

*CEN/TC 304
European Localization
Requirements*



EUROPEAN COMMITTEE FOR STANDARDIZATION
COMITE EUROPEEN DE NORMALISATION
EUROPAISCHES KOMITEE FÜR NORMUNG

ISSS

CEN/TC304 Project Team on
Coordination of Euro in IT Standardization –

DRAFT Report on the Euro Glyph in OCR-B

Source: Karl Ivar Larsson, editor of the Project Team (OCR-B)

Date: 1998-06-28

Status of this report: This report will be sent to members of CEN/ISSS forum, to CEN/TC304 National Standardization Body members and to the European Commission for comments and for information. It will also be sent to a number of font experts and users for comments. A revised version will form basis for further work on OCR-B standardisation.

Email Contact information:

Erkki Kolehmainen, Project Team leader: eik@tieke.fi
Karl Ivar Larsson, Project Team editor (OCR-B): kil@lwpc.a.se
Thorgeir Sigurdsson, Secretary of CEN/TC 304: thorgeir@stri.is



Contents

	Page
1 Scope of this report	3
2 ISO OCR-B standard situation	3
3 OCR-B characteristics	3
4 Euro glyph implementation	4
5 Euro glyph tentative design	4
Annex: Illustrations	5

CEN/ISSS Project Team on Coordination of Euro in IT Standardization – Euro Glyph in OCR-B (status report 1998-06-28)

1 Scope of this report

This report concerns the development of a Euro glyph, specifically intended for Optical Character Recognition (OCR). This task constitutes one of the items of work within the CEN/ISSS Project Team for the Coordination of Euro in IT Standardization (CEN/TC304 PT26).

Current OCR software, as used for text scanning, is capable of identifying glyphs from a large number of font designs. There however exist IT applications where a higher degree of confidence in recognition is required than can be expected from general-purpose OCR software. An example is the scanning of machine-readable passports. Also in the financial sector a number of these tasks exist.

For such higher-confidence applications, specialized scanning equipment and software is used, and a particular font design prescribed. Generally one of the two ISO-standardized fonts, OCR-A or OCR-B, is specified. The later one is preferred in Europe.

The present OCR-B glyph repertoire contains the dollar, the pound and the yen signs, as well as the international currency symbol. Although for most future banking transactions the EUR code will probably be used to designate Euro amounts, the Project Team has established – based on various input – that a requirement exists to complement the present OCR-B currency signs with the Euro.

2 ISO OCR-B standard situation

As the OCR technique developed in the sixties, separate work was carried out in the US and in Europe on standardized fonts for machine-reading, resulting in two differing schemes. These were collected in 1976 in the two-part standard ISO 1073 and named OCR-A and OCR-B, respectively.

In 1994 an extension to the OCR-B font was proposed, to better accommodate European national letters. A revision of part 2 of the standard was therefore started within ISO/IEC JTC 1/SC 2. It was however terminated in 1997, due to lack of industry and user support for the revision, particularly in the testing of the proposed new glyphs.

With the planning for the Euro introduction, renewed interest in an OCR-B revision appeared. Such a revision is however an issue where the initiative must be taken by CEN, not by ISO.

3 OCR-B characteristics

The OCR-B standard, ISO 1073/II-1976, specifies a total of 121 glyphs. (During the work started on the revision of the standard it was however established that four out of these are unnecessary in present-day OCR, and should be deleted.)

The font is specified for three sizes. Size I is the most commonly used, e.g. in machine-readable passports and on some European banknotes. Its size corresponds roughly to 12 typographical points.

The standard also defines two slightly differing versions of glyph shapes; the OCR-B standard therefore actually specifies two complete fonts.

The version originally intended for impact-type printers has the same thickness for all the lines – strokes – that the glyphs consist of. It is consequently termed "constant strokewidth". Also, the ends of most strokes are rounded. This version is the one most commonly used.

The second version was intended for more capable printing processes, and conforms to traditional typographical practices, with e.g. differing thickness of horizontal and vertical strokes. It is therefore termed "letterpress", and is useful as a font for "typewriter-class" applications. The ends of the strokes are squarely cut off, not rounded.

