



TESSERAE

AUTOMATING THE SEARCH FOR CROSS-LANGUAGE TEXT REUSE

James Gawley, Chris Forstall,
Konnor Clark

Digital Humanities 2014
Lausanne, Switzerland

Allusion Detection

tesserae.caset.buffalo.edu/cgi-bin/read_bin.p



TESSERA

LATIN GREEK ENGLISH OTHER TOOLS SOURCES

Sort by and format as

Show results at a time.

6125 results in 62 pages.

Go to page: [1](#) [2](#) [3](#) [4](#) [5](#) [\[next\]](#) [\[last\]](#)

| | target phrase | source phrase | matched on | score |
|----|--|--|------------------|-------|
| 1. | verg. aen. 4.652 accipite hanc animam, meque his exsolvite curis. | hom. il. 23.62 εὔτε τὸν ὕπνος ἔμαρπτε λύων μελεδήματα θυμοῦ | cura, exsoluo | 11 |
| 2. | verg. aen. 4.200 centum aras posuit , vigilemque sacraverat ignem, | hom. il. 10.46 Ἐκτορείοις ἄρα μάλλον ἐπὶ φρένα θῆχ' ἱερόισιν: | sacer, pono | 11 |
| 3. | verg. aen. 4.589 terque quaterque manu pectus percussa decorum, | hom. il. 13.282 ἐν δέ τε οἱ κραδίη μεγάλα στέρνοισι πατάσσει | percutio, pectus | 11 |
| 4. | verg. aen. 4.83 incubat, illum absens absentem auditque videtque; | hom. il. 18.53 εἶδετ' ἀκούουσαι ὄς ἐμῶ ἐνὶ κήδεα θυμῶ. | audio, uideo | 11 |
| 5. | verg. aen. 4.410 prospiceres arce ex summa , totumque videres | hom. il. 14.228 ἀκροτάτας κορυφάς: οὐδὲ χθόνα μάρπτε ποδοῖν: | superus, arx | 10 |

[hom. il. 23.62](#) εὔτε τὸν ὕπνος ἔμαρπτε **λύων μελεδήματα** θυμοῦ

[verg. aen. 4.652](#) accipite hanc animam, meque his **exsolvite curis.**

A Basis for Latin-Greek Matching

“Translate” the lookup indices for Greek texts



| | <i>Homer Iliad</i> | <i>Vergil Aeneid</i> | |
|------------------------------|--|--|-------|
| πατήρ • pater • parens | 1.98, 1.544 , 2.146, 3.50, 4.235 | 1.7, 1.60, 1.65 , 1.95, 1.345 | pater |
| άνήρ • vir • homo | 1.7, 1.544 , 2.554, 9.320, 12.57 | 1.65 , 1.229, 1.254 | homo |
| τε • atque • et | 1.5, 1.544 , 4.8, 17.677, 18.6, 24.729, | 1.40, 1.54, 1.58, 1.65 , 1.82 | atque |
| θεός • deus • divus | 1.8, 1.544 , 8.2, 16.443, 21.165, | 1.46, 1.65 , 1.79, 1.447, | divus |

Method 1: Parallel Texts

- Vulgate/Greek New Testament
- Aligned at verse level
- Probabilistic test for 1:1 correspondence between words

Method 1: Parallel Texts

<Matthew.19.7> λέγουσιν αὐτῷ Τί οὖν Μωυσῆς ἐνετείλατο δοῦναι **βιβλίον** ἀποστασίου καὶ ἀπολῦσαι ;

<Mark.10.4> οἱ δὲ εἶπαν Ἐπέτρεψεν Μωυσῆς **βιβλίον** ἀποστασίου γράψαι καὶ ἀπολῦσαι.

<Luke.4.17> καὶ ἐπέδοθη αὐτῷ **βιβλίον** τοῦ προφήτου Ἡσαίου, καὶ ἀνοίξας τὸ βιβλίον εὗρε [τὸν] τόπον οὗ ἦν γεγραμμένον

<Luke.4.20> καὶ πτύξας τὸ **βιβλίον** ἀποδοὺς τῷ ὑπηρέτῃ ἐκάθισεν: καὶ πάντων οἱ ὀφθαλμοὶ ἐν συναγωγῇ ἦσαν ἀτενίζοντες αὐτῷ.

<John.21.25> Ἔστιν δὲ καὶ ἄλλα πολλὰ ἃ ἐποίησεν ὁ Ἰησοῦς, ἅτινα ἐὰν γράφηται καθ' ἕνα, οὐδ' αὐτὸν οἶμαι τὸν κόσμον χωρήσειν τὰ γραφόμενα **βιβλία**.

