CRMtex

Textual Entities Model

Definition of the CRMtex
An Extension of CIDOC CRM to Model Ancient Textual Entities

Proposal for approval by CIDOC CRM - SIG

Version 1.0
June 2020

Currently maintained by Francesca Murano and Achille Felicetti.

Contributors: Martin Doerr, Francesca Murano, Achille Felicetti

Index

1 Introduction

1.1 Scope

This document presents CRM*tex*, an extension of the CIDOC CRM created to support the study of ancient documents from their material form on and to identify relevant textual entities involved in their study; furthermore, it proposes the use of the CIDOC CRM to encode them and to model the scientific process of investigation related to the study of ancient texts in order to foster 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 by its extensions, this extension introduces new specific classes more adequate to the specific needs of the various disciplines involved (including papyrology, palaeography, codicology and epigraphy). The effective application of IT to the study of ancient direct sources for expanding our knowledge of the past is the inspiring principle of this work.

What is said and what is written

Although every speech can be transposed into an equivalent written message, and vice versa, speech has a priority over writing, at least in four respects: phylogenetic, ontogenetic, functional and structural [1]. In fact, all languages are spoken but not necessarily written; every human being learns to speak naturally and spontaneously, the ability to write coming only later and through specific training; the spoken language is used in a wider and differentiated range of uses and functions; writing originated as a representation of speech. According to Ferdinand de Saussure [2], in fact, «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».
In this semiotic perspective, it is worth considering that even in writing, as in the analysis of the linguistic system, it is necessary to distinguish the concrete level of the personal performance (i.e. the real act of tracing signs on a surface) from the abstract level which all the single execution must be took back to, on the basis of a sameness principle (e.g. the identification of an "A", independently from the peculiar shape somebody gives to it).

Investigating written documentation

The first written documents date back to the IV millennium BC. 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 there is a single mechanism of production of written texts (i.e. supports, techniques, etc. are not relevant), 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 and epigraphy for inscriptions). Nevertheless, an interdisciplinary approach is essential, and the identification of common elements is paramount in order to confer uniformity and interoperability to all these disciplines, as well as to exploit complementary skills from different approaches.
The first and normally most obvious feature that holds our attention examining these documents is the text they bear, even though deterioration may make sophisticated methods necessary to reveal the textual nature. The second thing that should be observed, specifically in ancient texts, is the special relationship between the text and its support. In comparison to modern digital or printed texts, ancient ones are characterised by their uniqueness, being they the result of manual work rather than a mechanised process, as occurs with modern printing.

This and other 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 particular, personal writing styles, slurry writing, and deterioration make reading a complex task often evaluating possible interpretations of the symbols used against plausible senses. As with any human activity, writing also happens *hic et nunc*, therefore, even in the case of texts written by the same person on identical media and with identical technique, hand-writing is never completely identical with itself. This is the case of the *codices* produced by the amanuenses in European monasteries during the Middle Ages. *Au contraire*, starting from the invention of printing press, modern printed copies of books and documents are totally indistinguishable from one specimen to another, since the characters are etched from an identical matrix.

In the ancient world, nevertheless, some types of inscriptions were created through mechanised processes, such as the legends of coins, medals, stamps and seals. Also, the early printed texts, created before the invention of new industrial processes, are unique exemplars, since they were produced through typefaces created by hand, in the same style as manuscript. But even for these classes of objects it is fundamental to investigate the close relation linking the text with the ancient object that carries it. The uniqueness of the written text remains unchanged in this case also, since it is characterised by the peculiar history of the support.

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 handwritten texts and then to describe them by means of appropriate ontological instruments in a multidisciplinary perspective.

Since writing is an intellectual process aimed at the encoding of a language, it is absolutely necessary to understand and follow the underlying semiotic mechanism and, therefore, to distinguish – as we have seen – between the physical manifestation of the text, understood as a set of physical features shown on a given support through the use of a specific technique (e.g. scribbled with ink, painted, engraved, etc.), from its abstract dimension, i.e. from the set of mental images represented by these same physical features. In writing, as in any semiotic system, every component (sign) possesses a dual nature, one physical and another conceptual. 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 (see Figure 1).

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 text itself before moving to the investigation of their archaeological, palaeographic, linguistic and historical features (see Figure 2).

Written text segments

In designing the entities of CRMtex we began by thoroughly investigating the interconnections existing between the text and its various components. Some elements have proved to be absolutely essential for this purpose. On the side of the reading process (i.e. the decoding of the text), and therefore of the investigation of the text by the scholars, one in particular has strongly claimed the scene: the text segment element. Examples of text segments are: text columns, text fragments, sections, paragraphs, single words or letters, or other specific components of the written text.

Scholars of different disciplines, on the basis of the requirements of their study, need to identify and focus their attention on different types of segments, in order to describe their physical conditions (form, layout, etc.), to verify their legibility and particular phenomena (e.g. linguistic or palaeographic) connected to them, etc. It is important to unambiguously represent the logic and semantic relation of meronymy, connecting the entire text with its constituent parts, i.e the identified segments. In this way it is possible to assign specific issues to the individual segments, independently of the text as a whole. In fact, particular production (i.e. TX2 *Writing*) or destruction (E6) events can be associated to single segments, as in the case of letters or words damaged or worn by atmospheric agents or human interventions.

