SUNOIKISIS DC 2015

SESSION 15: TEXT ALIGNMENT IN ACTION

Neven Jovanović / neven.jovanovic@ffzg.hr

Petra Šoštarić
Mario Benedik
Dunja Ševerdija

University of Zagreb, Faculty of Humanities and Social Sciences
Department of Classical Philology

ALIGN

Pronunciation: /əˈlʌɪn/. Verb.

1.1 Put (things) into correct or appropriate relative positions: *the fan blades are carefully aligned*

figurative *aligning domestic prices with prices in world markets*

All this notwithstanding, the proposed Department is a positive move to align national form and function. How do you move them from one point to the other, without going through that point that characterizes that mind - correcting it and aligning it, to conform to the overall Classical understanding?

Origin

Late 17th century: from French *aligner*, from *à ligne* 'into line'.

THE PLAN

Introduction: what is alignment, and why should we care?
How do we do it?

Demonstration: what happens when we align **Greek** to **Latin**?
And when we align Latin (translation) to **English** (translation)?
And when we align Latin to **French**?
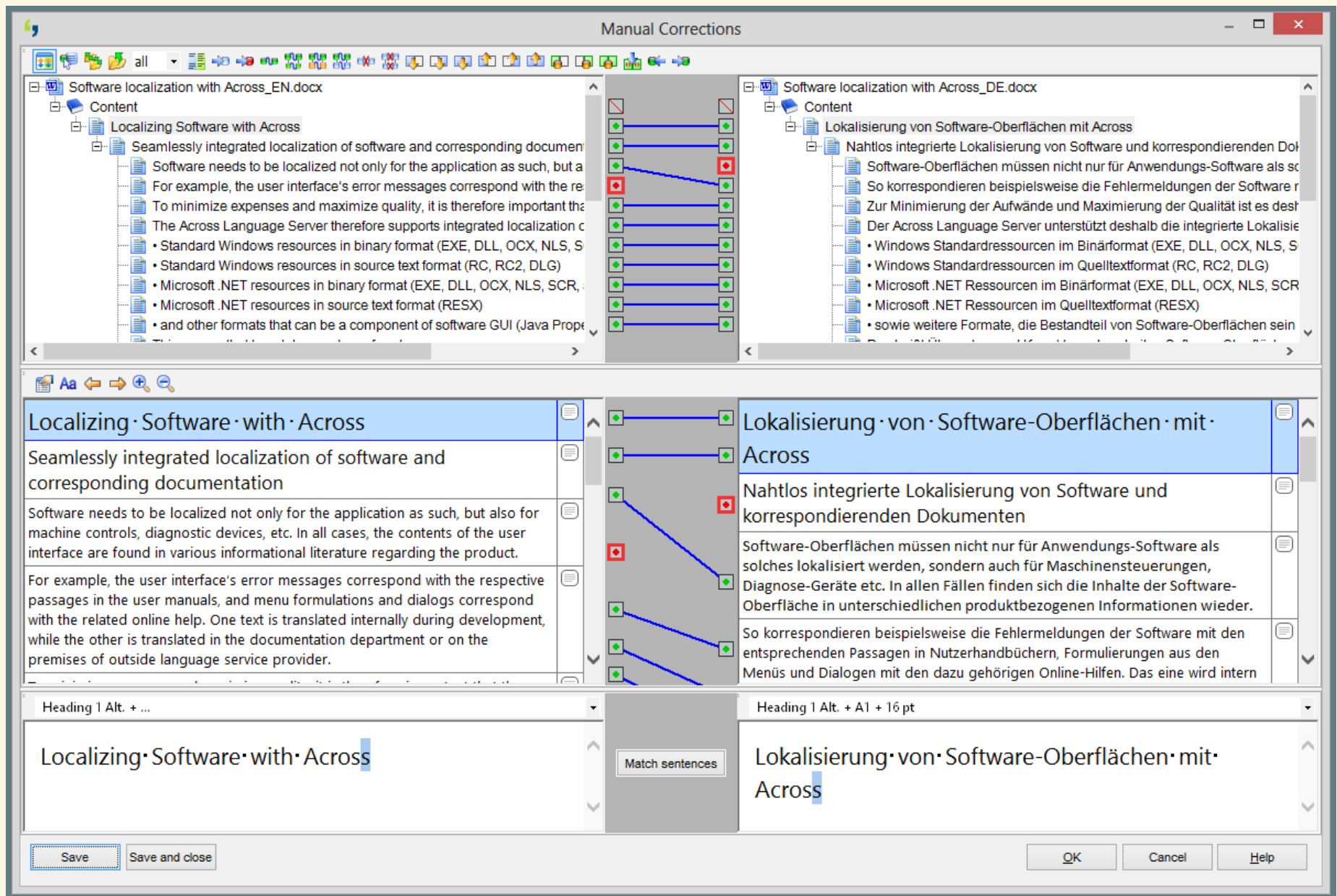And when we align **two Latin translations**?

