

# CRMtex

Textual Entities Model

## Definition of the CRMtex

An Extension of CIDOC CRM to Model Textual Entities

*Approved by the CIDOC CRM-SIG*

Version 2.0

June 2023

Currently maintained by Francesca Murano, Achille Felicetti, Pavlos Fafalios

Contributors:

Martin Doerr, Pavlos Fafalios, Achille Felicetti, Francesca Murano,  
Athanasios Velios, Eleni Tsouloucha



# Index

<b>Introduction</b>	<b>5</b>
Scope	5
Status	12
Naming Convention	12
<b>Classes and properties hierarchies</b>	<b>13</b>
CRMtex class hierarchy, aligned with portions from the CRMsci, CRMInf, FRBRoo and the CIDOC CRM class hierarchies	14
List of external classes used in CRMtex	15
CRMtex property hierarchy, aligned with portions from the CRMsci, CRMInf,FRBRoo and the CIDOC CRM class hierarchies	16
List of external classes used in CRMtex	17
<b>Graphical overview and instantiation examples</b>	<b>18</b>
<b>Classes and properties usage examples</b>	<b>21</b>
<b>CRMtex - Classes and properties</b>	<b>24</b>
<b>CRMtex Classes declarations</b>	<b>24</b>
TX1 Written Text	25
TX2 Writing	25
TX3 Writing System	26
TX4 Writing Field	26
TX5 Text Recognition	27
TX6 Transliteration	28
TX7 Written Text Segment	29
TX8 Grapheme	30
TX9 Glyph	30
TX10 Style	31
TX11 Grapheme Occurrence	32
TX12 Grapheme Sequence	32
TX13 Script	32
TX14 Reading	33
<b>CRMtex Properties declarations</b>	<b>34</b>
TXP1 used writing system (writing system used for)	35
TXP2 includes (is included within)	35
TXP4 has segment (is segment of)	36
TXP5 wrote (was written by)	37
TXP6 encodes (is encoding of)	37
TXP7 has item (is item of)	38
TXP8 has component (is component of)	38
TXP9 is encoded using (was used to encode)	39
TXP10 deciphered text (was deciphered by)	39
TXP11 transcribed (was transcribed by)	40

TXP12 has style (is style of)	41
TXP13 deciphered via the representation (was representation used for deciphering)	42
TXP14 used copy or representation of (was deciphered via copy or representation)	42
TXP15 recorded correspondence (was recorded by)	43
TXP16 employs script (is employed by)	44
TXP17 has part (forms part of)	44
TXP18 read (was read by):	45
<b>References</b>	<b>47</b>

# Introduction

## Scope

This document presents CRMt<sub>ext</sub>, an extension of the CIDOC CRM created to support the study of ancient and handwritten documents, i.e., texts characterised by uniqueness since they have been produced without the use of techniques typical of modern mechanized processes of production. CRMt<sub>ext</sub> aims to model information concerning ancient inscriptions (including coins, marks and stamps), papyri, medieval manuscripts, but also modern handwritten documents of any kind.

Furthermore, CRMt<sub>ext</sub> proposes the use of the CIDOC CRM to encode documents of this kind and to model their scientific process of investigation to foster the integration with other cultural heritage research fields. After identifying the key concepts, assessing the available technologies, and analysing the entities provided by CIDOC CRM and its extensions, CRMt<sub>ext</sub> introduces new classes and properties to address the needs of the disciplines involved (including epigraphy, papyrology, palaeography, and codicology).

### *Investigating written documentation*

On present archaeological evidence, full writing appeared in Mesopotamia and Egypt around the end of the IV millennium BC. (cf. Feldherr and Hardy 2011). With the evolution of this technology, humans began to write texts on different supports using different techniques: inscriptions, papyri, manuscripts, and other similar documents.

Although from the semiotic point of view (see below) the mechanism of production of written texts follows a unique approach (regardless of the supports, techniques, etc. used), traditionally, the study of ancient texts falls within different disciplines, generally grown around the specific characteristics of each class of documents (e.g., papyrology for the study of papyri, epigraphy for inscriptions and palaeography for the study of ancient manuscripts). Nevertheless, an interdisciplinary approach is essential, and the identification of common elements is paramount to confer uniformity and interoperability to all these disciplines, as well as to exploit complementary skills from different approaches.

What should be observed, specifically in texts for which this model was designed, is the relationship between the text and its support. In comparison to modern printed or digital texts, this kind of text is typically characterised by its uniqueness, being the result of manual work rather than a mechanised process, as occurs since the invention of modern printing techniques.

Such characteristics make the study and digitisation of this type of documentation particularly arduous: the close relationship between the text and its support requires careful analysis since they are inextricably linked to form a unique object of study.

In the ancient world, nevertheless, some types of inscriptions were created through mechanised processes, such as the legends of coins, medals, stamps, and seals. The uniqueness of the written text remains unchanged in this case also, since it is characterised by

the peculiar history of the support, which in most of the cases is a cultural object having significance also for other disciplines (e.g., numismatics, archaeology, etc.).

The first aim of this extension is therefore to identify and define in a clear and unambiguous way the main entities involved in the study and edition of ancient and other handwritten texts and then to describe them by means of appropriate ontological instruments in a multidisciplinary perspective.

In addition to dealing with text as an object, our model also focuses on the aspects of the research and provides classes and relationships to describe the typical operations that scholars from different disciplines put in place to gain knowledge about texts. It is evident, in this perspective, that the study of ancient texts typically starts from the analysis of the physical characteristics of the individual text itself before moving to the investigation of their archaeological, palaeographic, linguistic, and historical features.

#### *What is said and what is written*

“Writing is one of the most significant cultural accomplishments of human beings” (cf. Rogers 2005), since it offers humans new semiotic resources, allowing asynchronous communication to take place –i.e., the kind that occurs across different places and at different times –and, thus, to pass on memories of events and/or things, written texts being more enduring than spoken utterances.

Although every speech can be transposed into an equivalent written message, and vice versa, according to a common formulation of the relationship between speech and writing, the spoken language is prior to writing (cf. Lyons 1972), in the sense that writing results from the transference of the language from a primary phonic medium to a secondary graphic medium (cf. Lyons 1977: 65).

Writing originated as a representation of speech, “as the use of graphic marks to represent specific linguistic utterances” (Rogers 2005). According to Ferdinand de Saussure (1983), “a language and its written form constitute two separate systems of signs. The sole reason for the existence of the latter is to represent the former”.

Although writing can be examined from a variety of perspectives, being applicable to different human activities (Harris 1995), the theory on which CRMtex is based for the analysis of writing, is that of semiotics. In a semiotic perspective, language and writing are codes (i.e., systems of signs) and the transmission of a message is an encoding/decoding process: the formation of a message by the sender is an encoding and the interpretation of the message is a decoding by the receiver. Coding consists in assigning the appropriate expression to a given content; the decoding in identifying the content starting from the expression. In this theoretical framework (and desiring to simplify a very complex matter), writing is a secondary code, having as its content the expression of another code (i.e., the language).

Writing, therefore, appears as a code requiring an encoding process by the creator or writer and a decoding one by the receiver or reader to be properly understood. It is worth considering that in writing (characteristic in common with speech) every component (sign) possesses a dual nature, one physical and another conceptual: regardless of techniques and types of supports, writing “involves the physical production of variable tokens representing invariant types” (i.e., the ideal shapes of the signs of a writing system) (Coulmas 1999: 193).

Thus, for the analysis of written texts it is necessary to distinguish the concrete, physical, individual realization performed by a single person on a specific occasion (e.g., the specific unique sequence of marks I wrote on a paper with my pen to take down a note), and the abstract level concerning the mental knowledge and rules pertinent to a particular writing system, allowing the process of recognizability of the material mark with an “ideal” sign on the basis of a sameness principle. In brief, the semiotic process underlying the writing allows the identification of my personal “A” mark, independently from the peculiar shape I give to it, as the LATIN CAPITAL LETTER A.

### *Glyphs and Graphemes*

The physical elements (glyphs or graphs) composing a written text constitute the material manifestations of the graphemes, i.e., the abstract entities of a writing system (cf. Coulmas 1999). According to the common definitions, a graph is the minimal formal unit of written language on the level of handwriting or print and a grapheme is the minimal functional distinctive unit of writing on whatever structural level of language the writing system operates (Coulmas 1999; Pulgram 1976).

For a typological study of writing systems, scholars recognize a broad distinction between glottographic and non-glottographic (i.e. pure semasiographic) writing systems, “depending on whether the formation and interpretation of texts presupposes knowledge of a particular language” (Harris 1995: 95).

Concerning the glottographic systems (that non-glottographic writing systems in the narrower sense exist is very disputed), the “theory commonly adopted by linguists distinguishes different kinds of writing system according to which units in the spoken language appear to have been selected as the basic units for representation in writing” (Harris 1995: 95). According to Pulgram (1976: 2-3) a grapheme represents the minimal unit of some level: “in reducing a language to writing, that is, in making visible marks that evoke or recall linguistic performance, it would seem that each mark must represent a syntagme or a lexeme or a morpheme or a phoneme or whatever other kind of unit the inventor of the system may chose as his basis”.

In glottographic systems scholars recognize a difference “between logographic scripts, which assign distinct marks to meaningful units of a language, i.e., words or morphemes, and phonographic scripts which represent phonological units of one size or another” (Sampson 2016; cf. Sampson 1985 and Rogers 2005).

To better clarify: in principle, in an alphabetic writing system, e.g., the Latin alphabet, including the consonantal ones (i.e., the abjads as the Arabic alphabet), the basic unit of representation is the phoneme. Both in a syllabic and in an alphasyllabic writing system (i.e., respectively in syllabary as the Mycenaean or the Japanese systems, and abugida, as the Sanskrit or the Thai systems) basic unit is the syllable. In a logographic writing system, as (part of) the Egyptian hieroglyphic or the modern Chinese system the basic unit of representation is a grammatical/lexical unit (i.e., a morpheme or a word) (cf. Daniels and Bright 1996; Borgwaldt and Joyce 2013). We propose some examples. In a Latin inscription, each mark inscribed on the stone (i.e., each glyphs) represents a corresponding grapheme in

the Latin writing system (which in turn stands for a phoneme): e.g., the first five glyphs of the last line in Fig. 7 represent the graphemes <a>, <r>, <c>, <u> and <m> of the Latin alphabet, and in turn these graphemes codify the following sounds of the Latin language (phonemes): /a/, /r/, /k/, /u/ and /m/ (Lat. *arcum* ‘arch’ acc. sing.). In Mycenaean Linear B and in Old Persian cuneiform inscriptions, glyphs represent (for the most part) syllabograms, i.e., the graphemes representing a syllable, not a single sound. E.g., the first sequence visible on the inscription from the Palace of Darius the Great in Persepolis<sup>1</sup> represents the seven graphemes of the Old Persian writing system 𐎠 𐎡 𐎢 𐎣 𐎤 𐎥 𐎦 corresponding to the seven syllables /da/, /a/, /ra/, /ja/, /va/, /u/ and /ja/. In an Egyptian hieroglyphic text, glyphs may represent syllabic, alphabetic or ideographic elements, i.e., the elements standing for lexical/semantic units.

