

# Sunoikisis Digital Classics Spring 2020

## Session 7

# Text Alignment

Chiara Palladino  
Furman University

Tariq Yousef  
University of Leipzig



# Structure of the session

- Introduction: what is text alignment?
- Ugarit: a tool for text alignment
- Live demo of Ugarit
- Case studies: translation alignment in the classroom
- Applications: automatic translation alignment, graph databases, dynamic lexicon
- Presentation of the exercise: low stakes and high stakes

What is text alignment?

# What is Text Alignment?

- Text alignment is the comparison of two or more texts
- It tries to define similarities and divergences in texts in the same language, or relationships between texts in different languages
- One of the most important tasks in Natural Language Processing: it can be performed automatically
- It can be done at different levels (**paragraph-level, document-level, word-level, sentence-level**)

# Pairwise alignment

- Alignment of two texts
- We distinguish on the number of text because it determines differences in the use of the alignment algorithm

In	the	beginning			the	Word	already	existed	.			

In	the	beginning			was	the	Word					
												,

English - English text aligned automatically with *iAligner*

In	the	beginning	God	created	the	heavens	and	the	earth	.	
Im	Anfang	Anfang	Gott	schuf	den	Himmel	und	die	Erde	:	

English - German text aligned manually using Alpheios [Creation 1 - 1]

# 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”)

- είτα Δαμάστης ό Κιτιεύς τὰ πλεῖστα ἐκ τῶν Ἐκαταίου μεταγράφας περίπλουν ἔγραψεν.
- είτα Δαμάστης ό Σιγειεύς τὰ πλεῖστα ἐκ τῶν Ἐκαταίου μεταγράφας περίπλουν ἔγραψεν.
- είτα Δαμάστης ό Κητιεύς τὰ πλεῖστα ἐκ τῶν Ἐκαταίου μεταγράφας περίπλουν ἔγραψεν.
- είτα δαμάστης ό κητιεύς τὰ πλεῖστα ἐκ τῶν ἐκαταίου μεταγράφας, περίπλουν ἔγραψεν.

είτα	Δαμάστης	ό	Κιτιεύς	τὰ	πλεῖστα	ἐκ	τῶν	Ἐκαταίου	μεταγράφας	περίπλουν	ἔγραψεν	.
είτα	Δαμάστης	ό	Σιγειεύς	τὰ	πλεῖστα	ἐκ	τῶν	Ἐκαταίου	μεταγράφας	περίπλουν	ἔγραψεν	.
είτα	Δαμάστης	ό	Κητιεύς	τὰ	πλεῖστα	ἐκ	τῶν	Ἐκαταίου	μεταγράφας	περίπλουν	ἔγραψεν	.
είτα	δαμάστης	ό	κητιεύς	τὰ	πλεῖστα	ἐκ	τῶν	ἐκαταίου	μεταγράφας	, περίπλουν	ἔγραψεν	.

Four texts aligned with *iAligner*

# Cross-language alignment

- Alignment of texts in different languages
- Cross-language alignment is not easy to perform automatically and needs training data from manual alignment and bilingual dictionaries.

HYE

Յաղագս թէ որպէս կայսրն Վաղէս Հալածեալ զամենայն ուղղափառ  
հաւատացեալս ,և կամ որպէս կամեցաւ յանդիմանութիւն առնել կայսրն ի մէջ  
նշմարտութեանն հաւատացելոց ,և ի մէջ ձախող աղանդոյն արիանոսաց .

DEU

Darüber „wie der Kaiser Valēs alle rechtgläubigen Gläubigen verfolgte ,und noch wie  
der Kaiser beabsichtigte ,ein Streitgespräch unter den Gläubigen der Wahrheit und  
unter der bösen Sekte der Arianer zu machen .

( 29 ) 100% HYE - DEU

( 35 ) 100% HYE - DEU

Armenian-German aligned text, Alignment is done manually by Uta Koschmieder

# Word-level Alignment

- Word alignment is typically done after sentence alignment has already been completed.
- Needs training data in order to employ machine learning algorithms.
- It is not trivial:
  - It has to deal with **crossing dependencies**, where the order of sentences are changed.
  - Translators do not always translate one sentence in the input into one sentence in the output.
  - Cultural and linguistic differences condition a perfect alignment most of the time.



# What is a Parallel Text?

- A parallel text is a text placed alongside its translation or translations
- Large collections of parallel texts are called **parallel corpora**
- The alignment of parallel texts is also called **translation alignment**

# Examples of parallel Corpora [1]

The **Bible corpus** contains  
1169 unique translations

<http://paralleltext.info/data/all/>

Available translations

Translation	Language	Family
1 aai-x-bible	Arifama-Minifia	Austronesian, Malayo-Polynesian, Central-Eastern Malayo-Polynesian, Eastern Malayo-Polynesian
2 aak-x-bible	Ankave	Trans-New Guinea, Angan, Nuclear Angan
3 aau-x-bible	Abau	Sepik, Abau
4 abt-x-bible-maprik	Ambulas	Sepik, Ndu
5 abt-x-bible-wosera	Ambulas	Sepik, Ndu
6 aby-x-bible	Aneme Wake	Trans-New Guinea, Southeast Papuan, Yareban
7 acd-x-bible	Gikyode	Niger-Congo, Atlantic-Congo, Volta-Congo, Kwa, Nyo, Potou-Tano, Tano, Guang, North...
8 ace-x-bible	Aceh	Austronesian, Malayo-Polynesian, Malayo-Chamic, Chamic, Acehnese
9 acf-x-bible	Saint Lucian Creole French	Creole, French based
10 acn-x-bible	Achang	Sino-Tibetan, Tibeto-Burman, Ngwi-Burmese, Burmish, Northern
11 acr-x-bible-cubulcu	Rabinal Achí	Mayan, Yucatecan-Core Mayan, K'ichean-Mamean, K'ichean, Poqom-K'ichean, Core K'...
12 acr-x-bible-reformed	Rabinal Achí	Mayan, Yucatecan-Core Mayan, K'ichean-Mamean, K'ichean, Poqom-K'ichean, Core K'...
13 acr-x-bible-traditional	Rabinal Achí	Mayan, Yucatecan-Core Mayan, K'ichean-Mamean, K'ichean, Poqom-K'ichean, Core K'...
14 acu-x-bible	Achuar-Shiwiar	Jivaroan, Jivaro
15 ade-x-bible	Adele	Niger-Congo, Atlantic-Congo, Volta-Congo, Kwa, Nyo, Potou-Tano, Basila-Adele
16 adj-x-bible	Adioukrou	Niger-Congo, Atlantic-Congo, Volta-Congo, Kwa, Nyo, Agneby
17 aey-x-bible	Amele	Trans-New Guinea, Madang, Croisilles, Gum
18 afr-x-bible-1953	Afrikaans	Indo-European, Germanic, West, Low Saxon-Low Franconian, Low Franconian
19 afr-x-bible-1983	Afrikaans	Indo-European, Germanic, West, Low Saxon-Low Franconian, Low Franconian
20 afr-x-bible-boodskap	Afrikaans	Indo-European, Germanic, West, Low Saxon-Low Franconian, Low Franconian
21 afr-x-bible-lewende	Afrikaans	Indo-European, Germanic, West, Low Saxon-Low Franconian, Low Franconian
22 afr-x-bible-viralmal	Afrikaans	Indo-European, Germanic, West, Low Saxon-Low Franconian, Low Franconian
23 agd-x-bible	Agarabi	Trans-New Guinea, Kainantu-Goroka, Kainantu, Gadsup-Auyana-Awa
24 agy-x-bible	Angor	Senagi
25 agm-x-bible	Angaatiha	Trans-New Guinea, Angan, Angaatiha
26 agn-x-bible	Agutaynen	Austronesian, Malayo-Polynesian, Philippine, Kalamian
27 agr-x-bible	Aguaruna	Jivaroan
28 agu-x-bible	Aguacateco	Mayan, Yucatecan-Core Mayan, K'ichean-Mamean, Mamean, Awakateko-Ixil
29 ahk-x-bible	Akha	Sino-Tibetan, Tibeto-Burman, Ngwi-Burmese, Ngwi, Southern
30 aia-x-bible	Arosi	Austronesian, Malayo-Polynesian, Central-Eastern Malayo-Polynesian, Eastern Malayo-Polynesian

# Examples of parallel Corpora [3]

Goethe poems:

## An Luna

Schwester von dem ersten Licht,  
Bild der Zärtlichkeit und Trauer!  
Nebel schwimmt mit Silberschauer  
Um dein reizendes Gesicht;  
Deines leisen Fußes Lauf  
Weckt aus tagverschloßnen Höhlen  
Traurig abgeschiedne Seelen,  
Mich und nächtge Vögel auf.

## To Luna

Sister of the first-born light,  
Type of sorrowing gentleness!  
Quivering mists in silv'ry dress  
Float around thy features bright;  
When thy gentle foot is heard,  
From the day-closed caverns then  
Wake the mournful ghosts of men,  
I, too, wake, and each night-bird.

<http://goethe.lingvisto.org/>

## Al la luna

¡Oh tú, la hermana de la luz primera,  
símbolo del amor en la tristeza!  
Ciñe tu rostro encantador la bruma,  
orlada de argentados resplandores;  
Tu sigiloso paso de los antros  
durante el día cerrados cual sepulcros,  
a los tristes fantasmas despabila,  
y a mí también y a las nocturnas aves.

# Examples of parallel Corpora [5]

Films Subtitles:

<http://www.tvsubtitles.net/>

Friends season 1

#	Episode	Amount	Subtitles
1x24	The One Where Rachel Finds Out	8	
1x23	The One With The Birth	8	
1x22	The One With The Ick Factor	8	
1x21	The One With The Fake Monica	8	
1x20	The One With The Evil Orthodontist	8	
1x19	The One Where The Monkey Gets Away	8	
1x18	The One With All The Poker	8	
1x17	The One With Two Parts (2)	7	
1x16	The One With Two Parts (1)	8	
1x15	The One With The Stoned Guy	8	
1x14	The One With The Candy Hearts	8	
1x13	The One With The Boobies	8	
1x12	The One With The Dozen Lasagnas	8	
1x11	The One With Mrs. Bing	8	
1x10	The One With The Monkey	8	
1x09	The One Where Underdog Gets Away	8	
1x08	The One Where Nana Dies Twice	7	
1x07	The One With The Blackout	8	
1x06	The One With The Butt	8	
1x05	The One With The East German Laundry Detergent	8	
1x04	The One With George Stephanopoulos	8	
1x03	The One With The Thumb	8	
1x02	The One With The Sonogram At The End	8	
1x01	The One Where Monica Gets A Roommate	9	
All episodes		9	

## Films Subtitles:

<http://www.tvsubtitles.net/>

## Game Of Thrones

1  
00:04:18,100 --> 00:04:19,466  
Your Grace.  
The trial will be getting under way soon.

2  
00:04:55,721 --> 00:04:57,120  
You got my money?  
Later. Go away.

3  
00:05:10,254 --> 00:05:12,088  
Grand Maester.

4  
00:05:52,496 --> 00:05:54,964  
Ser Loras Tyrell.

5  
00:05:56,801 --> 00:05:59,534  
Are you prepared to stand trial

6  
00:05:59,637 --> 00:06:03,338  
and profess your guilt or  
innocence before the Seven?

7  
00:06:05,943 --> 00:06:08,343  
There'll be no need for a trial.

8  
00:06:10,848 --> 00:06:13,249  
I confess before the Seven  
and freely admit to my crimes.

1  
00:04:19,493 --> 00:04:23,953  
Euer Hoheit?  
Der Prozess beginnt in Kürze.

2  
00:04:57,114 --> 00:05:00,418  
- Habt Ihr mein Geld?  
- Später. Geh jetzt.

3  
00:05:11,670 --> 00:05:13,446  
Großmaester?

4  
00:05:54,171 --> 00:05:56,235  
Ser Loras Tyrell.

5  
00:05:58,175 --> 00:06:00,983  
Seid Ihr bereit,  
Euch dem Prozess zu stellen?