The constant-strokewidth version is primarily intended for use with a fixed horizontal spacing ("fixed pitch") of 2,54 mm, i.e. 10 characters per inch (10 cpi). The glyphs are however designed to permit a fixed spacing of 2,12 mm also, i.e. the 12 cpi more common in Europe. The letterpress version, although useable for fixed-pitch printing, is primarily intended for "proportional spacing" as normally used for text nowadays.

When output on medium- and high-quality printers the two versions can generally be distinguished by

the human eye. For most OCR readers, however, the two font versions are in practice identical; differentiating between them certainly requires OCR readers with better than 300 "dots per inch" resolution. Considering inevitable printing defects and tolerances, the two versions could therefore be seen as interchangeable from an OCR point of view.

The present OCR-B standard's glyph repertoire (with the four unnecessary glyphs removed) is depicted on page 5 at Size I, and also magnified from that four times (which does not correspond to any of the three sizes; this is for illustration only). The proposed tentative Euro glyph (see below) has also been included, in two versions, on the same line as used for the existing currency signs. An illustration of currency glyph use is also given.

Please note that the Size I illustrations may not be representative, depending on what printer was used for outputting this text; generally speaking the OCR-B standard's glyph definitions need printer-adapted implementations.

4 Euro glyph implementation

Although – as pointed out above – the constant-strokewidth style may not generally be distinguishable from the letterpress in machine-reading, an OCR-B Euro glyph should be defined in both styles.

The shape of the "Euro logotype" can not be directly taken over into OCR-B, due to its proportions. For optimum OCR performance the "bounding box" as applicable to the font should in principle be filled up by the glyph. This means that the Euro logotype should be transformed into a rather narrower shape.

Such a new design should be a) approved by typographical expertise and then b) accepted by the appropriate EU body.

However, the shape thus arrived at may still not be optimal for OCR. Testing of the recognition properties may reveal the need for detail changes. The initial approval/acceptance procedure may therefore need to be repeated.

Testing is the subject of a separate (and approved) CEN/TC304 Project Team.

After the final glyph shape has been decided it must be formally specified. For the present standard, the glyph shapes are defined by scale 100:1 reference drawings, produced on dimensionally-stable material. Such drawings will have to be produced for the Euro glyph also, for its constant-strokewidth and letterpress versions.

A different method, which should be investigated, would be to specify the Euro glyph by outline definitions according to standard ISO/IEC 9541-3. Possibly both methods should be used.

As mentioned above, the ISO revision of the OCR-B standard has been discontinued. It therefore appears that the way to ensure that the OCR-B Euro glyph is formally accommodated will be to produce an OCR-B EN, based on the ISO standard.

There exists some interest in ISO to have the OCR-B standard revised, even if the original purpose of the revision – i.e. the addition of a number of European national letters – is not met. Production of an EN might therefore include parallel ISO processing.

Development of an EN would require approval for a CEN/TC304 New Work Item. A ballot for this is planned. If the NWI is approved, development of the EN text would be within the scope of the Euro Project Team. The present financing may however not cover such EN work completely.

5 Euro glyph tentative design

Tentative shapes for the OCR-B Euro constant-strokewidth and letterpress styles have been produced.

These shapes are based on the design of the existing OCR-B capital letter C. The obvious extension to that character is to add two horizontal strokes. The exact positionings and lengths of those strokes would be the prime subject to investigate in a testing phase.

The tentative shapes are illustrated on pages 6 and 7 by scale 100:1 drawings.

It appears, however, that although these shapes may be acceptable from an OCR point of view, they are visually not completely satisfactory. Using a single horizontal stroke instead of two seems to result in better (human) readability. (It can be noted that the OCR-B glyph for the yen sign is also designed with a single stroke, although the conventional way of writing it is with two strokes.)

A single-stroke design is illustrated on pages 8–9. It seems this should be the preferred shape, if acceptable to the EU Euro coordination.

For comparison, the other OCR-B currency signs as well as the capital letter C are also illustrated, in pages 10–14.