Over time, writing systems tend to deviate from the ideal 1:1 correspondence between language units (whatever they be) and the units used in their written representation (graphemes). Spelling conventions are construed as resulting from changes to which linguistic systems are subject from a diachronic perspective.

This phenomenon is particularly evident in phonographic systems, because of the diachronic phonetic variations. From this follows that, in English, for example, many discrepancies appear between spelling and phonetic values: e.g., the grapheme <i> stands for various phonemes: /i/ (as in *him*), /ɪ/ (as in *time*), /i/ (as in *police*), /a/ (as in *timbre*); on the other hand, the phoneme /f/ can be represented with <f> (as in *film*), <ph> (as in *philology*) or <gh> (as in *enough*).

For scientific purposes, the International Phonetic Alphabet has been devised as a standardized representation of speech sounds in written form, having a 1:1 correspondence between phonological units and IPA symbols.

### *Recognising, reading and understanding the text*

Reading refers to the semiotic procedure of decoding a written text, and therefore of deriving meaning from it (i.e., understanding it). Reading is “a highly complex activity involving the interplay of visual-perceptual, linguistic and conceptual systems” (Coulmas 1999: 430).

From a semiotic point of view, according to communication theories, a complete retrieval of the information (i.e., reading of the written message) presupposes the code sharing by sender and receiver (Jakobson 1960). It is often the case that the code linking the written form to the meaning and sound combination that it stands for has been lost over time, and scholars need to recover it for scientific purposes.

Scholars propose various models of the reading process, based on the identification of “the perceptual and cognitive stages and activities leading from visual input to understanding the content of the written message” (Coulmas 1999: 432), and distinguish some stages, from the visual fixation to the character identification, to the word recognition, to the association of meanings and the application of linguistic rules, finally to the application of phonological rules and the assignment of a phonetic form.

---

<sup>1</sup> <https://www.livius.org/pictures/iran/persepolis/persepolis-palace-of-darius/dpa/> (accessed on 2023/06/06)



The reading process can be carried out for scientific purposes, to analyse and study the text according to different disciplinary perspectives. Although in principle, the written text is made to be read, its reading/comprehension depends on the degree of the initial knowledge on the part of the reader –for CRMtex, the intended reader would be the scientific community. In the case of languages and writing systems that are no longer in use, in fact, it is possible that scholars are unable to entirely decode the elements, i.e., to establish the value that those elements have within the system. A case of this kind is constituted by the Linear A and the writing of the Phaistos disc, of which the linguistic systems they represent are unknown.

According to the aim of the model, regardless of the cases in which the observation of visual items on a surface does not determine the recognition of a text and concerning only the cases of the observation of a text, we consider the following stages of the decoding process:

1. *character identification*: the process of identifying visual items as elements of a writing system; it is a necessary although not a sufficient condition of reading. Decoding processes that stop at this stage in the scientific field are due to the lack of knowledge of the language used (there is no shared code between sender/writer and receiver/reader). An example is the current state of knowledge of the writing of the Phaistos disc;

2. *signs recognition*: the process of identifying elements of a writing system known to the reader. At this stage the reader knows or can reconstruct the pronunciation and recognize the words, but the knowledge of the language is insufficient to have a complete linguistic comprehension of the text.

The deciphering of the signs can be achieved if the linguistic system represented is known; this is the case of Linear B, whose deciphering followed the understanding it represents a Greek language.

Since the writing systems have genealogical relationships with other known systems, it is possible that the writing systems do not present deciphering problems (so the scholar is able to attribute a rough value to the signs), even when the linguistic system is not yet known. This is the case of the Etruscan writing system, which was deciphered from the origins of Etruscology, the Etruscan alphabet deriving from the Euboean one, although knowledge of the language (i.e., the understanding of the texts) is the result of a long study process that still presents uncertainties;

3. *reading proper*: the process of associating the text with a complete linguistic meaning (cf. Coulmas 1999: 432).

On the level of the linguistic sounds, it will be the decoders (readers, including scholars), who from time to time, on the basis of the knowledge of the linguistic system, will attribute to each sign or group of signs the adequate (or reconstructed) phonetic value, also on the basis of spelling conventions in place in a given graphic system at a given historical moment, since the spelling rules can change over time, even if less quickly than the linguistic system does.

For the purposes of modelling the textual entity within the various disciplines for which CRMtex has been designed, within the model we distinguish two classes of text decoding depending on whether it is a proper reading or not.

For the goals of the study of texts, the reading activity requires a scientific autoptic examination of the text as preparatory action for the study. An autoptic examination consists of an accurate analysis of the surface and the signs and prescribes the use of specific tools and

procedures, for establishing as faithfully as possible the exact value of each sign drawn or applied on the physical feature.

### *Reproductions, transcriptions and transliterations of a text*

For research and scientific dissemination purposes, it is possible that there is a need to have a reproduction of text, also transposing it according to a writing system different from the original one.

According to their scientific purposes, scholars distinguish various stages:

1. an exact reproduction of the visual items recorded on a text (fac-simile). An example is the drawing of the inscription of Darius the Great in Persepolis<sup>2</sup> published through livius.org;
2. a reproduction of the recognised graphemes of a text using the same writing system (transcription in a broader sense). An example is the text of the Dreros Law from Crete published by the AXON project.<sup>3</sup>
3. a conversion (i.e., re-encoding) of the recognised graphemes of a text using a different writing system according to a 1:1 (i.e., unique and unambiguous) conversion (transliteration). Because of this 1:1 conversion this operation is reversible, allowing an “automatic” and unambiguous recreation of the original.

Since the purpose of transliteration is to enable those not familiar with a writing system in which a text is encoded to read it, commonly the Latin alphabet is used. An example is the text in Latin alphabet of the Ancient South Arabian inscription as-Sawdā' 49 published by the DASI project.<sup>4</sup>

Transliteration conventions for writing systems structurally identical (e.g., alphabets), do not pose difficulties; in turn, conventions for rendering written texts in a writing system of a structurally different type can be problematic.

In case of texts written in a non-alphabetic system, the conversion into the Latin alphabet can involve linguistic elements broader than a phoneme, notwithstanding the 1:1 relation between the graphemes of source writing system and the Latin encoding. An example is the transliteration <da-a-ra-ya-va-u-ša> of the first sequence of inscription of Darius the Great in Persepolis<sup>5</sup> published on livius.org, where a grapheme of the Persian syllabary corresponds to a syllable univocally and conventionally referred to in Latin script (e.g., <𐎠𐎡> □ <da>).

For scientific purposes competing systems are in use in different disciplines, but each transliteration is consistent for a specific field of study (e.g., Biblicists and linguists use different systems for transliterating Hebrew in Latin alphabet). Standards, such as the ISO<sup>6</sup> and BGN/PCGN, define the transliteration rules and are widely used to overcome these divergences.

4. a re-encoding of the recognised graphemes of a text using a different writing system according to a phonological (and even spelling) criterion (transcription in a narrower sense). For example, the name of the second largest Greek island, is rendered as 'Εύβοια' using the

---

<sup>2</sup> Inscription DPa on Livius.org:

<https://www.livius.org/sources/content/achaemenid-royal-inscriptions/dpa/> (accessed on 2023/06/06)

<sup>3</sup> AXON project: <http://doi.org/10.14277/2532-6848/Axon-1-2-17-01> (accessed on 2023/06/06)

<sup>4</sup> DASI project:

<http://dasi.cnr.it/index.php?id=146&prjId=1&corId=0&colId=29&navId=869660944&recId=2711>  
(accessed on 2023/06/06)

<sup>5</sup> <https://www.livius.org/sources/content/achaemenid-royal-inscriptions/dpa/> (accessed on 2023/06/06)

<sup>6</sup> ISO 01.140.10 Writing and transliteration. <https://www.iso.org/ics/01.140.10/x/> (accessed on 2023/06/06)

Greek alphabet. Its transliteration into Latin script is 'Euboia', but it can also be rendered as 'Evia', matching its Modern Greek pronunciation; notice that transcription is based on the phonetics, thus pronunciation problems can arise: for example an English speaker might read 'Evia' as ['ivia] instead of ['evia], and possibly transcribe 'Ivia'). Another example: the name of the Russian composer Чайкóвский is transliterated according to the modern transliterations of Russian ISO 9 standard Čajkovskij; in turn the name is anglicized (i.e., transcribed according to the English system) as *Tchaikovsky* or *Chajkovskij*, etc., while in German is more common the transcription *Tschaikowskyi* and in French *Tschaïkowsky*.

For scientific purposes a re-encoding of this type is useful in the case of text written in a non-alphabetic system, especially when, in composing words, the elements of the writing system do not match entirely with the actual phonetic structure of the represented word. An example is the transcription of the inscription of Darius the Great in Persepolis, published on livius.org using the Latin alphabet, where each word is re-encoded taking into account the actual pronunciation regardless of how it is written in the original text (e.g. the first sequence reported above is transcribed Dârayavauš).

A particular case is the conversion according to the International Phonetic Alphabet (IPA). Even though it consists of a change of the writing system employed, the use of the IPA has as specific purpose the reproduction of the exact pronunciation of the words.

#### *Written text segments*

Scholars of different disciplines, on the basis of the requirements of their study, need to identify and focus their attention on different types of text segments, in order to describe their physical conditions (form, layout, etc.), verify their legibility and particular phenomena (e.g., linguistic or palaeographic) connected to them, etc. For this reason, we created the class TX7 Written Text Segment, which allows one to investigate the interconnections existing between the text and its parts. Examples of text segments are columns, sections, paragraphs, but also single words or letters, or other specific components of the written text that scholars need for their purposes.

In this way it is possible to assign specific issues to the individual segments, independently of the text in its entirety. In fact, particular production (i.e. TX2 Writing) or destruction (E6) events can be associated with single segments, as in the case of letters or words damaged or worn out due to deterioration or human interventions.

Specifications about conditions (E3) for documenting the state of each textual part during the observation process (S4) can be easily stated as well. This allows scholars to document different events for the investigated parts in a more precise way and to assign observations and interpretations to them.

#### *Style and other palaeographic features*

Since the stylistic variations of hand-written texts are constitutive (e.g., an 'A' can appear as uppercase, lowercase, italics, round, printed or written by hand, or in different font families), a palaeographic study of stylistic variations has great importance in the description of written texts, using different styles for different purposes or at different times and places.

This approach is fundamental for the determination to date and determine the provenance of the texts, especially in reference to the styles developed in certain centres (for example, in the scriptoria of the monasteries). It is also relevant for the description of all the entities of a given epoch and place, e.g., the Ptolemaic cursive of Hellenistic Egypt, the capital uncial script (3rd-8th cent. AD), used both for Greek and Latin alphabets, or the more recent Carolingian minuscule, used from the beginning of the 8th cent. AD.