6  
00:06:01,011 --> 00:06:05,058  
Euch zu Eurer Schuld oder Unschuld  
vor den 7 zu bekennen?

7  
00:06:07,309 --> 00:06:09,622  
Ein Prozess ist nicht nötig.

8  
00:06:12,564 --> 00:06:16,570  
Ich gestehe vor den 7  
und bekenne mich zu meinen Verbrechen.

# Ugarit: a tool for text alignment

# Manual alignment

The **UGARIT** Translation  
Alignment Editor is a tool  
for manual cross-language  
alignment (of texts in  
different languages)  
[\(www.ugarit.ialigner.com\)](http://www.ugarit.ialigner.com)

## Matthew 2

Tariq Yousef / New Testament

Created on 2017-04-12 17:02:24

Modified on 2017-05-31 16:50:16

Aligned by Tariq Yousef

### English

#### Matthew 2

1 After Jesus was born in Bethlehem in Judea , during the time of King Herod , Magi [ a ] from the east came to Jerusalem

2 and asked , " Where is the one who has been born king of the Jews ? We saw his star when it rose and have come to worship him . "

( 15 ) 24% ENG

( 47 ) 76% ENG - DEU

### Deutsch

#### Matthaeus 2

1 Da Jesus geboren war zu Bethlehem im jüdischen Lande , zur Zeit des Königs Herodes , siehe , da kamen die Weisen vom Morgenland nach Jerusalem und sprachen :

2 Wo ist der neugeborene König der Juden ? Wir haben seinen Stern gesehen im Morgenland und sind gekommen , ihn anzubeten .

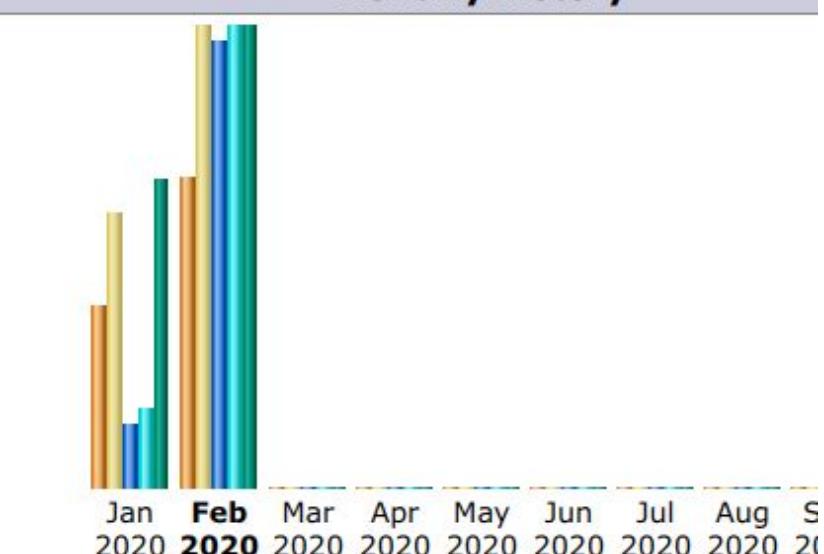
( 43 ) 78% ENG - DEU

( 12 ) 22% DEU

# What is UGARIT ?

- Web-based tool for manual alignment of parallel corpora.
- The first aim was to generate accurate training dataset for our automatic word alignment system.
- Translation pairs are downloadable in different formats.
- Crowdsourcing project (cumulative results).
- Ugarit serves also as reading environment and learning tool.

We started on  
March 2017



Month	Unique visitors	Number of visits	Pages
Jan 2020	3,376	5,092	13,088
<b>Feb 2020</b>	<b>5,747</b>	<b>8,494</b>	<b>88,936</b>
Mar 2020	0	0	0

Languages

36

Texts

23578

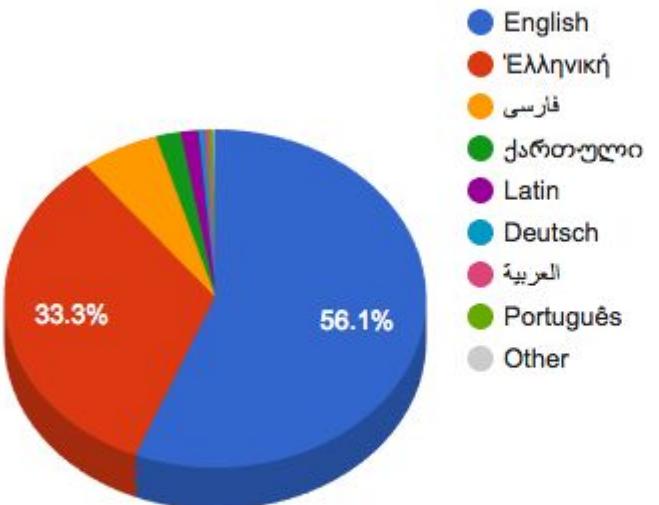
Translation Pairs

452085

Users

316

Languages by Number of Tokens

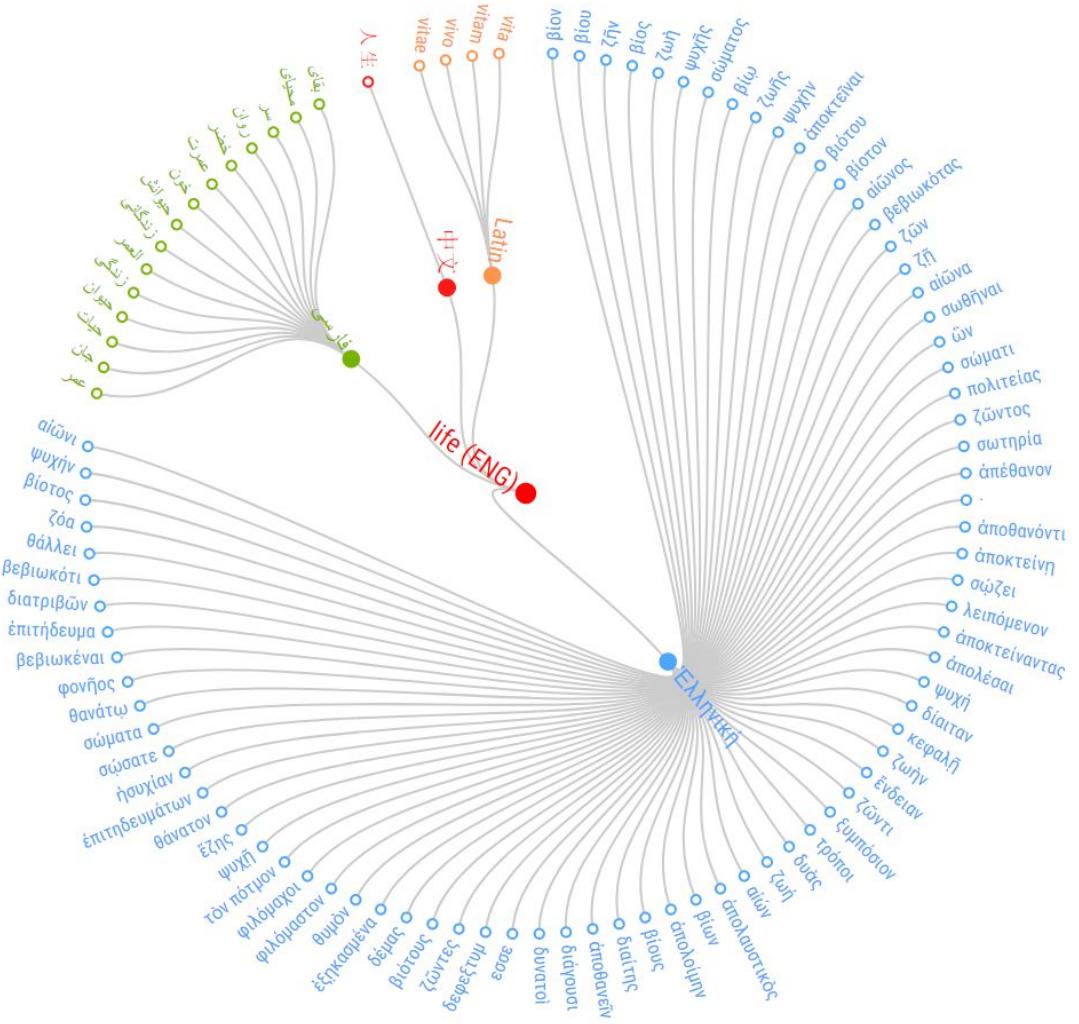


# life (English)

Frequency: 305

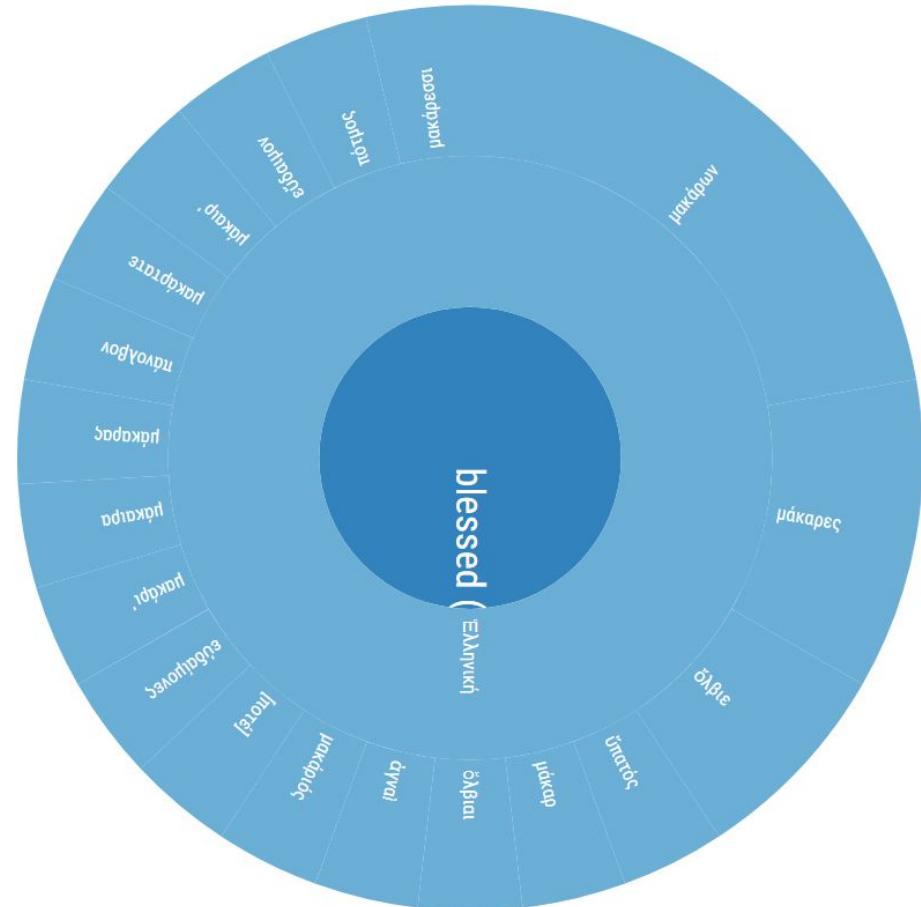
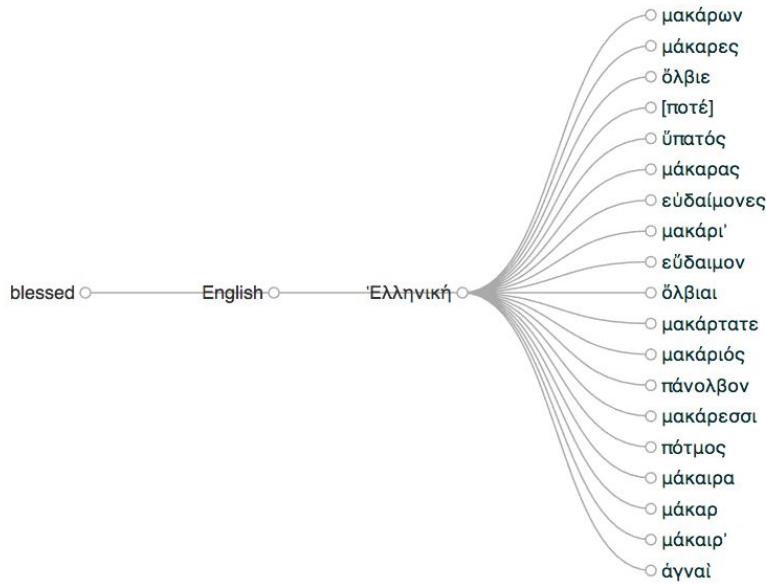
Translations:

- فارسی : حیات (7), حیوان (7), جان (30), عمر (31), سر (1), بقای (1), روان (1), زندگانی (2), عمر (1), خضر (1), محیای (1), خون (1), حیوانش (1)
- Ελληνική : βίου (25), βίος (9), βίον (40), τὸν πότμον (1), αἰών (1), φιλόμαχοι (1), φιλόμαστον (1), βιότου (3), ψυχῆς (7), αἰῶνα (2), θυμὸν (1), αἰῶνος (3), ἔξηκασμένα (1), δέμας (1), βιότους (1), ψυχῆ (1), ἔζης (1), αἰῶνι (1), βίοτον (3), ζῶν (2), ζῆν (16), ψυχήν (1), βίοτος (1), ζόα (1), θάλλει (1), βεβιωκότι (1), βεβιωκότας (2), διατριβῶν (1), ἐπιτήδευμα (1), βεβιωκέναι (1), βίω (5), ζῶντες (1), φονῆος (1), ζῆ (2), θανάτω (1), σώματος (6), σώματα (1), σώσατε (1), ὄν (2), σώματι (2), ἡσυχίαν (1), ἐπιτήδευμάτων (1), θάνατον (1), σωθῆναι (2), ζῶντος (1), πολιτείας (1), λειπόμενον (1), ἀποκτείναντας (1), ἀποκτεῖναι (3), ψυχὴν (4), ἀπολέσαι (1), · (1), σώζει (1), ἀποκτείνῃ (1), ἀποθανόντι (1), σωτηρία (1), ἀπέθανον (1), ψυχῆ (1), διαιταν (1), κεφαλῆ (1), ἀπολοίμην (1), διαιτης (1), ἀποθανεῖν (1), διάγουσι (1), δυνατοὶ (1), εσσε (1), δεφεξτυμ (1), ζωῆς (4), βίων (1), ἀπολαυστικός (1), βίους (1), ζωὴν (1), ζωὴ (7), ἔνδειαν (1), ζῶντι (1), ξυμπόσιον (1), τρόποι (1), δυὰς (1), ζωὴ (1)
- Latin : *vita* (1), *vitam* (1), *vitae* (2), *vivo* (1)
- 中文 : 人生 (2)



Frequency: 34

**Έλληνική** μακάρων 6 μάκαρες 3 ὅλβιες 2 ὅλβιαι 1 μάκαρ 1 ἀγναὶ 1  
 [ποτέ] 1 μακάρεσσι 1 μακάρι' 1 μακάριός 1 πάνολβον 1  
 εῦδαιμον 1 πότμος 1 μάκαιρ' 1 μακάρτατε 1 μάκαρας 1  
 ὑπατός 1 μάκαιρα 1



## Perseus-text-1999.01.0003-0-sen=225

Automatic Alignment (Giza++) / Greek - English / Aeschylus / Aeschylus, Agamemnon (1)

<http://ugarit.ialigner.com/search.php?q=Blessed&lang=>

Created on 2017-01-22 13:59:17

Aligned by Automatic Alignment (Giza++)

★ Add to Favorite

Ελληνική

Transliterate

Perseus-text-1999.01.0003-0-sen=225

οἴκων δ' ἄρ' εὐθυδίκων καλλίπαις πότμος αἰεὶ .

English

but when a house is righteous , the lot of its children is blessed  
always .

( 2 ) 25% GRC

( 6 ) 75% GRC - ENG

( 6 ) 37% GRC - ENG

( 10 ) 63% ENG



## Perseus-text-1999.01.0003-0-sen=495

Automatic Alignment (Giza++) / Greek - English / Aeschylus / Aeschylus, Agamemnon (1)

Created on 2017-01-22 13:59:17

Aligned by Automatic Alignment (Giza++)

★ Add to Favorite

Ελληνική

Transliterate

Perseus-text-1999.01.0003-0-sen=495

καὶ τῷδε πόλιν μὲν ἐλεῖν ἔδοσαν μάκαρες Πριάμου ·

English

" so the blessed ones have granted to our prince to capture priam ' s  
town ;

( 2 ) 22% GRC

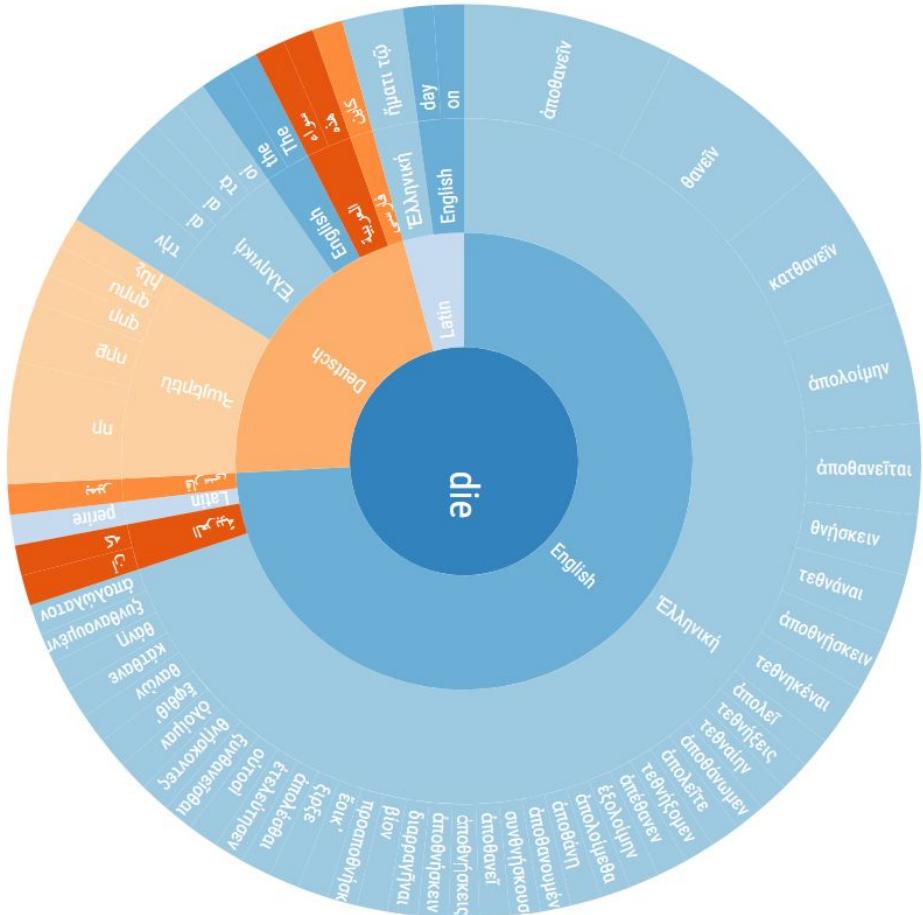
( 7 ) 78% GRC - ENG

( 10 ) 59% GRC - ENG

( 7 ) 41% ENG

188 results are found for *die*

Frequency: Latin 6 English 134 Deutsch 212



# Case Studies

# Why use translation alignment to study a language?

- To develop in-depth knowledge of the cultural concepts expressed in the source language, by systematically contrasting them with modern language translations
- To develop one's "translator's voice" by contrasting different translation strategies used by professional scholars
- To use skills in a third language as a bridge to convey similar constructs across two other languages
- To approach an original text even with limited knowledge of the source language

# Case Study 1: Students of the source language, advanced and intermediate level

Ελληνική Transliterate

**orthōs** ἀν μέχρι ἐνταῦθα πρός τὰ ἑρωτικὰ παιδαγωγῆτη, θεώμενος ἐφεξῆς τε καὶ ὄρθως τὰ καλά, πρὸς τέλος ἡδη ἴων τῶν ἑρωτικῶν ἔξαιρης κατόψεται τι θαυμαστὸν τὴν φύσιν καλόν, τοῦτο ἐκεῖνο, ὁ Σώκρατες, οὐδὲ δῆ ἔνεκεν καὶ οἱ ἔμπροσθεν πάντες πόνοι ἥσαν, πρώτον μὲν ἀει δῆν καὶ οὔτε γιγνόμενον οὔτε ἀπολλύμενον, οὔτε αὔξανόμενον οὔτε φθίνον, ἐπειτα οὐ τῇ μὲν καλόν, τῇ δὲ αἰσχρόν, οὐδὲ τοτὲ μὲν, τοτὲ δὲ οὐ, οὐδὲ πρὸς μὲν τὸ καλόν, πρὸς δὲ τὸ αἰσχρόν, οὐδὲ ἔνθα μὲν καλόν, ἔνθα δὲ αἰσχρόν, ὡς τιοὶ μὲν ὃν καλόν, τιοὶ δὲ αἰσχρόν: οὐδὲ αὖ φαντασθήσεται αὐτῷ τὸ καλόν οἷον πρόσωπον τι καλόν, οὐδὲ ἄλλο οὐδὲν ὃν σώμα μετέχει, οὐδὲ τις λόγος οὐδέ τις ἐπιστήμη, οὐδὲ ποῦ δην ἐτέρφα τινι, οἰον ἐν τῷ φύσῃ ἢ ἐν γῇ ἢ ἐν οὐρανῷ ἢ ἐν τῷ ἄλλῳ, ἄλλ’ αὐτὸς καθ’ αὐτὸν μετέχειν τοιούτοιν, οἷον γιγνομένων τε τῶν ἄλλων καὶ ἀπολλύμενων μηδὲν ἐκείνο μήτε τι πλέον μήτε ἔπαντον γίγνεσθαι μηδὲ πάσχειν μηδέν. ὅταν δῆ τις ἀπὸ τῶνδε διὰ τὸ ὄρθως παιδεραστεῖ ἐπανιώνει τὸν καλὸν ἄρχηται καθορᾶν, σχεδὸν ἄν τι διποτοι τοῦ τέλους. Τοῦτο γάρ δῆ ἔστι τὸ ὄρθως ἐπὶ τὰ ἑρωτικά ιέναι ἢ ὑπὸ ἀλλού ἀγεθθεῖ, ἀρχόμενον ἀπὸ τῶνδε τῶν καλῶν ἐκείνου ἔνεκα τοῦ καλοῦ ἀει ἐπανιέναι, ὥσπερ ἐπαναβασμοῖς χρώμενον, ἀπὸ ἐνός ἐπὶ δύο καὶ ἀπὸ δυοῖν ἐπὶ πάντα τὰ καλά σώματα, καὶ ἀπὸ τῶν καλῶν σωμάτων ἐπὶ τὰ καλὰ ἐπιτρέπειματα, καὶ ἀπὸ τῶν ἐπιτρέπειμάτων ἐπὶ τὰ καλὰ μαθήματα, καὶ ἀπὸ τῶν μαθήμάτων ἐπὶ τὸ μάθημα τελευτῆσαι, ὃ ἔστιν οὐκ ἄλλου ἢ αὐτοῦ ἐκείνου τοῦ καλοῦ μάθημα, καὶ γνῷ αὐτὸν τελευτῶν δῆ ἔστιν καλόν. ἐνταῦθα τοῦ βίου, ὁ φίλε Σώκρατες, ἐφη ἡ Μαντινική ἔνη, εἰπερ που ἀλλοθι, βιωτὸν ἀνθρώπων, θεωμένω αὐτῷ τὸ καλόν. ὃ ἔαν ποτε ίδης, οὐ κατὰ χροιν τε καὶ ἔσθητα καὶ τοὺς καλοὺς παιδάς τε καὶ νεανίσκους δόξει τοι εἶναι, οὓς νῦν ὄρων

