



A tool for syntax-based intra-language text alignment

Tariq Yousef, Chiara Palladino
University of Leipzig
Berlin Digital Classicist Seminars,
November 29, 2016

What is text alignment?

- Text alignment is the comparison of two or more parallel texts
- It tries to define correspondences/similarities and divergences/variants
- One of the most important tasks in Natural Language Processing: it can be performed automatically through algorithmic and dynamic programming methods

W1	This morning	the cat	observed	little	birds in the	trees	.	
W2		The cat	was observing	birds in the	little	trees	this morning, it observed birds for two hours	.

Intra-Language alignment: alignment of texts in the same language

iAligner [Home](#) [Instructions](#) [Bibliography](#) [Contact](#)

Alignment options: Ignore Nonalphabetical Case sensitive Ignore Diacritics Levenshtein Distance

- 1 The book of the generation of Jesus Christ, the son of David, the son of Abraham.
- 1 The book of the generation of Jesus Christ, the son of David, the son of Abraham.

1 The book of the generation of Jesus Christ , the son of David , the son of Abraham .
1 The book of the generation of Jesus Christ , the son of David , the son of Abraham .

Length: **21** Aligned-complete **21**

1 The book of the generation of Jesus Christ , the son of David , the son of Abraham .

The longest common substring

- 2 Abraham begat Isaac; and Isaac begat Jacob; and Jacob begat Judah and his brethren;
- 2 Abraham begat Isaac; and Isaac begat Jacob; and Jacob begat Judas and his brethren;

2 Abraham begat Isaac ; and Isaac begat Jacob ; and Jacob begat Judah and his brethren ;
2 Abraham begat Isaac ; and Isaac begat Jacob ; and Jacob begat Judas and his brethren ;

Length: **19** Aligned-complete **18** notAligned **1**

2 Abraham begat Isaac ; and Isaac begat Jacob ; and Jacob begat Judah and his brethren ;
2 Abraham begat Isaac ; and Isaac begat Jacob ; and Jacob begat Judas and his brethren ;

The longest common substring

2 Abraham begat Isaac ; and Isaac begat Jacob ; and Jacob begat
2 Abraham begat Isaac ; and Isaac begat Jacob ; and Jacob begat

- 3 and Judah begat Perez and Zerah of Tamar ; and Perez begat Hezron ; and Hezron begat Ram ;
- 3 And Judas begat Phares and Zara of Thamar; and Phares begat Esrom; and Esrom begat Aram;

3 and Judah begat Perez and Zerah of Tamar ; and Perez begat Hezron ; and Hezron begat Ram ;
3 And Judas begat Phares and Zara of Thamar ; and Phares begat Esrom ; and Esrom begat Aram ;

Length: **21** Aligned-complete **12** Aligned-case **1** notAligned **8**

3 and Judah begat Perez and Zerah of Tamar ; and Perez begat Hezron ; and Hezron begat Ram ;

The longest common substring

3 and
3 And

- 4 and Ram begat Amminadab; and Amminadab begat Nahshon; and Nahshon begat Salmon;
- 4 And Aram begat Aminadab; and Aminadab begat Naasson; and Naasson begat Salmon;

4 and Ram begat Amminadab ; and Amminadab begat Nahshon ; and Nahshon begat Salmon ;
4 And Aram begat Aminadab ; and Aminadab begat Naasson ; and Naasson begat Salmon ;

Length: **17** Aligned-complete **11** Aligned-case **1** notAligned **5**

4 and Ram begat Amminadab ; and Amminadab begat Nahshon ; and Nahshon begat Salmon ;
4 And Aram begat Aminadab ; and Aminadab begat Naasson ; and Naasson begat Salmon ;

The longest common substring

4 and
4 And

Cross-language alignment: alignment of texts in different languages

- Cross-language alignment is difficult to perform automatically
- It still needs training data from manual alignment

ALPHEIOS

Export XML

Show interlinear text

الا يا ايها الساقى ادر كاسا و ناولها ### كه عشق آسان نمود اول ولى افتاد مشكلها *** به بوى نافه‌اى كاخر صبا زان طره بگشايد ### ز تاب جعد مشكيش چه خون افتاد در دلها *** مرا در منزل جانان چه امن عيش چون هر دم ### جرس فرياد مى‌دارد كه برينديد محلها *** به مى سجاده رنگين كن كرت پير مغان گويد ### كه سالك بى‌خبر نبود ز راه و رسم منزلها *** شب تاريك و بيم موج و گردابى چنين هابل ### كجا دانند حال ما سبكاران ساحلها *** همه كارم ز خود كامى به بدنامى كشيد آخر ### نهان كى ماند آن رازى كز او سازند محفلا *** حضورى گر همى‌خواهى از او غايب مشو حافظ ### متى ما تلق من تهوى دع الدنيا و اهملها ***

Ho Saki , pass around and offer the bowl of love for God ### For the burden of love for God at first on the day of covenant appeared easy , but now difficulties have occurred . *** By reason of the perfume hope of the musk - pod , that , at the end of night the breeze displayeth from that knotted fore - lock , ### From the twist of its musky dark , fragrant curl , what blood of grief befell the hearts of the lovers of God *** With wine , becolour the prayer - mat if the Pir of the magians the perfect murshid bid thee ### For of the way and usage of the stages to God not without knowledge is the holy traveller the perfect murshid *** In the stage this world of the true Beloved , mine what ease and pleasure , when momentarily , ### The loud bell of the call of death giveth voice , saying Bind ye up the chattels of existence *** The dark night of the world and the fear of the wave of grief and the whirlpool so fearful the time of death ### The light - burdened ones of the shore ancestors who have passed the flood of death how know they our state *** By following my own fancy in hastening to union with God me only to ill fame all my work brought ### Secret , how remaineth that great mystery of love whereof great assemblies

A Persian poem manually aligned with an English translation, from the project Open Persian (<http://www.dh.uni-leipzig.de/wo/open-philology-project/open-persian/>)

...So, there is also manual alignment

- The Perseids Project and Alpheios Texts provide tools for manual alignment of texts in different languages
(<http://www.perseids.org/>,
<http://alpheios.net/>)

The screenshot shows the Alpheios web interface. At the top, there is a navigation bar with the logo and the name "ALPHEIOS". Below the navigation bar, there are several buttons: "1 <<" (previous), "1" (current), ">> 1" (next), "Go to sentence number" (input field), "Sentence list", "Save sentence", "< Undo", "Redo >", "Add Comment", "Export XML", and "Export Display". There is also a checkbox for "Show interlinear text".

The main content area is split into two columns. The left column contains the Greek text from Homer's *Iliad* (Book XXI), and the right column contains the English translation by A.T. Murray. The text is aligned line-by-line, with red vertical lines indicating the alignment points. The Greek text is in a serif font, and the English text is in a sans-serif font. The alignment is manual, as indicated by the red lines.

Αλλ' ὅτε δὴ πόρον ἴξον εὐροεῖος ποταμοῖο Ξάνθου
δινήεντος , ὄν ἀθάνατος τέκετο Ζεὺς , ἔνθα
διατμήξας τοὺς μὲν πεδίον δὲ δῖωκε πρὸς πόλιν , ἧ
περ Ἀχαιοὶ ἀτυζόμενοι φοβέοντο ἡματι τῷ προτέρῳ ,
ὅτε μαίνεται φαίδιμος Ἴκτωρ · τῆ ᾗ οἱ γέ
προχέοντο πεφυζότες , ἠέρα δ' Ἥρη πάντα πρόσθε
βαθειαν ἐρυσμένῃ · ἡμίσεες δὲ ἐς ποταμὸν εἰλεῦντο
βαθύρροον ἀργυροδίην , ἐν δ' ἔπεσον μεγάλῳ
πατάγῳ , βράχε δ' αἰπὰ ῥέεθρα , ὄχθαι δ' ἀμφί
περὶ μεγάλ' ἴαχον · οἱ δ' ἀλαλητῷ ἔννεον ἔνθα καὶ
ἔνθα ἔλισσόμενοι περὶ δίνας . ὡς δ' ὄθ' ὑπὸ ῥιπῆς
πυρὸς ἀκριδὲς ἠερέθονται φευγέμενα ποταμὸν δέ ·
τὸ δὲ φλέγει ἀκάματον πῦρ ὄρμενον ἐξαιφνης , ταὶ
δὲ πώσσοσι καθ' ὕδωρ · ὡς ὑπ' Ἀχιλλῆος Ξάνθου
βαθυδινήεντος πλήτο ῥόος κελάδων ἐπιμίξ ἴππων τε
καὶ ἀνδρῶν . Αὐτὰρ ὁ διογενὴς δόρυ μὲν λίπεν
αὐτοῦ ἐπ' ὄχθῃ κεκλιμένον μυρίζησιν , δ δ' ἔσθορε
δαίμονι ἴσος φάσανγον οἶον ἔχων , κακὰ δὲ φρεσὶ
μήδετο ἔργα , τύπτε δ' ἐπιστροφάδην · τῶν δὲ
στόνος ὄρνυτ' αἰκίης ἄορι θεινομένων , ἐρυθαίνετο
δ' αἵματι ὕδωρ . ὡς δ' ὑπὸ δελφίνος μεγακίητος
ἰχθύες ἄλλοι φεύγοντες πιμπλάσι μυχοὺς λιμένους