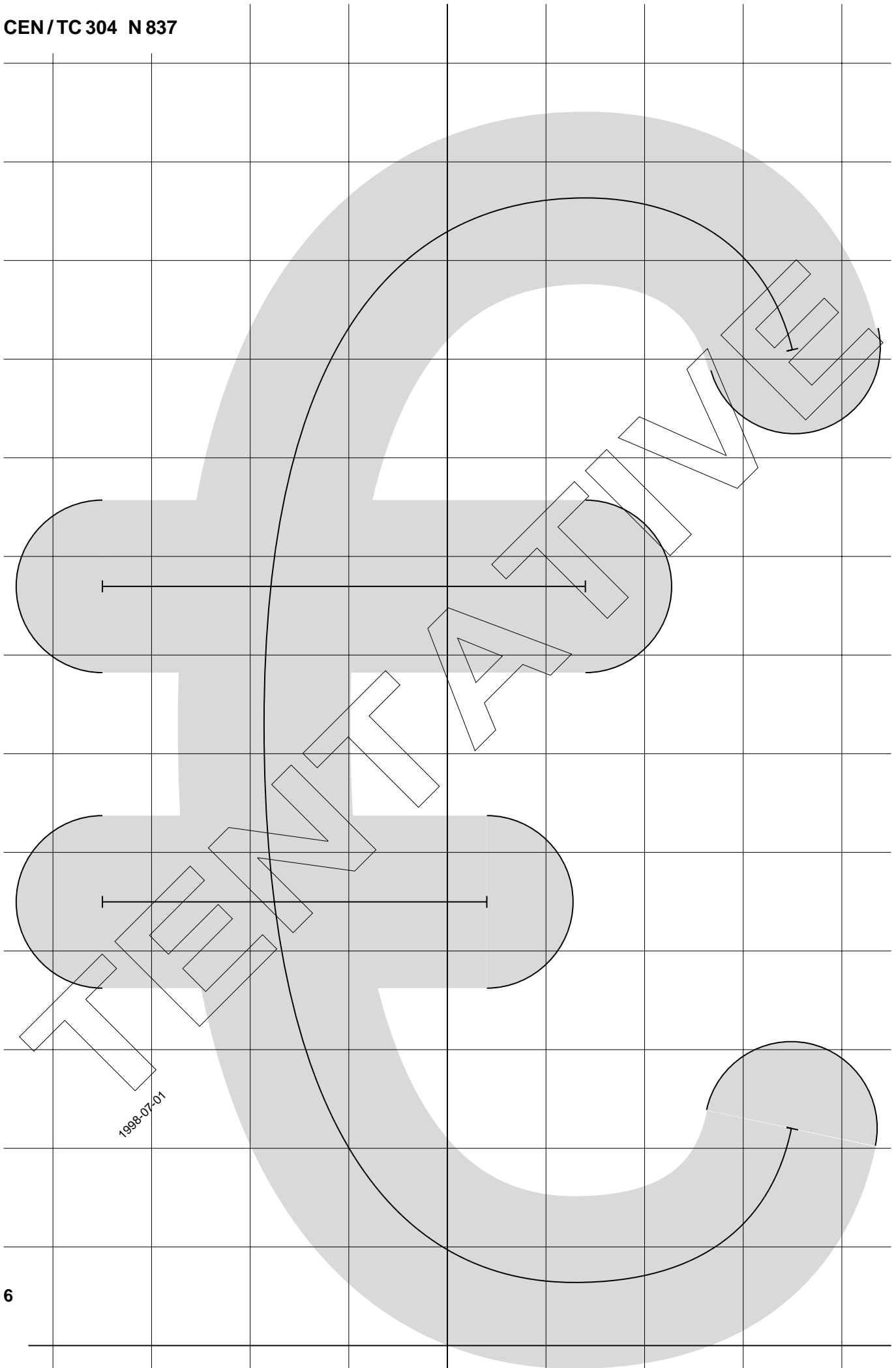
Annex

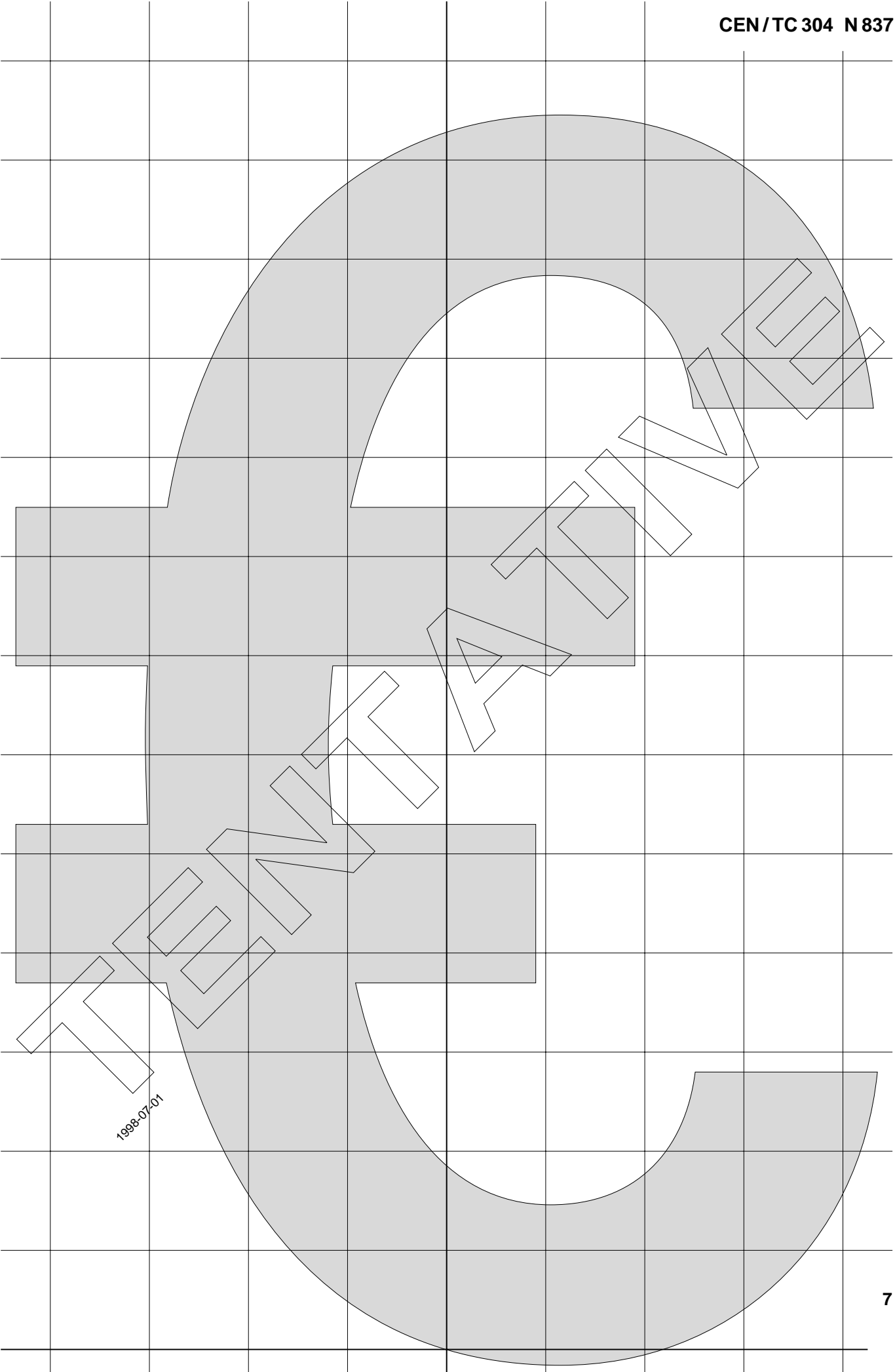
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
A B C D E F G H I J K L M	A B C D E F G H I J K L M
N O P Q R S T U V W X Y Z	N O P Q R S T U V W X Y Z
a b c d e f g h i j k l m	a b c d e f g h i j k l m
n o p q r s t u v w x y z	n o p q r s t u v w x y z
* + - = / . , : ; " ' _	* + - = / . , : ; " ' _
? ! () < > [] % # & @ ^	? ! () < > [] % # & @ ^
¤ £ \$ ¥ § \ { } m € €	¤ £ \$ ¥ § \ { } m € €
Ä Å Æ I J Ñ Ö Ø Ù à æ ij ø ß	Ä Å Æ I J Ñ Ö Ø Ù à æ ij ø ß
ˆ ˜ ˘ ˙ ˚	ˆ ˜ ˘ ˙ ˚
(SPACE)	(SPACE)

