interweb

Generated on 16/10/2026 16:41

KNOWLEDGE BASE · PUBLIC DOCUMENT

GSM-7 or Unicode in SMS: why one accent splits a message into more parts

WhatsApp & APIs

CODE	VERSION	COUNTRY · LANGUAGE
RLX-MZ-KB-2158	v2026.10.16	Mozambique · English

Interweb Mozambique
RLX Holdings, S.A. · Group RLX Holdings
Tax ID (NUIT) 401 191 607 · Legal entity no. (NUEL) 105 045 974 · Economic activity code (CAE) 62022

https://www.meu.interweb.co.mz/knowledgebase/2158

The short answer (for developers): an SMS carries a fixed number of bytes (140). With the **GSM-7** alphabet that is 160 characters. A **single** character outside that alphabet (ã, õ, á, ê, an emoji, a curly quote) switches the whole message to **UCS-2** (Unicode), where only 70 fit. A longer message is split into parts that the phone puts back together (concatenated SMS), and each part loses room to the header that joins them: 153 characters per part in GSM-7 and 67 in UCS-2. In Portuguese, everyday words like «não», «código» or «está» are enough to change the count.

What each character does to the message

What you write	What happens
Unaccented letters, digits, common punctuation, and é, è, à, ù, ì, ò, ñ, ü, Ç	GSM-7: 160 per message. These accented letters are part of the GSM alphabet.
ã, õ, á, í, ó, ú, â, ê, ô	Not in GSM-7: the whole message becomes UCS-2, 70 per message.
Lower-case ç	GSM-7 has Ç. Some providers turn ç into Ç, others switch the message to Unicode: it depends on the provider.
`€ [ ] { } ^ ~ \ |`	They live in the GSM-7 extension: each counts as two characters.
Emoji	UCS-2, and many take two positions, so they use twice the room.
Curly quotes, the long dash and non-breaking spaces pasted from a document	Outside GSM-7: they switch the message to Unicode without anyone noticing.

Keeping the size under control

1. **Choose the alphabet per message:** for short notices and codes you can write without the accents that force Unicode; for a message to customers, accept Unicode and write shorter. «nao» for «não» is fine in a code, but in a notice it looks careless.

2. **Count after filling in the variables**, not on the template. Check in code whether every character is in the GSM-7 alphabet; if one is not, the Unicode limit applies. Many provider APIs return the number of parts in the reply: read it and log it.

3. **Clean the text in code** before sending: curly quotes to straight ones, the long dash to a hyphen, non-breaking spaces to normal spaces. In PHP, for example:

```
$t = str_replace(
["\u{201C}", "\u{201D}", "\u{2018}", "\u{2019}", "\u{2014}",
"\u{00A0}"],
['"', '"', "'", "'", "-", " "],
$t);
```

4. **With SMPP, encoding is your job.** The **data_coding** field says how the text is sent: 0 is the SMS centre’s default alphabet (usually GSM-7) and 8 is UCS-2. Encode the text to match; a wrong data_coding shows question marks or scrambled letters on the phone. Long messages are split with a UDH header (flagged in esm_class) or sent in the message_payload field, whichever the provider prefers.

5	**With an HTTP API,** the provider splits and encodes for you; some ask you to flag the text as Unicode. Always send UTF-8 and read in their documentation what they do with characters they do not know.

6	**Test on a real phone**, with the real text and an accented customer name inside it.

An accented customer name in a template counted as GSM-7 turns one message into two or three. Names, addresses and product names are the variables that most often bring in Unicode without warning.

If the text already comes out of the database with question marks, the problem is before the SMS: **accents turning into strange symbols**. And the name shown as the sender has rules of its own: **SMS marketing and Sender ID**.

Accents already look broken on your site or in the database? Open a ticket and tell us where you see them.

Open a support ticket

SEE ALSO

SMS marketing: Sender ID, the rules here, and how to write one that works

Accents turning into strange symbols on your site: UTF-8 from file to database

Character sets: utf8mb4, collation, and the question marks in your text

SMPP or an HTTP SMS API: what each one is and which your application needs

Sending one-time codes (OTP) by SMS safely: expiry, attempt limits and logs

WhatsApp, e-mail or SMS: advantages, disadvantages, risks and who each is for

interweb

Online version of this document

Interweb is a leading internet services company, with a mission to help businesses to be born and prosper through technology.

2020	37,000+	130+
FOUNDED	CLIENTS	SIGN-UPS A DAY

Interweb Mozambique

RLX Holdings, S.A. · Group RLX Holdings

TAX ID (NUIT)	401 191 607
LEGAL ENTITY NO. (NUEL)	105 045 974
ECONOMIC ACTIVITY CODE (CAE)	62022
REGISTERED OFFICE	Bairro 7 de Setembro Vilankulo 1304 · CEP 0411-02, Inhambane, Mozambique
CONTACT	info@interweb.co.mz · +258 84 635 1111 · +258 86 635 1111 · www.interweb.co.mz
OPENING HOURS	Monday to Friday - 08 AM to 04 PM Saturday & Sunday - Closed

RLX-MZ-KB-2158 · v2026.10.16 · Generated on 16/10/2026 16:41