Therefore, in palaeography the concepts of stylistic class, style and canon are paramount to underline different meaningful observable aspects. The specific study of these stylistic variations needs to be properly addressed.

Palaeography uses different concepts, including aspects of the style, writing direction and other features related to the physical way the text is written and arranged.

## **Status**

CRMtex is the result of collaboration between scholars of many cultural heritage institutions. The first need that the model attempts to meet is to create a common ground for the integration and interoperability of records concerning ancient texts on every level, from the description of the supports and carried texts, to the management of the documentation produced by various institutions using national and institutional standards (e.g., TEI/EpiDoc). This document describes a community model, under approval by CRM SIG as being formally and methodologically compatible with CIDOC CRM. However, in a broader sense, it is always open to any possible integration and addition that may become necessary as a result of its practical use on real problems on a large scale. The model is intended to be maintained and promoted as an international standard.

## **Naming Convention**

All the classes declared were given both a name and an identifier constructed according to the conventions used in the CIDOC CRM model. For classes, the identifier consists of the letters TX followed by a number. Resulting properties were also given a name and an identifier, constructed according to the same conventions. That identifier consists of the letters TXP followed by a number, which in turn is followed by the letter “i” every time the property is mentioned “backwards”, i.e., from target to domain (inverse link). “TX” and “TXP” do not have any other meaning. They correspond respectively to letters “E” and “P” in the CIDOC CRM naming conventions, where “E” originally meant “entity” (although the CIDOC CRM “entities” are now consistently called “classes”), and “P” means “property”. Whenever CIDOC CRM classes are used in our model, they are named by the name they have in the original CIDOC CRM. CRMsci classes and properties are referred to with their respective names, classes denoted by S and properties by O.

## Classes and properties hierarchies

The CIDOC CRM model declares no “attributes” at all (except implicitly in its “scope notes” for classes), but regards any information element as a “property” (or “relationship”) between two classes. The semantics are therefore rendered as properties, according to the same principles as the CIDOC CRM model.

Although they do not provide comprehensive definitions, compact mono hierarchical presentations of the class and property IsA hierarchies have been found to significantly aid in the comprehension and navigation of the model and are therefore provided below.

The class hierarchy presented below has the following format:

- Each line begins with a unique class identifier, consisting of a number preceded by the appropriate letter “E”, “TX”, “S”
- A series of hyphens (“-”) follows the unique class identifier, indicating the hierarchical position of the class in the IsA hierarchy.
- The English name of the class appears to the right of the hyphens.
- The index is ordered by hierarchical level, in a “depth first” manner, from the smaller to the larger sub hierarchies.
- Classes that appear in more than one position in the class hierarchy as a result of multiple inheritance are shown in an italic typeface.

## CRMt<sub>ex</sub> class hierarchy, aligned with portions from the CRM<sub>sci</sub>, CRM<sub>inf</sub>, FRBR<sub>oo</sub> and the CIDOC CRM class hierarchies

This class hierarchy lists:

- all classes declared in CRM<sub>t<sub>ex</sub></sub>
- all classes declared in CIDOC CRM version 7.1.2, CRM<sub>sci</sub> version 2.0, CRM<sub>inf</sub> version 0.7, and FRBR<sub>oo</sub> version 2.4, that are declared as superclasses of classes declared in CRM<sub>t<sub>ex</sub></sub>,
- all classes declared in CIDOC CRM version 7.1.2, CRM<sub>sci</sub> version 2.0, CRM<sub>inf</sub> version 0.7, and FRBR<sub>oo</sub> version 2.4, that are either domain or range for a property declared in CRM<sub>t<sub>ex</sub></sub>,
- all classes declared in CIDOC CRM version 7.1.2, CRM<sub>sci</sub> version 2.0, CRM<sub>inf</sub> version 0.7, and FRBR<sub>oo</sub> version 2.4, that are either domain or range for a property declared in CIDOC CRM version 7.1.2, CRM<sub>sci</sub> version 2.0, CRM<sub>inf</sub> version 0.7, or FRBR<sub>oo</sub> version 2.4 that is declared as superproperty of a property declared in CRM<sub>t<sub>ex</sub></sub>,
- all classes declared in CIDOC CRM version 7.1.2, CRM<sub>sci</sub> version 2.0, CRM<sub>inf</sub> version 0.7, and FRBR<sub>oo</sub> version 2.4, that are either domain or range for a property that is part of a complete path of which a property declared in CRM<sub>t<sub>ex</sub></sub> is declared to be a shortcut.

**Table 1:** CRM<sub>t<sub>ex</sub></sub> Class Hierarchy

E1						CRM Entity
E77	-					Persistent Item
E70	-	-				Thing
E72	-	-	-			Legal Object
E18	-	-	-	-		Physical Thing
E19	-	-	-	-	-	Physical Object
E24	-	-	-	-	-	Physical Human-Made Thing
E25	-	-	-	-	-	Human-Made Feature
TX1	-	-	-	-	-	Written Text
TX7	-	-	-	-	-	Written Text Segment
TX9	-	-	-	-	-	Glyph
TX4	-	-	-	-	-	Writing Field
E26	-	-	-	-	-	Physical Feature
E90	-	-	-	-	-	Symbolic Object
E73	-	-	-	-	-	Information Object
E36	-	-	-	-	-	Visual Item
E29	-	-	-	-	-	Design or Procedure
TX3	-	-	-	-	-	Writing System
TX10	-	-	-	-	-	Style
TX12	-	-	-	-	-	Grapheme Sequence
TX11	-	-	-	-	-	Grapheme Occurrence

E71	-	-	-	Human-Made Thing
E24	-	-	-	Physical Human-Made Thing
E28	-	-	-	Conceptual Object
E55	-	-	-	Type
TX8	-	-	-	Grapheme
E56	-	-	-	Language
E89	-	-	-	Propositional Object
TX13	-	-	-	Script
E2	-	Temporal Entity		
E4	-	-	Period	
E5	-	-	-	Event
E7	-	-	-	Activity
E65	-	-	-	Creation
TX5	-	-	-	Text Recognition
TX6	-	-	-	Transliteration
E13	-	-	-	Attribute Assignment
S4	-	-	-	Observation
TX5	-	-	-	Text Recognition
I1	-	-	-	Argumentation
TX14	-	-	-	Reading
E11	-	-	-	Modification
E12	-	-	-	Production
F28	-	-	-	Expression Creation
TX2	-	-	-	Writing
S15	-	Observable Entity		
E5	-	-	Event	

## List of external classes used in CRMtex

*Table 2: List of external classes used by CRMtex, grouped by model and ordered by class identifier*

Class ID	Class name	Model	Version
E24	Physical Human-Made Thing	CIDOC CRM	7.1.2
E29	Design or Procedure	CIDOC CRM	7.1.2
E36	Visual Item	CIDOC CRM	7.1.2
E55	Type	CIDOC CRM	7.1.2
E56	Language	CIDOC CRM	7.1.2
E65	Creation	CIDOC CRM	7.1.2
E89	Propositional Object	CIDOC CRM	7.1.2
E90	Symbolic Object	CIDOC CRM	7.1.2
S4	Observation	CRMsci	2.0

I1	Argumentation	CRMinf	0.7
F28	Expression Creation	FRBRoo	2.4

## CRMtex property hierarchy, aligned with portions from the CRMsci, CRMinf, FRBRoo and the CIDOC CRM class hierarchies

This property hierarchy lists:

- all properties declared in CRMtex,
- all properties declared in CIDOC CRM version 7.1.2, CRMsci version 2.0, CRMinf version 0.7, and FRBRoo version 2.4 that are declared as superproperties of properties declared in CRMtex,
- all properties declared in CIDOC CRM version 7.1.2, CRMsci version 2.0, CRMinf version 0.7, and FRBRoo version 2.4 that are part of a complete path of which a property declared in CRMtex, is declared to be a shortcut.

**Table 3:** CRMtex Property Hierarchy

Prop. ID	Prop. Name	Entity - Domain	Entity - Range
P2	has type	E1 CRM Entity	E55 Type
TXP6	- encodes (is encoding of)	TX3 Writing System	E56 Language
P12	occurred in the presence of (was present at)	E5 Event	E77 Persistent Item
P16	- used specific object (was used for)	E7 Activity	E70 Thing
TXP11	- - transcribed (was transcribed by)	TX6 Transliteration	TX12 Grapheme sequence
TXP13	- - deciphered via the representation (was representation used for deciphering)	TX5 Text Recognition	E36 Visual Item
TXP14	- - used copy or representation of (was deciphered via copy or representation)	TX5 Text Recognition	TX1 Written Text
TXP18	- - read (was read by)	TX14 Reading	TX1 Written Text
P33	- - used specific technique (was used by)	E7 Activity	E29 Design or Procedure
TXP1	- - - used writing system (writing system used for)	TX2 Writing	TX3 Writing System
TXP12	- - - has style (is style of)	TX1 Written Text	TX10 Style
P31	- has modified (was modified by)	E11 Modification	E18 Physical Thing
P108	- - has produced (was produced by)	E12 Production	E24 Physical Human-Made Thing
TXP5	- - - wrote (was written by)	TX2 Writing	TX1 Written Text
P92	- brought into existence (was brought into existence by)	E63 Beginning of Existence	E77 Persistent Item
P94	- - has created (was created by)	E65 Creation	E28 Conceptual Object
TXP15	- - - recorded correspondence (was recorded by)	TX5 Text Recognition	TX12 Grapheme Sequence
P46	is composed of (forms part of)	E18 Physical Thing	E18 Physical Thing
TXP4	- has segment (is segment of)	TX1 Written Text	TX7 Written Text Segment
TXP8	- has component (is component of)	TX1 Written Text	TX9 Glyph
P56	- bears feature	E19 Physical Object	E26 Physical Feature
TXP2	- - includes (is included within)	TX4 Writing Field	TX1 Written Text
P67	refers to (is referred to by)	E89 Propositional Object	E1 CRM Entity
TXP7	- has item (is item of)	TX13 Script	TX8 Grapheme
P106	is composed of (forms part of)	E90 Symbolic Object	E90 Symbolic Object
TXP17	- has part (forms part of)	TX12 Grapheme Sequence	TX12 Grapheme Sequence



P132	spatiotemporally overlaps with	E92 Spacetime Volume	E92 Spacetime Volume
P10i	- contains (falls within)	E92 Spacetime Volume	E92 Spacetime Volume
P9	- - consists of (forms part of)	E4 Period	E4 Period
P140	assigned attribute to (was attributed by)	E13 Attribute Assignment	E1 CRM Entity
O8	- observed (was observed by)	S4 Observation	S15 Observable Entity
TXP10	- - deciphered text (was deciphered by)	TX5 Text Recognition	E24 Physical Human-Made Thing
P148	has component (is component of)	E89 Propositional Object	E89 Propositional Object
TXP16	- employs script (is employed by)	TX3 Writing System	TX13 Script
TXP9	is encoded using (was used to encode)	TX1 Written Text	TX3 Writing System