Conclusion: discuss!

And you try it out too!

WHAT IS ALIGNMENT?

A STRATEGY IN AUTOMATED TRANSLATION

reusable, machine-supported, used mostly for technical translation

Manual Corrections

all ▾

Software localization with Across_EN.docx
- Content
  - Localizing Software with Across
    - Seamlessly integrated localization of software and corresponding document
      - Software needs to be localized not only for the application as such, but a
      - For example, the user interface's error messages correspond with the re
      - To minimize expenses and maximize quality, it is therefore important tha
      - The Across Language Server therefore supports integrated localization o
      - • Standard Windows resources in binary format (EXE, DLL, OCX, NLS, S
      - • Standard Windows resources in source text format (RC, RC2, DLG)
      - • Microsoft .NET resources in binary format (EXE, DLL, OCX, SCR, ...
      - • Microsoft .NET resources in source text format (RESX)
      - • and other formats that can be a component of software GUI (Java Prope

Software localization with Across_DE.docx
- Content
  - Lokalisierung von Software-Oberflächen mit Across
    - Nahtlos integrierte Lokalisierung von Software und korrespondierenden Dok
      - Software-Oberflächen müssen nicht nur für Anwendungs-Software als so
      - So korrespondieren beispielsweise die Fehlermeldungen der Software m
      - Zur Minimierung der Aufwände und Maximierung der Qualität ist es desh
      - Der Across Language Server unterstützt deshalb die integrierte Lokalisie
      - • Windows Standardressourcen im Binärformat (EXE, DLL, OCX, NLS, S
      - • Windows Standardressourcen im Quelltextformat (RC, RC2, DLG)
      - • Microsoft .NET Ressourcen im Binärformat (EXE, DLL, OCX, NLS, SCR
      - • Microsoft .NET Ressourcen im Quelltextformat (RESX)
      - • sowie weitere Formate, die Bestandteil von Software-Oberflächen sein

---

Localizing · Software · with · Across

Seamlessly integrated localization of software and corresponding documentation

Software needs to be localized not only for the application as such, but also for machine controls, diagnostic devices, etc. In all cases, the contents of the user interface are found in various informational literature regarding the product.

For example, the user interface's error messages correspond with the respective passages in the user manuals, and menu formulations and dialogs correspond with the related online help. One text is translated internally during development, while the other is translated in the documentation department or on the premises of outside language service provider.

Lokalisierung · von · Software-Oberflächen · mit · Across

Nahtlos integrierte Lokalisierung von Software und korrespondierenden Dokumenten

Software-Oberflächen müssen nicht nur für Anwendungs-Software als solches lokalisiert werden, sondern auch für Maschinensteuerungen, Diagnose-Geräte etc. In allen Fällen finden sich die Inhalte der Software-Oberfläche in unterschiedlichen produktbezogenen Informationen wieder.

So korrespondieren beispielsweise die Fehlermeldungen der Software mit den entsprechenden Passagen in Nutzerhandbüchern, Formulierungen aus den Menüs und Dialogen mit den dazu gehörigen Online-Hilfen. Das eine wird intern

---

Heading 1 Alt. + ... ▾

Heading 1 Alt. + A1 + 16 pt ▾

Localizing · Software · with · Across

[Match sentences]

Lokalisierung · von · Software-Oberflächen · mit · Across

Save | Save and close

OK | Cancel | Help

"MOST COMMERCIAL TRANSLATION MEMORY APPLICATIONS NOW INCLUDE AN ALIGNER AS PART OF A SUITE OF TOOLS."

"ALIGNMENT IS A WAY OF MAKING USE OF PREVIOUS TRANSLATIONS."

"TRANSLATION MEMORY IS CREATED FROM EXISTING TRANSLATED MATERIALS."

