

Gnomology Database of Arabic and Syriac  
*Bridging Ancient Texts and Modern Language Models*

*Workshop „Exploring Formulaic Knowledge through  
Languages, Cultures and Time“*  
Trier, 2011/02/11

Marco BÜCHLER

Natural Language Processing Group  
Department of Computer Science  
University of Leipzig

Ute PIETRUSCHKA

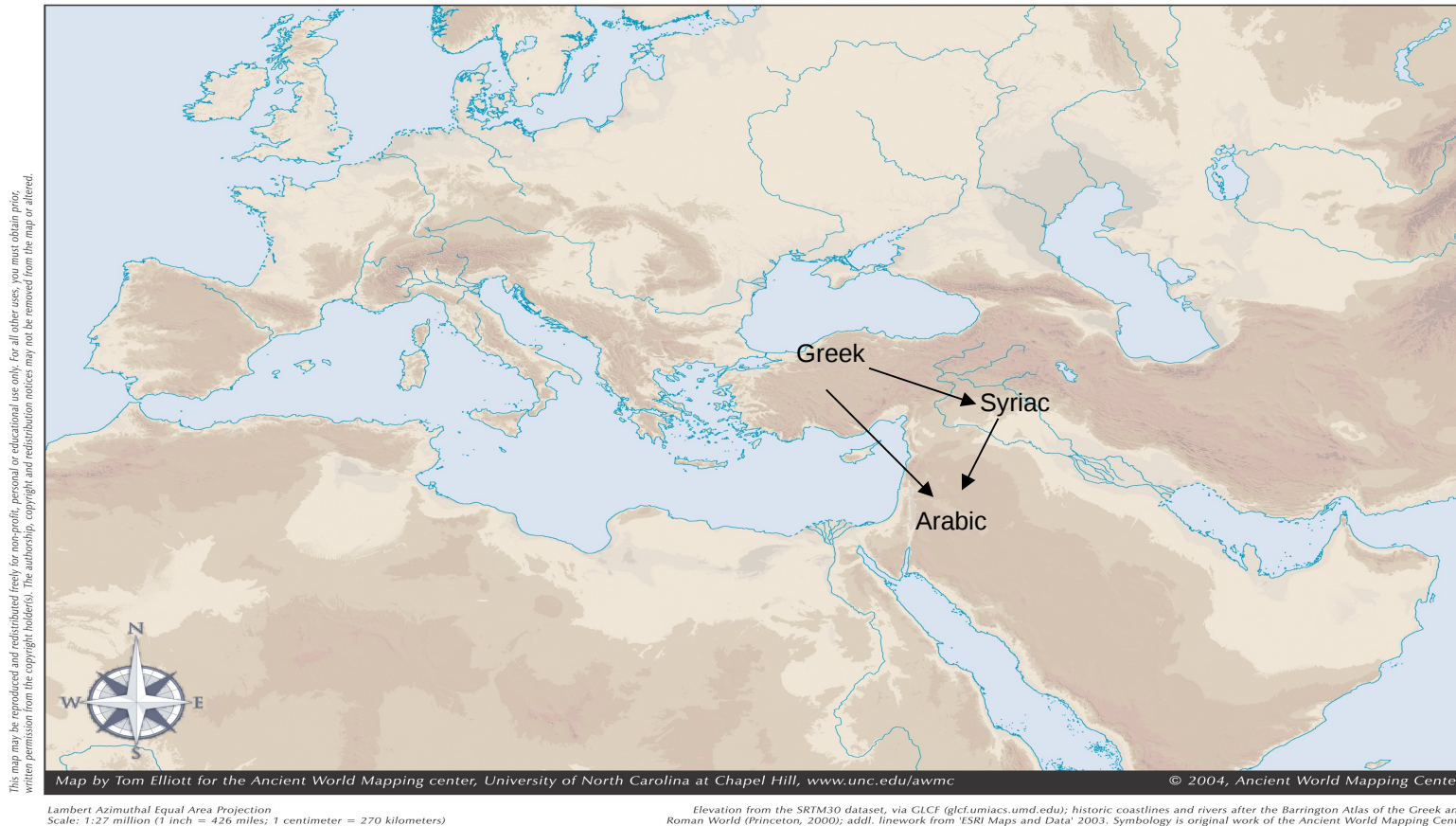
Corpus of Arabic and Syriac Gnomologia  
Oriental Institute  
Martin Luther University Halle-Wittenberg