Specifications about conditions (E3) for documenting its status during the observation process (S4) can be easily stated as well. This allows scholars to document different events for the investigated segments in a more precise and punctual way and to assign observations and interpretations to them (see Fig. 3).

Glyphs and Graphemes

The physical signs composing written text constitute the material manifestations (glyphs) of writing system units, i.e. the graphemes, the minimal functional distinctive units of writing. Ernst Pulgram stated that «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 syntagmeme 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» [3]. For

instance: in a Latin inscription, single alphabet letters (glyphs) represent graphemes, a grapheme corresponding to a letter only in alphabetic system of writing. In Mycenaean Linear B inscriptions and in Old Persian cuneiform inscriptions glyphs represent syllabograms (graphemes representing a syllable, not a single sound); in an Egyptian hieroglyphic text, glyphs represent syllabic, alphabetic and also ideographic elements, i.e. elements standing for lexical/semantic units.

Phonographic writing systems [4][5] represent phonological units of one size or another, but the 1:1 correspondence between sound (phoneme, syllables, *etc.*) and sign (grapheme) is lost in diachrony, obscured by spelling conventions and phonetic changes to which linguistic systems are subjected in history. Examples are provided, for instance, in English, where there are many discrepancies in spelling between writing and phonetic values: for example, the grapheme <i> stands for various phonemes: /ɪ/ (as in *him*), /ʌɪ/ (as in *time*), /i/ (as in *police*), /a/ (as in *timbre*); *vice versa*, the phoneme /f/ can be represented by <f> (as in *film*), <ph> (as in *philology*) or <gh> (as in *enough*).

Reading and understanding the text

Reading refers to the semiotic procedure of decoding and therefore understanding a written text.
This procedure can be carried out for scientific purposes, in order to analyse and study the text according to different disciplinary perspectives.

On the side of the message retrieval, since each grapheme is bound to a given linguistic unit of specific languages, reading the written message presupposes the ability to read the language of the writer.

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 phonetic value, also on the basis of spelling conventions in place in a given graphic system at a given historical moment, since the orthographic rules can change over time, even if less quickly than the linguistic system does.

In observing a text, thus, it is necessary to keep the deciphering procedure separate from the reading one. In the case of languages and writing systems that are no longer in use, in fact, it is possible that scholars are unable to decipher the signs, i.e to establish the value that those signs 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 not known.
The deciphering of the signs can be achieved if the linguistic system represented is known. This is the case of Linear B, whose deciphering came after 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 it is not yet known. This is the case of the Etruscan writing system, which was deciphered from the beginning, the Etruscan system deriving from the Greek one, although knowledge of the linguistic system has progressed gradually among the scholars.

The reading activity, thus, is intended as a specific observation (S4) in which the decoding of the signs is performed, i.e. the linguistic value is recognised and the message is understood. Cases in which decoding does not happen (e.g., the observer is able to describe the signs but not to assign a specific linguist value to them), the S4 class could be used as it is.

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 on the physical feature.

Transcription and transliteration

Transcription consists in representing the sounds of a language using a writing system (TX3) different from that of the original text (e.g., Latin letters to render a Mycenaean text). Although in the literature the current term of transcription is often found, on a scientific level this operation often properly consists in a 'transliteration', i.e. the conversion of the graphemes of a writing system into the graphemes of

another writing system. Thus, transliteration implies a 1:1 relation between the signs of the two writing systems, graphemes of the starting system always corresponding to the same ones of the target system. A transliteration is never ambiguous. For example, the name of Thebes is written Θήβα in Greek alphabet; the sequence has transliteration 'Thēba' in Latin script but has transcription 'Thìva', according with modern Greek pronunciation; notice that transcription is based on the phonetics, thus pronunciation problems can arise: for example an English speaker might read 'Thìva' as [ˈθʌɪva] instead of [ˈθiva], and possibly transcribe 'Theeva').

Style and other palaeographic features

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

This approach is fundamental for the determination of the dating and origin of the texts, especially in reference to the individual 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 the 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. The P15 was influenced by and the E83 Type Creation classes could constitute an optimal starting point for this activity.

Palaeography uses different concepts, including aspects of the style (TX10), ductus and other features related to the physical manner in which the text is written. Both individual and standardised styles (i.e., my personal handwriting and Carolingian minuscule) can be considered as subclasses of E29 Design or Procedure. TX1 Written Text is related to TX10 Style through the TXP12 has style property.

1.2 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.

1.3 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 that identifier consists of the letter 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 with their respective names, classes denoted by S and properties by O.

Letters in red colour in CRM Classes and properties are additions/extensions coming by the scientific observation model.

2 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.

