



Questioned Documents Section - 2015

J15 The Universal Coded Character Set (ISO/IEC 10464) and the Unicode Standard as Aids to Forensic Document Examination in an International and Transnational Environment

Peter V. Tytell, BA, Forensic Research, LLC, 15 Maiden Lane, Ste 308, New York, NY 10038-4017*

The goal of this presentation is to provide attendees with an awareness of developed standards, their potential benefit, and how to acquire and use them.

This presentation will impact the forensic science community by increasing awareness of, and the ability to deal with, non-English material.

Generations of American forensic document examiners have been well aware of the writing habits that successive waves of immigration brought to our shores and how these habits might carry over to the writing of these immigrants and even their descendants. These factors are still relevant today, but given the globalization of commerce, crime, and terrorism as well as the nature of contemporary internet society and social media, the 21st-century examiner can be confronted with documents from any part of the planet, perhaps in languages that use the Latin alphabet but with characters not seen in English, or even in languages with variants of completely different scripts. Such problems can arise in civil actions over a contract or a will; in criminal cases involving smuggled goods, drug records, or human trafficking; or in the context of a civil disaster like a train accident or airplane crash.

Signatures, handwritten text, or printed matter could be involved. What might seem to be an illegible, highly stylized, and personalized symbolic signature could actually be the completely legible and unremarkable writing of a name in an unfamiliar script. While it is sometimes appropriate to undertake the examination of a foreign text with the aid of a translator, the language must first be determined before the appropriate translator can be sought. A document examiner who can use a limited amount of “weird or foreign” text to provide a stymied investigator with information about the language of contraband packaging, the phone numbers in an address book, the message on a smartphone screen, or the charred remains of a receipt or newspaper clipping in a victim’s pocket might make a truly significant contribution to the resolution of the case.

The goal of this presentation is to familiarize examiners with the **International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) 10464**, the Universal Coded Character Set (UCS), and the Unicode Standard. These are standards developed for data exchange that provide unique names and code identifiers for the characters of the world’s written languages (including some long dead) as well as for many additional symbols.^{1,2} The responsible ISO Working Group has coordinated with the Unicode Consortium for more than 20 years and the two standards are essentially compatible, though Unicode includes much more information and some of the additional features are required for full implementation.

These standards provide internationally accepted naming conventions for characters and symbols, including diacritics, accented characters, digraphs, and new forms that extended the Latin alphabet for use with diverse languages worldwide. Searching the extensive Unicode Code Charts can provide useful leads regarding the language or language group using one of these modified characters or a combination of characters. Potentially useful information can also be found for languages using versions of the Cyrillic or Arabic alphabets and other writing systems.

Specific illustrations and internet search strategies will be provided.

References:

1. ISO/IEC 10646: Information Technology—Universal Coded Character Set (UCS). Geneva: International Organization for Standardization, 2012.
2. The Unicode Standard, Version 7.0.0. Mountain View, CA: The Unicode Consortium, 2014.

Unicode, Alphabets, Diacritics