## List of external classes used in CRMtex

*Table 4: List of external properties used by CRMtex, grouped by model and ordered by property identifier*

Property ID	Property name	Model	Version
P2	has type	CIDOC CRM	7.1.2
P9	consists of (forms part of)	CIDOC CRM	7.1.2
P10i	contains (falls within)	CIDOC CRM	7.1.2
P12	occurred in the presence of (was present at)	CIDOC CRM	7.1.2
P16	used specific object (was used for)	CIDOC CRM	7.1.2
P31	has modified (was modified by)	CIDOC CRM	7.1.2
P33	used specific technique (was used by)	CIDOC CRM	7.1.2
P46	is composed of (forms part of)	CIDOC CRM	7.1.2
P56	bears feature	CIDOC CRM	7.1.2
P67	refers to (is referred to by)	CIDOC CRM	7.1.2
P92	brought into existence (was brought into existence by)	CIDOC CRM	7.1.2
P94	has created (was created by)	CIDOC CRM	7.1.2
P106	is composed of (forms part of)	CIDOC CRM	7.1.2
P108	has produced (was produced by)	CIDOC CRM	7.1.2
P132	spatiotemporally overlaps with	CIDOC CRM	7.1.2
P140	assigned attribute to (was attributed by)	CIDOC CRM	7.1.2
P148	has component (is component of)	CIDOC CRM	7.1.2
O8	observed (was observed by)	CRMsci	2.0

# Graphical overview and instantiation examples

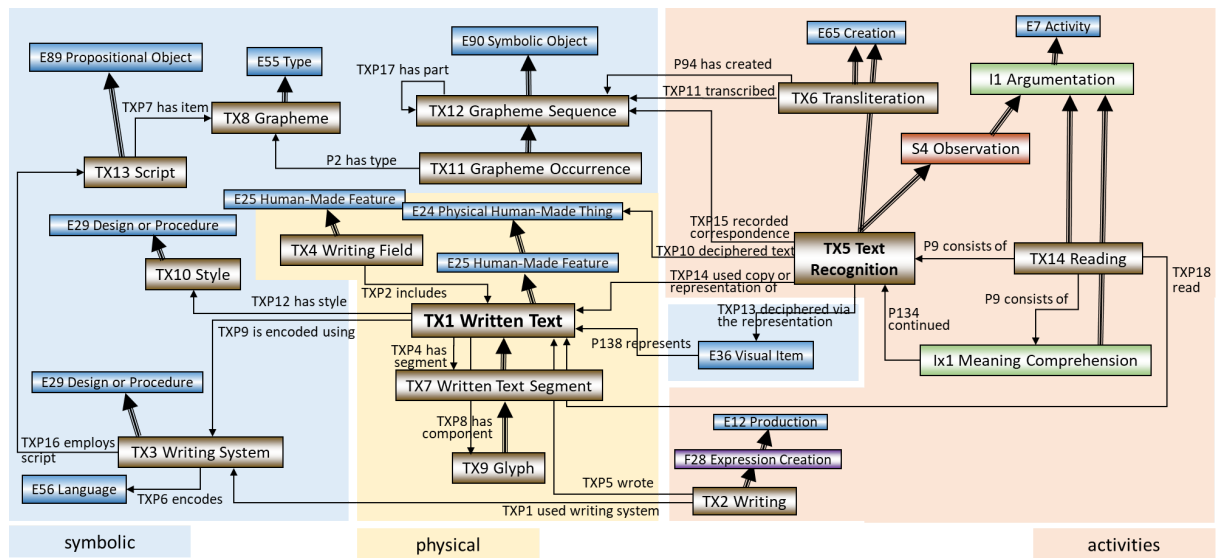


Figure 1: The CRMtex model

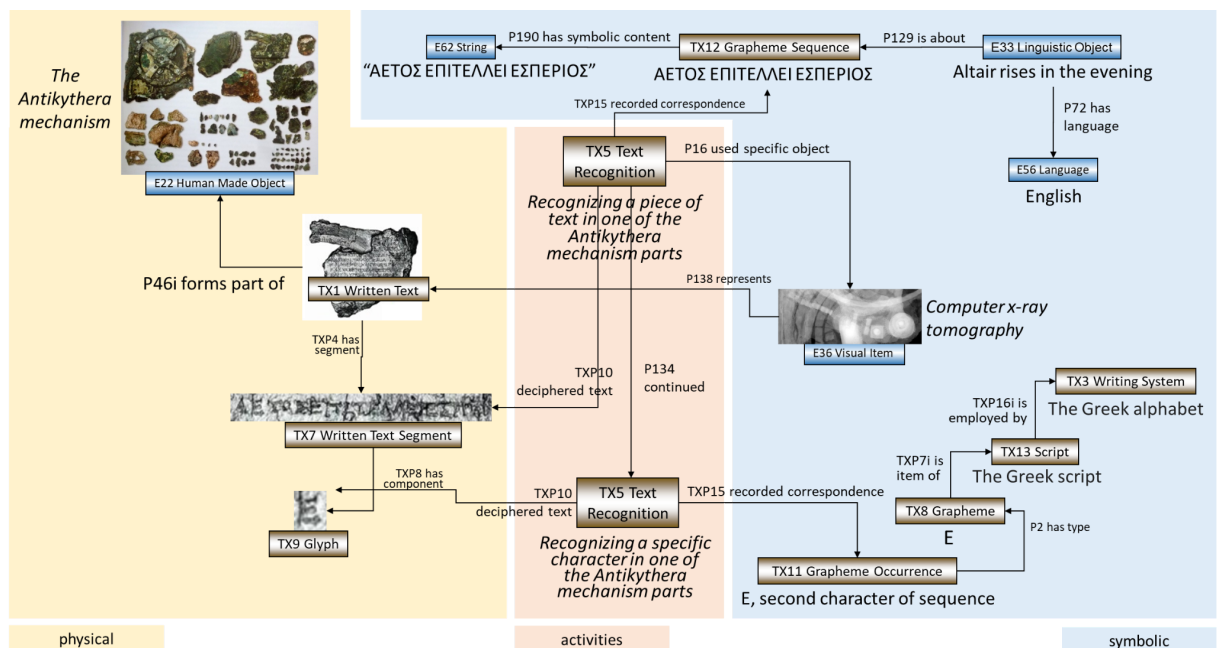


Figure 2: Instantiation example: recognizing text in the Antikythera mechanism.

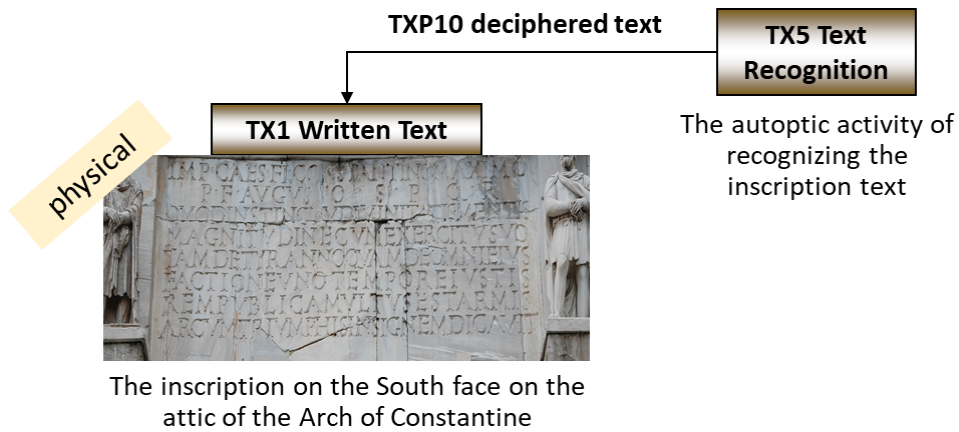


Figure 3: Text recognition through autoptic investigation only.

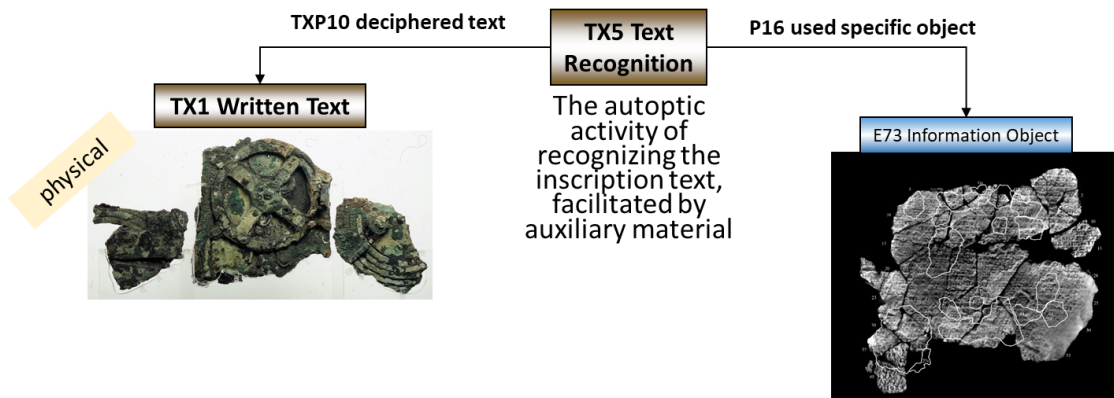


Figure 4: Text recognition through autoptic investigation and additional auxiliary material.