2.1 CRMtex class hierarchy, aligned with portions from the CRMsci, LRMoo and the CIDOC CRM class hierarchies

This class hierarchy lists:
• all classes declared in Ancient Text model (CRMtex)
• all classes declared in CRM*sci* and CIDOC CRM that are declared as superclasses of classes declared in the Ancient Text Model,
• all classes declared in CRM*sci* or CIDOC CRM that are either domain or range for a property declared in the Ancient Text Model,
• all classes declared in CRM*sci* and CIDOC CRM that are either domain or range for a property declared in Ancient Text Model or CIDOC CRM that is declared as superproperty of a property declared in the Ancient Text Model,
• all classes declared in CRM*sci* and CIDOC CRM that are either domain or range for a property that is part of a complete path of which a property declared in Ancient Text Model is declared to be a shortcut.

E1	CRM Entity								
S15	-	Observable Entity							
E2	-	-	Temporal Entity						
E5	-	-	-	Event					
E7	-	-	-	-	Activity				
TX6	-	-	-	-	-	Transcription			
E13	-	-	-	-	-	Attribute Assignment			
S4	-	-	-	-	-	-	Observation		
TX5	-	-	-	-	-	-	-	Reading	
E63	-	-	-	-	Beginning Of Existence				
E12	-	-	-	-	-	Production			
F28	-	-	-	-	-	-	Expression Creation		
TX2	-	-	-	-	-	-	-	Writing	
E77	-	-	Persistent Item						
E70	-	-	-	Thing					
E72	-	-	-	-	Legal Object				
E18	-	-	-	-	-	Physical Thing			
E26	-	-	-	-	-	-	Physical Feature		
E25	-	-	-	-	-	-	-	Man-made Feature	
TX1	-	-	-	-	-	-	-	-	Written Text
TX7	-	-	-	-	-	-	-	-	-	Written Text Segment
TX9	-	-	-	-	-	-	-	-	Glyph
TX4	-	-	-	-	-	-	-	-	Writing Field
E71	-	-	-	Man-made Thing					
E28	-	-	-	-	Conceptual Object				
E90	-	-	-	-	-	Symbolic Object			
TX8	-	-	-	-	-	-	Grapheme		
E73	-	-	-	-	-	-	Information Object		
E29	-	-	-	-	-	-	-	Design or Procedure	
TX3	-	-	-	-	-	-	-	-	Writing System
TX10	-	-	-	-	-	-	-	-	Style

2.2 CRMtex property hierarchy, aligned with portions from the CRMsci and the CIDOC CRM property hierarchies

This property hierarchy lists:
• all properties declared in Ancient Text Model,
• all properties declared in CRM*sci* and CIDOC CRM that are declared as superproperties of properties declared in Ancient Text Model,
• all properties declared in CRM*sci* and CIDOC CRM that are part of a complete path of which a property declared in Ancient Text Model, is declared to be a shortcut.

| Property id	Property Name	Entity - Domain	Entity - Range
TXP1	used writing system (writing system used for)	TX2 Writing	TX3 Writing System
TXP2	includes (is included within)	TX4 Writing Field	TX1 Written Text
TXP3	rendered (is rendered by)	TX6 Transcription	TX5 Reading
TXP4	has segment (is segment of)	TX1 Written Text	TX7 Written Text Segment
TXP5	wrote (was written by)	TX2 Writing	TX1 Written Text
TXP6	encodes (is encoding of)	TX3 Writing System	E33 Linguistic Object
TXP7	has item (is item of)	TX3 Writing System	TX8 Grapheme
TXP8	has component (is component of)	TX1 Written Text	TX9 Glyph
TXP9	is encoded using (was used to encode)	TX1 Written Text	TX3 Writing System
TXP10	read (was read by)	TX5 Reading	TX1 Written Text
TXP11	transcribed (was transcribed by)	TX6 Transcription	TX8 Grapheme
TXP12	has style (is style of)	TX1 Written Text	TX10 Style