But when they were now come to the ford of the
fair - flowing river , even eddying Xanthus that
immortal Zeus begat , there Achilles cleft them asunder
, and the one part he drave to the plain toward the
city , even where the Achaeans were fleeing in rout
the day before , what time glorious Hector was
raging—thitherward poured forth some in rout , and Hera
spread before them a thick mist to hinder them ; but
the half of them were pent into the deep - flowing
river with its silver eddies . Therein they flung
themselves with a great din , and the sheer - falling
streams resounded , and the banks round about rang
loudly ; and with noise of shouting swam they this
way and that , whirled about in the eddies . And as
when beneath the onrush of fire locusts take wing to
flee unto a river , and the unwearied fire burneth them
with its sudden oncoming , and they shrink down into
the water ; even so before Achilles was the sounding
stream of deep - eddying Xanthus filled confusedly with
chariots and with men . But the Zeus - begotten left
there his spear upon the bank , leaning against the
tamarisk bushes , and himself leapt in like a god with

Homer, *Iliad* XXI, aligned with the English translation by A.T. Murray

Pairwise alignment: alignment of two texts

We distinguish on the number of text because it determines differences in the use of the alignment algorithm

The screenshot displays the Versioning Machine (VM) interface for the poem "Faith is a fine invention". The interface includes a logo with "5.0" and a navigation bar with "7 Total Versions", "Line numbers", and "Bibliographic panel". Two side-by-side text windows are open, showing the following text:

Version a660: A 660, verse embedded in letter to	Version h201: H 201, fascicle version of poem.
1 "Faith" is a fine invention	1 Faith is a fine invention
2 When Gentlemen can see -	2 For Gentlemen who see -
3 But <u>Microscopes</u> are prudent	3 But <u>Microscopes</u> are prudent
4 In an Emergency .	4 In an Emergency !

Two versions of Emily Dickinson's *Faith is a fine invention*, aligned using the Versioning Machine (<http://v-machine.org/samples/faith.html>)

Multiple alignment: alignment of multiple texts (i.e. more than two)

The number of multiple texts is virtually unlimited: in an ideal world, you can align as many texts as you want (but you should be careful and avoid “alignment monsters”)

"Faith is a fine invention"

7 Total Versions Line numbers Bibliographic panel

Version h72: H 72, fascicle version of poem. X

- 1 "Faith" is a fine invention
- 2 For Gentlemen who see !
- 3 But **Microscopes** are prudent
- 4 In an Emergency !

Version a680: A 680, verse embedded in letter to X

- 1 "Faith" is a fine invention
- 2 When Gentlemen can see -
- 3 But **Microscopes** are prudent
- 4 In an Emergency .

Version h201: H 201, fascicle version of poem. X

- 1 Faith is a fine invention
- 2 For Gentlemen who see -
- 3 But **Microscopes** are prudent
- 4 In an Emergency !

Version p1891: Published as poem XXX in the second X

- 1 FAITH is a fine invention
- 2 For gentlemen who see ;
- 3 But **microscopes** are prudent
- 4 In an emergency !

Version f1894: Letter to Samuel Bowles published in X

- 1 Faith is a fine invention
- 2 For gentlemen who see ,
- 3 But **microscopes** are prudent
- 4 In an emergency .

Version cp32: Published as poem LVI in Martha X

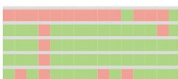
- 1 FAITH is a fine invention
- 2 For gentlemen who see ;
- 3 But **microscopes** are prudent
- 4 In an emergency !

Six versions of the same poem by Emily Dickinson

Four texts aligned with iAligner

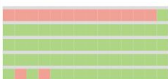
- εἶτα Δαμάστης ὁ Κιτιεύς τὰ πλείστα ἐκ τῶν Ἑκαταίου μεταγράψας περίπλου ἐγραψεν.
- εἶτα Δαμάστης ὁ Σιγυειεύς τὰ πλείστα ἐκ τῶν Ἑκαταίου μεταγράψας περίπλου ἐγραψεν·
- εἶτα Δαμάστης ὁ Κητιεύς τὰ πλείστα ἐκ τῶν Ἑκαταίου μεταγράψας περίπλου ἐγραψεν·
- εἶτα δαμάστης ὁ κητιεύς τὰ πλείστα ἐκ τῶν ἑκαταίου μεταγράψας, περίπλου ἐγραψεν·

εἶτα	Δαμάστης ὁ	Κιτιεύς	τὰ πλείστα ἐκ τῶν	Ἑκαταίου	μεταγράψας	περίπλου	ἐγραψεν .
εἶτα	Δαμάστης ὁ	Σιγυειεύς	τὰ πλείστα ἐκ τῶν	Ἑκαταίου	μεταγράψας	περίπλου	ἐγραψεν ·
εἶτα	Δαμάστης ὁ	Κητιεύς	τὰ πλείστα ἐκ τῶν	Ἑκαταίου	μεταγράψας	περίπλου	ἐγραψεν ·
εἶτα	δαμάστης ὁ	κητιεύς	τὰ πλείστα ἐκ τῶν	ἑκαταίου	μεταγράψας ,	περίπλου	ἐγραψεν ·



- ἐξῆς Δημόκριτος καὶ Εὐδοξος καὶ ἄλλοι τινὲς γῆς περιόδους καὶ περίπλους ἐπραγματεύσαντο.
- ἐξῆς Δημόκριτος καὶ Εὐδοξος καὶ ἄλλοι τινὲς γῆς περιόδους καὶ περίπλους ἐπραγματεύσαντο.
- ἐξῆς Δημόκριτος καὶ Εὐδοξος καὶ ἄλλοι τινὲς γῆς περιόδους καὶ περίπλους ἐπραγματεύσαντο.
- ἐξῆς δῆμόκριτος καὶ εὐδοξος καὶ ἄλλοι τινὲς γῆς περιόδους καὶ περίπλους ἐπραγματεύσαντο.

ἐξῆς	Δημόκριτος	καὶ	Εὐδοξος	καὶ	ἄλλοι τινὲς γῆς	περιόδους καὶ	περίπλους	ἐπραγματεύσαντο .
ἐξῆς	Δημόκριτος	καὶ	Εὐδοξος	καὶ	ἄλλοι τινὲς γῆς	περιόδους καὶ	περίπλους	ἐπραγματεύσαντο .
ἐξῆς	Δημόκριτος	καὶ	Εὐδοξος	καὶ	ἄλλοι τινὲς γῆς	περιόδους καὶ	περίπλους	ἐπραγματεύσαντο .
ἐξῆς	δῆμόκριτος	καὶ	εὐδοξος	καὶ	ἄλλοι τινὲς γῆς	περιόδους καὶ	περίπλους	ἐπραγματεύσαντο .