Illustration of OCR-B letterpress font size I, scales 4:1 and 1:1

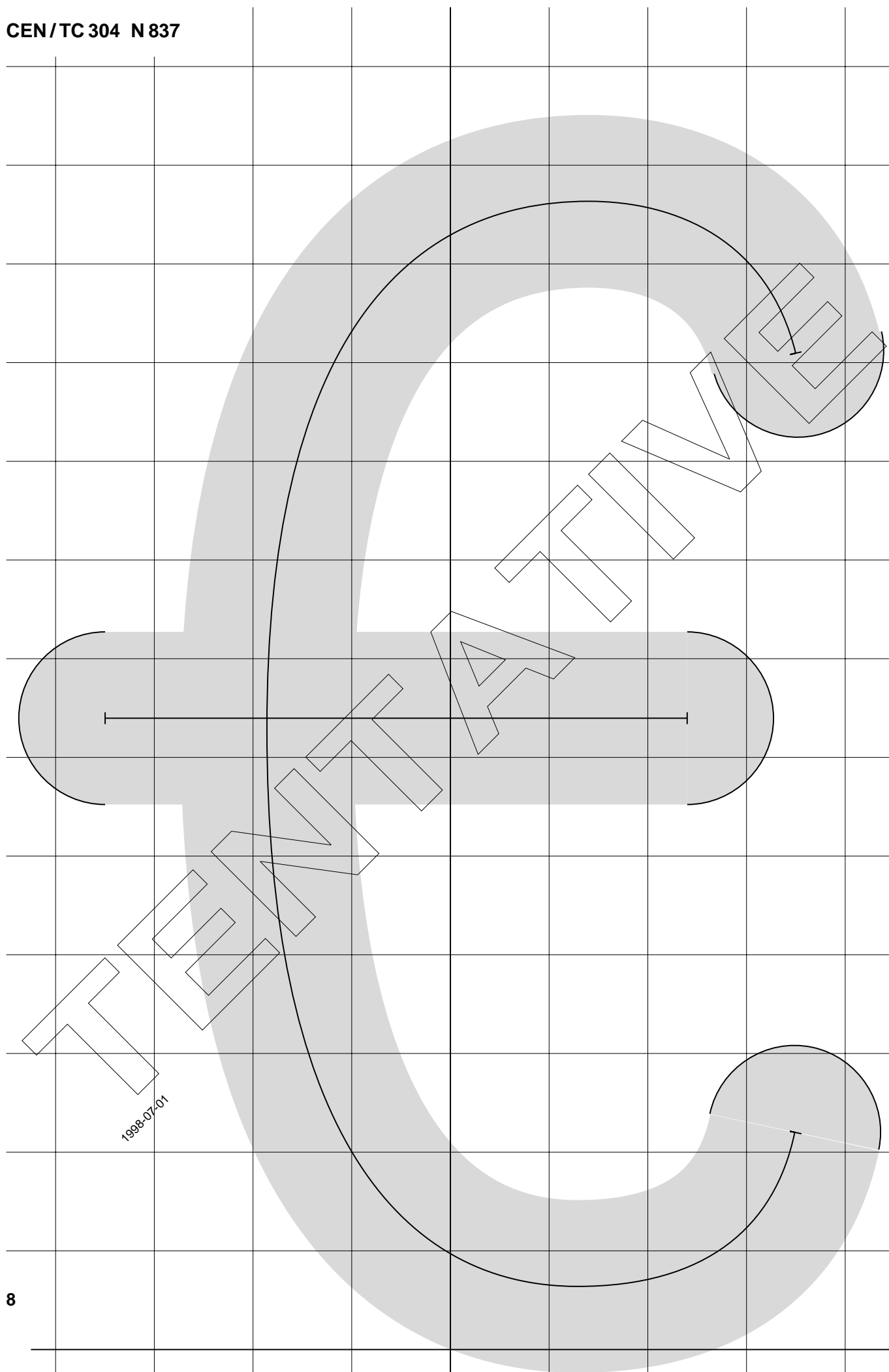
xyzxyz \$ 789 xyzxyz £ 789 xyzxyz ¥ 789 xyzxyz € 789 xyzxyz € 789 xyzxyz