3 Graphical overview
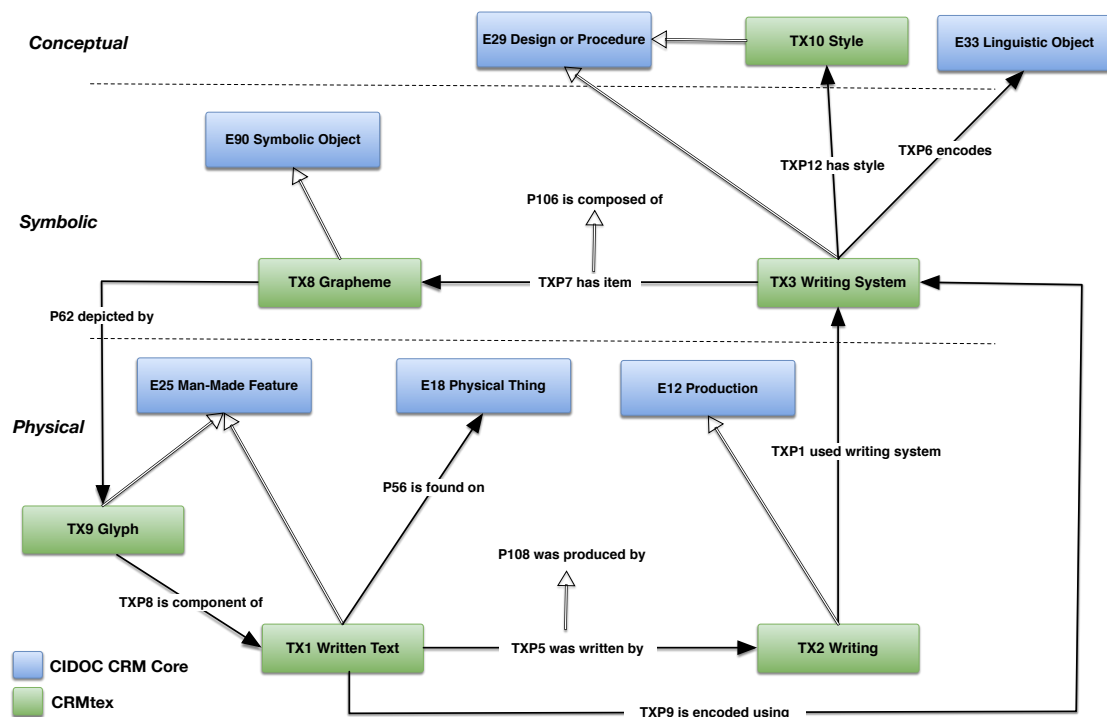
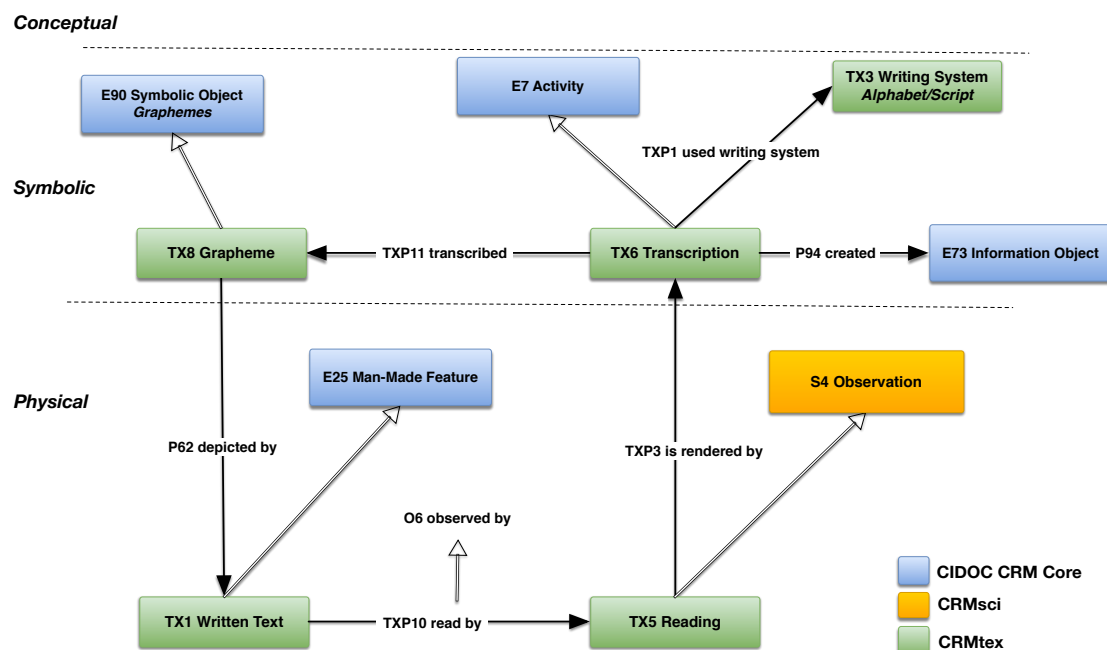**Figure 1: Text entities and text production in CRMtex**

Figure 2: The process of investigation of ancient texts in CRMtex
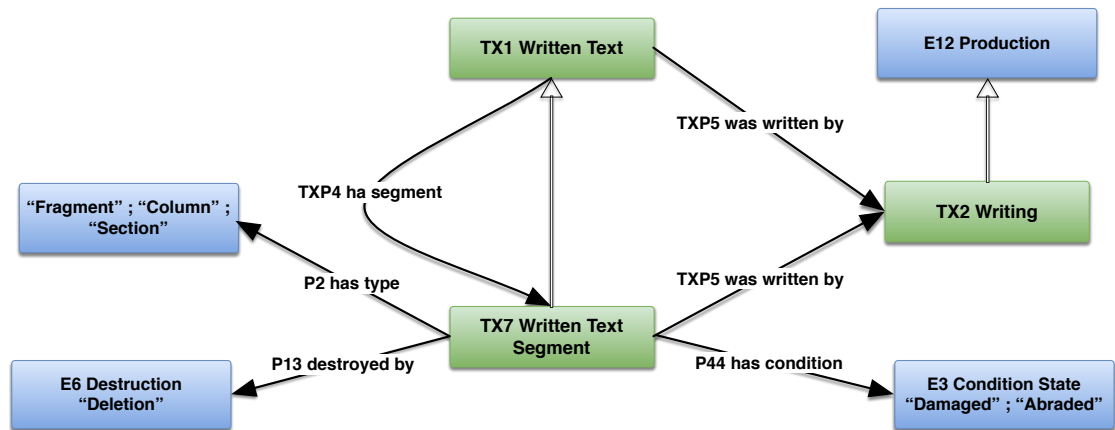
Figure 3: Written Text and Written Text Segments in CRMtex.