"CONVERT LEGACY TRANSLATIONS INTO MONEY SAVING TRANSLATION MEMORIES!"

PERHAPS MORE FAMILIAR TO PHILOLOGISTS:

A parallel text is a text placed alongside its translation or translations.

Parallel text alignment is the identification of the corresponding sentences in both halves of the parallel text.

LOEB CLASSICAL LIBRARY

Search

Advanced Search

BROWSE   ABOUT THE LIBRARY   USING THE LIBRARY

GELLIUS, *Attic Nights*   LCL 212: 78-79

Go to page: GO

Find in a Library   View cloth edition

ATTIC NIGHTS OF AULUS GELLIUS

tertium vitae annum cum periculo et clade aliqua venire aut
2 corporis morbique gravioris aut vitae interitus aut animi
aegritudinis. Propterea, qui rerum verborumque istiusmodi studio
tenentur eum aetatis annum appellant κλιμακτηρικόν.
3   Nocte quoque ista proxima superiore, cum librum *Epistularum*
divi Augusti, quas ad Gaium nepotem suum scripsit, legeremus
duceremurque elegantia orationis neque morosa neque anxia, sed
facili hercle et simplici, id ipsum in quadam epistula super eodem
anno scriptum offendimus; eiusque epistula exemplum hoc est:

    "IX. Kal. Octobris.

"Ave, mi Gai, meus asellus iucundissimus, quem semper medius
fidius desidero, cum a me abes. Set praecipue diebus talibus, qualis
est hodiernus, oculi mei requirunt meum Gaium, quem, ubicumque
hoc die fuisti, spero laetum et bene valentem celebrasse quartum et
sexagesimum natalem meum. Nam, ut vides, κλιμακτῆρα
communem seniorum omnium tertium et sexagesimum annum
evasimus. Deos autem oro ut mihi quantumcumque superest

BOOK XV. VII.

is attended with danger, and with some disaster involving either
serious bodily illness, or loss of life, or mental suffering. Therefore
those who are engaged in the study of matters and terms of that
kind call that period of life the "climacteric."[1]

    Night before last, too, when I was reading a volume of letters of
the deified Augustus, written to his grandson Gaius, and was led on
by the elegance of the style, which was easy and simple, by Heaven!
without mannerisms or effort, in one of the letters I ran upon a
reference to this very belief about that same year. I give a copy of
the letter:[2]

    "The ninth day before the Kalends of October.[3]

    "Greeting, my dear Gaius, my dearest little donkey,[4] whom, so
help me! I constantly miss whenever you are away from me. But
especially on such days as to-day my eyes are eager for my Gaius,
and wherever you have been to-day, I hope you have celebrated my
sixty-fourth birthday in health and happiness. For, as you see, I have
passed the climacteric common to all old men, the sixty-third year.
And I pray the gods that whatever time is left to me I may pass with

you safe and well, with my country in a flourishing condition, while
you are playing the man and preparing to succeed to my position."

Conrad Celtis

Ad Wilhelmum Mommerlochum civem Coloniensem et philosophum

Wilhelme Agrippae moenibus editus,
per prisca nobis stemmata cognitus,
    Romana quae quondam tuae urbis
      gloria contulerat colonis

in urbe tecum hac condidici vagas
inferre fraudes per συλλογιστικοὺς
    nexus, quod et contentioso
      tradiderat dialexis ore.

primaeque tecum hac prendideram sacros
libros sophiae, tunc mihi cognitum
    Albertus et quid Thomas alti
      in physicis docuere rebus.

nemo hic Latinam grammaticam docet,
nec expolitis rhetoribus studet,
    mathesis ignota est, figuris
      quidque sacris numeris recludit.

nemo hic per axem candida sidera
inquirit, aut quae cardinibus vagis
    moventur, aut quid doctus alta
      contineat Ptolemaeus arte.

ridentur illic docta poemata,
Maronianos et Ciceronios
    libros verentur tanquam Apella
      carne timet stomacho suilla.

hos inter unus tu nitidus sedes
mathesin alto pectore comprobans
    et quod vetustas diligenti
      posteritatis amore scripsit.

Conrad Celtis

An Wilhelm Mommerloch aus Köln, einen Philologen

Freund Wilhelm, der du den Mauern von Köln entstammst,
bekannt durch altehrwürdigen Adel mir,
    den einst den Bürgern deiner Heimat
      römische Glorie verliehen hatte:

Hier hab mit dir zusammen ich einst gelernt,
wie trüglich man verknüpft syllogistische
    Sophisterein und das Gezänke
      von dialektischen Künsteleien.