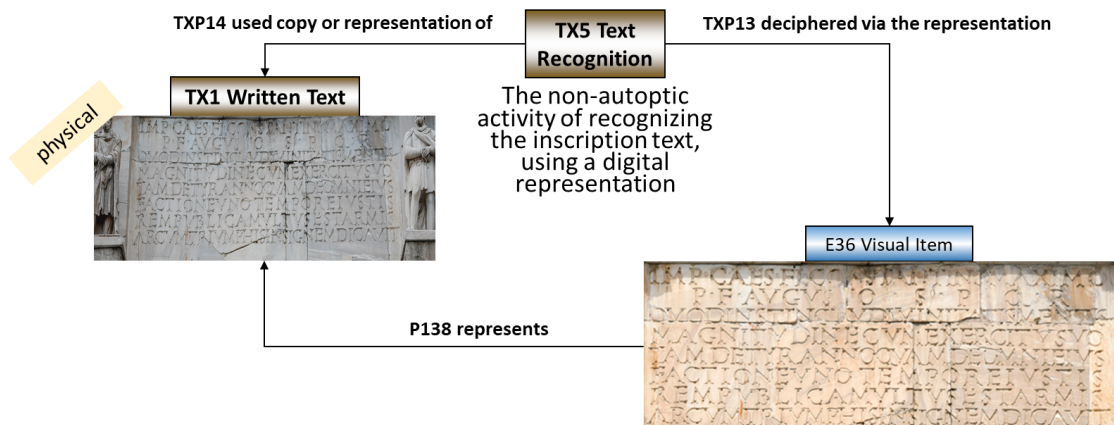
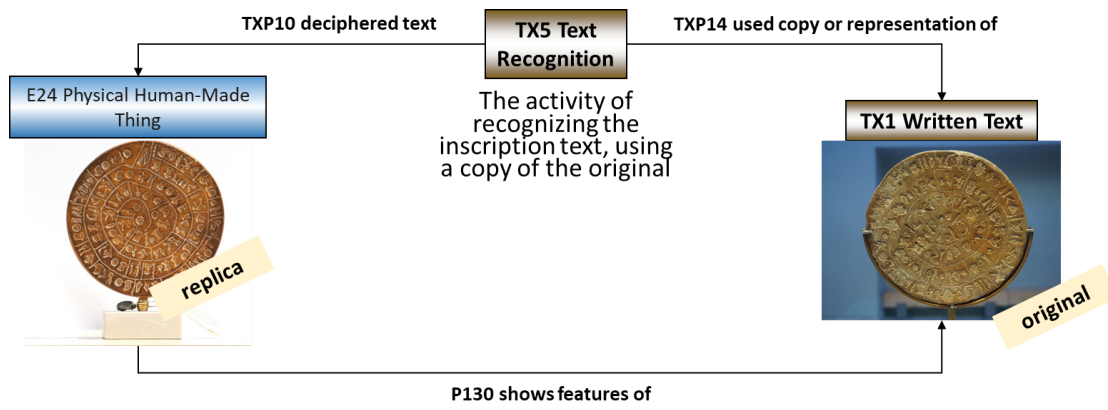


Figure 5: Text recognition through non-autoptic investigation, using a digital representation only.



**Figure 6:** Text recognition using a copy/replica of the original thing.

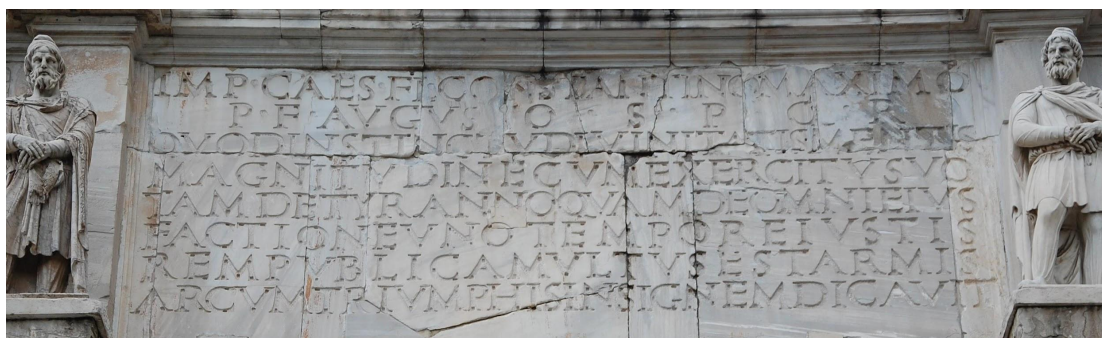
## Classes and properties usage examples

The following example is intended to illustrate how CRMtex classes and properties could be used to encode, for instance, epigraphic information. The inscriptions on the Arch of Constantine, one of the most famous ancient monuments in Rome, have been chosen as examples of an ancient text occurring on a physical carrier, in order to show how they can be semantically described in relation to the archaeological object carrying them.

The monument, still located in its original position between the Colosseum and the Roman Forum, is a triumphal marble arch – the largest monument of this kind in the Roman Empire – dedicated in 315/316 A.D. by the Roman Senate to the emperor Constantine, after his victory over Maxentius in the Battle of the Milvian Bridge in 312 A.D.

Among other decorations (including statues, panels, reliefs and similar decorative material), the arch carries, on its attic, two identical inscriptions (reference number: CIL VI 1139), originally inlaid with gilded bronze letters, explaining the reason for its construction.

As of today, the bronze letters are lost and only the large cuttings in the marble, in which the bronze letters sat, remain. The inscription (Fig. 7) is repeated, identically, on the South and North faces of the arch's attic. A transcription and a translation in English of the same inscription are presented below.



**Figure 7:** The inscription on the South face on the attic of the Arch of Constantine.

### Inscription Transcription

IMP(ERATORI) · CAES(ARI) · FL(AVIO) · CONSTANTINO · MAXIMO · P(IO) ·  
F(ELICI) · AVGUSTO · S(ENATUS) · P(OPULUS) · Q(UE) · R(OMANUS) · QVOD ·  
INSTINCTV · DIVINITATIS · MENTIS · MAGNITVDINE · CVM · EXERCITV · SVO ·  
TAM · DE · TYRANNO · QVAM · DE · OMNI · EIVS · FACTIONE · VNO · TEMPORE ·  
IVSTIS · REMPVBLICAM · VLTIVS · EST · ARMIS · ARCVM · TRIVMPHIS ·  
INSIGNEM · DICAUIT

### Inscription Translation

*To the Emperor Caesar Flavius Constantine, the Greatest, Pius, Felix, Augustus: inspired by (a) divinity, in the greatness of his mind, he used his army to save the state by the just force of arms from a tyrant on the one hand and every kind of factionalism on the other; therefore, the Senate and the People of Rome have dedicated this exceptional arch to his triumphs.*

## CRMtex description of the text

The Arch is an archaeological object and according to the CIDOC CRM it can be represented as an instance of the E22 Human-made Object class. The monument, made of marble, was overall intended to commemorate the emperor. A writing event (TX2) can be assigned to the inscriptions, thus it is always possible to distinguish the production event of the monument from the one of the inscriptions when it is needed.

CRMtex can be used to describe the two inscriptions appearing on the arch and relate them to the monument via the P56 bears feature (is found on) property. Each of the two inscriptions can be rendered as a TX1 Written Text, being the physical features intended to carry a particular significance. A TX2 Writing event can be specified for each TX1 via the *TXP5 was written by* property to render the production of the cuttings made to host the bronze letters. Since there are two inscriptions, we have the opportunity, this way, to distinguish the two processes that led to the production of each of them.

The TX4 Writing Field class can be used to describe the portion of the surface of the arch reserved by the builders and appositely arranged for accommodating the inscription, in order to highlight it from the other parts of the object and to enhance its readability. Thus, the CRMtex encoding in this case will include two instances of TX4.

The linguistic message to be conveyed (E33 Linguistic Object) encoded by means of a language (E56 Language) and by means of the writing system (TX3 Writing System) this language uses. From this follows that the TX1 Written Text class is the concrete graphical manifestation (i.e. a set of signs – in this case the engraved letters – we can read on the stone) of the content of an expression encoded in language through the semiotic activity of writing (TX2 Writing), by means of a TX3 Writing System (in this case, Latin alphabet) and of the graphemes (TX8) composing it.

The reading of a text, from a semiotic point of view, is a decoding activity. In CRMtex a reading – specially carried out for scientific purposes – can be documented using the TX14 Reading class, underlying the scientific nature of the investigation.

In fact, over the centuries, the arch of Constantine has been investigated thousands of times by scholars from all over the world and also reproduced by famous illustrators such as Giovan Battista Piranesi. Its inscriptions have been studied and transcribed several times in order to understand its nature, clarify the meaning of each section and improve its historical comprehension so as to put it in direct relation with the events that determined its creation. For this type of activity, specific classes and properties. The transcription of the text(s) present in *Corpus Inscriptionum Latinarum* (CIL VI 1139), for instance, can be represented via the TX6 Transliteration class, while the analysis of the same inscription(s) carried out by Rodolfo Lanciani in 1892 [6] can be documented using the TX14 Reading class.

The TX7 Written Text Segment class can be used to highlight specific portions of text on which the study focuses, on which specific phenomena appear or from which it is possible to derive special meanings.

# CRMtex - Classes and properties

## CRMtex Classes declarations

The classes are comprehensively declared in this section using the following format:

- Class names are presented as headings in bold face, preceded by the class's unique identifier;
- The line "Subclass of:" declares the superclass of the class from which it inherits properties;
- The line "Superclass of:" is a cross-reference to the subclasses of this class;
- The line "Scope note:" contains the textual definition of the concept the class represents;
- The line "Examples:" contains a bulleted list of examples of instances of this class;
- The line "In first-order logic:" expresses the formal constraints of the class in terms of logical axioms in a first-order logic notation;
- The line "Properties:" declares the list of the class's properties. Each property is represented by its unique identifier, its forward name, and the range class that it links to, separated by colons;
- Inherited properties are not represented;
- Properties of properties, if they exist, are provided indented and in parentheses beneath their respective domain property.



## TX1 Written Text

Subclass of:

E25 Human-Made Feature

Superclass of:

[TX7](#) Written Text Segment

Scope Note:

This class comprises visible or tactile marks (called glyphs or graphs), which relate in a systematic way to units of speech, intentionally traced (i.e., “written”) on some kind of physical support by using specific techniques and tools, with the purpose of conveying a message towards a given receiver or group of receivers.

Examples:

- The signs composing the inscription engraved on the South side of the attic of the Arch of Constantine (E22) in Rome (see section 1.3.1).
- The signs composing the text written on papyrus PSI XIII 1304 containing the so-called *Hellenica Oxyrhynchia* (TM 59482<sup>7</sup>).

In First Order Logic:

$TX1(x) \Rightarrow E25(x)$

Properties:

[TXP4](#) has segment (is segment of): [TX7](#) Written Text Segment

[TXP8](#) has component (is component of): [TX9](#) Glyph

[TXP9](#) is encoded using (was used to encode): [TX3](#) Writing System

[TXP12](#) has style (is style of): [TX10](#) Style

(TXP12.1 has type: E55 Type)

## TX2 Writing

Subclass of:

F28 Expression Creation

Scope Note:

This class describes the activity of communicating information by means of permanent, visible marks in a non-mechanical way, using various techniques (painting, sculpture, etc.) and by means of specific tools, on a given support.

Examples:

- The process of engraving in the marble of the inscription (TX1) placed on the south attic of the Arch of Constantine (E22) in Rome.

In First Order Logic:

$TX2(x) \Rightarrow F28(x)$

---

<sup>7</sup> <https://www.trismegistos.org/text/59482> (accessed on 2023/06/06)

Properties:

[TXP1](#) used writing system (writing system used by): [TX3](#) Writing System  
[TXP5](#) wrote (was written by): [TX1](#) Written Text

## TX3 Writing System

Subclass of:

E29 Design or Procedure

Scope Note:

This class represents a conventional symbolic system designed to represent units of a natural language with the purpose of recording and transmitting information. A writing system consists of a set of symbols (graphemes, TX8), instantiated through physical signs of a visual or tactile nature (glyphs, TX9) representing linguistic units of any kind and the related syntactic (i.e., graphotactic) rules.