4 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 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 the 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 of its construction.
As of today, the bronze letters are lost and only the large cuttings remain in the marble, in which the bronze letters were fixed. The inscription (Figure 5) 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 is presented below.
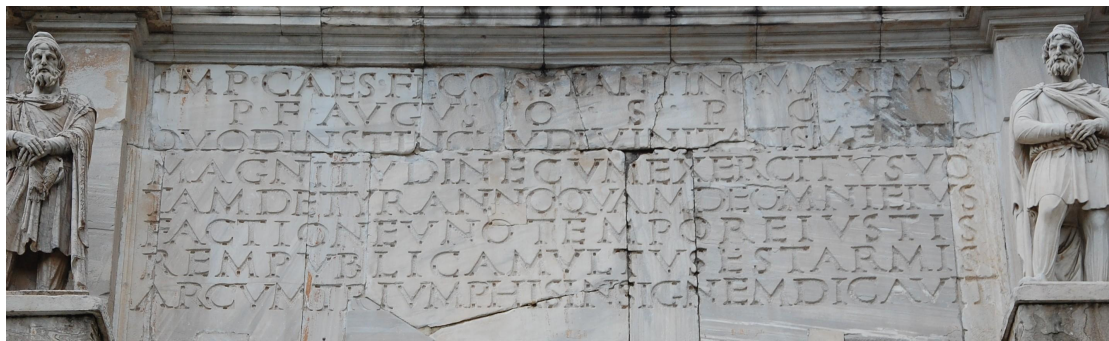

Figure 4: The inscription on the South face on the attic of the Arch of Constantine.

Inscription Transcription

IMP(ERATORI) · CAES(ARI) · FL(AVIO) · CONSTANTINO · MAXIMO · P(IO) · F(ELICI) · AVGVSTO · 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 · VLTVS · EST · ARMIS · ARCVM · TRIVMPHIS · INSIGNEM · DICAVIT

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 *Man-made Object* class. The monument, made of marble, was overall intended to commemorate the emperor and not to carry the various the inscriptions present on it. A writing event (TX2) can be assigned to the inscriptions, thus it is always possible to distinguish the production event of the monument from that 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.

A 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 TX4s instances.

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., the signs – in this case the engraved letters – that we can read on the stone) of the conceptual level of encoding a linguistic expression 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 TX5 *Reading* class, a subclass of the S4 *Observation* CRMsci 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. Also, the 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 *Transcription* class, while the analysis of the same inscription(s) carried out by Rodolfo Lanciani in 1892 [6] can be documented using the reading (TX5) class. Reading and transcription (TX6) activities can be related via the TXP3 *rendered* property, inherited by CIDOC CRM core.

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. Figure 5 shows a CRMtex conceptualisation of the South inscriptions on the Arch of Constantine.
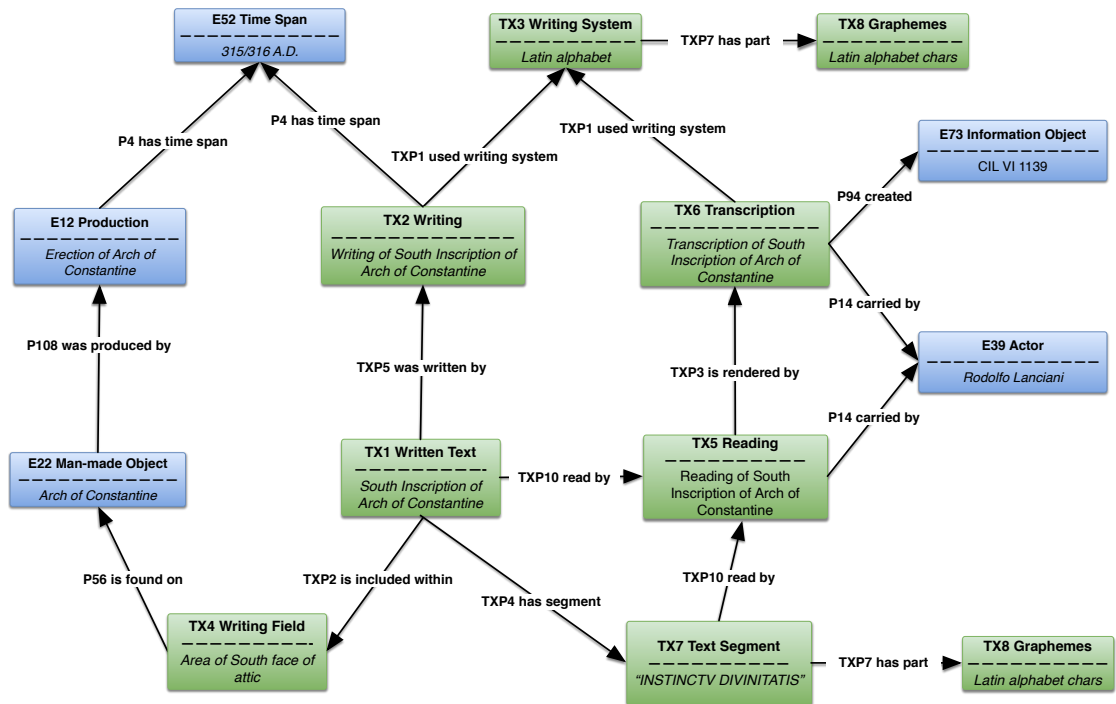
Figure 5: CRMtex encoding of one of the inscriptions (South) on the Arch of Constantine.