Alignment can be
visualized in
different ways

As a table

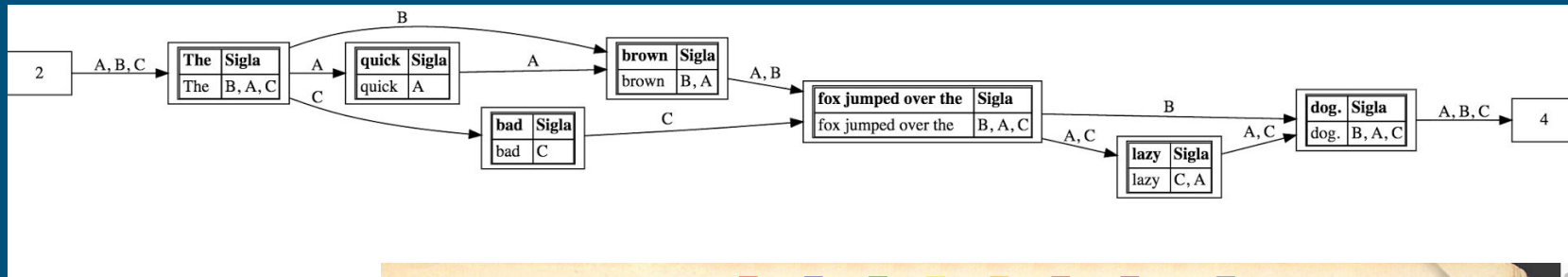
A	The	quick	brown fox jumps over the	-	dog.
B	The	-	brown fox jumps over the	lazy	dog.

Alignment Table

W1	The black			cat
W2	The black	and	white	cat
W3	The black	and	green	cat
W4	The black		very special	cat
W5	The black	not	very special	cat

A	B	C
The	The	The
quick	-	bad
brown	brown	-
fox	fox	fox
jumped	jumped	jumped
over	over	over
the	the	the
lazy	-	lazy
dog	dog	dog
.	.	.

As a graph



Alignment graph using Collatex
(<http://collatex.net/>)

Main Branch: ASV BBE DV KJV WR WEB YLT Majority

Genesis 1:1 Genesis 1:5

1 in the beginning of gods preparing the heavens and the earth

2 and the earth was without form and void and darkness

was upon the face of the deep and the spirit of god

Alignment graph using TRAViz
(<http://www.traviz.viccovery.org/>)

As matching segments in aligned sentences

Alignment of three
sample texts on
CATView

(<http://catview.uzi.uni-halle.de/overview.html>)

Overview Usage Demo (t) Demo (r) Imprint Disclaimer

W1
W2
W3

0 5 10 15 20 25 30 35 40 45 50

In egestas lacus vitae enim hendrerit scelerisque.

In egestas lacus vitae enim hendrerit scelerisque. Suspendisse non ornare nisi.

In egestas lacus vitae enim hendrerit scelerisque. Suspendisse non ornare nisi.

Ut id vulputate odio, vel tempor erat. Sed sed ornare justo.

missing

Ut id vulputate odio, vel tempor erat.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

Ut dignissim odio nulla, ac cursus diam tempus non. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

Ut dignissim odio nulla, ac cursus diam tempus non. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas.

As a dynamic visualization

(<http://www.digitalvariants.org/variants/valerio-magrelli>)

GENETIC MACHINE

I TESTI

I sei componenti :
riposo; Il corpo è
apparizione nel ma
raccolta di poesia di

La recensione ha ri
pubblicazioni su riv
Inoltre, Magrelli si è
prime in ordine di a
composito e stratific

Il materiale autogra
l'avantesto del libr
l'area d'indagine pa
dopo una o due ste
aggiustamento, app

I motivi che hanno
eccezionalmente coi

LA "GENETIC M

All'apertura la Mach
manoscritto (A), op
comparazione incro
trascrizioni della pri
legenda in basso a d

Sono inoltre disponi
in pannelli scorrev

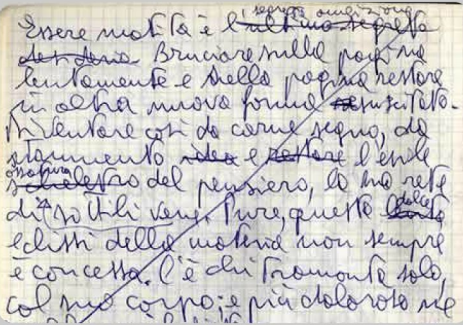
Genetic Machine - Ora serrata retinae

ESSERE MATITA

Valerio Magrelli, Notebook VI (cm 15,5x20,5), p. 19. Text handwritten with blue pen; "SP" added with violet felt-tip pen at the end of paragraph, before the following poem [Fermare la realtà in parole]

book-cover fade transcription zoom effect

A B C



Il corpo è eclissi come
è anzi un porro
Ma' ~~mi~~ dare al q'ito del
di se: q'oce nullo, nullo
00

A B C D E

A
Essere matita è segreta ambizione.
Bruciare sulla pagina lentamente
e nella pagina restare
in altra nuova forma suscitato.
Diventare così da carne segno, e da strumento
esile ossatura del pensiero
la sua rete di vene sottili.

B
Però questa dolce eclissi della materia
non sempre è concessa. C'è
chi tramonta solo col suo corpo;
più doloroso ne è allora il distacco.

Valerio Magrelli, *Primi esperimenti*,
in "Periodo ipotetico", gennaio 1977, p. 89.

Transcription of the autograph

Essere matita è [[ultimo segreto desiderio]
<segreta ambizione> //
Bruciare sulla pagina //
lentamente e sulla pagina restare //

Why do we align
texts?

To highlight
correspondences
in different
versions of a text

The screenshot shows the V-Machine 5.0 interface for the text "Faith is a fine invention". The interface includes a logo with "VM 5.0", a title bar, and three tabs for different versions of the text. Each tab displays the text with yellow highlights indicating correspondences between versions.

7 Total Versions ▾ Line numbers Bibliographic panel

Version h72: H 72, fascicle version of poem. ✕	Version a660: A 660, verse embedded in letter to ✕	Version h201: H 201, fascicle version of poem. ✕
1 "Faith" is a fine invention	1 "Faith" is a fine invention	1 Faith is a fine invention
2 For Gentlemen who see !	2 When Gentlemen can see -	2 For Gentlemen who see -
3 But Microscopes are prudent	3 But Microscopes are prudent	3 But Microscopes are prudent
4 In an Emergency !	4 In an Emergency .	4 In an Emergency !

To highlight
divergences
across various
versions of the
same text

Shared By: Kristin H Jensen
Description: Thomas Jefferson's rough draft of the Declaration of Independence (from the Papers of...)

Need Help? Click here

Declaration of Independence

POWERED BY VM4.0 Declaration of Independence has 2 versions

New Version | Bibliographic Info

Bibliographic Information

Declaration of Independence

Witness List

- Witness wit-47: Declaration - final - Project Gutenberg
- Witness wit-46: Declaration - rough draft - Princeton

Witness 1: wit-47

IN CONGRESS, July 4, 1776

The unanimous Declaration of the thirteen UNITED STATES OF AMERICA, .

When in the course of human events it becomes necessary for one people to dissolve the political bands which have **connected them with another, and**, & to assume among the powers of the earth the separate and equal & station to which the laws of nature and & of nature's god entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the separation .

We hold these truths to be self-evident ; that all men are created equal & , that they are endowed by their Creator with certain unalienable rights that , among these are life, & liberty and & the pursuit of happiness; that to secure these rights , governments are instituted among men, deriving their just powers from the consent of the governed; that whenever any form of government becomes destructive of these ends, it is the right of the people to alter or to abolish it and & to institute new government, laying its foundation on such

Witness 2: wit-46

A Declaration of the Representatives of the UNITED STATES OF AMERICA, in General Congress assembled .

When in the course of human events it becomes necessary for a people to advance from that subordination in which they have **hitherto remained**, & to assume among the powers of the earth the equal & independant station to which the laws of nature & of nature's god entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the change .

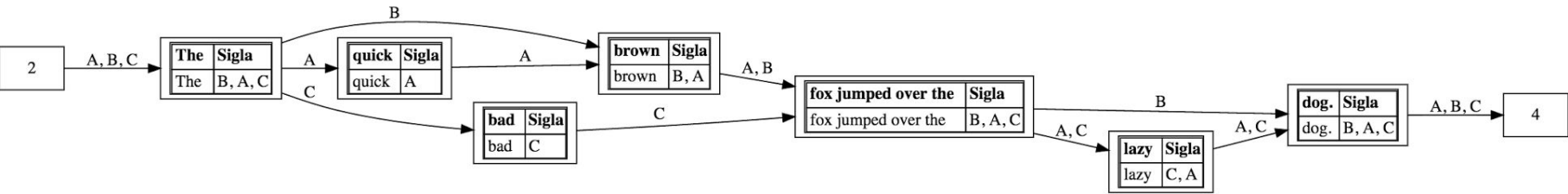
We hold these truths to be sacred & undeniable ; that all men are created equal & independant , that from that equal creation they derive rights inherent & inalienable , among which are the preservation of life, & liberty , & the pursuit of happiness; that to secure these ends , governments are instituted among men, deriving their just powers from the consent of the governed; that whenever any form of government shall become destructive of these ends, it is the right of the people to alter or to abolish it , & to institute new government, laying it's foundation on such principles & organising it's powers in such form, as to them shall seem most likely to effect their safety & happiness. prudence

Like Share 0 G+ Convidivi 0 Tweet

Search: Current Document

(<http://juxtacommons.org/>)

A	The	quick	brown fox jumps over the	-	dog.
B	The	-	brown fox jumps over the	lazy	dog.