---

- Introduction
  - Main goals
  - Relationship between Ancient Greek, Arabic, Syriac
  - 'Construction Kit'
  - By way of example
- User Interface
- Existent data
- Challenges / lessons learnt
- Current & further work

- Since 2009, supported by *Fritz Thyssen Stiftung*
- Concrete task:
  - Creation of gnomes database with focus on Arabic and Syriac
  - Gnomes and
    - General meta data
    - Parallel texts
    - Keywords regarding names, topics etc.
    - Translations
- Goal: 10,000 data entries (gnome + quotation, translation, keywords)
- End of 2011: open access for researchers of gnomologia

**EUROPE, NORTH AFRICA AND WEST ASIA: PHYSICAL GEOGRAPHY**



- detecting lines of transmission
- approach of the compiler (such as translation technique or situational context)
- sources (original works, other collections of sayings)

Do Syriac collections go back to Greek gnomologia or are they original collections?

Can we reconstruct the original collections from the existing fragments in Syriac?

Is there an archetype of Syriac gnomologia?

Many of the overlapping sayings in the Arabic collections are found in different versions

Reasons:

- different wording/ several traditions already in the Greek sources
- rewording of an original Arabic translation within the Arabic tradition
- going back to a Syriac translation of a Greek saying

Arabic collections:

Diogenes saw a woman who was carried away by the **flood**. He said: "The evil is destroyed by the evil."  
(Ḥunayn 9)

Diogenes saw a woman who was carried away by the **flood**. He said: "**She adds to the nuisance a nuisance**. The evil is destroyed by the evil."  
(Ibn Hindu 480)

Diogenes saw a woman who was carried away by the **flood**. He said: "**Let** the evil destroy the evil."  
(Ibn Hindu 509 = griech.Cod. V 8 (Giannantoni, Socratis et Socraticorum reliquiae), 206)

Diogenes saw a woman who was carried away by the **water**. He said: "**It fits the saying: let** the evil **wash away** the evil."  
(Muḥtaṣar Ṣiwān al-ḥikma 42 = Shahrastani 17)

Example for simple rewording of the Arabic tradition  
or another translation?

→ translation technique Greek-Syriac / Syriac-Arabic

**Names/ persons as „placeholders“**

**Diogenes** saw a **woman** who was carried away  
by the flood. He said: **“The evil is destroyed  
by the evil.”**

**Keyword: quintessence  
of the saying**

- Names, persons, content as an unstable element
- the gnome itself as a stable element