English

You see, the man who has been thus far guided in matters of Love, who has beheld beautiful things in the right order and correctly, is coming now to the goal of Loving: all of a sudden he will catch sight of something wonderfully beautiful in its nature; that, Socrates, is the reason for all his earlier labors: First, it always is and neither comes to be nor passes away, neither waxes nor wanes. Second, it is not beautiful this way and ugly that way, nor beautiful at one time and ugly at another, nor beautiful in relation to one thing and ugly in relation to another; nor is it beautiful here but ugly there, as it would be if it were beautiful for some people and ugly for others. Nor will the beautiful appear to him in the guise of a face or hands or anything else that belongs to the body. It will not appear to him as a speech or a kind of knowledge does. It is not anywhere in another thing, as in an animal, or in earth, or in heaven, or in anything else, but itself by itself with itself, it is always one in form; and all the other beautiful things share in that, in such a way that when those others come to be or pass away, this does not become the least bit smaller or greater nor suffer any change. So when someone rises by these stages, through loving boys correctly, and begins to see this beauty, he has almost grasped his goal. This is what it is to go aright, or be led by another, into the art of Love: one goes always upwards for the sake of this Beauty, starting out from beautiful things and using them like rising stairs: from one body to two and from two to all beautiful bodies, then from beautiful bodies to beautiful customs, and from customs to learning beautiful things, and from these lessons he arrives in the end at this lesson, which is learning of this very Beauty, so that in the end he comes to know just what it is to be beautiful. And there in life, Socrates, my friend, "said the woman from Mantinea, "there if anywhere should a person live his life, beholding

Plato's *Symposium*, Greek text aligned with the translation by Nehamas and Woodruff (1989), in the context of a research project that integrated different levels of linguistic digital annotation, with the goal of enabling readers with no Greek knowledge to explore the original through the dynamic reading of the aligned parallel texts.

The student created a consistent alignment strategy that only matched words corresponding in meaning and grammatical function (with the partial exception of verbs for which a perfect match would not be possible, such as subjunctives or participles with predicative use).

# Case Study 1: Students of the source language, advanced and intermediate level

Latin

**Interea** quidam senex de summo colle prospectat , quem circum capellae pascentes opilonem esse profecto clamabant . Eum rogavit unus e nostris haberetne venu lactem vel adhuc liquidum vel in caseum recentem inchoatum . At ille diu capite quassanti 'Vos autem ' inquit 'De cibo vel poculo ) vel omnino ulla refectione nunc cogitatis ? An nulli scitis quo loco comederitis ? ' Et cum dicto conductus oviculus conversus longe recessit . Quae vox eius et fuga **pastoribus** nostris non mediocrem pavorem incussit : ac dum perterriti de loci qualitate sciscitari gestiunt nec est qui doceat , senex alius , magnus ille quidem , gravatus annis , totus in baculum pronus et lassum trahens vestigium , ubertim lacrimans per viam proximat , visisque nobis cum fletu maximo singulorum iuvenum genua contingens sic adorabat :

English

Now an old man appeared , gazing down on us from a summit at hand ; a goat-herd he was , as could be seen by the she-goats browsing round him . One of us asked him if he'd any milk or curds for sale . He shook his head several times before replying : 'How can you dream of food and drink , or anything else right now ? Don't you know where you are ? ' Then he gathered his goats , and made off into the distance . His words and his sudden flight filled us all with no little dread . We wondered what was wrong with the place , but there was no one the others could ask , till a second old man approached on the road , tall and bent with the years , hunched over his staff , wearily dragging his feet , and weeping copiously . Meeting with us he clasped the knees of all the young men in turn , wracked by tears .

English

An old man appeared at the top of the hill with goats feeding around him . One of our people hailed him and asked whether he had any milk or fresh cheese for sale . He shook his head two or three times before answering : 'How can you think of food or drink or anything else of the sort ? Don't you know in what sort of a place you are camping ? ' He turned his back and went off with his goats . His question and the abrupt way he left us alarmed our people . They all began wondering what was wrong with the place . But there was nobody to enlighten them until another old man appeared , a tall , bent old man , dragging his feet wearily towards us and leaning heavily on a stick . When he reached the glade where we had halted , he fell down on his knees , his eyes streaming with tears , embraced our people one after the other and groaned

*Metamorphoses* by Apuleius, aligned with two chosen English translations (A.S. Kline 2016 and R. Graves 1950), to discover differences in translation

strategies in a very complex and rhetorically skilful literary work. The focus of the observation was the semantics of different words and expressions with potential ambiguity , but also the different ways the translators addressed the challenges of rendering Latin syntactical structures in a different linguistic system.

The student then created a new English translation, discussing where it distanced itself from the original and which aspects were instead retained. The comparison between two very different authorial translations proved useful to understand different strategies of addressing relevant linguistic and semantic problems, but also to help the student to find their own “translator's voice”.

<http://ugarit.ialigner.com/text.php?id=25919>

# Case study 2: Students of a third language

English

Germany is separated from the Galli , the Rhæti , and Pannonii , by the rivers Rhine and Danube ; mountain ranges , or the fear which each feels for the other , divide it from the Sarmatæ and Daci . Elsewhere ocean girds it , embracing broad peninsulas and islands of unexplored extent , where certain tribes and kingdoms are newly known to us , revealed by war . The Rhine springs from a precipitous and inaccessible height of the Rhaetian Alps , bends slightly westward , and mingles with the Northern Ocean . The Danube pours down from the gradual and gently rising slope of Mount Abnoba , and visits many nations , to force its way at last through six channels into the Pontus ; a seventh mouth is lost in marshes .

(15) 11%

( 121 ) 89% ENG - DEU

Deutsch

Ganz Germanien wird von den Galliern und Rätern und von den Pannoniern durch die Flüsse Rhein und Donau , von den Sarmaten und Daziern durch wechselseitige Furcht oder Gebirge getrennt ; das Übrige umgibt der Ozean , weite Buchten und unermessliche Inselräume umfassend , wobei man vor Kurzem von gewissen Völkerschaften und Königen erfahren hat , die der Krieg ans Licht brachte . Der Rhein , entsprungen aus dem unerreichbaren und steilen Gipfel der rätischen Alpen , mischt sich , nachdem er sich mit einer mäßigen Beugung gen Westen gewandt hat , mit dem nördlichen Ozean . Die Donau , aus dem sanften und mild herausragenden Gipfel des Schwarzwaldes herausströmend , sucht mehrere Völker auf , bis sie sich über sechs Läufe ins Schwarze Meer ergießt : die siebte Mündung wird von Sumpfen verschlungen .

( 125 ) 93% ENG - DEU

Latin

Germania omnis a Gallis Raeticisque et Pannoniis Reno et Danubio fluminibus , a Sarmatis Dacisq[ue] mutuo metu aut montibus separatur : cetera Oceanus ambit , latos sinus et insularum immensa spatia complectens , nuper cognitis quibusdam gentibus ac regibus , quos bellum aperuit . Rhenus , Raeticarum Alpium inaccesso ac praecipiiti vertice ortus , modico flexu in occidentem versus septentrionali Oceano miscetur . Danubius molli et clementer edito montis Abnoba iugo effusus plures populos adit , donec in Ponticum mare sex meatus erumpat ; septimum os paludibus hauritur .

( 125 ) 93% ENG - DEU

( 10 )

Trilingual alignment of Tacitus' *Germania*, against an English and a German translation.  
Higher matching rate between the two inflected languages than with English: comparing German and Latin resulted easier than English and Latin.

# Case study 2: Students of a third language

Ελληνικά

ΑΙσ

διανθ' ο μαρκός κανθαρίζητος χρόνος φυεὶ τὸν ἀδηλού καὶ φωνέτα κρίπτει : κούν' ἔστιν ἀλλάντον οὐδέν , ὅλας ἀλισκεταί χώ δενός δρός καὶ περισκελές φρέσει .

κάγω νέφος ὃς τὸν ἐκπρέπειον τότε ,  
μερι τούρης ός ἐθέλησθη στήρου  
πρὸς τῆρας τὴν γενεῖαν : οὐδέτε δέ νιν  
χρήσιν παρ' ἔγροις ποιεῖ τὸν θελόντα λημένι

ἄλλα εἴην πρὸς τε λοιποὺς καὶ παροικούς λευκόνες , ώς ὁν λαμψι ἀγνός εμά μηντι θεραιν εξαλύσεις έθει :  
μολύν τε ψόφον ἐνὸν ἀστιθι κίρη ,  
κρημόν τοῦς ἔγροις τούρην , ἔθιστον θελόν ,  
γιατὶς ὄρθεται ἐνὸν μὲν τὰ δρεπαὶ ,  
ἄλλα αὐτὸν νῦν Ἀΐδης τε σφιχτῶν κάτω .

ἔγω γάρ οὐ καὶ τοῦτο ἔδειμανην παρ' Ἐκτορὸς δύοτην διμονεντάτου .  
ούσια τι κενῶν ἔχοντας Ἀργείων πάρα .  
ἄλλα ἔστιν ὀληρῆς η βροτῶν παροια .  
ἔχοντας δύοτην καὶ σύνομα .

τοιαύτην τὸ λοιπὸν εἰσίμεθα μὲν θεοῖς εἰκαν , μαθησόμεθα δὲ Ατρεΐδες αὐτεῖν .

διρρότες εἰσον , οὐδὲν μηκέτεον , τι μήν ;  
και γάρ τὸ δενάν καὶ τὸ καρτερότατο τημένιοι : τοῦτο μὲν νιαροπήβες χειρῶνες ἔχωρούσιν εὐάρπηθερει :  
έξισται δὲ νικότης κύκλος τῆς λευκοπόλεως φέγγος ήμερος φλέγειν :  
δενόν τι δημο πνευμάτων ἐκούμενος στένοντα πόντον : ἐν δὲ ὁ παγκατής υπόνοιας λιπει πεθήσας , οὐδὲν οὖτε λαβον ἔχει .

中文

埃 同 斯

漫 长 数 不 尽 的 岁 月 揭 示 了 所 有 混 沌 ,  
然 后 在 它 显 现 后 , 再 次 跑 离 了 它 ,  
没 有 什 么 是 无 法 预 料 的 ,  
庄 严 的 誓 言 与 最 倔 强 的 心 灵 都 被 征 服 了 。

同 样 的 , 我 曾 经 尽 疾 艰 , 如 常 火 的 钢 铁 般 坚 硬 ,  
如 今 我 的 刀 也 被 这 个 女 人 软 化 了 ,  
我 于 心 不 忍 让 她 或 为 敌 人 埋 里 的 婴 儿 ,  
或 让 我 的 孩 子 成 为 酋 儿 。

但 是 我 走 向 沐 浴 之 处 和 海 岸 边 的 草 地 ,  
为 了 造 脱 女 神 的 羞 力 ,  
我 洗 净 了 我 的 污 罴 ,  
在 抵 达 一 个 不 被 打 扰 的 地 方 后 ,  
我 将 在 没 人 能 发 现 的 土 里 藏 起 我 的 剑 ,  
这 最 可 憎 的 武 器 。  
但 就 让 黑 夜 与 王 堂 哈 德 斯 将 它 保 存 在 地 下 。

因 为 自 从 我 从 最 大 的 手  
赫 克 托 耳 那 里 接 受 了 这 个 礼 物 ,  
我 就 再 也 没 有 从 同 伴 互 处 得 到 任 何 好 处 ,  
但 使 人 们 的 这 句 谚 语 是 真 的 ,  
敌 人 的 礼 物 并 非 礼 物 而 且 无 益 。

所 以 今 后 我 们 会 明 白 向 神 朝 屈 服 ,  
我 们 会 学 会 去 尊 敬 阿 特 柔 丝 的 儿 子 们 。

他 们 是 首 领 , 所 以 旁 人 必 须 屈 服 于 他 们 。 毋 庸 置 疑 ,  
即 使 是 赫 可 布 烈 壮 力 的  
也 要 屈 服 于 狙 戟 。  
积 雪 的 寒 冬 要 为 多 产 的 炎 夏 赢 得 位 置 ,  
光 明 代 替 了 无 尽 长 夜 的 轮 回 ,  
在 白 色 骏 马 飞 跃 的 白 日 里 闪 耀 ,  
骇 人 狂 风 的 劲 吹 供 给 息 息 之 海 入 睡 ,  
全 能 的 睡 慊 放 了 被 捆 绑 的 ,  
它 不 会 永 远 捆 绑 它 所 捆 绑 的 。

English

Ajax

All things the long and countless years first draw from darkness , and then bury from light ;  
and there is nothing which man should not expect : the dread power of oath is conquered , as  
is unyielding will .