To establish relations between witnesses of a text and see where they overlap and diverge

Comparing texts as philological practice

Collatio

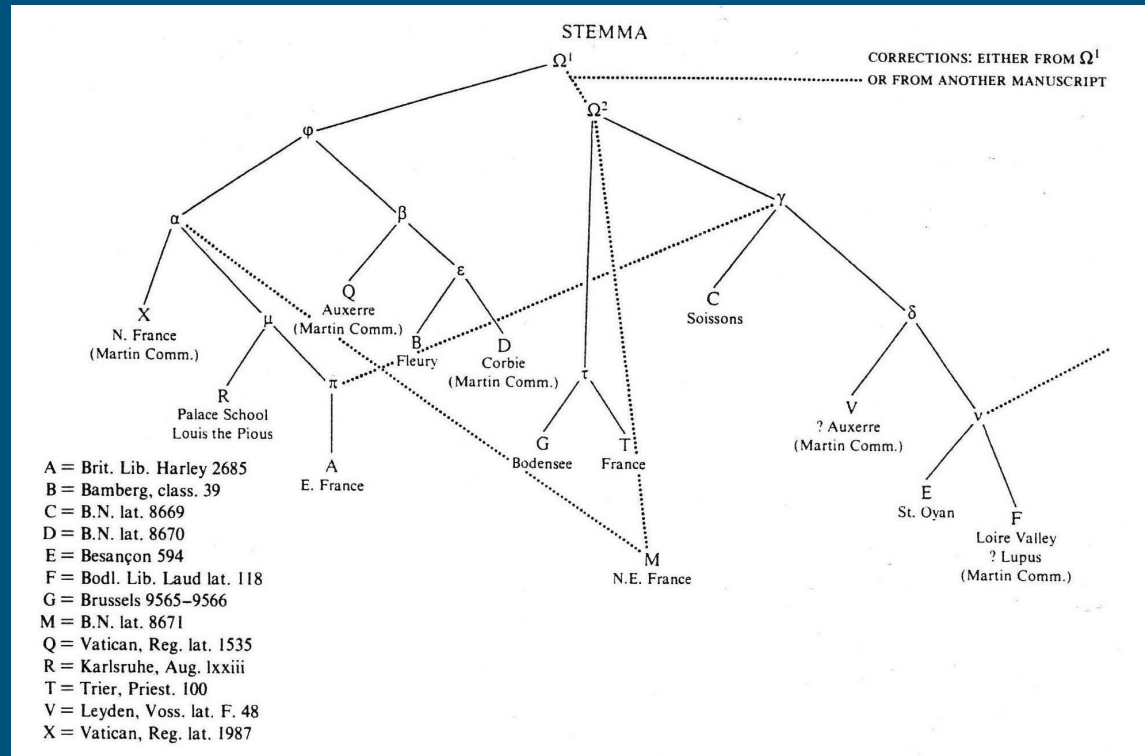
- Detection and transcription of variants in witnesses
- It is made by close reading each witness and comparing the texts with each other
- Evaluation of the variants and of the witnesses bearing them

....and yes, it is usually done manually.



Recensio

- To establish relationships between witnesses and which ones bear the “best text”
- To establish an organic scheme the transmission of a text, often represented as a genealogical tree of witnesses (*stemma*)



Example of a stemma. Stemma for *De nuptiis Philologiae et Mercurii* by Martianus Capella proposed by Danuta Shanzer (1986, p. 62-81).

Critical editions

- Usually display textual variants in the form of *apparatus criticus*
- The *apparatus* is a choice in itself: it does not collect all the variants found through collation, but only those that the editor had judged significant for the reconstruction of the text
- The *apparatus* can be very complex to understand in large textual traditions

Critical text

Critical apparatus

C. SALLUSTI CRISPI
~~CATILINAE CONIURATIO~~

1 Omnis homines, qui sese student praestare ceteris animalibus, summa ope niti decet, ne vitam silentio transeant veluti pecora, quae natura prona atque ventri 5
2 oboedientia finxit. sed nostra omnis vis in animo et corpore sita est: animi imperio, corporis servitio magis utimur; alterum nobis cum dis, alterum cum beluis commune est. quo mihi rectius videtur ingeni quam virium opibus gloriam quaerere, et quoniam vita ipsa qua fruimur brevis est, memoriam nostri quam maxime longam efficere. nam divitiarum et formae gloria fluxa atque fragilis est, virtus clara aeternaque habetur.

5 Sed diu magnum inter mortalis certamen fuit, vine corporis an virtute animi res militaris magis procederet. nam et prius quam incipias consulto et ubi consulueris mature facto opus est. ita utrumque per-

2 De titulo vide Ag 155 sqq. 3 omnis homines *Char. gramm. I 149, 17 Diom. gramm. I 305, 29 omnis... student Prisc. gramm. II 358, 15 omnis... praestare Non. p. 371, 11 omnis... animalibus Char. gramm. I 140, 1 Eugraph. Ter. Eun. 232 omneis Char. omnes Eugraph. qui... animalibus Arus. gramm. VII 508, 4 praestare ceteris animalibus Diom. gramm. I 313, 11 5 pecora... finxit Arus. gramm. VII 496, 27 quae... finxit Non. p. 309, 11 Victorin. rhet. p. 160, 36 Prisc. gramm. III 370, 18 ventri oboedientia Sen. epist. 8(60), 4 oboedientes Sen. 6 sed... sita est Seru. Aen. 2, 452 georg. 1, 198 sed... utimur Lact. inst. 2, 12, 12 7 animi... utimur Hier. ad Gal. 5, 16 p. 410 ad Eph. 5, 33 p. 537 animi... commune est* Hier. adv. Iovin. 2, 10 Aug. civ. 9, 9 animae Hier. adv. Iovin. 8 utimur] vivere Hier. ad Gal. alterum nobis... commune est Seru. Aen. 5, 81 9 videtur] esse videtur XNMTm videtur esse BKHDF1sn 10 et... efficere Victorin. rhet. p. 160, 33 17 nam... opus est Don. Ter. Andr. 334 Prisc. gramm. III 226 3 288 .17

Now we can do some of
these things automatically

iAligner

<http://i-alignment.com/>

https://github.com/OpenGreekAndLatin/ILA_python



iAligner

Home

Get Started

Bibliography

Contact

The tool performs automatic syntax-based intra-language alignment. It performs automatic alignment of different versions of a text. Its concept is based on a modified version of the Needleman-Wunsch algorithm (for more information, see the Bibliography).

The tool allows automatic alignment between parallel texts in the same language. Its purpose is to display various degrees of textual variants based on syntactic alignment.

Upload your file ?

Select File

Scegli file Nessun file selezionato

Supported formats: **txt and csv**

- Ignore Nonalphabetical Ignore Diacritics Case sensitive
 Levenshtein Distance

[View on Github \(Python\)](#)

[View on Github \(Php\)](#)

Enter Your Text ?

- Ignore Nonalphabetical Ignore Diacritics Case sensitive
 Levenshtein Distance

A tool for automatic syntax-based intra-language alignment

- **Automatic:** it is performed with algorithmic methods to reduce human intervention in the mechanical process of comparison.
- **Syntax-based:** in programming language, defines the order of the characters and the order of the words in a sentence.
- **Intra-language:** works with texts in the same language.
- **Pairwise or multiple:** works with two texts or with an unlimited number of multiple texts.

Algorithmic methods to produce alignment

The Needleman-Wunsch algorithm

- used in bioinformatics to align protein or nucleotide sequences.
- it uses Dynamic Programming to find the optimal alignment.
- divides a large problem into a series of smaller problems and uses the solutions to the smaller problems to reconstruct a solution to the larger problem.
- uses a score function and similarity matrix to represent all possible combinations of tokens and their resulting score.