It is used to produce a TX1 Written Text during a TX2 Writing event.

Examples:

- The Latin alphabet used to encode the signs (TX1) composing the text (E33) of the inscription in Latin language occurring on the Arch of Constantine (E22).
- The Roman Latin writing system for creating public inscriptions.
- The Cypriot syllabary<sup>8</sup> used in Iron Age Cyprus for codifying the Arcado-Cypriot dialect.
- The Chinese (Han) script used by Wang Xizhi to write the manuscript *Lanting Xu* (“Orchid Pavilion Preface”).

In First Order Logic:

$TX3(x) \Rightarrow E29(x)$

Properties:

[TXP6](#) encodes (is encoding of): E56 Language  
[TXP16](#) employs script (is employed by): TX13 Script

## TX4 Writing Field

Subclass of:

E25 Human-Made Feature

Scope Note:

This class describes the portion of the physical carrier arranged and usually reserved and delimited for the purpose of accommodating a written text,

---

<sup>8</sup> <https://www.worldswritingsystems.org/> (accessed on 2023/06/06)

highlighting and isolating it from the other parts of the object to which it belongs, enhancing and guaranteeing its readability. This entity is paramount specially in epigraphy, in which a specific element called “epigraphic field” has been defined by the discipline itself. Its importance is also evident in papyrology and codicology, where a clear distinction between area(s) containing the written text and empty parts of the support (margins, *intercolumnia*, etc.) is significant for the definition of styles and periods of the document.

Examples:

- The portion of the marble tombstone<sup>9</sup> (E22) of M. Helvius Geminus from Ephesus reserved for accommodating the inscription (TX1).

In First Order Logic:

$TX4(x) \Rightarrow E25(x)$

Properties:

[TXP2](#) includes (is included within): [TX1](#) Written Text

## TX5 Text Recognition

Subclass of:

S4 Observation

E65 Creation

Scope Note:

This class comprises activities of recognizing physical features on some surface, often an instance of TX4 Writing Field, as an arrangement of a series of identifiable glyphs of some known script, deciphered or not, in an order characteristic for a text.

For study purposes, the text recognition procedure requires a scientific autoptic examination of the text. An autoptic examination consists of an accurate analysis of the surface and the signs, and prescribes the use of specific tools and procedures for establishing the exact value of each sign on the physical feature. Deterioration of the original medium or “sloppy” writing may render parts of the original text as undecipherable or ambiguous, which may be annotated in the transcript following epigraphic standards; a text recognition typically results in a record of an equivalent sequence of graphemes on another persistent medium in a scholarly established form of representation of the respective graphemes, often called a “transcript”.

An instance of TX5 Text Recognition may in particular apply even to a single glyph, typically forming part of an instance of TX5 Text Recognition applying to a larger sequence of glyphs containing the former glyph.

---

<sup>9</sup> [https://www.britishmuseum.org/collection/object/G\\_1867-1122-415](https://www.britishmuseum.org/collection/object/G_1867-1122-415) (accessed on 2023/06/06)

The recognition process may be assisted by mechanical means, imaging technology, or a traditional squeeze for incised glyphs. In case the recognition process is solely based on the latter, the observation concerns only the representations on the latter as present to the researcher in some physical form or projection and should unambiguously be documented as such.

In case the recognized text has not been documented in a transcript, text recognition may constitute an implicit part of an overarching reading process, instance of TX14 Reading, which has resulted in other noteworthy propositions related to the content of the recognized text. On the other side, recognition of single glyphs or contracted parts of texts, as they are characteristic for the use of ligatured scripts, may quite well be implicitly supported by the reader's comprehension of the text and the creator of the transcript may have chosen not to annotate parts that the reader regarded as unambiguous. Since these cases can often hardly be separated from the shape recognition of the glyphs in isolation, documenting such implicit comprehension as a separate process may not be relevant. It is however regarded as good practice to document explicitly the reading process and associated interpretative reasoning for any non-trivial resolution of ambiguity or gaps in the recognized text that has a bearing on the transcript or further completion of the transcript.

Examples:

- The autoptic investigation of the South inscription ([TX1](#)) on the Arch of Constantine (E22) made by Rodolfo Lanciani between 1893 and 1901.

In First Order Logic:

$TX5(x) \Rightarrow S4(x)$

$TX5(x) \Rightarrow E65(x)$

Properties:

[TXP10](#) deciphered text (was deciphered by): E24 Physical Human-Made Thing

[TXP13](#) deciphered via the representation (was representation used for deciphering): E36 Visual Item

[TXP14](#) used copy or representation of (was deciphered via copy or representation): [TX1](#) Written Text

[TXP15](#) recorded correspondence (was recorded by): [TX12](#) Grapheme Sequence

## **TX6 Transliteration**

Subclass of:

E65 Creation


Scope Note:

This class comprises activities of exactly re-writing (i.e., re-encoding) an instance of TX12 Grapheme Sequence, i.e., the characters of a text, a contiguous part or a single character of it, by using a writing system (TX3)

different from that of the original text, without changing the order of characters or words, by using standard correspondences.

This operation may apply a 1:1 relation between the signs of the two writing systems, a “transliteration” in the narrower sense (e.g., the ALA-LC Romanization of Greek to Latin). It may also apply an approximation of the sounds of a language, as defined by the source writing system, by that of the target writing system, normally called a “transcription” (e.g., the “rōmaji” Romanization of Japanese), or a mixture of both (e.g. the ELOT 743 Type 2 – transcription of Greek to Latin letters). In a broader sense, the term “transcription” also applies to the activity of re-encoding a text using the same writing system (see example 1). The *P16 used specific object (was used for)* property can be used to specify the applied method of correspondence.

Examples:

- Transcription, in Latin letters, of the Latin inscription(s) (TX1) on the Arch of Constantine (E22) reported in *Corpus Inscriptionum Latinarum* (CIL VI 1139).
- The transliteration and the transcription of the ancient Persian name of king Darius I, written in Persian cuneiform , into Latin script as respectively ‘da-a-ra-ya-va-u-ša’ and ‘Dârayavauš’.

In First Order Logic:

$TX6(x) \Rightarrow E65(x)$

Properties:

[TXP11](#) transcribed (was transcribed by): [TX12](#) Grapheme Sequence

## TX7 Written Text Segment

Subclass of:

[TX1](#) Written Text

Superclass of:

[TX9](#) Glyph

Scope Note:

This class describes portions of text considered to be of particular significance by scholars, as witnesses of a certain meaning or bearers of a particular phenomenon relevant to the investigation, study and understanding of a text. Examples of such text portions are columns, fragments, sections, paragraphs, as well as single words or signs, or other components of a written text. To each of these entities can be associated a single production event (TX2) or destruction event (E6), as in the case of letters or words damaged or worn by atmospheric agents or human interventions, as well as specific conditions (E3) for documenting its status during the text recognition process (TX5). The relationship between a written text (TX1) and its components is documented through the *TXP4 has segment* property.

Examples:

- The “INSTINCTV DIVINITATIS” text portion of the inscription (TX1) on the Arch of Constantine (E22), commented by Rodolfo Lanciani in 1892, in his book *Pagan and Christian Rome* (see section 1.3.1).
- The first paragraph of the Darius I’s inscription (TX1) in Bagistan.

In First Order Logic:

$$\text{TX7}(x) \Rightarrow \text{TX1}(x)$$

## TX8 Grapheme

Subclass of:

E55 Type


Scope Note:

This class comprises symbols used as kinds of atomic units with distinctive value in a given writing system in order to represent linguistic units of some level to encode elements of a message. According to the typology of the writing system, the represented linguistic units can be phonemes (as in Latin), syllables (as in Mycenaean Linear B), up to complete words (as in Chinese and Sumerian scripts).

A writing system also provides the conventions determining how the graphemes are to be used to write a language (orthographic rules).

In some writing systems, graphemes may also be used as auxiliary signs, for instance, for disambiguating senses of homonyms, as in the Japanese writing system, or to mark the semantic categories of the words, as in the ancient Egyptian determinatives.

Examples:

- The abstract unit “S” of the Latin alphabet, used to represent the /s/ sound
- The abstract unit  of the ancient Persian syllabary, used to represent the /da/ syllable.
- The abstract unit “安” of the Han script, used to represent the meaning “peace”.
- 行きます, ching, gyo, iku, zuku.

In First Order Logic:

$$\text{TX8}(x) \Rightarrow \text{E55}(x)$$

## TX9 Glyph

Subclass of:

[TX7](#) Written Text Segment

Scope Note:

This class describes the physical, concrete features traced by a writer, representing the material manifestations of the graphemes needed to codify a linguistic expression. Glyphs are typically observed by the scholars during a text recognition activity (TX5) carried out to decode and recognise the graphemes (TX8) they represent.

Examples:

- The S-shaped feature engraved on the second line of the South inscription on the Arch of Constantine, representing the letter (grapheme) “S” of the Latin writing system used to render the sound of the /s/ phoneme (see section 1.3.1).
- The first feature engraved on the first line of Darius I’s inscription (TX1) in Bagistan, representing the ideal syllabogram  $\overline{\text{𐎧}}$  of the ancient Persian syllabary, used to render the /da/ syllable.

In First Order Logic:

$\text{TX9}(x) \Rightarrow \text{TX7}(x)$

## TX10 Style

Subclass of:

E29 Design or Procedure

Scope Note:

This class describes stylistic variations of texts, including local script styles (as the Carolingian minuscule for the Latin script) and individual scribal hands. It includes: the general appearance of the script, in terms of general design, aspects related to a bilinear system (i.e., upper- and lowercase), measures (i.e., large, medium or small), shape and number of strokes forming a character, its order and direction. A style includes also information about ductus (the direction the text), ligatures and *nexi* (i.e., the connection between characters obtained by tracing them without detaching the writing instrument from the support and using one or more strokes in common), and the writing angle (i.e., the position the writing instrument is located with respect to the support). The style corresponds to fonts and their variations in the modern printing process.

Examples:

- The *Roman square capitals* style, also called *capitalis monumentalis*, or *capitalis quadrata* used to write the inscription on the Arch of Constantine.
- The “Carolingian minuscule” style used in the Carolingian Gospel Book identified as “British Library, Add MS 11848”.

In First Order Logic:

$\text{TX10}(x) \Rightarrow \text{E29}(x)$

## TX11 Grapheme Occurrence


Subclass of:

[TX12](#) Grapheme Sequence

Scope Note:

This class comprises single occurrences of a Grapheme used as an atomic unit at a particular position in the abstract form of a given particular piece of text.

Examples:

- The ideal letter “S” of the Latin alphabet, used to represent the /s/ sound, rendered by the specific S-shaped feature engraved on the second line of the South inscription on the attic of the Arch of Constantine (see section 1.3.1)
- The ideal ‘da’ syllabogram of the ancient Persian syllabary, used to represent the /da/ syllable rendered by the cuneiform sign  engraved on the first line of Darius I’s inscription ([TX1](#)) in Bagistan.