For even I , who used to be so tremendously strong – yes , like tempered iron–feit my tongue  
’s sharp edge emasculated by this woman ’s words , and I feel the pity of leaving her a  
widow and the boy an orphan among my enemies .

But I will go to the bathing-place and the meadows by the shore so that by purging my  
defilements I may escape the heavy anger of the goddess . Then I will find some isolated spot  
, and bury this sword of mine , most hateful weapon , digging down in the earth where none  
can see . Let Night and Hades keep it underground !

For ever since I took into my hand this gift from Hector , my greatest enemy , I have gotten  
no good from the Greeks . Yes , men ’s proverb is true : the gifts of enemies are no gifts  
and bring no good .

And so hereafter I shall , first , know how to yield to the gods , and , second , learn to  
revere the Aeteadre .

They are rulers , so we must submit . How could it be otherwise ? Things of awe and might  
submit to authority . So it is that winter with its snow-covered paths gives place to fruitful  
summer ; night ’s dark orbit makes room for day with her white horse to kindle her radiance ;  
the blast of dreadful winds allows the groaning sea to rest ; and among them all , almighty  
Sleep releases the fettered sleeper , and does not hold him in a perpetual grasp .

And we men—must we not learn self-restraint ? I , at least , will learn it , since I am newly  
aware that an enemy is to be hated only as far as suits one who will in turn become a friend .  
Similarly to a friend I would wish to give only so much help and service as suits him who  
will not forever remain friendly . For the masses regard the haven of comradeship as  
treacherous .

But concerning these things it will be well . You , wife , go inside and pray to the gods that  
the desires of my heart be completed to the very end . You also , my comrades , honor my  
wishes just as she does , and command Teucer , when he comes , to take care of us , and  
to be kind to you at the same time .

Trilingual alignment of Sophocles' Ajax, aligning the Greek against the English translation by Sir R.C. Jebb (1896), and a Chinese version based on the English.

While the Chinese and the English obviously displayed a higher matching rate, the student was able to use the knowledge of both languages to individuate correspondences with the Greek, and also to establish where there was no available match for Greek words or concepts.

# Case 3: no prior knowledge of the source language

1. Individuate **single words in the source text**, such as “holy”, “honor”, or “love”, and align them with the corresponding words in a translation, **to develop an understanding of the depth of cultural concepts**, by assessing the many different ways in which the same word in the source text could be translated.
2. Compare how **specific expressions** were rendered in two translations, and then align them against the source text. The students focused on different ways of conveying morphosyntactic constructs across languages, critically approaching different translation strategies.
3. Complete bilingual or trilingual alignment of a short passage, with intensive and systematic analysis of the linguistic and expressive content, to let students assess the linguistic differences between the languages in question, and partly to demonstrate to them that they could, in fact, approach a text in a language that they did not know at all, by means of translation alignment.

# Case 3: no prior knowledge of the source language

Ελληνική

Ἐγχοτος δέ οι ἐν χερι τοῦ παιδὸς τὸν γάμον ἀπικνέεται ἐξ τὰς Σάρδις ἀνὴρ συμφορῇ ἔχόμενος καὶ οὐ καθαρὸς χεῖρας , ἐών Φύρξ μὲν γενεῖ , γένεος δὲ τοῦ βασιλήου . Παρελθὼν δὲ ὡῦτος ἐξ τὰ Κροίσου οἰκία κατὰ νόμους τοὺς ἐπιχωρίους καθαροῖσιν ἔδεστο κυρῆσαι , Κροίσος δὲ μιν ἐκάθηπε . "Εστι δέ παραπλασίη ἡ κάθαρσις τοῖσι Λυδοῖσι καὶ τοῖσι "Ἐλλησι . "Ἐπείτε δὲ τὰ νομιζόμενα ἐποίησε ὁ Κροίσος , ἐπυνθάνετο ὄκοθεν τε καὶ τις εἴη , λέγων τάδε . " "Ωνθώστε , τίς τε ἐών καὶ κόβεν τῆς Φρυγίης ἥκων ἐπίστιος μοι ἔγένεο ; Τίνα τε ἀνδρῶν ἡ γυναικῶν ἐφόνευας ; " "Ο δέ ὀμειβετό . " "Ω βασιλεῦ , Γορδίεω μὲν τοῦ Μίδεω εἰμὶ παῖς , ὀνομάζομαι δὲ Ἀδράστος , φονεύας δὲ ὀδελφεόν ἐμεωτοῦ ἀέκων πάρειμ ἔξεληλαμένος τε ὑπὸ τοῦ πατρὸς καὶ ἐστερημένος πάντων . " Κροίσος δὲ μιν ἀμειβετο τοῖσθε . " Ανδρῶν τε φίλων τυγχάνεις ἔκγονος ἐών καὶ ἐλήλυθας ἐς φίλους , ἐνθα ἀμπχανίσεις χρήματα ὡς οὐδενὸς μένων ἐν ἡμετέρου συμφορήν τε τάυτην κουφότατα φέρων κερδανέεις πλείστον . "

English

Then while he was engaged about the marriage of his son , there came to Sardis a man under a misfortune and with hands not clean , a Phrygian by birth and of the royal house . This man came to the house of Croesus , and according to the customs which prevail in that land made request that he might have cleansing ; and Croesus gave him cleansing : now the manner of cleansing among the Lydians is the same almost as that which the Hellenes use . So when Croesus had done that which was customary , he asked of him whence he came and who he was , saying as follows : " Man , who art thou , and from what region of Phrygia didst thou come to sit upon my hearth ? And whom of men or women didst thou slay ? " And he replied : " O king , I am the son of Gordias , the son of Midas , and I am called Adrastos ; and I slew my own brother against my will , and therefore am I here , having been driven forth by my father and deprived of all that I had . " And Croesus answered thus : " Thou art , as it chances , the offshoot of men who are our friends and thou hast come to friends , among whom thou shalt want of nothing so long as thou shalt remain in our land : and thou wilt find it most for thy profit to bear this misfortune as lightly as may be . " So he had his abode with Croesus .

Latin

Interim , dum nuptias filii parat Croesus , advenit Sardes vir calamitate obstricetus , cui manus non erant purae , natione Phryx , regio de genere . Qui quum ad aedes Croesi venisset , orassetque ut ritu patrio expiari sibi contingaret , expiavit eum Croesus : est autem modus expiandi apud Lydos similis ei , quo Graeci utuntur . Peractis legitimis , percontatur Croesus hominem , unde veniret , quisve esset , his usus verbis : " Quis tu es , homo , et quoniam e Phrygiae loco veniens , ad meos lares supplex te receperisti ? Cui ille respondit : " O rex , Gordiae sum filius , Midae nepos ; est autem nomen mihi Adrasto . Fratrem meum occidi invitus ; unde ejectus a patre adsum , et rebus omnibus destitutus . Tum Croesus : " Ex viris amicis , " inquit , " oriundus es , et ad amicos venisti , ubi nullius rei indigebis quoad apud nos manseris . Istam autem calamitatem quam poteris levissime ferendo , plurimum lucri feceris , Ita ille in Croesi aedibus vitam agebat . "

Trilingual alignment of Herodotus' *Histories*, Ancient Greek, English and Latin. Knowledge of Greek and English was used to leverage the unfamiliarity with a text in Latin.

The student was able to recognize **cognate words** and match them with the corresponding Greek, and to use the knowledge of inflected languages to compare parts of speech between Latin and Greek (for example, to identify the accusative *fratre-m* based on the similar ending with the Ionic Greek ἀδελφεό-v).

# Case 3: no prior knowledge of the source language

The screenshot shows a trilingual alignment interface with three columns: Ancient Greek (left), English (middle), and Chinese (right). Each column contains a text snippet and a progress bar at the bottom.

**Ancient Greek:**

τό σοφόν δ' ού σοφία  
τό τε μὴ θητὰ φρονεῖν .  
βραχὺς αἰών : ἐπὶ τούτῳ  
δέ τις ἀν μεγάλα διώκων  
τὰ παρόντ' οὐχί φέροι .

**English:**

But being clever isn't wisdom .  
And thinking **deeply about things**  
isn't **suitable for mortal men** .  
**Our life is brief** – that's why the man **who chases**  
**greatness**  
fails to grasp what's near at hand .

**Chinese:**

但 显 得 聪 明 不 是 智 慧 。  
且 凡 人 不 适 合  
对 事 物 追 究 过 深  
人 生 稍 纵 即 逝 - 因 此 追 求 伟 大 之 人  
会 让 睡 手 可 得 之 物 溜 走 。

Progress bars at the bottom indicate the percentage of alignment for each row:

- (8) 31% GRC
- (18) 69% GRC - ENG
- (25) 68% GRC - ENG
- (12) 32% ENG
- (25) 68% GRC - ENG
- (12) 32% ENG

<http://ugarit.ialigner.com/text.php?id=2134>

Trilingual alignment of Euripides' *Bacchae*, comparing the English translation by I. Johnson (<http://jelks.nu/libri/classics/bacchae.html>) and a manually revised Chinese version of it generated by Google Translate (as there were no available Chinese translations of *Bacchae*).

As a **Chinese native speaker**, the student used that language as a bridge to establish accurate correspondences between the English and the Ancient Greek words that had never been seen before. The result was an investigation into the **meaning of Ancient Greek concepts** through a meaningful comparison with similar Chinese terms, often with an emphasis on imperfect **grammatical and semantic correspondences** between the Greek and its translation (such as τό σοφόν, “cleverness”, translated as “being clever”).

# Applications and exercise

# Alignment Levels

- **Document Alignment:** documents are aligned if they contain similar contents. meta-information, file size, Mutual Information and Maximum-Entropy based models, etc.
- **Paragraph-level Alignment:**  
Uses similarity measures and dynamic programming methods to perform the alignment at paragraph level.