The Needleman-Wunsch algorithm

- Aligning Bible Text John 1:1

NLT: In the beginning the Word already existed.

KJB: In the beginning was the Word

$$M[0][j] = j * \text{indel} \quad \text{where } j \in [0, n],$$

$$M[i][0] = i * \text{indel} \quad \text{where } i \in [0, m]$$

$$M[i][j] = \max \begin{cases} M[i-1][j-1] + m \searrow & \text{Matching} \\ M[i-1][j-1] + \text{mis} \searrow & \text{Mismatching} \\ M[i][j-1] + \text{indel} \downarrow & \text{Gap in S1} \\ M[i-1][j] + \text{indel} \rightarrow & \text{Gap in S2} \end{cases}$$

$$\text{Where } 0 < i \leq m, 0 < j \leq n$$

The used score function (Matching = 5, Mismatching = -5, In/Del = -2)

		In	the	beginning	the	Word	already	existed	.
	0	-2→	-4→	-6→	-8→	-10→	-12→	-14→	-16→
In	-2↓								
the	-4↓								
beginning	-6↓								
was	-8↓								
the	-10↓								
Word	-12↓								
,	-14↓								

The used score function (Matching = 5, Mismatching = -5, In/Del = -2)

		In	the	beginning	the	Word	already	existed	.
	0	-2→	-4→	-6→	-8→	-10→	-12→	-14→	-16→
In	-2↓	5 ↘	3→	1 →	-1 →	-3 →	-5 →	-7 →	-9 →
the	-4↓	0↓							
beginning	-6↓	-2↓							
was	-8↓	-4↓							
the	-10↓	-8↓							
Word	-12↓	-10↓							
,	-14↓	-12↓							

The used score function (Matching = 5, Mismatching = -5, In/Del = -2)

		In	the	beginning	the	Word	already	existed	.
	0	-2→	-4→	-6→	-8→	-10→	-12→	-14→	-16→
In	-2↓	5 ↘	3→	1 →	-1 →	-3 →	-5 →	-7 →	-9 →
the	-4↓	0↓	10 ↘	8 →	13 →	11 →	9 →	7 →	-5 →
beginning	-6↓	-2↓	8 ↓						
was	-8↓	-4↓	6 ↓						
the	-10↓	-8↓	4 ↓						
Word	-12↓	-10↓	2 ↓						
,	-14↓	-12↓	-0 ↓						

The used score function (Matching = 5, Mismatching = -5, In/Del = -2)

		In	the	beginning	the	Word	already	existed	.
	0	-2→	-4→	-6→	-8→	-10→	-12→	-14→	-16→
In	-2↓	5 ↘	3→	1 →	-1 →	-3 →	-5 →	-7 →	-9 →
the	-4↓	0↓	10 ↘	8 →	13 →	11 →	9 →	7 →	-5 →
beginning	-6↓	-2↓	8 ↓	15 ↘	13 →	11 →	9 →	7 →	-5 →
was	-8↓	-4↓	6 ↓	8 ↓					
the	-10↓	-8↓	4 ↓	5 ↓					
Word	-12↓	-10↓	2 ↓	0 ↓					
,	-14↓	-12↓	-0 ↓	-5 ↓					

The used score function (Matching = 5, Mismatching = -5, In/Del = -2)

		In	the	beginning	the	Word	already	existed	.
	0	-2→	-4→	-6→	-8→	-10→	-12→	-14→	-16→
In	-2↓	5 ↘	3→	1 →	-1 →	-3 →	-5 →	-7 →	-9 →
the	-4↓	0↓	10 ↘	8 →	13 →	11 →	9 →	7 →	-5 →
beginning	-6↓	-2↓	8 ↓	15 ↘	13 →	11 →	9 →	7 →	-5 →
was	-8↓	-4↓	6 ↓	8 ↓	11↓	9→	7→	5→	3→
the	-10↓	-8↓	4 ↓	5 ↓	13↘	11→	9→	7→	5→
Word	-12↓	-10↓	2 ↓	0 ↓	11↓	18↘	16→	14→	12→
,	-14↓	-12↓	-0 ↓	-5 ↓	9↓	16↓	14→	12→	10→

The used score function (Matching = 5, Mismatching = -5, In/Del = -2)

		In	the	beginning	the	Word	already	existed	.
	0	-2→	-4→	-6→	-8→	-10→	-12→	-14→	-16→
In	-2↓	5 ↘	3→	1 →	-1 →	-3 →	-5 →	-7 →	-9 →
the	-4↓	0↓	10 ↘	8 →	13 →	11 →	9 →	7 →	-5 →
beginning	-6↓	-2↓	8 ↓	15 ↘	13 →	11 →	9 →	7 →	-5 →
was	-8↓	-4↓	6 ↓	8 ↓	11↓	9→	7→	5→	3→
the	-10↓	-8↓	4 ↓	5 ↓	13↘	11→	9→	7→	5→
Word	-12↓	-10↓	2 ↓	0 ↓	11↓	18↘	16→	14→	12→
,	-14↓	-12↓	-0 ↓	-5 ↓	9↓	16↓	14→	12→	10→

The used score function (Matching = 5, Mismatching = -5, In/Del = -2)

The Needleman-Wunsch algorithm

John 1:1

New Living Translation	In	the	beginning		the	Word	already	existed	.
King James Bible	In	the	beginning	was	the	Word			,

The modification to the algorithm

$$k = 14, n = 157, m = 134$$

$$\text{Search space} = m * n = 21038$$

after modification

$$(2k+1) * m = 3886$$

1In the beginning God created the heavens and the earth. 2The earth was formless and empty, and darkness covered the deep waters. And the Spirit of God was hovering over the surface of the waters. 3 Then God said, "Let there be light," and **there** was light. 4And God saw that the light was good. Then he separated the light from the darkness. 5God called the light "day" and the darkness "night." And evening passed and morning came, marking the first day.

6 Then God said, "Let there be a space between the waters, to separate the waters of the heavens from the waters of the earth." 7And that is what happened. God made this space to separate the waters of the earth from the waters of the heavens. 8God called the space "sky."

1In the beginning God created the heavens and the earth. 2The earth was without form, and void; and darkness was on the face of the deep. And the Spirit of God was hovering over the face of the waters. 3 **Then God said, "Let there be light"; and there was light.** 4**And God saw the light, that it was good; and God divided the light from the darkness.** 5 God called the light Day, and the darkness He called Night. So the evening and the morning were the first day.

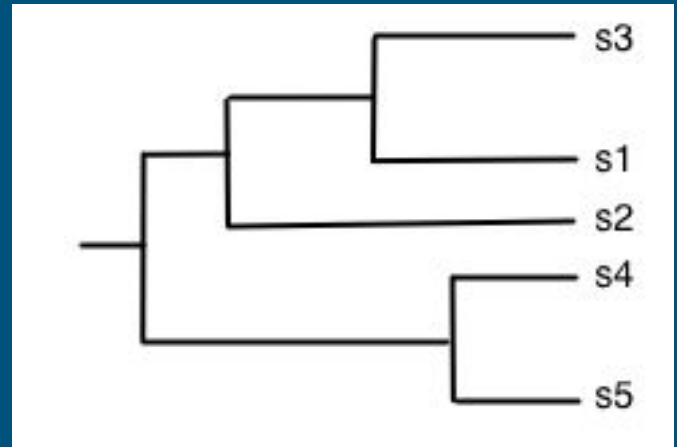
6 Then God said, "Let there be a firmament in the midst of the waters, and let it divide the waters from the waters." 7 Thus God made the firmament, and divided the waters which were under the firmament from the waters which were above the firmament; and it was so. 8And God called the firmament Heaven. So the evening and the morning were the second day.

