

Mining Paper Catalogues

A Multilingual Solution to Reduce Verbose Fields to Consistent Terminology

Felix Kußmaul



University of Cologne

Dr Tim Evans

UNIVERSITY *of York*

1 September 2017

European Association of Archaeologists 2017

MOTIVATION

Form 23 Conical cup with smooth vertical rim
Konische Schale mit glattem Steilrand
Coppa troncoconica con orlo verticale
Coupe tronconique à rebord vertical lisse

Conical cup representing the further evolution of Form 22. The floor is now always flat or biconical (meeting the wall at a sharp angle on the inside), usually with a low foot.

23.1: Plain tapering rim, inclined slightly inwards, sometimes bearing applied decoration.

23.2: Rim with flat outer face bearing applied decoration bounded above and below by simple convex mouldings; inner face plain or with a groove at lip.

Production

Subform 23.1 is probably made in many parts of Italy; examples in Padana ware do not show applied decoration. Subform 23.2 is made in Italy but apparently not in the Padana region.

Date

Subform 23.2 belongs to the second and third quarters of the first century A.D.: it is common in the South Stoa deposits at Corinth and at Pompeii. Subform 23.1 is less readily datable as it may occur as a simplified version of Form 22 or Form 23: other features of the vessel (e.g. foot-profile, decoration) may provide a clearer indication of date than the shape of the rim.

Distribution

Subform 23.2 is very common throughout the Mediterranean region, with sporadic examples found in the North and in North-Italy; Subform 23.1 is relatively uncommon.

References

- 23.1.1 Karthago K 78/172a, unpublished. Stamp L.MA, O.-C. -, Italy.
- 23.1.2 Berenice B210.2. Anepigraphic stamp. Italy.
- 23.2.1 Corinth 1973 pl.84,70. Stamp CAMVRI, O.-C. 397. Arezzo.
- 23.2.2 Berenice B216.2. Italy.

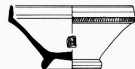
Other findspots

- 23.1 Asciburgium, Bologna, Bolsena, Conimbriga, Köln, Luni, Magdalensberg, Ordona, Pollentia, Roma.
- 23.2 Not separately listed.

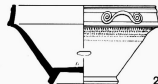
Concordance

- 23.1: Goudineau 20c; 25a; 37a. – Berenice B210.2.
 - 23.2: Goudineau 40. – Barocelli 11. – Berenice B216. – Hayes 23.
- Pieces described as Haltern 9 sometimes belong to this form.

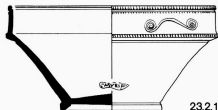
P. M. K.



23.1.1



23.1.2



23.2.1



23.2.2

Figure 1: Sample from *Conspectus* catalog.

What we have:

Production

Subform 23.1 is probably made in many parts of Italy; examples in Padana ware do not show applied decoration. Subform 23.2 is made in Italy but apparently not in the Padana region.

Distribution

Subform 23.2 is very common throughout the Mediterranean region, with sporadic examples found in the North and in North-Italy; Subform 23.1 is relatively uncommon.

What we have:

Production

Subform 23.1 is probably made in many parts of Italy; examples in Padana ware do not show applied decoration. Subform 23.2 is made in Italy but apparently not in the Padana region.

Distribution

Subform 23.2 is very common throughout the Mediterranean region, with sporadic examples found in the North and in North-Italy; Subform 23.1 is relatively uncommon.

What we want:

```
{  
  "form": "23.1",  
  "origin": "Italy",  
  "decoration": "none",  
  "occurs": "uncommon"  
},  
{  
  "form": "23.2",  
  "origin": "Italy, not Padana",  
  "occurs": "Mediterranean region;  
             North-Italy"  
}
```

What we have:

Production

Subform 23.1 is probably made in many parts of Italy; examples in Padana ware do not show applied decoration. Subform 23.2 is made in Italy but apparently not in the Padana region.

Distribution

Subform 23.2 is very common throughout the Mediterranean region, with sporadic examples found in the North and in North-Italy; Subform 23.1 is relatively uncommon.

UNSTRUCTURED DATA

What we want:

```
{  
  "form": "23.1",  
  "origin": "Italy",  
  "decoration": "none",  
  "occurs": "uncommon"  
},  
{  
  "form": "23.2",  
  "origin": "Italy, not Padana",  
  "occurs": "Mediterranean region;  
             North-Italy"  
}
```

STRUCTURED DATA

Problem: Running texts contain a lot of irrelevant information.
Information scientists call them **redundant**.



DEATH

The world's leading cause of death

TEXT MINING: THEORY

Text Mining is a **general term** covering several different ideas.

Text Mining is a **general term** covering several different ideas.

For instance:

- Information retrieval

Text Mining is a **general term** covering several different ideas.

For instance:

- Information retrieval
- Cluster analysis

Text Mining is a **general term** covering several different ideas.

For instance:

- Information retrieval
- Cluster analysis
- Statistical (lexical) analysis

Text Mining is a **general term** covering several different ideas.

For instance:

- Information retrieval
- Cluster analysis
- Statistical (lexical) analysis
- Predictive analytics

Text Mining is a **general term** covering several different ideas.

For instance:

- Information retrieval
- Cluster analysis
- Statistical (lexical) analysis
- Predictive analytics
- **Information extraction**

Text Mining is a **general term** covering several different ideas.

For instance:

- Information retrieval
- Cluster analysis
- Statistical (lexical) analysis
- Predictive analytics
- **Information extraction**
- ...

The most common approach is to look for certain **relation triples** in the text.

TEXT MINING: PRACTICAL

Subform 20.1 is never common, but has a long history from the Augustan period until the late Tiberian or early Claudian period.

results in

have history until: Subform 20.1, late tiberian

have history until: Subform, tiberian

have history until: Subform 20.1, tiberian

have history until: Subform, late tiberian

Mining Paper Catalogues

A Multilingual Solution to Reduce Verbose Fields to Consistent Terminology

Felix Kußmaul



University of Cologne

Dr Tim Evans

UNIVERSITY *of York*

1 September 2017

European Association of Archaeologists 2017