In First Order Logic:

$TX11(x) \Rightarrow TX12(x)$

## TX12 Grapheme Sequence

Subclass of:

E90 Symbolic Object

Superclass of:

[TX11](#) Grapheme Occurrence

Scope Note:

This class comprises particular sequences of Graphemes used for representing the abstract written form of a section of a given particular text.

Examples:

- The grapheme sequence ‘INSTINCTV DIVINITATIS’ [as recognised by the autoptic investigation of the Arch of Constantine, carried out by Rodolfo Lanciani]

In First Order Logic:

$TX12(x) \Rightarrow E90(x)$

Properties:

[TXP17](#) has part (forms part of): [TX12](#) Grapheme Sequence

## TX13 Script

Subclass of:



## E89 Propositional Object

### Scope Note:

This class comprises functionally complete sets of mutually different graphemes employed by one or more languages, regardless of the specific operating rules in a particular language. A writing system, on the other hand, also refers to the set of relations between symbols and linguistic units they represent. The same language may be written using different scripts.

### Examples:

- The Latin script used by the Italian and English writing systems.
- The Latin and the Greek scripts used for the encoding of the Oscan language, creating the Oscan-Greek and Oscan-Latin writing systems.

### In First Order Logic:

$TX13(x) \Rightarrow E89(x)$

### Properties:

[TXP7](#) has item (is item of): [TX8](#) Grapheme

## TX14 Reading

### Subclass of:

I1 Argumentation

### Scope Note:

This class describes the complete intellectual activity, involving the interaction of visual-perceptual, linguistic, and conceptual systems, leading from text recognition (TX5) until its association with a complete linguistic meaning.

### Examples:

- The reading of the South inscription (TX1) on the Arch of Constantine (E22) made by Rodolfo Lanciani between 1893 and 1901.
- The reading of the Greek text present on the Derveni papyrus (E22).

### In First Order Logic:

$TX5(x) \Rightarrow I1(x)$

### Properties:

[TXP18](#) read (was read by): [TX1](#) Written Text

## CRMtex Properties declarations

The properties are comprehensively declared in this section using the following format:

- Property names are presented as headings in bold face, preceded by unique property identifiers;
- The line “Domain:” declares the class for which the property is defined;
- The line “Range:” declares the class to which the property points, or that provides the values for the property;
- The line “Subproperty of:” is a cross-reference to any superproperties the property may have;
- The line “Superproperty of:” is a cross-reference to any subproperties the property may have;
- The line “Quantification:” declares the possible number of occurrences for domain and range class instances for the property;
- The line “Scope note:” contains the textual definition of the concept the property represents;
- The line “Examples:” contains a bulleted list of examples of instances of this property.
- The line “In first-order logic:” expresses the formal constraints of the property in terms of logical axioms in a first-order logic notation.

## **TXP1 used writing system (writing system used for)**

Domain:

[TX2](#) Writing

Range:

[TX3](#) Writing System

Subproperty of:

P33 used specific technique (was used by)

Quantification:

many to many (0,n:0,n)

Scope note:

This property identifies the specific instance of TX3 Writing System employed during the writing event (TX2) that led to the creation of a written text (TX1).

Examples:

- The Roman stonecutter *used writing system* “Latin” (TX3) for the engraving (TX2) of the inscription on the Arch of Constantine (TX1) (see section 1.3.1)
- The Greek scribe *used writing system* “Greek” (TX3) to trace (TX2) in ink the letters that compose the text of the Papyrus of Derveni (TX1).

In First Order Logic:

$TXP1(x,y) \Rightarrow TX2(x)$

$TXP1(x,y) \Rightarrow TX3(y)$

$TXP1(x,y) \Rightarrow P33(x,y)$

## **TXP2 includes (is included within)**

Domain:

[TX4](#) Writing Field

Range:

[TX1](#) Written Text

Subproperty of:

P56 bears feature (is found on)

Quantification:

one to many (0,n:0,1)

Scope note:

This property describes the relation existing between a TX1 Written Text and the TX4 Writing Field, specifically created to accommodate the text, within

which it is inscribed. This relation becomes quite relevant in the very frequent case where more than a single text is found on different areas of a specific support.

Examples:

- The South framework (TX4) carved by the Roman stonecutter on top of the Arch *includes* the inscription on the South face of the Arch of Constantine (TX1).

In First Order Logic:

$TXP2(x,y) \Rightarrow TX1(x)$

$TXP2(x,y) \Rightarrow TX4(y)$

$TXP2(x,y) \Rightarrow P56(x,y)$

### **TXP4 has segment (is segment of)**

Domain:

[TX1](#) Written Text

Range:

[TX7](#) Written Text Segment

Subproperty of:

P46 is composed of (forms part of)

Quantification:

one to many (0,n:0,1)

Scope note:

This property is intended to correlate a text and the different parts of it which a scholar can identify, such as: letters, words, lines, columns, pages, or any other scan that can be made by scholars because it is considered to have a particular relevance for the investigation of the text itself.

Examples:

- The “INSTINCTV DIVINITATIS” text portion *is segment of* the inscription (TX1) on the Arch of Constantine reported and commented by Rodolfo Lanciani in 1892 in his book *Pagan and Christian Rome* (see section 1.3.1).

In First Order Logic:

$TXP4(x,y) \Rightarrow TX1(x)$

$TXP4(x,y) \Rightarrow TX7(y)$

$TXP4(x,y) \Rightarrow P46(x,y)$

$TXP4(x,y) \wedge TX9(x) \Rightarrow \neg TX7(y)$

## **TXP5 wrote (was written by)**

Domain:

[TX2](#) Writing

Range:

[TX1](#) Written Text

Subproperty of:

P108 has produced (was produced by)

Quantification:

one to one (0,1:1,1)

Scope note:

This property is used to describe in detail the close relationship between a text and the writing event that led to its production.

Examples:

- The activity (TX2) carried out by the Greek stonecutters *wrote* the Gortyn Law inscription (TX1) on the wall of the Amphitheatre of Gortyn, Crete.

In First Order Logic:

$TXP5(x,y) \Rightarrow TX2(x)$

$TXP5(x,y) \Rightarrow TX1(y)$

$TXP5(x,y) \Rightarrow P108(x,y)$

## **TXP6 encodes (is encoding of)**

Domain:

[TX3](#) Writing System

Range:

E56 Language

Subproperty of:

P2 has type (is type of)

Quantification:

many to many (0,n:0,n)

Scope note:

This property is used to indicate the language encoded by the TX3 Writing System and used for writing, reading or rendering (i.e. transcribing) a TX1 Written Text.

Examples:

- The Latin alphabet (TX3), used to encode the identical inscriptions (TX1) on the Arch of Constantine, *encodes* the Latin language (E56) used to convey the message of the inscriptions.

In First Order Logic:

$\text{TXP6}(x,y) \Rightarrow \text{TX3}(x)$

$\text{TXP6}(x,y) \Rightarrow \text{E56}(y)$

$\text{TXP6}(x,y) \Rightarrow \text{P2}(x,y)$

### **TXP7 has item (is item of)**

Domain:

[TX13](#) Script

Range:

[TX8](#) Grapheme

Subproperty of:

P67 refers to (is referred to by)

Quantification:

many to many (0,n:0,n)

Scope note:

This property associates an instance of TX13 Script with an instance of TX8 Grapheme employed by this script. Different instances of TX13 Script may have some graphemes in common.

Examples:

- The Latin script (TX13) *has item* the ideal capital letter “S”.

In First Order Logic:

$\text{TXP7}(x,y) \Rightarrow \text{TX3}(x)$

$\text{TXP7}(x,y) \Rightarrow \text{TX8}(y)$

### **TXP8 has component (is component of)**

Domain:

[TX1](#) Written Text

Range:

[TX9](#) Glyph

Subproperty of:

P46 is composed of (forms part of)

Quantification:

one to many (0,n:0,1)

Scope note:

This property is used to state the (physical) belonging of a glyph to a given TX1 Written Text.

Examples:

- The inscription (TX1) on the South face of the Arch of Constantine, *has component* the S-shaped glyph (TX9) engraved on the second line, representing the letter (TX8) “S” of the Latin writing system (TX3).

In First Order Logic:

$\text{TXP8}(x,y) \Rightarrow \text{TX1}(x)$

$\text{TXP8}(x,y) \Rightarrow \text{TX9}(y)$

$\text{TXP8}(x,y) \Rightarrow \text{P46}(x,y)$

### **TXP9 is encoded using (was used to encode)**

Domain:

[TX1](#) Written Text

Range:

[TX3](#) Writing System

Quantification:

many to many (0,n:0,n)

Scope note:

This property has the purpose of directly associating a TX1 Written Text with the TX3 Writing System used for encoding it. It is a shortcut of the more fully articulated path from TX1 Written Text through *TXP5i was written by*, TX2 Writing, *TXP1 used writing system* to TX3 Writing System.

Examples:

- The Gortyn Law inscriptions (TX1), engraved on the wall of the Amphitheatre of Gortyn (Crete), *is encoded using* the Greek alphabet (TX3).

In First Order Logic:

$\text{TXP9}(x,y) \Rightarrow \text{TX1}(x)$

$\text{TXP9}(x,y) \Rightarrow \text{TX3}(y)$

$\text{TXP9}(x,y) \Leftrightarrow (\exists z)[\text{TX2}(z)] \wedge \text{TXP5}(z, x) \wedge \text{TXP1}(z, y)$

### **TXP10 deciphered text (was deciphered by)**

Domain:

[TX5](#) Text Recognition

Range:

E24 Physical Human-Made Thing

Subproperty of:

O8 observed (was observed by)

Quantification:

one to one (0,1:0,n)

Scope note:

This property associates an instance of TX5 Text Recognition with an instance of E24 Physical Human-Made Thing carrying a glyph or a sequence of glyphs that was recognized in the respective activity of text recognition. Typically, the associated instance of E24 Physical Human-Made Thing is more specifically an instance of TX1 Written Text, however, a text may also be recognized from a mechanical copy, a photograph, squeeze or other form of material copy of a written original, which would not by itself constitute an instance of TX1 Written Text. In the latter case, the material copy should be associated with the original written text using the property ‘P130 shows features’.

If the text was actually recognized only from a digital representation, this property should not be used, rather the property *TXP13 deciphered via the representation* should be used instead.

Examples:

- The autoptic investigation (TX5) carried out by Rodolfo Lanciani between 1893 and 1901, *deciphered* the South inscription (TX1) on the Arch of Constantine.