Multiple Sequence Alignment (In progress)

- Progressive alignment

builds up a final MSA by combining pairwise alignments beginning with the most similar pair and progressing to the most distantly related, it requires two stages:

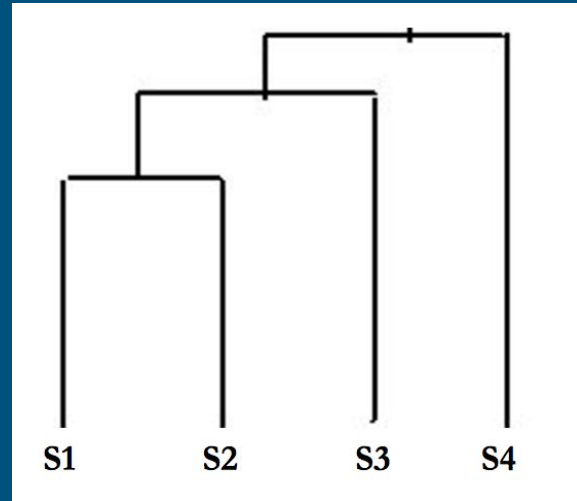
- creating the guide tree (clustering)
- adding the sequences sequentially to the growing MSA according to the guide tree



Multiple Sequence Alignment (In progress)

- Iterative alignment

The aim is to reduce the problem of a multiple alignment to an iteration of pairwise alignments.



How to align your texts with iAligner: copy your text on the editor

1 The book of the generation of Jesus Christ, the son of Davi
1 The book of the generation of Jesus Christ, the son of Davi

2 Abraham begat Isaac; and Isaac begat Jacob; and Jacob begat
2 Abraham begat Isaac; and Isaac begat Jacob; and Jacob begat

3 and Judah begat Perez and Zerah of Tamar; and Perez begat H
3 And Judas begat Phares and Zara of Thamar; and Phares begat

4 and Ram begat Amminadab; and Amminadab begat Nahshon; and N
4 And Aram begat Aminadab; and Aminadab begat Naasson; and Na

5 and Salmon begat Boaz of Rahab; and Boaz begat Obed of Ruth
5 And Salmon begat Booz of Rachab; and Booz begat Obed of Rut

6 and Jesse begat David the king. And David begat Solomon of
6 And Jesse begat David the king; and David the king begat So

7 and Solomon begat Rehoboam; and Rehoboam begat Abijah; and
7 And Solomon begat Roboam; and Roboam begat Abia; and Abia b

8 and Asa begat Jehoshaphat; and Jehoshaphat begat Joram; and
8 And Asa begat Josaphat; and Josaphat begat Joram; and Joram

9 and Uzziah begat Jotham; and Jotham begat Ahaz; and Ahaz be
9 And Ozias begat Joatham; and Joatham begat Achaz; and Achaz

The text has to be parsed in sentences first

Enter Your Text

1 The book of the generation of Jesus Christ, the son of David, the son of Abraham.
1 The book of the generation of Jesus Christ, the son of David, the son of Abraham.

2 Abraham begat Isaac; and Isaac begat Jacob; and Jacob begat Judah and his brethren;
2 Abraham begat Isaac; and Isaac begat Jacob; and Jacob begat Judas and his brethren;

3 and Judah begat Perez and Zerah of Tamar; and Perez begat Hezron; and Hezron begat Ram;
3 And Judas begat Phares and Zara of Thamar; and Phares begat Esrom; and Esrom begat Aram;

4 and Ram begat Amminadab; and Amminadab begat Nahshon; and Nahshon begat Salmon;

Ignore Nonalphabetical Ignore Diacritics Case sensitive Levenshtein Distance

Reset

Align

...Or upload it

Currently supports .txt and .csv files

Upload your file

Select File

bible test.txt

Supported formats: **txt and csv**

- Ignore Nonalphabetical
- Ignore Diacritics
- Case sensitive
- Levenshtein Distance

Refinement criteria

Enter Your Text

1 The book of the generation of Jesus Christ, the son of David, the son of Abraham.
1 The book of the generation of Jesus Christ, the son of David, the son of Abraham.

2 Abraham begat Isaac; and Isaac begat Jacob; and Jacob begat Judah and his brethren;
2 Abraham begat Isaac; and Isaac begat Jacob; and Jacob begat Judas and his brethren;

3 and Judah begat Perez and Zerah of Tamar; and Perez begat Hezron; and Hezron begat Ram;
3 And Judas begat Phares and Zara of Thamar; and Phares begat Esrom; and Esrom begat Aram;

4 and Ram begat Amminadab; and Amminadab begat Nahshon; and Nahshon begat Salmon;

Ignore Nonalphabetical Ignore Diacritics Case sensitive Levenshtein Distance

Reset

Align

Upload your file

Select File

Scegli file bible test.txt

Supported formats: **txt** and **csv**

Upload

Ignore Nonalphabetical Ignore Diacritics Case sensitive
 Levenshtein Distance

- **Ignore non-alphabetical:** ignores symbols, such as punctuation and numbers, anything that is not an alphabetical character
- **Case sensitive:** if activated, detects variation across words according to the case
- **Ignore diacritics:** ignores any type of diacritical character (including punctuation marks)
- **Levenshtein distance:** applies a revised version of the Levenshtein algorithm and increases the tolerance threshold on the alignment of similar words.

The Levenshtein distance

The Levenshtein distance between two words is the minimum number of single-character edits (i.e. insertions, deletions or substitutions) required to change one word into the other. e.g

$$\text{lev}(\text{Hellani}\text{kos}, \text{Hellani}\text{cus}) = 2$$

Mathematically, the Levenshtein distance between two strings $\{a,b\}$ (of length $|a|$ and $|b|$ respectively) is given by $\text{lev}_{a,b}(|a|, |b|)$

$$\text{lev}_{a,b}(i, j) = \begin{cases} \max(i, j) & \text{if } \min(i, j) = 0, \\ \min \begin{cases} \text{lev}_{a,b}(i-1, j) + 1 \\ \text{lev}_{a,b}(i, j-1) + 1 \\ \text{lev}_{a,b}(i-1, j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

Modified Levenshtein Distance

Levenshtein distance is not very helpful in our case, because it is binary and there is no tolerance with errors produced by OCR or Transcription.

the distance between letters is not binary, but it is on scale. The cost of insertion or deletion depends on:

- Letter position
- Letter type (vowel or consonant)

`lev(Hellanikos, Hellanicus) = 0.3`

Group elements	Distance within group elements
E, e Ē, ē I, i Y, y	0.15
Ō, ō O, o U, u Y, y	0.15
C, c K, k Q, q	0.15
S, s C, c	0.25
M, m N, n	0.35

A Greek text with no refinement criteria

τῆς δ' ἁσίας ἀπὸ κανώβου ἕως ταναΐδος ποταμοῦ μετὰ τῶν κόλπων ὁ παράπλους, σταδίων μυριάδων δ καὶ ρια·
τῆς δὲ ἁσίας, ἀπὸ κανώβου, ἕως ταναΐδος ποταμοῦ· μετὰ τῶν κόλπων ὁ παράπλους· στάδια Δ ρια·

τῆς	δ'	ἁσίας	ἀπὸ	κανώβου	ἕως	ταναΐδος	ποταμοῦ	μετὰ	τῶν	κόλπων	ὁ	παράπλους	,	σταδίων	μυριάδων	δ	καὶ	ρια	·	
τῆς	δὲ	ἁσίας	·	ἀπὸ	κανώβου	·	ἕως	ταναΐδος	ποταμοῦ	·	μετὰ	τῶν	κόλπων	ὁ	παράπλους	·	στάδια	Δ	ρια	·

Length: 24 Aligned-complete 13 notAligned 5 Gap 5 Aligned-case 1

The longest common substring

μετὰ τῶν κόλπων ὁ
μετὰ τῶν κόλπων ὁ

The same text with additional refinement criteria applied

τῆς δ' ἁσίας ἀπὸ κανώβου ἕως ταναΐδος ποταμοῦ μετὰ τῶν κόλπων ὁ παράπλους, σταδίων μυριάδων δ καὶ ρια·
τῆς δὲ ἁσίας, ἀπὸ κανώβου, ἕως ταναΐδος ποταμοῦ· μετὰ τῶν κόλπων ὁ παράπλους· στάδια Δ ρια·

τῆς	δ'	ἁσίας	ἀπὸ	κανώβου	ἕως	ταναΐδος	ποταμοῦ	μετὰ	τῶν	κόλπων	ὁ	παράπλους	,	σταδίων	μυριάδων	δ	καὶ	ρια	·	
τῆς	δὲ	ἁσίας	·	ἀπὸ	κανώβου	·	ἕως	ταναΐδος	ποταμοῦ	·	μετὰ	τῶν	κόλπων	ὁ	παράπλους	·	στάδια	Δ	ρια	·

Length: 24 Aligned-complete 13 notAligned 1 Gap 5 Aligned-levenshtein 2 Aligned-removedDiacritics 1 Aligned-removedNonAlphanumeric 1 Aligned-case 1