5 CRMtex - Classes and properties

5.1 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 "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 Man-Made Feature

Scope Note: This class comprises visible set of symbols (glyphs or graphs) intentionally traced (i.e. "written") on some kind of physical support by using specific techniques and having semiotic significance and the purpose of conveying a message towards a given addressee or group of addressees.

Examples:
- The inscription engraved on the South side of the attic of the Arch of Constantine (E22) in Rome (see section 1.3.1).

In First Order Logic:
TX1(x) ⊃ 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

TX2 Writing

Subclass of: F28 Expression Creation
Superclass of:
Scope Note: Subclass of F28 Expression Creation indicating the activity of physically creating a text, using various techniques (painting, sculpture, etc.) and by means of specific tools, on a given carrier in a non-mechanical way.
.
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 (see section 1.3.1).

In First Order Logic:
TX2(x) ⊃ F28(x)

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
Superclass of:
Scope Note: Subclass of E29 Design or Procedure refers to a conventional system consisting of a set of signs (graphemes, TX8) used to codify a natural language, by means of specific rules in the combination and phonological value assignment of the chosen graphemes. As each code, a writing systems requires shared understanding between writers and readers allowing the transmission of the message. It is used to produce a TX1 Written Text during a TX2 Writing event.

Examples:
- The Latin alphabet used for codifying the inscriptions (TX1) in Latin language occurring on the Arch of Constantine (E22).

- The Cypriot syllabary used in Iron Age Cyprus for codifying the Arcado-Cypriot dialect.

In First Order Logic:

> TX3(x) ⊃ E29(x)

Properties:

> TXP6 encodes (is encoding of): E33 Linguistic Object
> TXP7 has item (is item of): TX8 Grapheme

TX4 Writing Field

Subclass of: E25 Man-Made Feature
Superclass of:
Scope Note: Subclass of E25 Man-Made Feature, it is the portion of the physical support arranged and usually reserved and delimited for the purpose of accommodating a written text, 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 surface of the South part of the attic of the Arch of Constantine (E22) reserved by the builders for accommodating the South inscription (TX1) (see section 1.3.1).
- The areas reserved to writing by the scribe in a papyrus scroll.

In First Order Logic:

> TX4(x) ⊃ E25(x)

Properties:

> TXP2 includes (is included within): TX1 Written Text

TX5 Reading

Subclass of: S4 Observation
Superclass of:
Scope Note: Subclass of the S4 Observation, referring to the semiotic procedure of decoding (and therefore understanding) a written text. This procedure can be carried out for scientific purposes, in order to analyse and study the text according to different disciplinary perspectives. The reading activity is thus intended as a specific observation (S4) in which the decoding of the signs is performed, i.e. the linguistic value is recognised, and the message is understood. Cases in which decoding does not happen (e.g., the observer is able to describe the signs but not to assign a specific linguist value to them), the S4 class could be used as it is. For study purposes, the reading procedure 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 on the physical feature.

Examples:

- The autoptic investigation 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) ⊃ S4(x)

Properties:

    TXP10 read (was read by): TX1 Written Text

## TX6 Transcription

Subclass of: E7 Activity
Superclass of:
Scope Note: Subclass of E7 Activity, referring to the activity of re-writing the text conducted by an editor. This operation could involve a writing system (TX3) different from that of the original text, implying a transposition of the sounds of a language from a writing system to another one (e.g., Latin letters to render a Mycenaean text). For scientific purposes, this operation often consists in a 'transliteration', because it implies a 1:1 relation between the signs of the two writing systems. The P16 used specific object (was used for) property can be used to specify the role of the original graphemes during the commuting operations.

Examples:

- Transcription, in Latin letters, of the 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 𐎭𐎠𐎼𐎹𐎺𐎢𐏁, into Latin script as respectively 'da-a-ra-ya-va-u-ša' and 'Dârayavauš'.

In First Order Logic:

    TX6(x) ⊃ E7(x)

Properties:

    TXP3 renders (is rendered by): TX5 Reading
    TXP11 transcribed (was transcribed by): TX8 Grapheme

TX7 Written Text Segment

Subclass of: TX1 Written Text
Superclass of:
Scope Note: Subclass of TX1 Written Text intended to identify 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 the ancient text. Examples of such text portions are: columns, fragments, sections, paragraphs, as well as single words or letters, or other components of the 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 observation process (S4). 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 long Darius I's inscription (TX1) in Bagistan.

