



SMS encoding and message length

Why one extra character can double the cost of a message, and how to keep SMS text within the length you expect.

Educational guide. General information about messaging integrations. It does not describe a live Wise Network API, and examples are illustrative only.

Online version: <https://wisenetwork.co/support/sms-encoding/> · Content updated 28 October 2026

Two main encodings

A single SMS carries up to 140 bytes of user data. How many characters fit depends on the encoding, which is defined in 3GPP specification [TS 23.038](#).

Characters per SMS by encoding

Encoding	Used for	Single SMS	Per segment when split
GSM-7 (GSM default alphabet)	Latin letters, digits, common punctuation and some accented letters	160 characters	153 characters
UCS-2 (16-bit)	Any other script, such as Arabic, Chinese, Cyrillic, Hindi, and emoji	70 characters	67 characters
8-bit data	Binary content, rarely used for text	140 bytes	134 bytes

If any character in a message is not in the GSM-7 alphabet, the whole message is usually sent as UCS-2. One curly quotation mark or emoji in an otherwise plain English message can cut the limit from 160 to 70 characters.

Extension characters count twice

The GSM-7 alphabet has an extension table reached with an escape character. Each of these characters uses two positions: ^ { } \ [~] | € and the form feed. A message of 160 characters that includes one euro sign no longer fits in one SMS.

The basic GSM-7 alphabet includes the Latin letters, digits, common punctuation, a few accented letters such as é, è, ü and ñ, and some Greek capital letters. It does not include many common characters, such as curly quotes, the en dash or most accented capitals. Some countries can use national language shift tables defined in TS 23.038; support for them depends on the networks and provider.

Long messages and segments

Longer text is split into segments that the handset joins back together. Each segment carries a small header (the user data header) identifying the part, which is why the per-segment limit is lower: 153 GSM-7 characters or 67 UCS-2 characters. Providers usually charge per segment. An extension character is never split across two segments.

Examples of segment counts

Message	Encoding	Segments
160 plain Latin characters	GSM-7	1
161 plain Latin characters	GSM-7	2
306 plain Latin characters	GSM-7	2
70 Arabic characters	UCS-2	1
71 Arabic characters	UCS-2	2

Over SMPP, long messages are sent either as several `submit_sm` requests with a user data header, as one request using the `message_payload` field, or with the SAR fields, depending on what the provider supports. The `data_coding` value states the encoding, for example 0 for the default alphabet and 8 for UCS-2.

Tips to control length

- Replace curly quotes, long dashes and ellipsis characters with plain ' , " , - and . . . when the language allows.
- Keep one-time passcode messages short and free of emoji.
- Watch for characters pasted from word processors, which often add curly quotes and invisible characters.
- Test the exact text, including personalised values, before a campaign.

Use the [SMS length calculator](#) to check a message. It runs in your browser and estimates the encoding and segments; the final count is decided by the provider and networks that carry the message.

Reference

- 3GPP [TS 23.038](#): alphabets and language-specific information.
- 3GPP [TS 23.040](#): technical realisation of SMS, including concatenation.

Questions? Email support@wisenet.co. This PDF is generated from the web page above; the online version is the reference.