Insert

☒ Insert
 ☐ Search

Author

Source/Edition

Gnome-ID

Page

Line

Content

+

-

Hints

+

-

Person

+

-

Keywords

+

-

Original

Griechisch

▼

Reset

Translation

Englisch

▼

-

+

Search Window

Quotation

Arabisch

▼

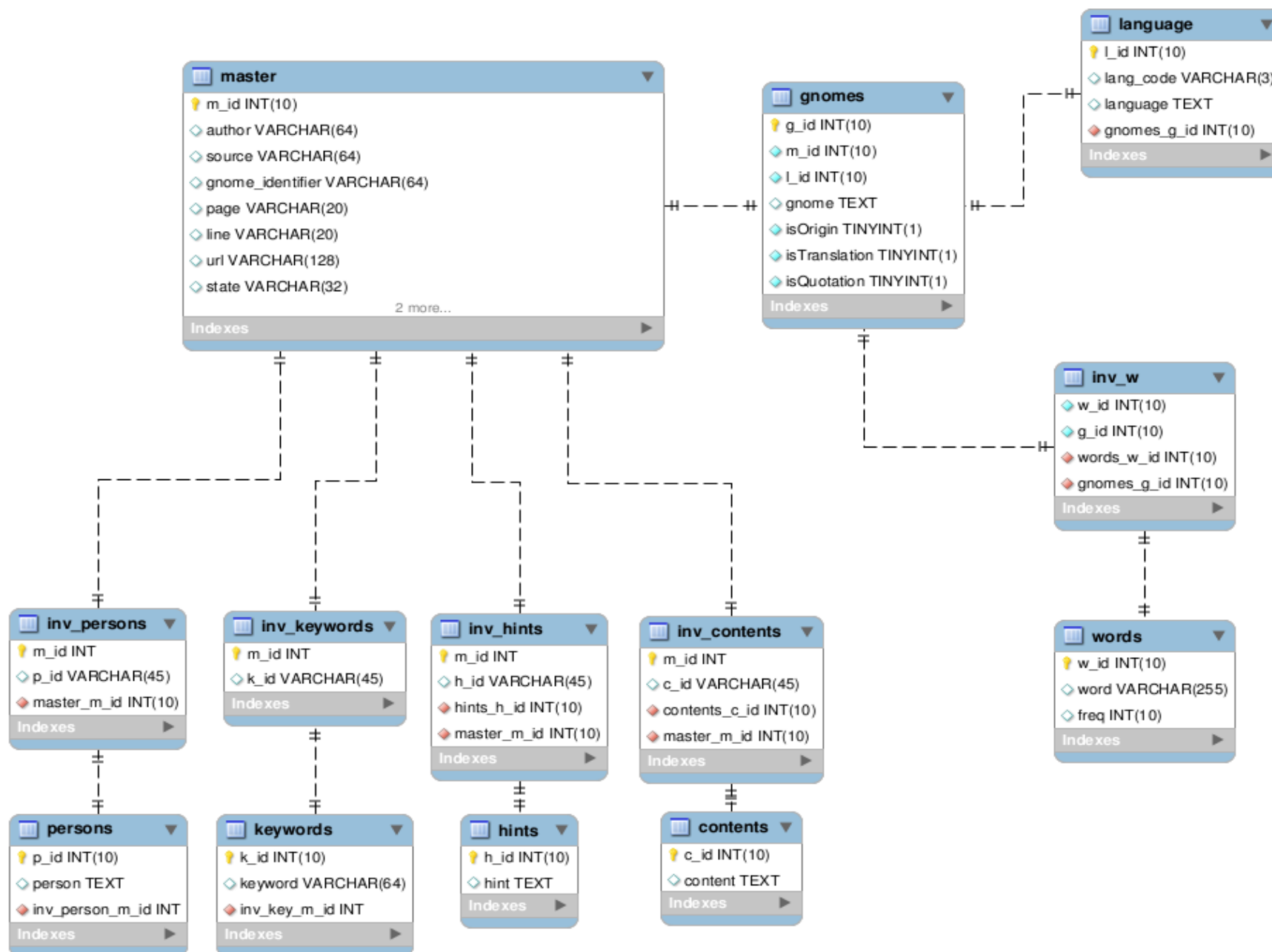
-

+

Insert

Link

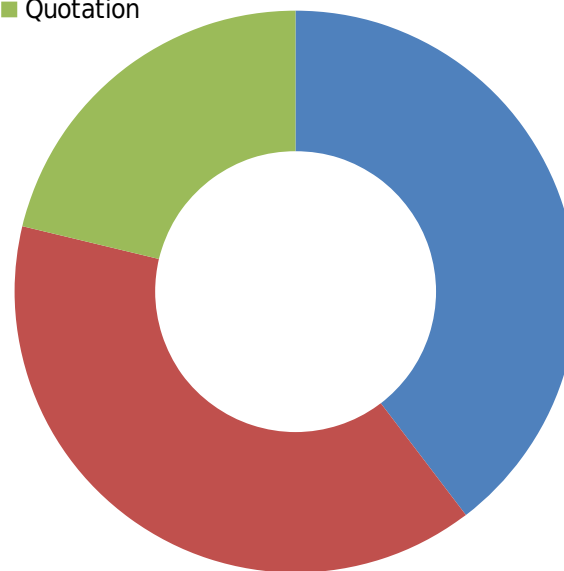
State



Gnome ID	Author	Page	Original	Translation	Content	Keywords
Source	Line					
38.	E 30	Aristoteles PQ	?? ??	وقال من اراد ان ينظر الى صورة نفسه مجردة فليجعل الحكمة مرآته	He said: Let him who wants to look at the form of his self laid bare make philosophy his mirror. (Gutas 1975)	Weisheit Gestalt Spiegel Philosophie Seele
39.	E 34	Aristoteles PQ	?? ??	وقال لاصحابه لكن عنايتكم برياضة النفس فائتاً برياضة البدن فلتعنوا بها لما يدعو اليه الاضطراب واهربوا عن اللذات فائتها تنزف النفوس الضعيفة ولا قوة بها على القوة	He said to his disciples: Concern yourselves with the training of the soul; as for the training of the body, concern yourselves with it when necessity calls for it. Flee pleasures, for they drain the weak souls but have no power over the strong ones. (Gutas 1975)	ausbilden Genuss Seele Körper
40.	E 41	Aristoteles PQ	?? ??	وقال عودوا النفس الآداب لأن منها وفيها تظهر عجائب الفكر ولطائف النظر	He said: Accustom the soul to [all] the parts of education because they are the source and the repository of the marvels of thought and the subtleties of speculation. (Gutas 1975)	Wunder denken Feinheit Spekulation Seele Bildung
41.	E 55	Aristoteles PQ	?? ??	وقال له احفظ عني ما اقول لك اذا كنت في مجلس الشراب فلنكن مذكرتك في الغزل فان النفس انسة بذلك واذا جلست الى خاصمتك فاذكر الحكمة فائهم لها افهم واذا خلوت لليوم فاذكر الحق فائتها تمنعك ان تضع نطفك فيما لا معنى له	He said to Alexander: remember that I have said what i am telling you now: When you are at a drinking party, let your conversation be about love, because the soul takes the greatest delight in this matter. When you are sitting together with your intimate friends, talk about philosophy, for they understand it best. And when you retire to sleep, keep your thoughts on chastity because it will prevent you from depositing your sperm where it will serve no purpose. (Gutas 1975)	Liebe Philosophie Weisheit Züchtigkeit Seele Konversation