In First Order Logic:

$TXP10(x,y) \Rightarrow TX5(x)$

$TXP10(x,y) \Rightarrow E24(y)$

$TXP10(x,y) \Rightarrow O8(x,y)$

$TXP10(x, z1) \wedge TXP14(x, z2) \Rightarrow P130(z1, z2)$

## **TXP11 transcribed (was transcribed by)**

Domain:

[TX6](#) Transliteration

Range:

[TX12](#) Grapheme sequence

Subproperty of:

P16 used specific object (was used for)

Quantification:

many to many (0,n:0,n)

Scope note:

This property describes the relation between an activity of TX6 Transliteration and the identified sequence of graphemes (TX12) represented in an instance of TX1 Written Text.

Examples:



- The transcription (TX6) of the S-shaped feature engraved on the second line of the South inscription of the Arch of Constantine *transcribed* the prototypical letter “S” (TX8) of the Latin writing system (TX3).

In First Order Logic:

$TXP11(x,y) \Rightarrow TX6(x)$

$TXP11(x,y) \Rightarrow TX12(y)$

$TXP11(x,y) \Rightarrow P16(x,y)$

## TXP12 has style (is style of)

Domain:

[TX1](#) Written Text

Range:

[TX10](#) Style

Subproperty of:

P33 used specific technique (was used by)

Quantification:

many to many (0,n:0,n)

Scope note:

This property describes information about the style used for the realization of the written text (TX1). The property *TXP12.1 has type* allows the nature of the style to be specified, for example to record the direction, orientation or the linear system of the text.

Examples:

- The Latin text in the Carolingian Gospel Book identified as “British Library, Add MS 11848”<sup>10</sup>, *has style* “Carolingian minuscule”
- The inscription (TX1) on the Arch of Constantine *has style* ductus (TX10) *has type* dextroverse (E55)

In First Order Logic:

$TXP12(x,y) \Rightarrow TX1(x)$

$TXP12(x,y) \Rightarrow TX10(y)$

$TXP12(x,y) \Rightarrow P33(x,y)$

Properties:

TXP12.1 has type: E55 Type

---

<sup>10</sup> [https://www.bl.uk/manuscripts/FullDisplay.aspx?ref=Add\\_MS\\_11848](https://www.bl.uk/manuscripts/FullDisplay.aspx?ref=Add_MS_11848) (accessed on 2023/06/06)

### **TXP13 deciphered via the representation (was representation used for deciphering)**

Domain:

[TX5](#) Text Recognition

Range:

E36 Visual Item

Subproperty of:

P16 used specific object (was used for)

Quantification:

one to one (0,1:0,n)

Scope note:

This property associates an instance of TX5 Text Recognition with an instance of E36 Visual Item, capturing the optical impression of an instance of TX1 Written Text by some mechanical method, that was used for recognizing the text without access to the original text and without an explicitly documented material copy or electronic display device that was used for the process.

If the text was actually recognized from an autoptic recognition or from a material reproduction, this property may not be used but the property “TXP10 deciphered text (was deciphered by)” should be used instead.

This property should also not be used, if the recognition of the text was actually carried out from the original text or a material copy of it together with an auxiliary instance of E36 Visual Item. In this case, the use of the auxiliary material should be documented with the more general property *P16 used specific object*.

Examples:

- The recognition of text in the Antikythera mechanism (TX5) *deciphered via the representation* produced using BTI imaging (E36).

In First Order Logic:

$TXP13(x,y) \Rightarrow TX5(x)$

$TXP13(x,y) \Rightarrow E36(y)$

$TXP13(x,y) \Rightarrow P16(x,y)$

$TXP13(x, y) \Rightarrow (\exists z) [TXP14(x, z) \wedge P138(y, z) \wedge \neg TXP10(x, z)]$

### **TXP14 used copy or representation of (was deciphered via copy or representation)**

Domain:

[TX5](#) Text Recognition

Range:

## [TX1](#) Written Text

Subproperty of:

P16 used specific object (was used for)

Quantification:

many to many (0,n:0,n)

Scope note:

This property associates an instance of TX5 Text Recognition, carried out only via copies or representations of a text, with the original instance of TX1 Written Text that was represented on the used copies or digital surrogates.

This property is to be used only for non-autoptic recognition. If this particular recognition of the text was actually carried out from the original text, the property *TXP10 deciphered text* should be used for associating the instance of TX5 Text Recognition with the original instance of TX1 Written Text.

If some form of material copy of the written text was used for the text recognition, then this material copy should be associated with the original written text using the property 'P130 shows features of'.

Examples:

- The non-autoptic recognition of the inscription text on the Arch of Constantine (TX5) *used a copy or representation of* the written text (TX1) on the Arch of Constantine [performed using a photo of the arch].

In First Order Logic:

$TXP14(x,y) \Rightarrow TX5(x)$

$TXP14(x,y) \Rightarrow TX1(y)$

$TXP14(x,y) \Rightarrow P16(x,y)$

$TPX14(x, z2) \wedge TXP10(x, z1) \Rightarrow P130(z2, z1)$

## **TXP15 recorded correspondence (was recorded by)**

Domain:

[TX5](#) Text Recognition

Range:

[TX12](#) Grapheme Sequence

Subproperty of:

P94 has created (was created by)

Quantification:

one to one (0,1:1,1)

Scope note:

This property associates an instance of TX5 Text Recognition with an instance of TX12 Grapheme Sequence that was created by this activity of

text recognition for recording and representing as faithfully as possible the exact value of each sign on the physical material of the recognized instance of TX1 Written Text.

Examples:

- The autoptic investigation carried out by Rodolfo Lanciani (TX5) *recorded correspondence* the grapheme sequence ‘INSTINCTV DIVINITATIS’ on the Arch of Constantine (TX12) .

In First Order Logic:

$TXP15(x,y) \Rightarrow TX5(x)$

$TXP15(x,y) \Rightarrow TX12(y)$

$TXP15(x,y) \Rightarrow P94(x,y)$

### **TXP16 employs script (is employed by)**

Domain:

[TX3](#) Writing System

Range:

[TX13](#) Script

Subproperty of:

P148 has component (is component of)

Quantification:

many to many (0,n:0,n)

Scope note:

This property associates an instance of TX3 Writing System with one of the instances of the script (TX13) it employs.

Examples:

- The Latin writing system used in the inscription of the Arch of Constantine (TX3) *employs script* the Latin script (TX13).
- The Oscan writing system used in the inscription of the Tabula Bantina (TX3) *employs script* the Latin script (TX13).
- The Oscan writing system (TX3) used in the inscription of the Arch of Constantine *employs script* the Greek script (TX13).

In First Order Logic:

$TXP16(x,y) \Rightarrow TX3(x)$

$TXP16(x,y) \Rightarrow TX13(y)$

$TXP16(x,y) \Rightarrow P148(x,y)$

### **TXP17 has part (forms part of)**

Domain:

[TX12](#) Grapheme Sequence

Range:

[TX12](#) Grapheme Sequence

Subproperty of:

P106 is composed of (forms part of)

Quantification:

one to many (0,n:0,1)

Scope note:

This property associates an instance of TX12 Grapheme Sequence with another instance of TX12 Grapheme Sequence appearing at a particular position of the sequence. The property can be also used by an instance of TX11 Grapheme Occurrence (subclass of TX12 Grapheme Sequence) for denoting that a grapheme occurrence has part another grapheme occurrence. Note that a grapheme occurrence may be a symbolic composite containing another grapheme occurrence, such as the minute character “e” on top of the character “u” in former German writing systems denoting the symbol for “ü”.

Examples:

- The “DIVINITATIS” grapheme sequence (TX12), corresponding to the glyph sequence of the inscription (TX1) on the Arch of Constantine, *has part* the “AT” grapheme sequence (TX12) [which appears to be damaged].

In First Order Logic:

$TXP17(x,y) \Rightarrow TX12(x)$

$TXP17(x,y) \Rightarrow TX12(y)$

$TXP17(x,y) \Rightarrow P106(x,y)$

$TXP17(x,y) \wedge TX11(x) \Rightarrow \neg TX12(y)$

**TXP18 read (was read by):**

Domain:

[TX14](#) Reading

Range:

[TX1](#) Written Text

Subproperty of:

P16 used specific object (was used for)

Quantification:

many to many (0,n:0,n)

Scope note:

This property associates an instance of TX14 Reading with an instance of TX1 Written Text whose linguistic meaning was interpreted/understood through the reading process. It is a shortcut of the fully developed path from TX14 Reading through *P9 consists of*, TX5 Text Recognition, *TXP10 deciphered text*, to TX1 Written Text.

Examples:

- Reading the Greek text present on the Derveni papyrus (TX14) *read* the papyrus (TX1) [interpreted the linguistic meaning that was carried by it]

In First Order Logic:

$TXP18(x,y) \Rightarrow TX14(x)$

$TXP18(x,y) \Rightarrow TX1(y)$

$TXP18(x,y) \Rightarrow P16(x,y)$

$TXP18(x,y) \Rightarrow (\exists z) [TX5(z) \wedge P9(x,z) \wedge TXP10(z, y)]$

## References

- Borgwaldt S. R., and T. Joyce (eds), 2013. *Typology of writing systems*, Amsterdam – Philadelphia PA, John Benjamins.
- Coulmas F., 1999. *The Blackwell Encyclopedia of Writing Systems*, Oxford – Cambridge MA, Blackwell Publishing.
- Daniels, P. T. and W. Bright (eds), 1996. *The World's Writing Systems*, New York – Oxford, Oxford University Press. 1996.
- Feldherr, A., and G. Hardy (eds), 2006. *The Oxford History of Historical Writing. Volume 1: Beginnings to AD 600*, Oxford, Oxford University Press.
- Harris, R., 2005. *Rethinking Writing*, London – New York, Continuum.
- Jakobson, R., 1960. “Linguistics and Poetics”, in: *Style in Language*, T. Sebeok, ed., Cambridge MA, M.I.T. Press, 350–377.
- Lanciani, R., 1892. *The significance of the inscription on the Arch of Constantine*, in *Pagan and Christian Rome*, Houghton, Mifflin and Company, Boston and New York.
- Lyons, J., 1972. *Human Language*, in: *Non-Verbal Communication*, R. A. Hinde, ed., Cambridge – New York, Cambridge University Press, 49–85.
- Lyons, J., 1977. *Semantics. Volume I*, Cambridge, Cambridge University Press.
- Pulgram, E., 1976. *The Typologies of Writing-Systems*, in: *Writing Without Letters*, W. Haas, ed., Manchester, Manchester University Press, 1–28.
- Rogers, H., 2005. *Writing Systems: A Linguistic Approach*, Oxford, Blackwell.
- Sampson, G., 1985. *Writing Systems: A Linguistic Introduction*, London, Hutchinson (2nd ed. Sheffield, Equinox, 2015).
- Sampson, G., 2016. “Typology and the study of writing systems”, *Linguistic Typology* 20(3), 561–7.
- Saussure, F. de, 1983. *Course in General Linguistics*, Lausanne – Paris, Payot, 1916, translated and annotated by Roy Harris, London, Duckworth, 1983.