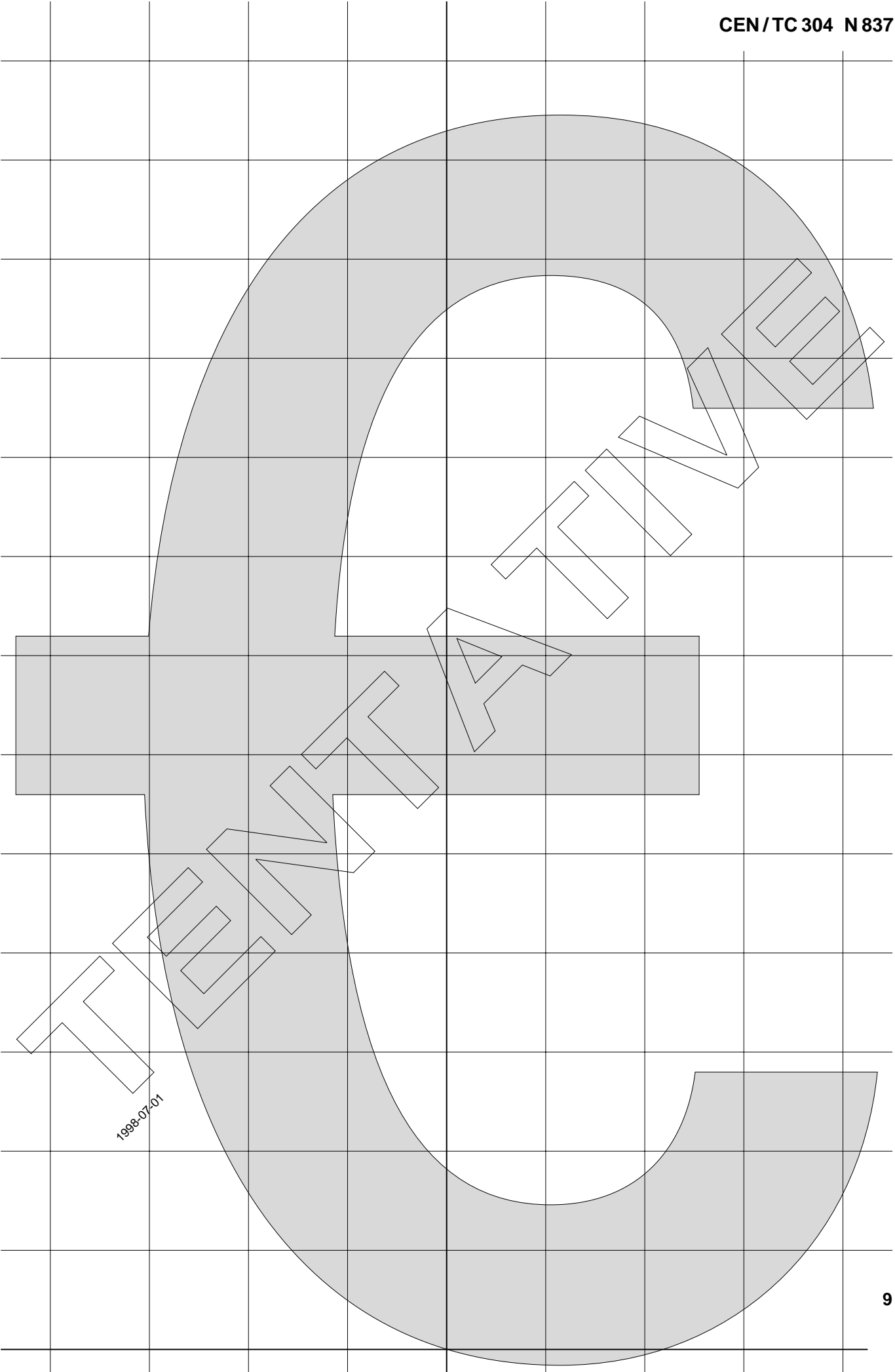
Illustration of currency sign use





1998-07-01





1998-07-01

