carefully selected so that receiving entities, which do not support the specific table, display symbols (glyphs) similar to the wanted symbols (glyphs) as far as possible.

B.2 Template for Single Shift Language Tables

The format and structure of the table below shall be used to document the Language specific character codes used in the National Language selection mechanism.

It is recommended that a National Language Single Shift Table includes the characters represented in the GSM 7 bit default alphabet extension table (as defined in subclause 6.2.1.1) in the same character position. This ensures the availability of these characters in case when the single shift mechanism is used.

Language – (Note. The actual Country and table content will be annotated when the country is known).

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0								
0	0	0	1	1								
0	0	1	0	2								
0	0	1	1	3								
0	1	0	0	4								
0	1	0	1	5								
0	1	1	0	6								
0	1	1	1	7								
1	0	0	0	8								
1	0	0	1	9								
1	0	1	0	10	3)							
1	0	1	1	11		1)						
1	1	0	0	12								
1	1	0	1	13	4)							
1	1	1	0	14								
1	1	1	1	15								

NOTE 1): This code is reserved for the extension to another extension table. On receipt of this code, a receiving entity shall display a space until another extension table is defined.
NOTE 2): Void
NOTE 3): This code is defined as a Page Break character and may be used for example in compressed CBS messages. Any mobile station which does not understand the GSM 7 bit default alphabet table extension mechanism will treat this character as Line Feed.
NOTE 4): This code represents a control character and therefore must not be used for language specific characters.

B.3 Template for Locking Shift Language Tables

The format and structure of the table below shall be used to document the Language specific character codes used in the National Language selection mechanism.

Language – (Note. The actual Country and table content will be annotated when the country is known).

				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1		0	1	2	3	4	5	6	7
0	0	0	0	0			SP					
0	0	0	1	1								
0	0	1	0	2								
0	0	1	1	3								
0	1	0	0	4								
0	1	0	1	5								
0	1	1	0	6								
0	1	1	1	7								
1	0	0	0	8								
1	0	0	1	9								
1	0	1	0	10	LF							
1	0	1	1	11		1)						
1	1	0	0	12								
1	1	0	1	13	CR							
1	1	1	0	14								
1	1	1	1	15								

NOTE 1): This code is an escape to an extension of this table (either to the GSM 7 bit default alphabet extension table, see subclause 6.2.1.1, or a National Language Single Shift Table, see subclause 6.2.1.2.2). A receiving entity which does not understand the meaning of this escape mechanism shall display it as a space character.

Annex C (Informative): Example for locking shift and single shift mechanisms

C.1 Introduction

This annex gives an overview on how the national language extension mechanism of the GSM 7 bit default alphabet works. This annex shows how a message with an indication of the Turkish National Language Identifier is decoded, but the same principles apply to other languages.

C.2 Example of single shift

This example outlines the behaviour of both supporting and non-supporting receiving entities where the Turkish National Language Single Shift Table is indicated in the received message. In this example there is no locking shift mechanism used in parallel.

A non-supporting receiving entity will ignore the National Language Single Shift IE, and decode the message contents using the GSM 7 bit default alphabet table defined in subclause 6.2.1, including possible escape characters to the GSM 7 bit default alphabet extension table specified in subclause 6.2.1.1. For example the Turkish word "Türkçe" will be displayed as "Türkce".

A receiving entity that supports the Turkish National Language Single Shift Table will detect a National Language Single Shift IE in a TP User Data Header. This IE tells the receiving entity that the single shift mechanism is used.

A supporting receiving entity will notice the language code, in this example coded as '0000 0001', and therefore use the Turkish National Language Single Shift Table defined in subclause A.2.1 instead of the GSM 7 bit default alphabet extension table defined in subclause 6.2.1.

If the next character is any character except <escape>, then the GSM 7 bit default alphabet table is used for the decode. If the next character is <escape> then the Turkish language specific table is used for the decode of the one character that follows the <escape>. This process will be repeated until the end of the received message, or until the end of the current segment of a concatenated message.

The Language selection at the start of a message takes 4 octets which correspond to five 7 bit characters which reduces the maximum number of characters per single message to 155.

Thereafter, the number of characters within that single message will be dependent upon the number of times a character is used that is within the National Language Single Shift Table.

Every character used from the National Language Single Shift Table will need an additional character to identify the escape to the National Language Single Shift Table. The available 155 character capacity of a single message will therefore be reduced accordingly. This reduction of overall message length also applies when using characters from the GSM 7 bit default alphabet extension table (see subclause 6.2.1.1) when the National Language Single Shift IE is not used.

C.3 Example of locking shift

This example outlines the behaviour of both supporting and non-supporting receiving entities where the Turkish National Language Locking Shift Table is indicated in the received message.

A non-supporting receiving entity will ignore the National Language Locking Shift IE, and decode the message contents using the GSM 7 bit default alphabet defined in subclause 6.2.1, including possible escape characters to the GSM 7 bit default alphabet extensions specified in subclause 6.2.1.1.