Hier las zuerst mit dir ich die heiligen
Weltweisheitsbücher: hier auch ward mir bekannt,   10
    was Albert und der große Thomas
      über des Weltalls Natur dozierten.

Latein jedoch – Grammatik – lehrt niemand hier,
es kümmert niemand sich um gepflegten Stil;
    was über Zahlen und Figuren
      Mathematik lehrt, ist unbekannt hier.

Am Himmel forscht hier keiner den Sternen nach,
noch den Planeten, wie sie exzentrisch sich
    bewegen, noch was Ptolemäus
      uns in erhabener Weisheit lehrte.   20

Was heut Gelehrte dichten, wird hier verhöhnt,
und für die Werke Ciceros und Vergils
    hat man den Abscheu, den ein Jude
      zeigt, den's vorm Fleische des Schweines ekelt.

Allein strahlst du hervor unter solchem Volk,
der hochgemut die Mathematik du pflegst
    und das verehrst, was uns die Alten,
      liebend der Nachwelt gedenkend, schrieben.

Harry C. Schnur (ed.). Lateinische Gedichte deutscher Humanisten. Lateinisch und Deutsch, Stuttgart 1967, 44f.

WHY SHOULD WE CARE?

The task of aligning makes us

CONNECT ALL EQUIVALENTS.

❧

This leads to

DEEPER UNDERSTANDING AND CLOSER EXPERIENCE

of both the source text and the target text as its equivalent.

EXEMPLO DISCIMUS!

More pragmatically,
turn to our pedagogical advantage the fact that a lot of Greek
and Latin is today read in translation,
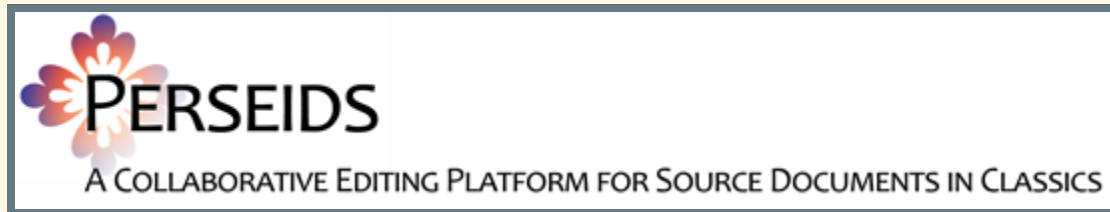or with translation at hand.

Also,

CONTRIBUTE TO BUILDING

A DIGITAL COLLECTION OF

PARALLEL TEXTS

(and thus help the machines help us read Greek and Latin)

HOW DO WE DO IT?

USE PERSEIDS!

PERSEIDS

A COLLABORATIVE EDITING PLATFORM FOR SOURCE DOCUMENTS IN CLASSICS

perseids.org

"New Text Alignment"

Powered by Son of Suda Online

Boards   Communities

Select a publication to work with:

pilots ⌄   click to select... ⌄   ⌄   ⌄   Emend

Create Edition

New Treebank Annotation   New Text Alignment   Import Annotations

-New

Publication	Passage Text	Transcription Text (EpiDoc)	Translation Text (EpiDoc)	Text Annotations	Commentary Annotation	Treebank Annotation	Alignment Annotation
Greek Treebank Collection/201657/3	-	-	-	-	-	grafousi - parsing problems	-

DEMONSTRATIONS

GREEK AND LATIN

Homer, the Iliad in Greek — and translated into Latin by Andreas Divus Iustinopolitanus, Venice 1537

Iustinopolis / Capodistria / Koper / Kopar / www.wikidata.org/wiki/Q1015

"Lie quiet Divus. I mean, that is Andreas Divus, In officina Wecheli, 1538, out of Homer."
Ezra Pound, *Canto I* (1917)

ILIAS HOMERI;
ANDREA DIVO IVSTINOPO.
LITANO INTERPRETE,
ad verbum translata.

ILIADIS. A. HOMERI
COMPOSITIONIS.

Alpha, preces Chryſæ, peſtem exercitus, odium Regũ.