# Alignment Levels

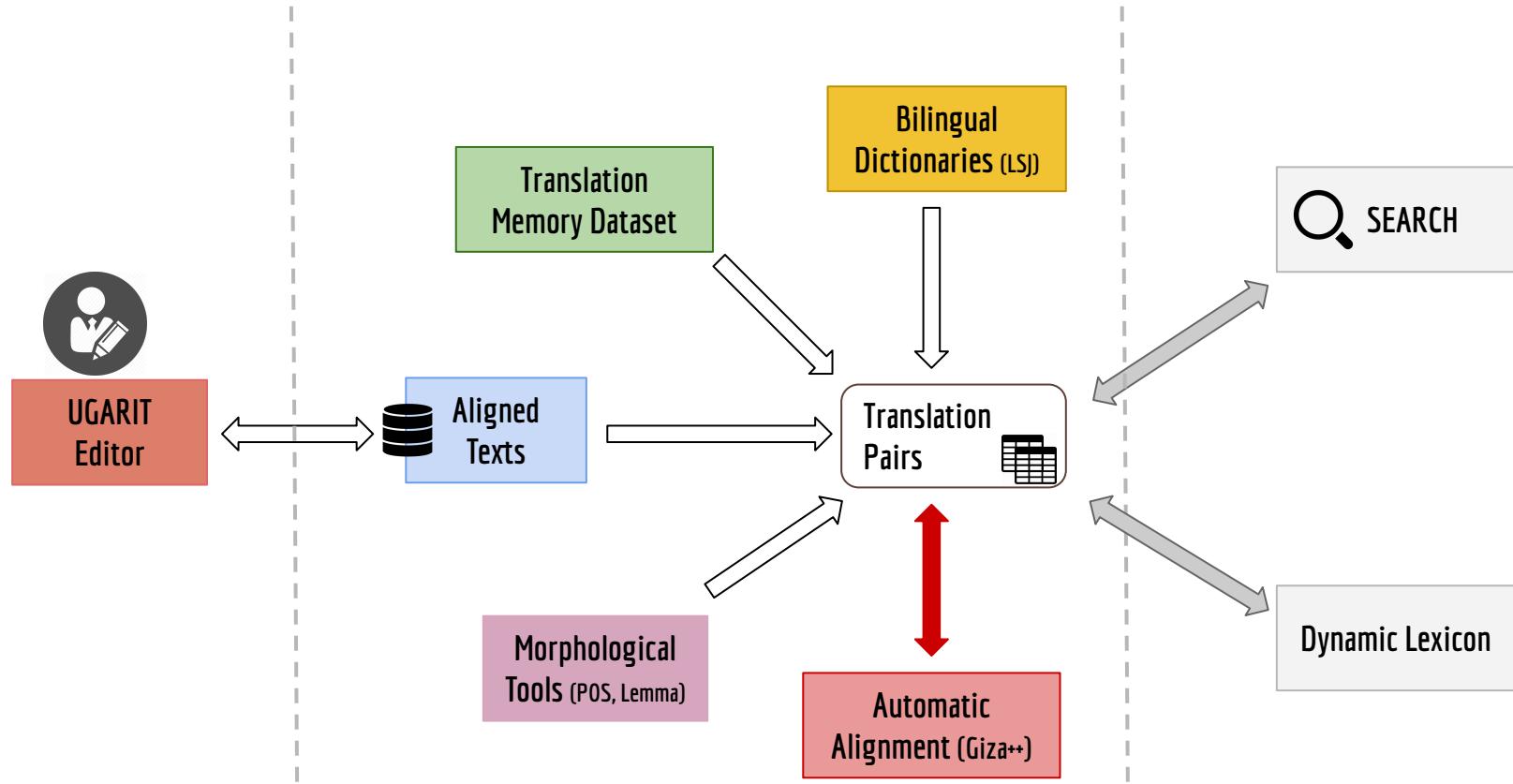
- Sentence-Level Alignment

- **Length-based models:**

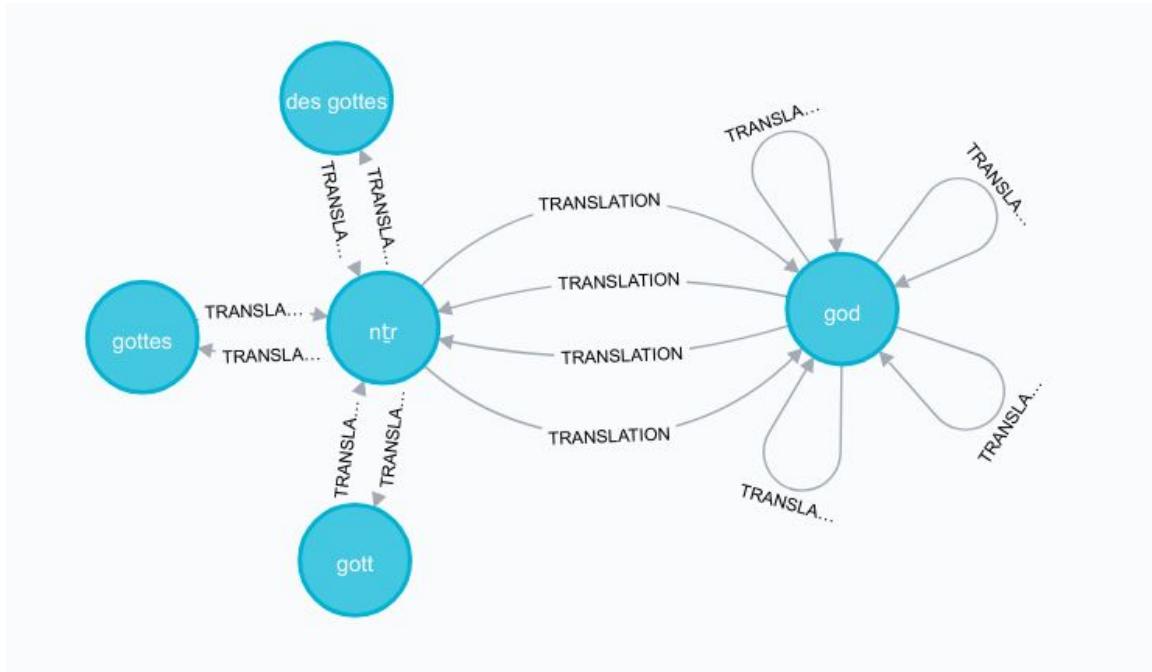
longer sentences in one language tend to be translated into longer sentences in the other language, and that shorter sentences tend to be translated into shorter sentences.

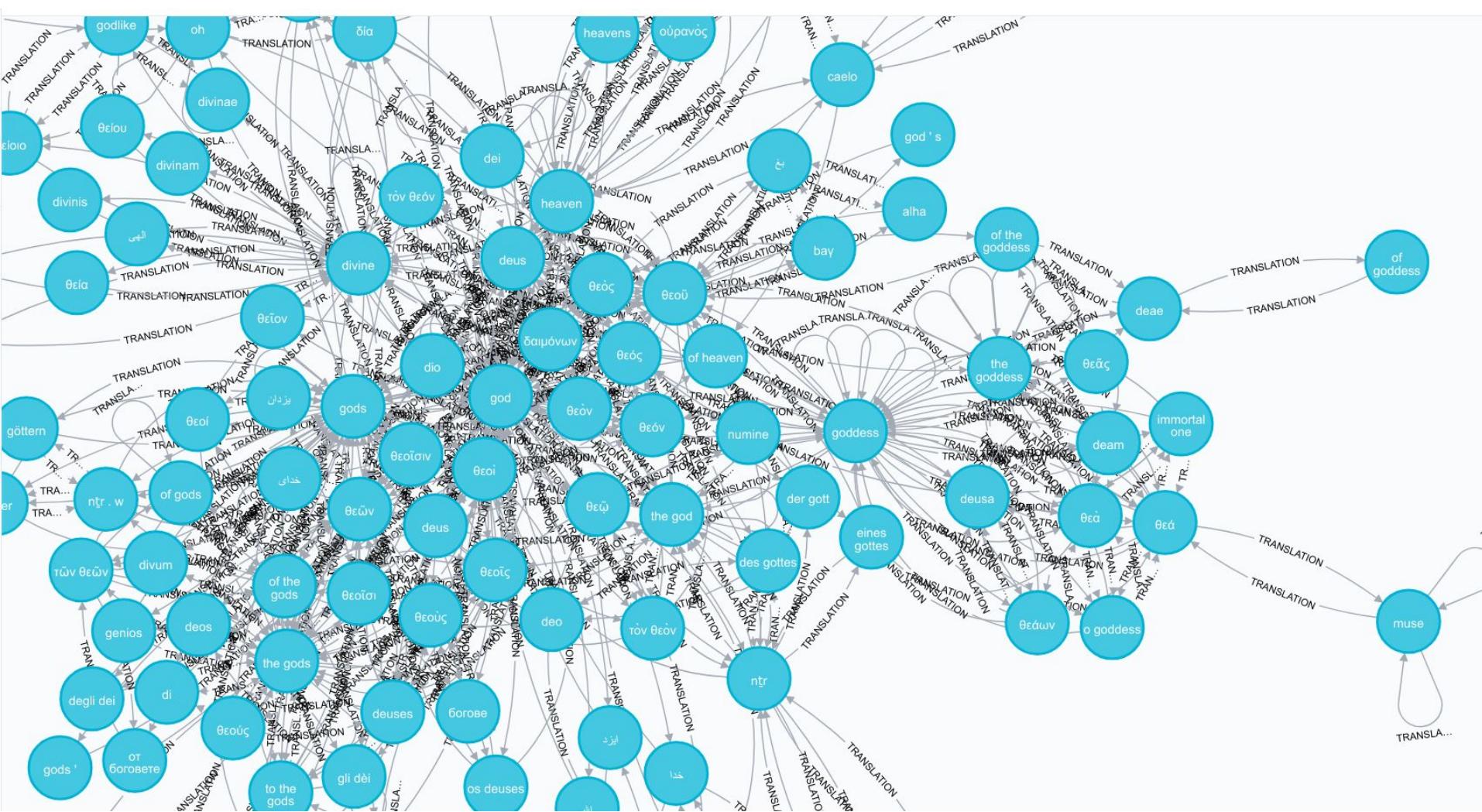
- **Word-correspondence methods:**

a pair of sentences containing an aligned pair of words must themselves be aligned. It follows that a partial alignment on the word level could induce a much more complete alignment on the sentence level.



# Translation Pairs as Graph

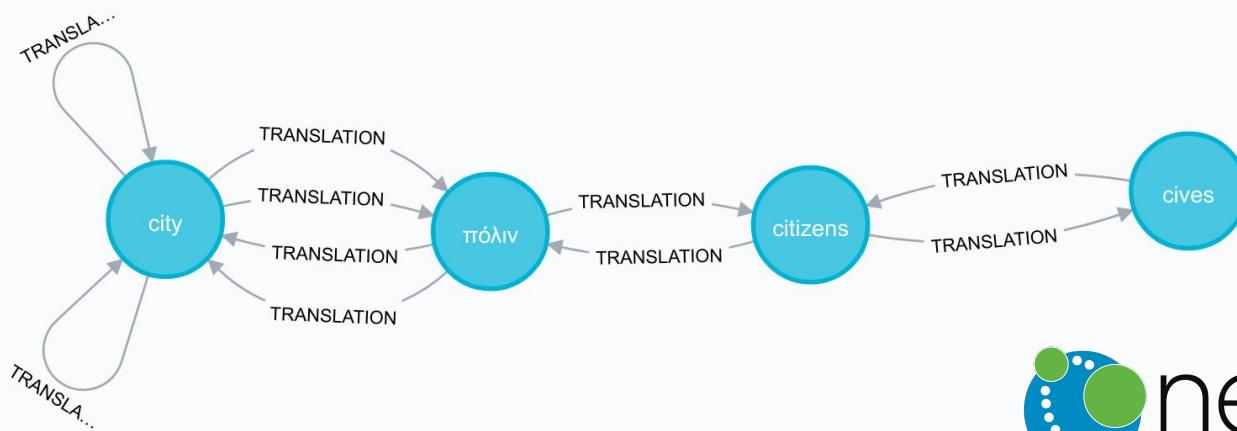




```
$ match pa=shortestPath( (n{text:"city", lang:"eng"})-[:TRANSLATION*1..5]→(p{text:"cives", lang:"lat"}))    return pa
```

\*(4) phrase(4)

\*(10) TRANSLATION(10)