A receiving entity that supports the scheme will detect a National Language Locking Shift IE in a TP User Data Header. This IE tells the receiving entity that the locking shift mechanism is used. If no National Language Single Shift IE is indicated additionally to the National Language Locking Shift IE, then the whole message is decoded using the National Language Locking Shift Table defined for Turkish language in subclause 6.2.1.2.4.1.

If, in addition to the National Language Locking Shift IE (which may be for Turkish or another language), a National Language Single Shift IE for the Turkish language is indicated, then <escape> makes an exception to the use of the National Language Locking Shift Table for the Turkish or another language. In that case a character following <escape> is decoded using the National Language Single Shift Table for the Turkish language, after which the use of the National Language Locking Shift Table for the Turkish or another language is resumed until the next <escape> or the end of the message is met.

The Language selection at the start of a message takes 4 octets which corresponds to five 7 bit characters which reduces the maximum number of characters per single message to 155, unless the National Language Single Shift IE has also been included, in which case there is a further reduction of 3 octets making 7 octets in total, which corresponds to eight 7 bit characters, which reduces the maximum number of characters per single message to 152.

Thereafter, if the single shift mechanism is used additionally to the locking shift mechanism, the number of characters within that single message will be dependent upon the number of times a character is used that is within the National Language Single Shift Table.

Every character in the National Language Single Shift Table will use an additional character. The available 152 character single message length will therefore be reduced accordingly. This reduction of overall message length also applies when using characters from the GSM 7 bit default alphabet extension table (see subclause 6.2.1.1) when the National Language Single Shift IE is not used.

Annex D (informative): Document change history

TSG#	TDoc	VERS	NEW_ VERS	CR	REV	Rel	CAT	WORK ITEM	SUBJECT
T#4			3.0.0	New					Creation of 3GPP TS 23.038 v1.0.0 out of GSM 03.38 v7.1.0
T#4	TP-99124	3.0.0	3.1.0	001		R99	A	MExE	Data Coding Scheme for WAP over USSD and CB
T#5	TP-99177	3.1.0	3.2.0	002		R99	B	TEI	Language codes for Hebrew,Arabic and Russian
T#6	TP-99237	3.2.0	3.3.0	003		R99	F	TEI	Adaptations for UMTS
T#8	TP-000074	3.3.0	4.0.0	004		Rel4	B	TEI	Automatic removal of 'read' SMS
T#10	TP-000195	4.0.0	4.1.0	005		Rel4	B	TEI	Data coding scheme value for the Icelandic language
T#11	TP-010029	4.1.0	4.2.0	006		Rel4	C	UICC1-CPHS	Message Waiting Indication Status storage on the USIM
T#13	TP-010194	4.2.0	4.3.0	007		Rel4	F	TEI4	Support to UCS2 and editorial corrections
T#14	TP-010280	4.3.0	4.4.0	008		Rel4	F	TEI4	Deletion of GSM 01.04 reference
T#15	TP-020015	4.4.0	5.0.0	009		Rel5	F	TEI5	User Data Header support over CBS
T#21	TP-030173	5.0.0	6.0.0	010		Rel6	C	TEI6	Additional Indications in SMS DCS
T#25	TP-040205	6.0.0	6.1.0	013	2	Rel6	F	TEI6	Message Waiting Indication – how to handle Multiple Subscriber Profiles
T#25	TP-040171	6.0.0	6.1.0	014		Rel6	F	TEI6	Enhanced Voice Mail Information – not applicable for CBS
CT#31	CP-060126	6.1.0	7.0.0	015r1		Rel-7	F	TEI7	CBS – Reference removal
CT-39	CP-080223	7.0.0	8.0.0	0017	5	Rel-8	B	TEI8	SMS default alphabet. Generic solution for all languages
CT-39	CP-080138	7.0.0	8.0.0	0019		Rel-8	B	TEI8	SMS-addition of turkish national language locking shift table
CT-40	CP-080361	8.0.0	8.1.0	0020		Rel-8	F	TEI8	Corrections to single shift language tables for Turkish and Spanish
CT-40	CP-080361	8.0.0	8.1.0	0021	3	Rel-8	F	TEI8	Addition of national language tables for Portuguese
CT-41	CP-080536	8.1.0	8.2.0	0229	1	Rel-8	F	TEI8	Clarification of Locking / Single shift IE's for different languages in a single SM
CT-45	CP-090682	8.2.0	9.0.0	0231	3	Rel-9	B	TEI9	Addition of language tables for India
CT-45	CP-090682	8.2.0	9.0.0	0232		Rel-9	F	TEI9	CBS Message Class 0
CT-46	CP-090912	9.0.0	9.1.0	0235	3	Rel-9	F	ETWS	Define the use of data coding scheme for the ETWS warning message
CT-46	CP-090925	9.0.0	9.1.0	0236		Rel-9	B	IMS_SCC-ICS_I1	Uniquely identify the I1 protocol in USSD
		9.1.0	9.1.1			Rel-9			Correction of a typo error in the change history table (wrong version number)
CT-51		9.1.1	10.0.0			Rel-10			Upgrade to Rel-10
CT-57		10.0.0	11.0.0			Rel-11			Upgrade to Rel-11