42.	E 7	Aristoteles PQ	?? ??	وكان يقول الدواء طب الأبدان والحكمة طب الأنفس فان كان الناس يمدحون المشطبين على اصلاح الأبدان فانما احق بان امدح على اصلاح الأنفس	He used to say: Drugs cure bodies and philosophy cures souls; so, if people are in the habit of praising doctors for healing bodies, I am more deserving of praise for healing souls. (Gutas 1975)	Medizin Philosophy Arzt preisen Körper Seele
43.	E 74	Aristoteles PQ	?? ??	وقال النفس ليست في البدن بل البدن في النفس لانها اوسع منه وابسط	He said: The soul is not in the body but rather the body is in the soul because the soul is wider and more spacious than the body. (Gutas 1975)	Seele Körper
44.	16	Sextus Lagarde	3 ??	روحانية الروحانية. روحانية الروحانية.	Das, was der Seele ist, eigne dir an, weil es dauert.	dauern Seele aneignen
45.	81	Sextus Lagarde	6 ??	لأن الروحانية لا تتغير مع تغير الجسمانية.	Denn die Seele der Gläubigen hat in ihrer Reinheit keine Übersättigung von der Gottesfurcht.	Gläubige Gottesfurcht Übersättigung Seele Reinheit
46.	107	Sextus Lagarde	7 ??	والتواضع لله وحده.	Und nimm in Acht mehr als deine Seele das Wort über Gott.	Gott achten Seele Wort
47.	111	Sextus Lagarde	7 ??	ولأن الروحانية لا تتغير مع تغير الجسمانية.	Und nicht ist es möglich Gott zu fürchten ohne Mitleid mit den Leiden der in der Seele Betrüben.	Gott Mitleid fürchten Leid Seele

Original	1821
Translation	1796
Quotation	977

■ Original  
■ Translation  
■ Quotation



■ Arabic  
■ Syriac  
■ Greek  
■ German  
■ English  
■ French



Language	Entries
Arabic	490
Syriac	831
Greek	1480
German	785
English	945
French	63

## Arabic version:

When he was being taken out to be **executed**, he said to his wife whom he had seen crying, „What makes you cry?“ „How can I help crying,“ she replied, „seeing that you are going to be killed unjustly.“ „Why, said **Socrates**, „ would you rather wish that I were killed justly?“

(Philosophical Quartet, Socr. 23; Mubaššir Socrates 311= Bocados 198.1-5)

Arabic version follows the Greek tradition (Gnomologium Vaticanum 478)