The longest common substring

μετὰ τῶν κόλπων ὁ παράπλους , σταδίων
μετὰ τῶν κόλπων ὁ παράπλους · στάδια

Alignment output: a table-graph

iAligner displays all the nuances of variants according to a color-key:

- Completely aligned tokens (deep green)
- Tokens aligned by excluding case sensitivity or punctuation detection (light green)
- Gaps (yellow)
- Divergences (red)
- Tokens aligned by applying Levenshtein distance (blue-green)

της δ' άσιας άπό κανώβου έως τανάιδος ποταμού μετά τών κόλπων ό παράπλους, σταδίων μυριάδων δ και ρια ·
της δέ άσιας. άπό κανώβου. έως ταναΐδος ποταμού· μετά τών κόλπων ό παράπλους· στάδια Δ ρια·

της	δ'	άσιας	·	άπό	κανώβου	·	έως	τανάιδος	ποταμού	·	μετά	τών	κόλπων	ό	παράπλους	,	σταδίων	μυριάδων	δ	και	ρια	·	
της	δέ	άσιας	·	άπό	κανώβου	·	έως	ταναΐδος	ποταμού	·	μετά	τών	κόλπων	ό	παράπλους	·	στάδια			Δ		ρια	·

Length: 24 Aligned-complete 13 notAligned 1 Gap 5 Aligned-Levenshtein 2 Aligned-removedDiacritics 1 Aligned-removedNonAlphanumeric 1 Aligned-case 1

The longest common substring

μετά	τών	κόλπων	ό	παράπλους	,	σταδίων
μετά	τών	κόλπων	ό	παράπλους	·	στάδια

What can you do
with iAligner?
Some case studies

Manuscript alignment

Manuscript Alignment

Plato's *Crito*

Clark: A digital encoding of Ms Clark 39, 20v-26r, Oxford, University Bodleian Library

Paris1808: A digital encoding of Ms Grec 1808, 17r-21v, Paris, Bibliothèque Nationale

Tuebingen: A digital encoding of Ms Gr Mb 14, 21-38, Tübingen, Universität, Bibliothek

43 ▾ 44 ▾ 45 ▾ 46 ▾ 47 ▾ 48 ▾ 49 ▾ 50 ▾ 51 ▾ 52 ▾ 53 ▾ 54 ▾

46C

CLARK

#Σωκράτης καὶ τιμῶ, οὔσπερ καὶ πρότερον· ὦν ἐὰν μὴ βελτίω ἔχωμεν λέγειν ἐν τῷ παρόντι, εὐ ἴσθι. ὅτι οὐ μὴ σοὶ συγχωρήσω· οὐδ' ἂν πλείω τῶν νῦν παρόντων, ἢ τῶν πολλῶν δύναμις, ὥσπερ παῖδας ἡμᾶς μορμολύττηται· δεσμούς· καὶ θανάτου ἐπιπέμπουσα· καὶ χρημάτων ἀφαιρέσεις· πῶς οὖν ἂν μετριώτατα σκοποίμεθα αὐτά; εἰ πρῶτον μὲν τοῦτον τὸν λόγον ἀναλάβοιμεν, ὃν σὺ λέγεις περὶ τῶν δοξῶν· πότερον καλῶς ἐλέγετο ἐκάστοτε, ἢ οὐ· ὅτι ταῖς μὲν δεῖ τῶν δοξῶν προσέχειν τὸν νοῦν.

Paris1808

#Σωκράτης καὶ τιμῶ οὔσπερ καὶ πρότερον· ὦν ἐὰν μὴ βελτίω ἔχωμεν λέγειν ἐν τῷ παρόντι· εὐ ἴσθι ὅτι οὐ μὴ σοὶ συγχωρήσω. οὐδ' ἂν πλείω τῶν νῦν παρόντων ἢ τῶν πολλῶν δύναμις; ὥσπερ παῖδας ἡμᾶς μορμολύττηται δεσμούς· καὶ θανάτου ἐπιπέμπουσα· καὶ χρημάτων ἀφαιρέσεις· πῶς οὖν μετριώτατα σκοποίμεθα αὐτά· εἰ πρῶτον μὲν τοῦτον τὸν λόγον ἀναλάβοιμεν· ὃν σὺ λέγεις περὶ τῶν δοξῶν· πότερον· καλῶς ἐλέγετο ἐκάστοτε ἢ οὐ· ὅτι ταῖς . δεῖ τῶν δοξῶν προσέχειν τὸν νοῦν·

Tuebingen

#Σωκράτης καὶ τιμῶ· οὔσπερ καὶ πρότερον, ὦν ἐὰν μὴ βελτίω ἔχωμεν λέγειν ἐν τῷ παρόντι. εὐ ἴσθι. ὅτι οὐ μὴ σοὶ συγχωρήσω· οὐδ' ἂν πλείω τῶν νῦν παρόντων· ἢ τῶν πολλῶν δύναμις, ὥσπερ παῖδας ἡμᾶς μορμολύττηται· δεσμούς· καὶ θανάτου ἐπιπέμπουσα· καὶ χρημάτων ἀφαιρέσεις· πῶς οὖν ἂν μετριώτατα, σκοποίμεθα αὐτά, εἰ πρῶτον μὲν τοῦτον τὸν λόγον ἀναλάβοιμεν· ὃν σὺ λέγεις περὶ τῶν δοξῶν· πότερον καλῶς ἐλέγετο ἐκάστοτε, ἢ οὐ· ὅτι ταῖς μὲν, δεῖ τῶν δοξῶν προσέχειν τὸν νοῦν·

#Σωκράτης καὶ τιμῶ , οὔσπερ καὶ πρότερον · ὦν ἐὰν μὴ βελτίω ἔχωμεν λέγειν ἐν τῷ παρόντι , εὐ ἴσθι . ὅτι οὐ μὴ σοὶ συγχωρήσω

#Σωκράτης καὶ τιμῶ οὔσπερ καὶ πρότερον · ὦν ἐὰν μὴ βελτίω ἔχωμεν λέγειν ἐν τῷ παρόντι · εὐ ἴσθι ὅτι οὐ μὴ σοὶ συγχωρήσω

#Σωκράτης καὶ τιμῶ · οὔσπερ καὶ πρότερον , ὦν ἐὰν μὴ βελτίω ἔχωμεν λέγειν ἐν τῷ παρόντι . εὐ ἴσθι . ὅτι οὐ μὴ σοὶ συγχωρήσω

συγχωρήσω · οὐδ' ἂν πλείω τῶν νῦν παρόντων , ἢ τῶν πολλῶν δύναμις , ὥσπερ παῖδας ἡμᾶς μορμολύττηται · δεσμούς · καὶ θανάτου

συγχωρήσω . οὐδ' ἂν πλείω τῶν νῦν παρόντων ἢ τῶν πολλῶν δύναμις : ὥσπερ παῖδας ἡμᾶς μορμολύττηται δεσμούς καὶ θανάτου

συγχωρήσω · οὐδ' ἂν πλείω τῶν νῦν παρόντων · ἢ τῶν πολλῶν δύναμις , ὥσπερ παῖδας ἡμᾶς μορμολύττηται · δεσμούς καὶ θανάτου

ἐπιπέμπουσα · καὶ χρημάτων ἀφαιρέσεις · πῶς οὖν ἂν μετριώτατα σκοποίμεθα αὐτά ; εἰ πρῶτον μὲν τοῦτον τὸν λόγον ἀναλάβοιμεν ,

ἐπιπέμπουσα · καὶ χρημάτων ἀφαιρέσεις · πῶς οὖν μετριώτατα σκοποίμεθα αὐτά · εἰ πρῶτον μὲν τοῦτον τὸν λόγον ἀναλάβοιμεν ·

ἐπιπέμπουσα · καὶ χρημάτων ἀφαιρέσεις · πῶς οὖν ἂν μετριώτατα , σκοποίμεθα αὐτά , εἰ πρῶτον μὲν τοῦτον τὸν λόγον ἀναλάβοιμεν ·

Three manuscripts of Plato's *Crito*
aligned (<http://i-alignment.com/crito/>)