RAM cane Dea Pelidæ Achillis
Pernitioſam: quæ infinitos Achiuis
   dolores inflixit
Multas át fortes animas iſeris miſię
Heroũ, ipſos át laniãta fecit canibꝰ
Auibusꝗ; oĩbus. Iouis aũt pficiebat

Ex quo ſanè primũ diuiſi ſunt contédentes   (volũtas
Atridesꝗ; Rex uirorum, ⁊ diuus Achilles.
Quis nam ipſos Deorum contétioni commiſit pugnare?
Latonæ, ⁊ Iouis filius.hic enim regi iratus

Morbũ p exercitũ ſuſcitauit malũ:peribant aũt populi,
Quoniam Chryſem inhonorauit ſacerdotem
Atrides.hic enim uenit celeres ad naues Achiuorum.
Liberaturusꝗ; filiam, ferensꝗ; infinita dona;
Coronas habens in manibus longeiaculantis Apollinis
Aureo cum ſceptro:⁊ obſecrabat omnes Achiuos.
Atridas autem maxime duos,principes populorum.
Atridæꝗ; ⁊ alij beneocreati Achiui,
Vobis quidem Dij dent cœleſtes domos habentes
Expugnare Priami ciuitatem:bene uero domũ reuerti
Filiam autem mihi ſoluite dilectam: hæc dona accipite:
Venerantes Iouis filium longeiaculantem Apollinem.
Tunc alij quidem omnes laudauerunt Achiui
Reuerédũꝗ; eſſe ſacerdotẽ et ſplédida accipiéda dona.

LATIN AND FRENCH

Hom. Il. 3
(Perseus, urn:cts:greekLit:tlg0012.tlg001.perseus-grc1:3)
Latin: Andreas Divus, Venice 1537.
French: d'après Eugène LASSERRE, Homère, Iliade. Paris,
Classiques Garnier, 1955. (via Hodoi elektronikai)

LATIN AND ENGLISH

Hom. Il. 3
(Perseus, urn:cts:greekLit:tlg0012.tlg001.perseus-grc1:3)
Latin: Andreas Divus, Venice 1537.
English: Alexander Pope, The Iliad (1715).

LATIN AND LATIN

Hom. Il. 3 (Perseus, urn:cts:greekLit:tlg0012.tlg001:3)
Latin: Laurentius Valla, 1444.
Latin: Andreas Divus, Venice 1537.

DISCUSSION

MODELLING TRANSLATION

WITH RDF

```
http://croala.ffzg.unizg.hr/ghlh/translation-source/10344
    croala:wordform "λέγεις"@grc ;
    saws:isDirectTranslationOf http://croala.ffzg.unizg.hr/ghlh/translati
    dcterms:isPartOf "10331" .

http://croala.ffzg.unizg.hr/ghlh/translation-source/10356
    croala:wordform "πότερον"@grc ;
    saws:isLooseTranslationOf http://croala.ffzg.unizg.hr/ghlh/translatio
    dcterms:isPartOf "10331" .
```

QUERYING RDF OF ALIGNED TEXTS

WITH **SPARQL**

Query: show all Croatian translation equivalents of καὶ, counting occurrences of each equivalent.

```
arq --data ghlh-rdf/ghlh-rdf.ttl --query kai4.rq

# contents of file kai4.rq:

SELECT ?hrvr (COUNT(?hrv) AS ?brhrvr)
WHERE
{
?id ?p "καὶ"@grc .
?id ?pp ?hrv .
?hrv c:wordform ?hrvr .
}
GROUP BY ?hrvr
ORDER BY ?brhvr
```

A SPARQL query in file kai4.rq (using arq).

```
------------------------
| hrvr          | brhrvr |
========================
| "i"           | 59     |
| "a"           | 6      |
| "te"          | 4      |
| ","           | 2      |
| "."           | 1      |
| "A"           | 1      |
| "I"           | 1      |
| "No"          | 1      |
| "djeda"       | 1      |
| "drukčije"    | 1      |
| "između"      | 1      |
| "kako"        | 1      |
| "naslonjaču"  | 1      |
| "oca"         | 1      |
```

Result of SPARQL query from file kai4.rq.

HOMEWORK

Choose a sentence (preferably in Greek or Latin) and its translation (existing one, or the one you make yourself). Using Perseids ("New Text Alignment"), align words from the source with equivalents in the target. Submit to the *Zagreb Homer 2015* board for review and credit points.