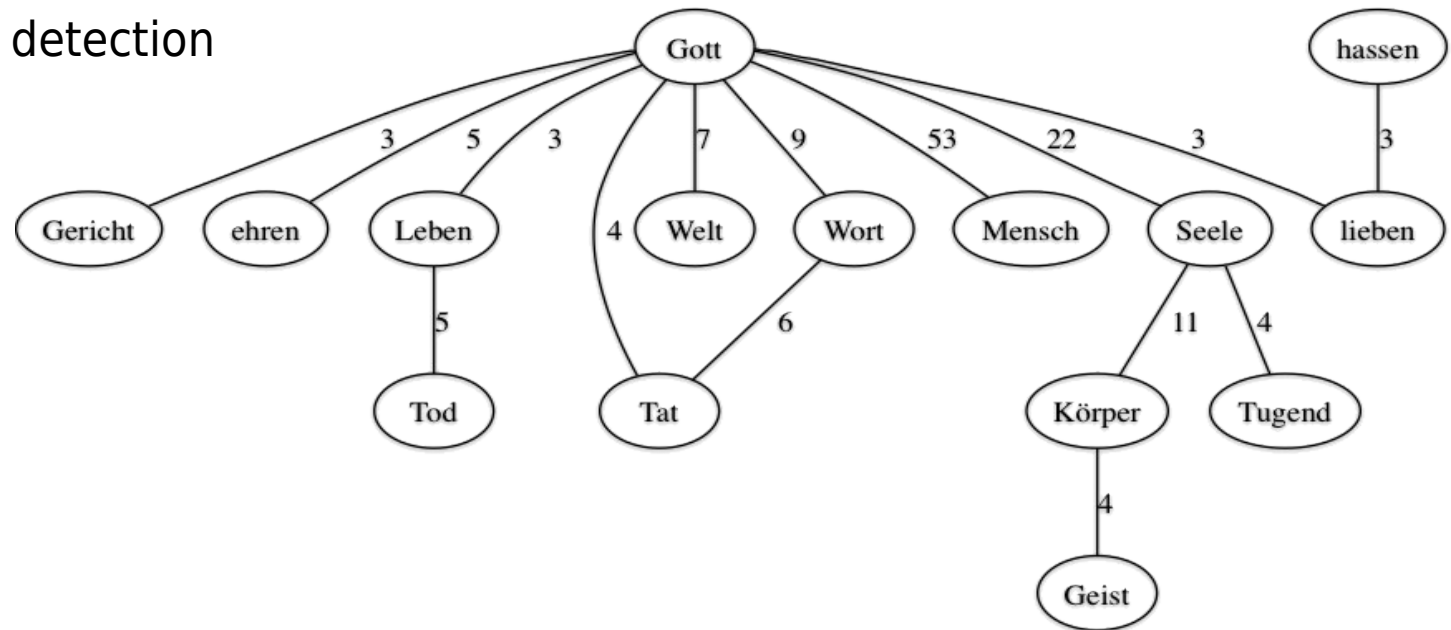
## Syriac version:

A certain woman saw **Socrates** as they were carrying him along to **crucify** him, and she wept and said: „Woe is me, for they are about to slay you without having committed any offence. And Socrates gave an answer to her, saying: „O foolish woman, would you have also commit some crime that I might be punished like a criminal?“ (Bar Hebraeus, Laughable Stories, 5)

Is the different wording in the Syriac tradition result of another tradition in Greek or a inner-Syriac development?

- Humanists:
  - Most of the text is not digital available. Much needs to be typed in manually.
  - Based on available data: identifying different translation schools by comparing chosen vocabulary
- Computer scientists:
  - Lack of huge amounts of text as basis for statistical processing
  - Avoiding inconsistencies of keywords (synonyms of keywords for the same topic, spell errors)
  - Avoiding redundancies of parallel texts and translations.
  - Current work on both topics by a diploma thesis

- Validation of database: as more the database increase as more probable inconsistencies can be observed. For this reason:
  - Keyword validation
  - Duplicate detection



- Availability of digital texts
- Automatic detection of parallel texts

- **Extraction of fragmentary authors**
- **String approaches:**
  - GST
  - Letter n-grams
- **Syntactic approaches:**
  - *Longest Common Consecutive Words*
  - Word n-grams
  - Distance based co-occurrences
- **Semantic approaches:**
  - Semantic clustering
  - *Semantic graph based approach(es)*
  - Contrastive semantics
  - *Radius retrieval*
- *More complex approaches:*
  - *DCT*