In First Order Logic:

TX7(x) ⊃ TX1(x)

TX8 Grapheme

Subclass of: E90 Symbolic Object
Superclass of:
Scope Note: Subclass of E90 Symbolic Object used to represent the abstract units with distinctive values in a given writing system. A grapheme is a character or sequence of characters that functions as a distinct unit within an orthography. It may be a single character, a multigraph, or a diacritic, but in all cases graphemes are defined in relation to the particular orthography. Most graphemes represent a single phoneme, but some represent a sequence of phonemes. For instance, the character sequence ‹ch› is often used to represent the phoneme /tʃ/ in English, while the single letter ‹x› usually represents the phoneme sequence /ks/. In a highly phonemic writing system, there is a close correspondence between graphemes and phonemes. English (written with Latin script) is an example of a writing system that is not highly phonemic, and therefore the mappings between graphemes and phonemes are more complex. Graphemes are often written enclosed in angle brackets (‹›).

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 𒁕 engraved on the first line of Darius I's inscription (TX1) in Bagistan.

In First Order Logic:

TX8(x) ⊃ E90(x)

TX9 Glyph

Subclass of: E25 Man-Made Feature
Superclass of:
Scope Note: Subclass of E25 Man-Made Feature intended to represent the concrete manifestation of single signs traced by the writer while codifying a linguistic expression. Glyphs are typically observed by the scholar during a reading 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 feature 𒁕 engraved on the first line of Darius I's inscription (TX1) in Bagistan, representing the ideal syllabogram 'da' of the ancient Persian syllabary, used to render the /da/ syllable.

In First Order Logic:

TX9(x) ⊃ E25(x)

TX10 Style

Subclass of: E29 Design or Procedure
Superclass of:
Scope Note: Subclass of E29 Design or Procedure that describes styles and variations of texts, including local script styles (Carolingian minuscule) 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 lowercases), 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 is written), ligatures and *nexi* (i.e. the connection between characters obtained 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 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:
TX10(x) ⊃ E29(x)

5.2 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 "Superproperty of:" is a cross-reference to any subproperties the property may have;
- 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.

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 is intended to identify the specific instance of TX3 Writing System employed during the writing event that led to the creation of a TX1 Written Text.

Examples:

- The Roman stonecutter *used* the Latin *writing system* (TX3) for the engraving (TX2) of the inscription on the Arch of Constantine (TX1) (see section 1.3.1)

- The Greek scribe *used* the Greek *writing system* (TX3) to trace (TX2) in ink the letters that compose the text of the Papyrus of Derveni (TX1).

In First Order Logic:
$$\text{TXP1}(x,y) \supset \text{TX2}(x)$$
$$\text{TXP1}(x,y) \supset \text{TX3}(y)$$
$$\text{TXP1}(x,y) \supset \text{P33}(x,y)$$

TXP2 includes (is included within)

Domain: TX4 Writing Field
Range: TX1 Written Text
Subproperty of: P56 bears feature

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

Scope note: This property is intended to describe 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:
$$\text{TXP2}(x,y) \supset \text{TX1}(x)$$
$$\text{TXP2}(x,y) \supset \text{TX4}(y)$$
$$\text{TXP2}(x,y) \supset \text{P56}(x,y)$$

TXP3 rendered (is rendered by)

Domain: TX6 Transcription
Range: TX5 Reading
Subproperty of: P20 had specific purpose (was purpose of)

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

Scope note: This property is intended to emphasize the close connection for scientific purposes between the TX5 Reading and TX6 Transcription activities, by outlining, in particular,

the accurate observation required by scholars in order to perform a valid transcription of a the text being investigated.

Examples:

- The transcription ([TX6)](#)) of the inscription on South Attic of the Arch of Constantine, carried out by Rodolfo Lanciani, *rendered* its reading ([TX5](#)) of the same inscription done by him in 1892 (see section 1.3.1).

In First Order Logic:

$$TXP3(x,y) \supset TX5(x)$$
$$TXP3(x,y) \supset TX6(y)$$
$$TXP3(x,y) \supset P20(x,y)$$

TXP4 ha 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 which a scholar can identify, such as: letters, words, lines, columns, pages, or any other scan that can be made by scholars because 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).
- The phrase "xšâyathiya xšâyâthiânâm" ("King of Kings") recognised by Grotefend in1802 on the Achaemenid inscription from Persia.

In First Order Logic:

$$TXP4(x,y) \supset TX1(x)$$
$$TXP4(x,y) \supset TX7(y)$$
$$TXP4(x,y) \supset P46(x,y)$$

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:

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

TXP5(x,y) ⊃ TX1(y)
TXP5(x,y) ⊃ P108(x,y)

TXP6 encodes (is encoding of)

Domain: TX3 Writing System
Range: E33 Linguistic Object
Subproperty of: P2 has type

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. The specification of the language is fundamental for understanding the writing phenomena and the rules of the writing system used for the production or transcription of a specific text.

Examples:
  ▪ The Latin alphabet (TX3), used to encode the identical inscriptions (TX1) on the Arch of Constantine, *encodes* the Latin language (E33) used to convey the message of the inscriptions.

In First Order Logic:
TXP6(x,y) ⊃ TX3(x)
TXP6(x,y) ⊃ E33(y)
TXP6(x,y) ⊃ P2(x,y)

TXP7 has item (is item of)

Domain: TX3 Writing System
Range: TX8 Grapheme
Subproperty of: P106 is composed of (forms part of)

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

Scope note: This property is used to state the (conceptual) belonging of a TX8 Grapheme to a given TX3 Writing System.