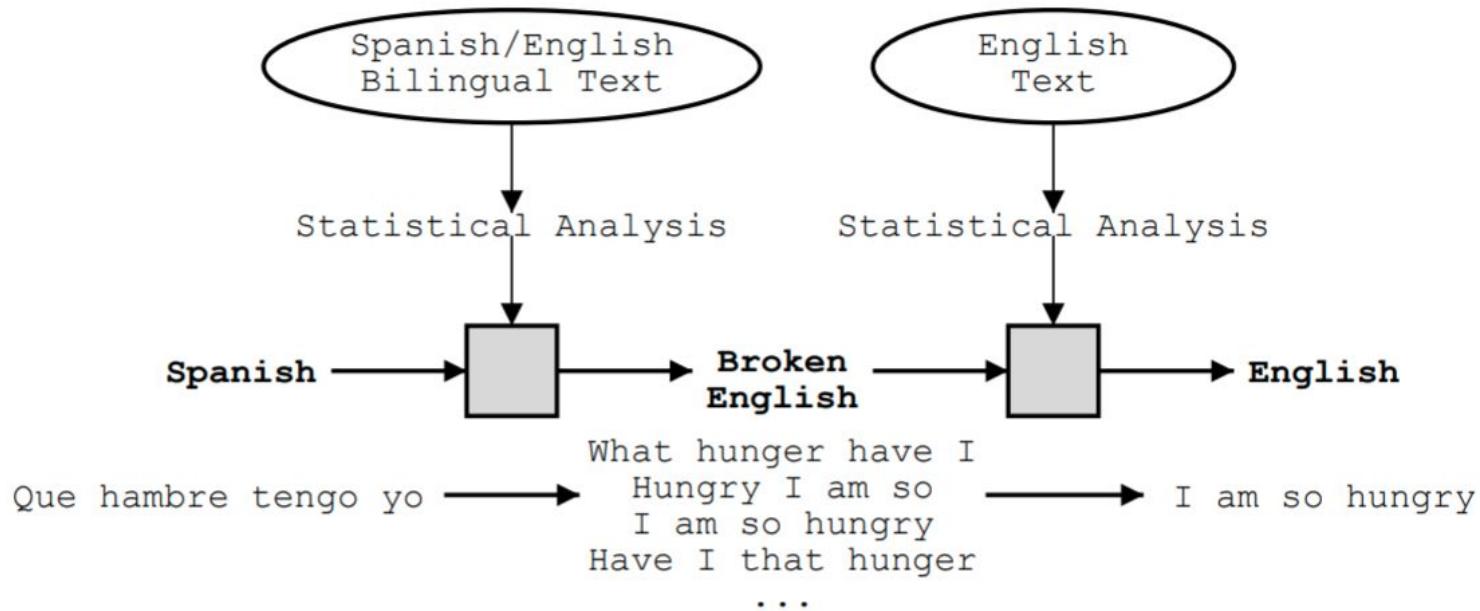
# Translations Cloud

Level: 3

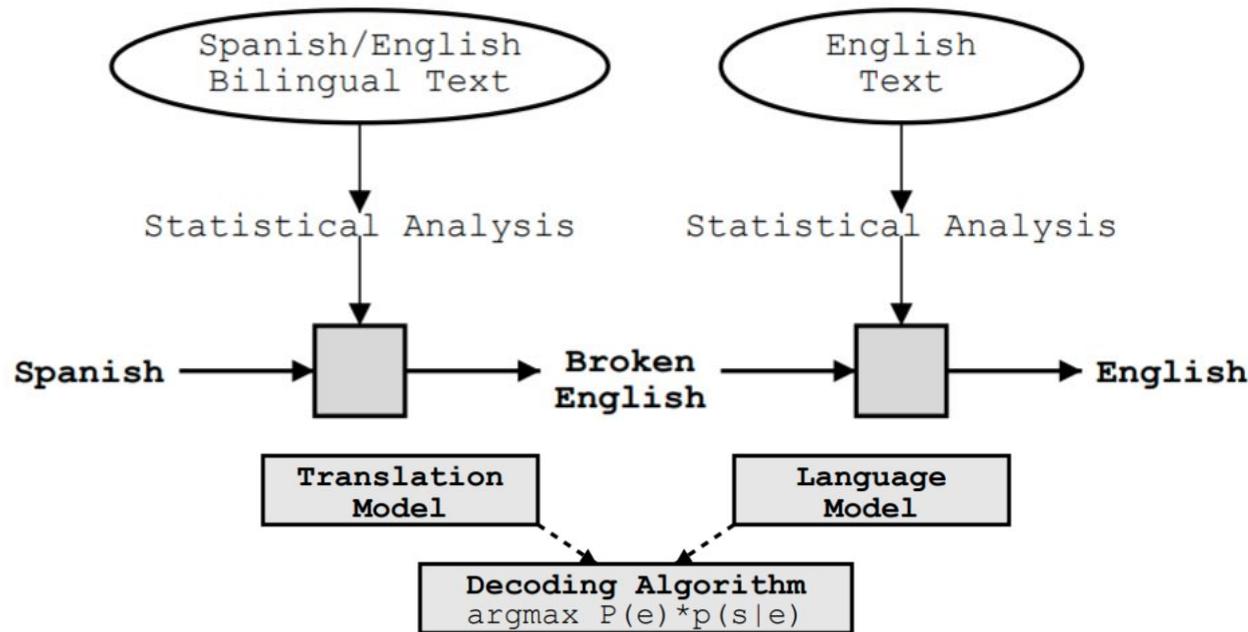
τοῦτον τούτον  
αὐτὸν κατ' περὶ πολίτην n n<sup>7</sup> αὐτὸν  
εσος que δημοσίων ἀθηναῖος civium n<sup>7</sup>  
n<sup>7</sup> χώραν πολίταις ἀρχῆς  
n<sup>7</sup> οππόδιο citizens ἀθάναις χώρα χώρας r n<sup>7</sup> على  
politeias حل t ٤٣ public in a ἐνθάδε die ἐπ' ٤٣ اَسْتَوْنَ τούς  
nti iw=w δημοσία شہر ٤٣ city - state citadel p3 stadt  
πτόλιν ἀνὰ O quae πόλεώς das civitatem country دیار FAR  
αὐτῶν λέγουσιν p<sup>7</sup> quae crepue una urbem ntj اَثَηْنَوْنَ n<sup>7</sup> p<sup>7</sup>  
κοινῶν made the government , urbe athenas ἄστυ اَسْتَوْنَ  
περὶ τοῦ πόλις città ἐπ' πόλισμι urbis πτόλιος la populi της  
tautia ἀγρῶν t<sup>7</sup> ἔς πόλεος the city of πόλεως ἐφ' πόλει n<sup>7</sup> . w  
qui λέγεται a city πόλιν =w acropolis града pr civitate δημοσίους  
άστεως attica tō τὴν cives тѣн πόλιν oppidum πόλεις attica πολιτῶν  
на града ἀθήνας ٤٣ πόλεως δημοσίαν ἀθηναῖος τὸν ἀττικῆς  
super ρήτορεν lo que διστοῖς πόλεων πόλεως δημοσίους  
iw=w αἰκροπόλει nti urbs γαλανάς πόλισμα κοινὸν da ἀθηναῖοι n<sup>7</sup> ab urbe  
πόλεων of city che civitas αἱ athens publica δημόσιον di cui  
quos δημόσια μν̄ τὸν town γῆς πτολειέθρον ρήτορα πολις τὸ  
αὐτὸν ἀθηναῖων ἀθήναις upon γῆ the town  
τὴν ἀκρόπολιν πολίτας πατρίδα ἐγχωρίων ἀπὸ<sup>7</sup>  
n<sup>7</sup>. w ἀρχὴν τὰ ἀστοῖσιν κατὰ  
έκ της

- city (ENG)
- πόλις (GRC)
- country (ENG)
- دیار (FAR)

# Statistical Machine Translation



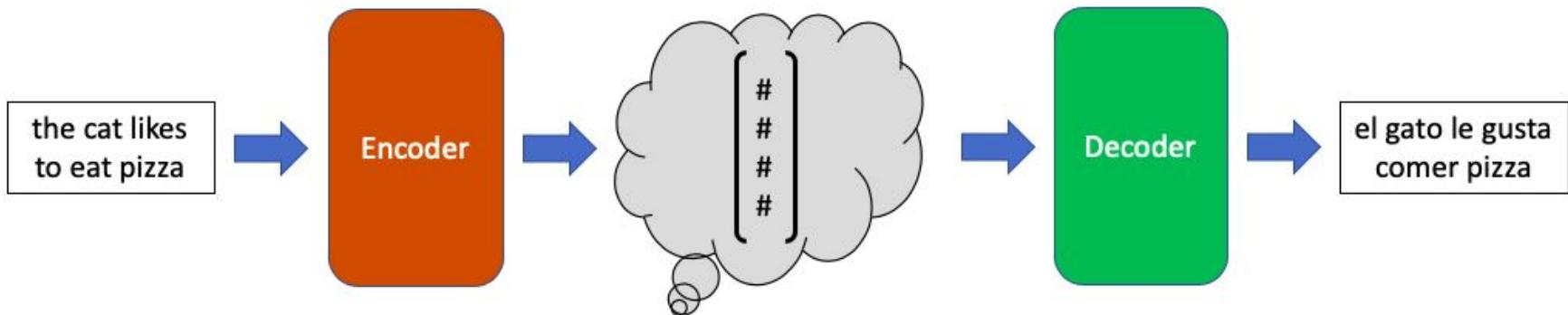
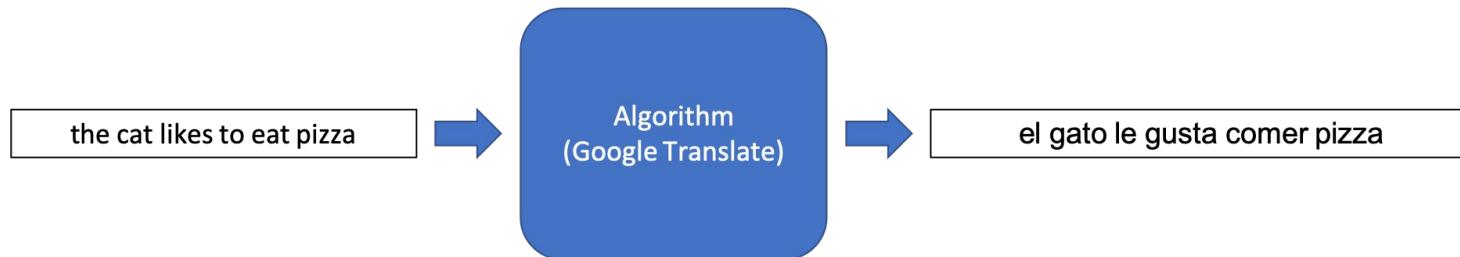
# Statistical Machine Translation



# Software

- **GIZA++** <http://www.fjoch.com/GIZA++.html>
- **Moses** <http://statmt.org/moses/>
- **UCAM-SMT** (The Cambridge Statistical Machine Translation system)  
<http://ucam-smt.github.io/>
- **Stanford Phrasal** (Phrase-Based Translation System)  
<https://nlp.stanford.edu/phrasal/>

# Neural Machine Translation



# IBM Models

IBM alignment models are a sequence of increasingly complex models used in SMT to train a translation model and an alignment model.

The original work on statistical machine translation at IBM proposed five models:

- **Model 1:** lexical translation
- **Model 2:** additional absolute alignment model
- **Model 3:** extra fertility model
- **Model 4:** added relative alignment model
- **Model 5:** fixed deficiency problem.

# IBM Model 1

Generative model: break up translation process into smaller steps

- IBM Model 1 only uses lexical translation

Translation probability:

- for a foreign sentence  $\mathbf{f} = (f_1, \dots, f_{l_f})$  of length  $l_f$
- to an English sentence  $\mathbf{e} = (e_1, \dots, e_{l_e})$  of length  $l_e$
- with an alignment of each English word  $e_j$  to a foreign word  $f_i$  according to the alignment function  $a : j \rightarrow i$
- Parameter  $\epsilon$  is a normalization constant

$$p(\mathbf{e}, a | \mathbf{f}) = \frac{\epsilon}{(l_f + 1)^{l_e}} \prod_{j=1}^{l_e} t(e_j | f_{a(j)})$$

# Example

das	
<i>e</i>	$t(e f)$
the	0.7
that	0.15
which	0.075
who	0.05
this	0.025

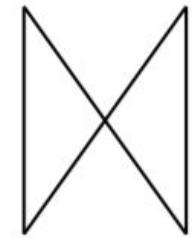
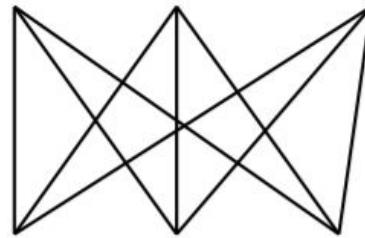
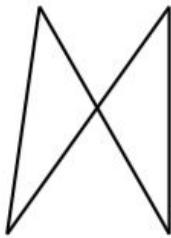
Haus	
<i>e</i>	$t(e f)$
house	0.8
building	0.16
home	0.02
household	0.015
shell	0.005

ist	
<i>e</i>	$t(e f)$
is	0.8
's	0.16
exists	0.02
has	0.015
are	0.005

klein	
<i>e</i>	$t(e f)$
small	0.4
little	0.4
short	0.1
minor	0.06
petty	0.04

$$\begin{aligned}
 p(e, a|f) &= \frac{\epsilon}{4^3} \times t(\text{the}| \text{das}) \times t(\text{house}| \text{Haus}) \times t(\text{is}| \text{ist}) \times t(\text{small}| \text{klein}) \\
 &= \frac{\epsilon}{4^3} \times 0.7 \times 0.8 \times 0.8 \times 0.4 \\
 &= 0.0028\epsilon
 \end{aligned}$$

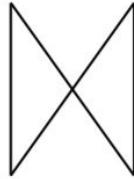
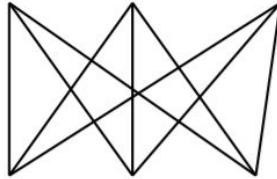
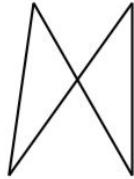
... la maison ... la maison blue ... la fleur ...