**- Level 1: Pre-processing**

- Capitalisation (e. g. all letters to lowercase)
- Normalisation (e. g. removing all diacritics)
- Lemmatisation (e. g. replace inflected words by baseform)
- Synonym replacements (e. g. replace a word by the most common (most frequent) synonym)
- String similarity (words that are similar written)

**- Level 2: Feature training**

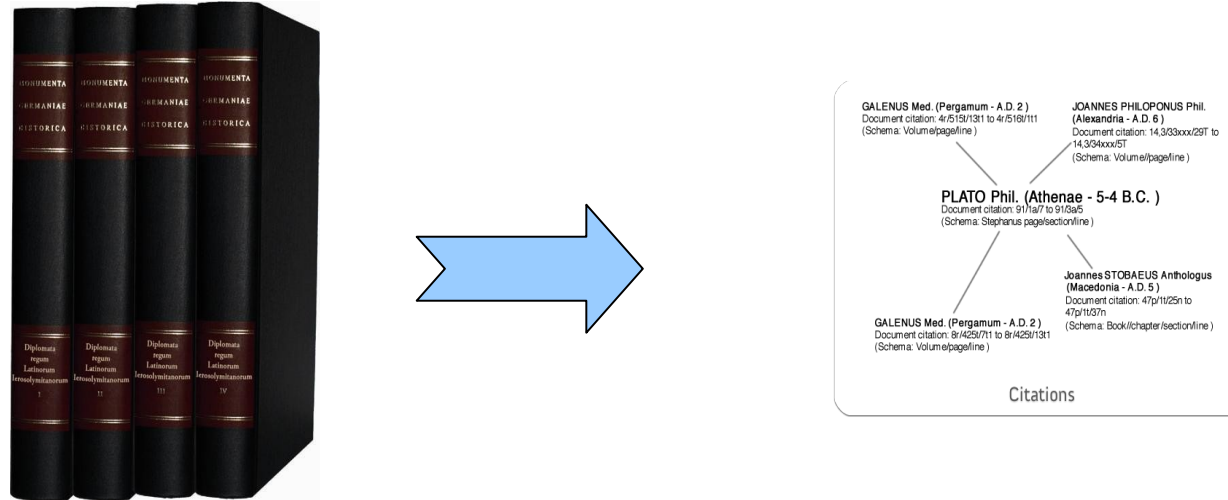
**- Level 3: Feature selection (Fingerprinting)**

- *Level 4: Linking*

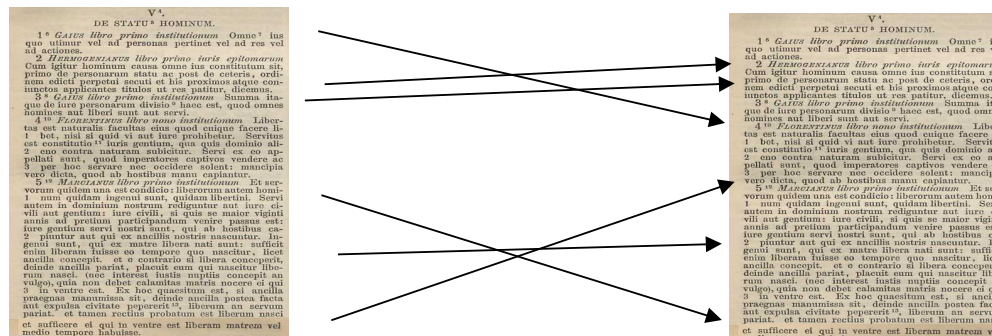
**- Level 5: Scoring**

- *Level 6: Post-processing*

## Intra corpus detection (Text reuse):



## Inter corpus detection (Modern: Plagiarism, Ancient: e.g. bible):





John Lee: A Computational Model of Text Reuse in Ancient Literary Texts, 2009.

# Marco Büchler

**A text re-use from a document with a high text re-use coverage is more trustworthy than from a less frequently re-used text.**

**A text re-use from a section of a document with a high text re-use temperature is more trustworthy than from a less frequently re-used part of a document.**

