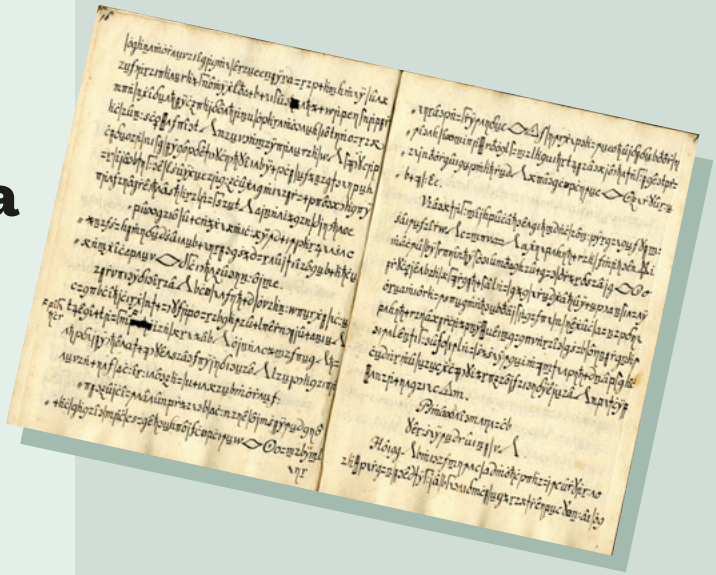
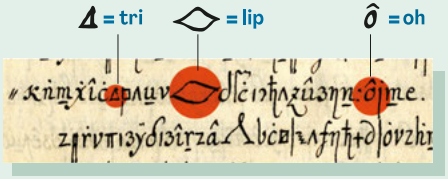
Så blir historiska chiffer läsbara

AV JOHAN JARNESTAD OCH LINA WENNERSTEN-GRADERT



1 DIGITALISERING OCH TRANSKRIBERING

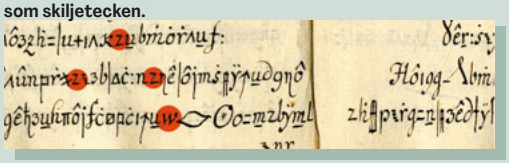
När ett chiffer har hittats är det första steget att digitalisera det. Bilder av sidorna läggs in i databasen Decode tillsammans med metadata, alltså information om chiffret. Sedan behöver bilden eller bilderna av manuskriptet omsättas till datorläsbar text. Det kan göras helt manuellt eller halvautomatiskt med hjälp av ett AI-verktyg för handskriftsläsning och symbolöversättning: Transcript tool. Här är tre exempel på transkribering av symboler i det så kallade Copiale-chiffret:



COPIALE-CHIFFRET

Bakgrunden till att de tvärvetenskapliga internationella forskningsprojekten Decode och Decrypt föddes var avkodningen av Copiale-chiffret. På första uppslaget i manuskriptet stod det: ”Philipp 1866”, vilket antogs vara en ägarmärkning, och på sista sidan stod det ”Copiales 3”.

Allt annat i dokumentet var krypterat och ursprungsspråket i den kodade texten förmodades vara tyska. Chiffret visade sig vara ett så kallat homofoniskt substitutionschiffer, alltså att vissa bokstäver kodades med flera olika symboler. Latinska bokstäver användes som skiljetecken.



De tre forskarna Kevin Knight, Beáta Megyesi och Christiane Schaefer lyckades knäcka koden 2011. Manuskriptet daterades till omkring 1730–1760 och härrör från en hemlig orden: okulisterna.

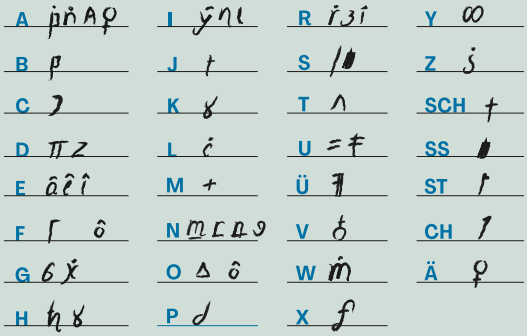
2 ANALYS

För att identifiera vilket språk som chiffret utgår från används språkmodeller baserade på historiska språkdata. Uppbyggnaden av den krypterade texten kan då jämföras med olika språks uppbyggnad. En svårighet är att stavning och vokabulär varierar mycket över tid, vilket AI-verktygens algoritmer måste ta hänsyn till. Dessutom kan chiffrens ursprungstexter växla mellan flera språk – några ord kan till exempel vara skrivna på latin och andra på franska, medan huvudspråket kan vara ett tredje. Forskarna använder verktyg för att göra frekvensanalys, alltså att räkna hur ofta olika tecken förekommer, för att kunna jämföra med hur ofta en viss bokstav förekommer i olika språk. I Copiale-chiffret var exempelvis ⋀ vanligast.

3 DEKRYPTERING

När forskarna ska försöka dekryptera ett historiskt chiffer är utgångspunkten att det är ett substitutionschiffer, eftersom det är vanligast, alltså att en bokstav eller ett ord har ersatts med vissa tecken, siffror eller symboler. Men substitutionschiffren kan vara uppbyggda på sinsemellan olika sätt. Forskarna har skapat verktyg för att gruppera chiffer som finns i databasen och som liknar varandra. Med hjälp av verktyget CrypTool 2 kan till exempel de flesta krypterade texter från Vatikanens arkiv avkodas.

CHIFFERNYCKEL



Forskning & Framsteg

Bild: Johan Jarnestad