Examples:
  ▪ The Latin alphabet (TX3), used to encode the inscription (TX1) on South face of the Arch of Constantine, *has item* the idel letter "S" (TX8) used in this alphabet to represent the /s/ sound.

In First Order Logic:
TXP7(x,y) ⊃ TX3(x)
TXP7(x,y) ⊃ TX8(y)
TXP7(x,y) ⊃ P106(x,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 TX9 Gliph to a given TX1 Written Text.

Examples:

▪ The inscription (TX1) on South face of the Arch of Constantine, *contains* 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:
    TXP8(x,y) ⊃ TX1(x)
    TXP8(x,y) ⊃ TX9(y)
    TXP8(x,y) ⊃ 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 from which the signs used for its writing and therefore incorporated in the text have been taken. It is a shortcut of the more fully articulated path from TX1 Written Text, through TXP5 wrote (was written by), TX2 Writing, TXP1 used writing system (writing system used for) to TX3 Writing System.

Examples:

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

In First Order Logic:
    TXP9(x,y) ⊃ TX1(x)
    TXP9(x,y) ⊃ TX3(y)

TXP10 read (was read by)

Domain: TX5 Reading
Range: TX1 Written Text
Subproperty of   O8 observed (was observed by)

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

Scope note: This property links an instance of TX1 Written Text with a TX5 Reading event carried out to investigate its intrinsic characteristics and to perform its decoding.

Examples:

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

In First Order Logic:
    TXP10(x,y) ⊃ TX5(x)
    TXP10(x,y) ⊃ TX1(y)
    TXP10(x,y) ⊃ O8(x,y)

TXP11 transcribed (was transcribed by)

Domain: TX6 Transcription
Range: TX8 Grapheme
Subproperty of: P16 used specific object (was used for)

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

Scope note: This property highlights the specific way in which an activity of TX6 Transcription results in the rendering of the specific TX8 Grapheme(s) of which an instance of TX1 Written Text is composed.

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) ⊃ TX6(x)
TXP11(x,y) ⊃ TX8(y)
TXP11(x,y) ⊃ 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 assigns information to a certain script in order to accurately determine the style used for the realization of the written text. The *TPX12.1 has typ*e property of *TXP12 has style* allows the nature of the style to be specified reading from domain to range, for example to record the ductus, orientation or the linear system of the text.

Examples:
- The Latin text in the Carolingian Gospel Book identified as "British Library, Add MS 11848", has style "Carolingian minuscule"

- The inscription on the Arch of Constantine has ductus (TPX12 has style + TPX12.1 has type "ductus") *dextroverse*

In First Order Logic:
TXP12(x,y) ⊃ TX1(x)
TXP12(x,y) ⊃ TX10(y)
TXP12(x,y) ⊃ P33(x,y)

Properties: TXP12.1 has type: E55 Type

6 Referred CIDOC CRM version 6.2.1 Classes and Properties

Classes	Properties
E1 CRM Entity	
E2 Temporal Entity	
E5 Event	
E6 Destruction	
E7 Activity	**P16 used specific object (was used for)**
	P20 had specific purpose (was purpose of)
	P33 used specific technique (was used by)
E12 Production	**P108 has produced (was produced by)**
E13 Attribute Assignment	
E18 Physical Thing	**P46 is composed of (forms part of)**
E19 Physical Object	**P56 bears feature (is found on)**
E22 Man-Made Object	
E24 Physical Human-Made Thing	**P62 depicts (is depicted by)**
E25 Man-Made Feature	
E26 Physical Feature	
E28 Conceptual Object	
E29 Design or Procedure	
E33 Linguistic Object	
E56 Language	
E63 Beginning of Existence	
E65 Creation	**P94 has created (was created by)**
E70 Thing	
E71 Man-Made Thing	
E72 Legal Object	
E73 Information Object	
E77 Persistent Item	
E89 Propositional Object	**P67 refers to (is referred to by)**
E90 Symbolic Object	**P106 is composed of (forms part of)**

7 Referred CRMsci version 1.2.8 Classes and Properties

Classes	Properties
S4 Observation	**O8 observed (was observed by)**
S15 Observable Entity	

8 LRMoo version 06 Classes and Properties

Classes	Properties
F28 Expression Creation	

References

[1] J. Lyons, Human Language, in: Non-Verbal Communication, R. A. Hinde, ed., Cambridge University Press, New York, 1972, 49– 85.

[2] F. de Saussure, Course in General Linguistics, Lausanne – Paris, Payot, 1916, translated and annotated by Roy Harris, London, Duckworth, 1983.

[3] E. Pulgram, The Typologies of Writing-Systems, in: Writing Without Letters, W. Haas, ed., Manchester, Manchester University Press, 1976, 1–28.

[4] G. Sampson, Writing Systems: A Linguistic Introduction, London, Hutchinson, 1985 (2nd ed. Sheffield, Equinox, 2015).

[5] H. Rogers H., Writing Systems: A Linguistic Approach, Oxford, Blackwell, 2005.

[6] Lanciani, R.: *The significance of the inscription on the Arch of Constantine,* in *Pagan and Christian Rome,* Houghton, Mifflin and Company, Boston and New York, 1892.