<Matthew.19.7> dicunt illi quid ergo Moses mandavit dari **libellum** repudii et dimittere

<Mark.10.4> qui dixerunt Moses permisit **libellum** repudii scribere et dimittere

<Luke.4.17> et traditus est illi liber prophetae isaiae et ut revolvit **librum** invenit locum ubi scriptum erat

<Luke.4.20> et cum plicuisset **librum** reddidit ministro et sedit et omnium in synagoga oculi erant intendentes in eum

<John.21.25> sunt autem et alia multa quae fecit Iesus quae si scribantur per singula nec ipsum arbitror mundum capere eos qui scribendi sunt **libros** amen

Method 2: Pivot Language*

- Collect English definitions for Greek and Latin lemmata
- Calculate similarity of definitions
 - based on TF-IDF scores for English words
 - using Python package *Gensim*
- Assume lemmata with similar definitions are semantically related

*With thanks to Jeff Rydberg-Cox

Method 2: Pivot Language

μένω

stay, wait; stand fast; **stay** at home, **stay** where one is; lodge, **stay; stay** away, be absent; to be a shirker; **stay, tarry**; loiter, be idle; to be **lasting, remain, stand**; is fixed; having no proper motion; are stable, **permanent; remain; remain** as one was; **stand** ...

maneo

perf.; a.; Neutr., to **stay, remain**; Impers. pass.; To **stay, tarry**, stop, continue, abide, pass the night; to **remain**, last, endure, continue; dat.; Part. act. fut.; that which will abide; endure; **lasting, permanent**; part. pres.; Act., to **wait** for, await, expect; to await ...

sto

to **stand**; to **stand** still; **remain** standing; **stand** upright.; Pass. impers.: Ps.; Sim.; Gn.; Pa.; to **stand** firm; immovable; to last; **remain**; continue; To **remain; tarry**; linger; remaining in the city; To **stand** in the ranks; under arms; to fight; fin.; to **stand** firm; **stand** ...

Comparison of Methods

Parallel Texts

- 7,000 stems
- Less Classical vocabulary
- Good for names
- Good for common words
- Easily expanded to other languages

Pivot Language

- 27,000 stems
- More Classical vocabulary
- Not as good with names
- Privileges words with concise definitions
- Easily expanded to intra-language synonymy

Results

Latin translation candidates

| Parallel Texts | Pivot |
|----------------|-------|
| 20% | 76% |

Successful translations

| Parallel Texts | Pivot |
|----------------|-------|
| 35% | 67% |

Benchmark:

- G. N. Knauer (1965) *Der Aeneis und Homer*

| | Parallel Texts | Pivot |
|---------------------|----------------|-------|
| All allusions (282) | 14% | 21% |
| Bigrams only (264) | 18% | 27% |
| Bigrams, combined | 34% | |

Allusion Detection

Two storms, one real, one metaphorical:

Incubuere **mari**, totumque a sedibus imis
una **Eurusque Notusque** ruunt creberque procellis
Africus, et **vastos** volvunt ad litora **fluctus**. (Aen. 1.85)

They press upon the sea, and rake it all up from the deepest places, both Eurus and Notus together, and Africus, thick with storms, and they roll huge waves toward the shores

κινήθη δ' ἀγορὴ φη̃ **κύματα μακρὰ θαλάσσης**
πόντου Ἰκαρίοιο, τὰ μὲν τ' **Εὐρὸς τε Νότος τε**
ᾠρορ' ἐπαίξας πατρὸς Διὸς ἐκ νεφελᾶων. (Iliad 2.145)

And the assembly was set in motion, like the long waves of the sea, the Icarian sea, which both Eurus and Notus arouse, rushing out of the clouds of Father Zeus.

Results II: Allusion Detection

Parallel Texts Method

Incubuere **mari**, totumque a **sedibus** imis
una **Eurusque Notusque** ruunt creberque **procellis**
Africus, et **vastos** volvunt ad litora **fluctus**. (Aen. 1.85)

They press upon the sea, and rake it all up from the deepest places, both Eurus and Notus together, and Africus, thick with storms, and they roll huge waves toward the shores

κινήθη δ' ἀγορὴ φη' **κύματα μακρὰ θαλάσσης**
πόντου Ἰκαρίοιο, τὰ μὲν τ' **Εὐρὸς τε Νότος τε**
ᾠρορ' ἐπαίξας πατρὸς Διὸς ἐκ **νεφελάων**. (Iliad 2.145)

And the assembly was set in motion, like the long waves of the sea, the Icarian sea, which both Eurus and Notus arouse, rushing out of the clouds of Father Zeus.

Results II: Allusion Detection

Pivot Method

Incubuere **mari**, totumque a sedibus imis
una **Eurusque Notusque** ruunt creberque procellis
Africus, et **vastos** volvunt ad litora **fluctus**. (Aen. 1.85)

They press upon the sea, and rake it all up from the deepest places, both Eurus and Notus together, and Africus, thick with storms, and they roll huge waves toward the shores

κινήθη δ' ἀγορὴ φῆ **κύματα μακρὰ θαλάσσης**
πόντου Ἰκαρίοιο, τὰ μὲν τ' **Εὐρὸς τε Νότος τε**
ᾠρορ' ἐπαΐξας πατρὸς Διὸς ἐκ νεφελῶν. (Iliad 2.145)

And the assembly was set in motion, like the long waves of the sea, the Icarian sea, which both Eurus and Notus arouse, rushing out of the clouds of Father Zeus.

Future Work

- More sophisticated combination of dictionaries
- Compare against independent machine standard:
 - Bizzoni *et al.* (2014):
 - pivot method based on Princeton Wordnet
 - small, hand-checked benchmark
- Combine with other dictionaries
- Expand to new languages



TESSERAE

THANK YOU!

<http://tesseract.caset.buffalo.edu>

<https://github.com/tesseract>

cforstall@gmail.com

james.gawley@gmail.com