OCR output alignment

Example of Aligning OCR-outputs:

ooo.31924054869700_ocr/024.txt	hvd.32044019207893_ocr/024.txt	Alignment
δυνάμει πεπυργωμένην, καὶ οἶον ἀμάχῳ φλογί ταῖς	δυνάμει πεπυργωμένην, καὶ οἶον ἀμάχῳ φλογί ταῖς	δυνάμει πεπυργωμένην, καὶ οἶον ἀμάχῳ φλογί ταῖς
ἄνωθεν εὐμενεῖαις διεζωσμένην. Τρέχει τοῖνον ὁ ἄγ-	ἄνωθεν εὐμενεῖαις διεζωσμένην. Τρέχει τοῖνον ὁ ἄγ-	ἄνωθεν εὐμενεῖαις διεζωσμένην. Τρέχει τοῖνον ὁ ἄγ-
γελος ἐκμετρῆσαι τὴν Ἱερουσαλήμ, καὶ οἶον ὀρίσαι	γελος ἐκμετρῆσαι τὴν Ἱερουσαλήμ, καὶ οἶον ὀρίσαι	γελος ἐκμετρῆσαι τὴν Ἱερουσαλήμ, καὶ οἶον ὀρίσαι
πλατεῖαν αὐτὴν καὶ εὐμηκεστάτην. Καὶ ταυτί μὲν	πλατεῖαν αὐτὴν καὶ εὐμηκεστάτην. Καὶ [®] ταυτί μὲν	πλατεῖαν αὐτὴν καὶ εὐμηκεστάτην. <input type="text" value="Καὶ"/> μὲν
ἱπορικῶς.	ἱστορικῶς. η	<input type="text" value="ἱπορικῶς."/> <input type="text" value="ἱ"/>
Φαίην δ' ἂν ὅτι καὶ ἐπ' αὐτῆς τῆς Ἐκκλησίας	. Φαίην δ' ἂν ὅτι καὶ ἐπ' αὐτῆς τῆς Ἐκκλησίας	<input type="text" value="ἱ"/> Φαίην δ' <input type="text" value="ἂν"/> ὅτι καὶ ἐπ' αὐτῆς τῆς Ἐκκλησίας <input type="text" value="ἱ"/>
Χριστοῦ νοοῖτο ἂν εἰκότως ἡ ὄρασις. Τετυραννεύκει	Χριστοῦ νοοῖτο ἂν εἰκότως ἡ ὄρασις. Τετυραννεύκει	Χριστοῦ νοοῖτο ἂν εἰκότως <input type="text" value="ἡ"/> ὄρασις. Τετυραννεύκει
μὲν γὰρ κατὰ πάντων ὁ Σατανᾶς τῶν ὄντων ἐπὶ τῆς	μὲν γὰρ κατὰ πάντων ὁ Σατανᾶς τῶν ὄντων ἐπὶ τῆς	<input type="text" value="μὲν"/> γὰρ <input type="text" value="κατὰ"/> πάντων <input type="text" value="ὁ"/> Σατανᾶς τῶν ὄντων ἐπὶ τῆς
γῆς, καὶ γεγόναμεν αἰχμάλωτοι, τοῖς ἐκείνου ζυγοῖς	γῆς, καὶ γεγόναμεν αἰχμάλωτοι, τοῖς ἐκείνου ζυγοῖς	γῆς, καὶ γεγόναμεν αἰχμάλωτοι, τοῖς ἐκείνου

Alignment of two OCR outputs from the *Patrologia Graeca*. The third column shows the overlapping sections and offers the user the choice between two variants where the two texts diverge.

OCR output alignment

auctoritatem , sub significatione nativitatis , proprietas naturalis ostensa
sit : Patrem suum dicebat Deum , aequalem se faciens Deo . Anne
naturalis nativitas non est , ubi per nomen patris proprii , naturae
aequalitas demonstratur ? Non enim ambigitur , quin aequalitas nihil
differat . Quis porro dubitabit , quin indifferentem naturam nativitas
consequatur ? Hinc enim est sola illa quae vere esse possit aequalitas :
quia naturae aequalitatem sola possit praestare nativitas . Aequalitas
vero numquam ibi esse credetur , ubi unio est : nec tamen illic reperietur ,
ubi differt . Ita similitudinis aequalitas nec solitudinem habet , nec
diversitatem : quia omnis aequalitas nec diversa , nec sola sit . 16 . Filii
nativitas et cum Patre aequalitas ex ipsius

auctoritatem , sub significatione nativitatis , proprietas naturalis ostensa
sit : Patrem suum dicebat Deum , qualem se faciens Deo . Anne naturalis
nativitas non est , ubi per nomen patris proprii , naturae
demonstratur ? Non enim ambigitur , quin aequalitas nihil differat . Quis
porro dubitabit , quin indifferentem naturam nativitas consequatur ? Hinc
enim est sola illa quae vere esse possit aequalitas : quia naturae aequalitatem
sola possit praestare nativitas . Aequalitas vero numquam ibi esse credetur , ubi
unio est : nec (c) tamen illic reperietur , ubi differt . Ita similitudinis
aequalitas nec solitudinem habet , nec diversitatem : quia omnis aequalitas
nec diversa , nec sola sit . 16 . Filii nativitas et cum Patre aequalitas ex
ipsius

Length: 127 || Aligned-complete (109) || Gap (3) || Aligned-levenshtein (14) || notAligned (1) ||

Patrologia Latina: OCR output vs. correct version: www.i-alignment.com/pl/

Alignment of editions

- Ζεὺς μὲν ἀφίκτωρ ἐπίδοι προφρόνως
- Ζεὺς μὲν ἀφίκτωρ ἐπίδοι προφρόνως

Ζεὺς	μὲν	ἀφίκτωρ	ἐπίδοι	προφρόνως
Ζεὺς	μὲν	ἀφίκτωρ	ἐπίδοι	προφρόνως

Length: 5 Aligned-complete 5

- στόλον ἡμέτερον νάιον ἀρθένι'
- στόλον ἡμέτερον νάιον ἀρθένι'

στόλον	ἡμέτερον	νάιον	ἀρθένι'
στόλον	ἡμέτερον	νάιον	ἀρθένι'

Length: 4 Aligned-complete 4

- ἀπὸ προσομίων λεπτοψαμάθων
- ἀπὸ προσομίων λεπτοψαμάθων

ἀπὸ	προσομίων	λεπτοψαμάθων
ἀπὸ	προσομίων	λεπτοψαμάθων

Length: 3 Aligned-complete 3

Vv. 1-3

- καὶ στασίαρχος τάδε πεσσονομῶν
- καὶ στασίαρχος τάδε πεσσονομῶν,

καὶ	στασίαρχος	τάδε	πεσσονομῶν
καὶ	στασίαρχος	τάδε	πεσσονομῶν

Length: 6 Aligned-complete 4 Gap 2

- κύδιστ' ἀχέων ἐπέκρανε,
- κύδιστ' ἀχέων ἐπέκρανε,

κύδιστ'	ἀχέων	ἐπέκρανε	,
κύδιστ'	ἀχέων	ἐπέκρανε	,

Length: 5 Aligned-complete 4 notAligned 1

Vv. 12-13

- πνεύματι χώρας ἄρσενοπληθῆ δ'
- πνεύματι χώρας ἄρσενοπληθῆ δ'

πνεύματι	χώρας	ἄρσενοπληθῆ	δ'
πνεύματι	χώρας	ἄρσενοπληθῆ	δ'

Length: 5 Aligned-complete 4 notAligned 1

- ἔσμον ὕβριστήν Αἴγυπτογενῆ,
- ἔσμον ὕβριστήν Αἴγυπτογενῆ,

ἔσμον	ὕβριστήν	Αἴγυπτογενῆ	,
ἔσμον	ὕβριστήν	Αἴγυπτογενῆ	,

Length: 5 Aligned-complete 5

- πρὶν πόδα χέρσω τῆ δ' ἐν ἀσώδει
- πρὶν πόδα χέρσω τῆ δ' ἐν ἀσώδει

πρὶν	πόδα	χέρσω	τῆ	δ'	ἐν	ἀσώδει
πρὶν	πόδα	χέρσω	τῆ δ'	ἐν	ἀσώδει	

Length: 7 Aligned-complete 5 Gap 1 notAligned 1

Vv. 29-31

Three excerpted editions of Aeschylus' *Supplices* aligned.
www.i-alignment.com/Aeschylus

Future work

Import and export options

Language dependent options for Latin, Greek, Arabic

Handling crossings and transpositions

Thanks for the attention!