... the house ... the blue house ... the flower ...

	the	house
la	1	1
maison	1	1

... la maison ... la maison blue ... la fleur ...



... the house ... the blue house ... the flower ...

	the	blue	house
la	1	1	1
maison	1	1	1
bleu	1	1	1

	the	flower
la	1	1
fleur	1	1

	the	house	blue	flower	
la	3	2	1	1	7
maison					
bleu					
fleur					

	the	house
la	1	1
maison	1	1

	the	blue	house
la	1	1	1
maison	1	1	1
bleu	1	1	1

	the	flower
la	1	1
fleur	1	1

	the	house	blue	flower	
la	3	2	1	1	7
maison	2	2	1	0	5
bleu					
fleur					

	the	house
la	1	1
maison	1	1

	the	blue	house
la	1	1	1
maison	1	1	1
bleu	1	1	1

	the	flower
la	1	1
fleur	1	1

	the	house	blue	flower	
la	3	2	1	1	7
maison	2	2	1	0	5
bleu	1	1	1	0	3
fleur					

	the	house
la	1	1
maison	1	1

	the	blue	house
la	1	1	1
maison	1	1	1
bleu	1	1	1

	the	flower
la	1	1
fleur	1	1

	the	house	blue	flower	
la	3	2	1	1	7
maison	2	2	1	0	5
bleu	1	1	1	0	3
fleur	1	0	0	1	2

	the	house
la	3/7	2/7
maison	2/5	2/5

	the	blue	house
la	3/7	1/7	2/7
maison	2/5	1/5	2/5
bleu	1/3	1/3	1/3

	the	flower
la	3/7	1/7
fleur	1/2	1/2

	the	house	blue	flower	
la	3/7	2/7	1/7	1/7	1
maison	2/5	2/5	1/5	0	1
bleu	1/3	1/3	1/3	0	1
fleur	1/2	0	0	1/2	1

	the	house
la	3/7	2/7
maison	2/5	2/5

	the	house
la	0,6	0,4
maison	0,5	0,5

	the	blue	house
la	3/7	1/7	2/7
maison	2/5	1/5	2/5
bleu	1/3	1/3	1/3

	the	blue	house
la	0,5	0,16	0,33
maison	0,4	0,2	0,4
bleu	0,33	0,33	0,33

	the	flower
la	3/7	1/7
fleur	1/2	1/2

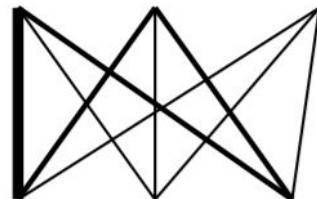
	the	flower
la	0,75	0,25
fleur	0,5	0,5

... la maison ... la maison blue ... la fleur ...

	the	house
la	0,6	0,4
maison	0,5	0,5



... the house ...



... the blue house ...



... the flower ...

	the	blue	house
la	0,5	0,16	0,33
maison	0,4	0,2	0,4
bleu	0,33	0,33	0,33

	the	flower
la	0,75	0,25
fleur	0,5	0,5

	the	house	blue	flower	
la	1,85	0,73	0,16	0,25	3
maison	0,9	0,9	0,2	0	2
bleu	0,33	0,33	0,33	0	1
fleur	0,5	0	0	0,5	1

	the	house
la	0,6	0,4
maison	0,5	0,5

	the	blue	house
la	0,5	0,16	0,33
maison	0,4	0,2	0,4
bleu	0,33	0,33	0,33

	the	flower
la	0,75	0,25
fleur	0,5	0,5

	the	house	blue	flower	
la	3/7	2/7	1/7	1/7	1
maison	2/5	2/5	1/5	0	1
bleu	1/3	1/3	1/3	0	1
fleur	1/2	0	0	1/2	1

	the	house	blue	flower	
la	0,61	0,24	0,05	0,08	1
maison	0,45	0,45	0,1	0	1
bleu	0,33	0,33	0,33	0	1
fleur	0,5	0	0	0,5	1

	the	house	blue	flower	
la	0,43	0,29	0,14	0,14	1
maison	0,4	0,4	0,2	0	1
bleu	0,33	0,33	0,33	0	1
fleur	0,5	0	0	0,5	1

	the	house	blue	flower	
la	0,61	0,24	0,05	0,08	1
maison	0,45	0,45	0,1	0	1
bleu	0,33	0,33	0,33	0	1
fleur	0,5	0	0	0,5	1

	the	house
la	0,61	0,24
maison	0,45	0,45

## NORMALIZATION

	the	blue	house
la	0,61	0,05	0,24
maison	0,45	0,1	0,45
bleu	0,33	0,33	0,33

	the	flower
la	0,61	0,08
fleur	0,5	0,5

	the	house	blue	flower	
la	0,61	0,24	0,05	0,08	1
maison	0,45	0,45	0,1	0	1
bleu	0,33	0,33	0,33	0	1
fleur	0,5	0	0	0,5	1

	the	house
la	0,72	0,28
maison	0,5	0,5

## NORMALIZATION

	the	blue	house
la	0,67	0,05	0,27
maison	0,45	0,1	0,45
bleu	0,33	0,33	0,33

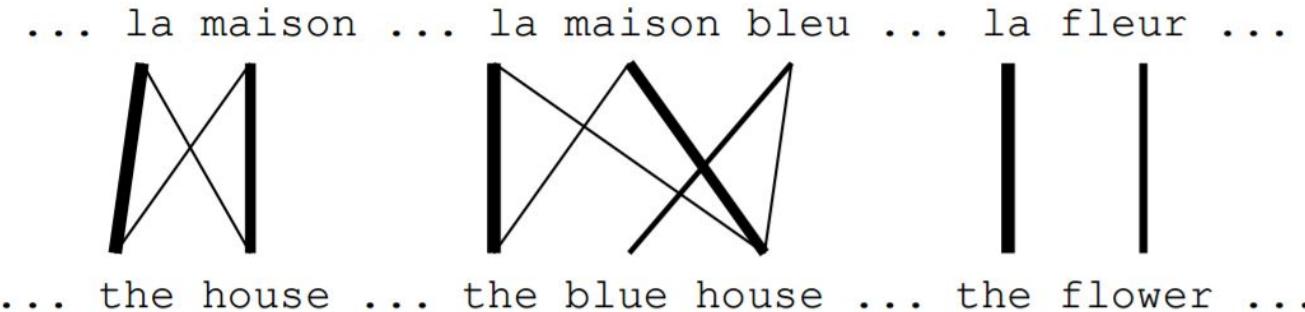
	the	flower
la	0,88	0,12
fleur	0,5	0,5

	the	house	blue	flower	
la	2,27	0,55	0,05	0,12	3
maison	0,95	0,95	0,1	0	2
bleu	0,33	0,33	0,33	0	1
fleur	0,5	0	0	0,5	1

	the	house	blue	flower	
la	0,61	0,24	0,05	0,08	1
maison	0,45	0,45	0,1	0	1
bleu	0,33	0,33	0,33	0	1
fleur	0,5	0	0	0,5	1

	the	house	blue	flower	
la	0,76	0,18	0,02	0,04	1
maison	0,47	0,47	0,05	0	1
bleu	0,33	0,33	0,33	0	1
fleur	0,5	0	0	0,5	1

	the	house
la	0,76	0,18
maison	0,47	0,47



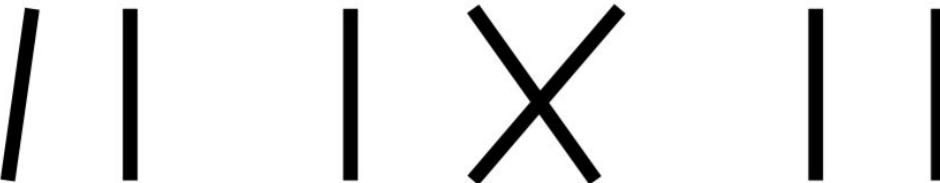
	the	blue	house
la	0,76	0,02	0,18
maison	0,47	0,05	0,47
bleu	0,33	0,33	0,33

	the	flower
la	0,76	0,04
fleur	0,5	0,5

	the	house	blue	flower	
la	0,76	0,18	0,02	0,04	1
maison	0,47	0,47	0,05	0	1
bleu	0,33	0,33	0,33	0	1
fleur	0,5	0	0	0,5	1

	the	house
la	0,81	0,19
maison	0,5	0,5

... la maison ... la maison bleu ... la fleur ...



... the house ... the blue house ... the flower ...

	the	blue	house
la	0,79	0,02	0,19
maison	0,47	0,05	0,47
bleu	0,33	0,33	0,33

	the	flower
la	0,95	0,05
fleur	0,5	0,5

	the	house	blue	flower	
la	0,85	0,12	0,01	0,02	1
maison	0,485	0,485	0,03	0	1
bleu	0,33	0,33	0,33	0	1
fleur	0,5	0	0	0,5	1

# Exercise: Low Stakes

# Align a language you don't know

- Go on Ugarit and create a bilingual alignment of a parallel corpus of your choice (or use our [provided corpus of the Bible](#)). Choose two languages that you are familiar with and focus on the differences across translation:
  - what words align perfectly?
  - What words align imperfectly, or not at all?
  - What words are missing across the two texts?
  - What is the overall percentage of matches?
- After you have completed the bilingual alignment, choose a parallel text in a third language that you do not know and perform a trilingual alignment. See how much of the third language you can align, by using the two other languages as an aid for better understanding.

Herodotus, The Histories  
Maryam Foradi /

Created on 2018-06-05 08:27:06 Modified on 2018-06-05 08:33:09 Translated by Godley, Alfred Denis Aligned by Maryam Foradi

Ελληνική English فارسی

urn:cts:greekLit:tlg0016.tlg001.perseus-grc2.1.1  
urn:cts:greekLit:tlg0016.tlg001.perseus-grc2.1.1 histories

این نمای تاریخ هرودوت هالیکارناسی است که نه آنچه زاده اندیمی است در کثر زمان رنگ باخته شود نه کارهای بزرگ و خارق العادهای که برخی از یک سو توسط پیونانهای و برخی توسط پیربرها به نمایش درآمده و پیوناونشان شوید و دیگر آنچه باعث چنگ با دیگران شد .

This is the display of the inquiry of Herodotus of Halicarnassus , so that things done by man not be forgotten in time , and that great and marvelous deeds , some displayed by the Hellenes , some by the barbarians , not lose their glory , including among others what was the cause of their waging war on each other .

( 9 ) 20% GRC ( 36 ) 80% GRC - ENG ( 46 ) 74% GRC - ENG ( 16 ) 26% ENG ( 46 ) 74% GRC - ENG ( 16 ) 26% ENG

# Exercise: High Stakes