



Translation Alignment

Theory, Technical Development, Application

Tariq Yousef, Saman Rezaie, Gabriel Bordad, Maryam Foradi



Translation: History

Translation is taking the meaning from one language and transferring it to another one

Multilingual Inscriptions of the ancient world shows the age of translation

There are different debates in the history of translation studies



Translation: different approaches

Philosophical Approach, 1920s: deconstruction and abusive fidelity

Walter Benjamin: translation serves to express the central reciprocal relationship between languages

Machine translation, 1950s: producing automatic translation

Analysis the product and process of translation, 1950s

Jakobson and the concept of equivalence, 1958: interlingual, intralingual and intersemiotic



Translation: different approaches

1960s: more focus on equivalence

Formal and dynamic equivalence: based on Chomsky's generative transformational grammar by Nida

1972: translation has been recognized as an independent discipline by Holmes

1977: emphasizing on text and discourse

Skopos theory by Hans Vermeer shifted the attention from equivalence to purpose of translation



Translation: different approaches

Major development in translation studies by Snell-Hornby: Cultural Turn

Other models and approaches: descriptive translation studies, translational action, cognitive models, post-colonial translation studies, corpus-based translation studies



Translation: different types

General translation: general texts, everyday speeches etc.

Legal translation: contracts, legal documents

Commercial translation: advertisements, business texts

Technical translation: manuals, handouts

Literary translation: literature, poetry



Literary translation

Is the poetry translatable?

Poetry includes issues such as:

Equivalence, idiomatic meaning, metaphorical expressions, rhyme and rhythm of the poem, etc.

Some believe that style should be sacrificed for the meaning and for conveying the message

Mona Baker discusses the transability of poems



Literary translation: viewpoints

Translation of poetry should be poetry

Poetry is cryptic and ambiguous and exhibits special relation between form and meaning

Translator of poems needs to have critical as well as writing abilities

Some believe that translator should create a new version

others state that translator should preserve the original text



Technical translation

Consistency is important

Repetition is usual

Manuals for different models of a device are mostly the same

Computer-aided translation is very useful



Computer-aided translation

CAT-Tools are widely used (commercial and free tools)

They work with translation memories

Translation memory is a sentence databank in form of language pairs

Source and Target translation units (segments) are aligned automatically during translation

A sentence is not always the translation unit



Application of translation memory

No need to translate repeated sentences, keep the translation consistent, although translated by different translators

Bilingual concordance search is possible

Translator finds the meaning of the terms in the context

It is still not a dynamic lexicon

Translator doesn't exactly know, which word is the translation of the searched word



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



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.



Word-level Alignment

- Word alignment is typically done after sentence alignment has already been done.
- 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.



There are different types of alignment

—



Intra-Language alignment

- alignment of texts in the same language

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	.

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

Թաղազս թէ որպէս կայսրն Վաղէս հալածեալ զամենայն ուղղափառ հաւատացեալս ,և կամ որպէս կամեցաւ յանդիմանութիւն առնել կայսրն ի մէջ հշմարտութեանն հաւատացելոց ,և ի մէջ ծախող աղանդոյն արիանոսաց .

(29) 100% HYE - DEU

DEU

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

(35) 100% HYE - DEU

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

Manual alignment

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

The UGARIT Translation Alignment Editor is a tool for manual cross-language alignment (of texts in different languages)

(www.ugarit.ialigner.com)

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

Pairwise alignment

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

<ul style="list-style-type: none">• In the beginning the Word already existed .• In the beginning was the Word ,									
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”)

- εἶτα Δαμάστης ὁ Κιτιεὺς τὰ πλείστα ἐκ τῶν Ἑκαταίου μεταγράψας περίπλουν ἔγραψεν.
- εἶτα Δαμάστης ὁ Σιγιεὺς τὰ πλείστα ἐκ τῶν Ἑκαταίου μεταγράψας περίπλουν ἔγραψεν·
- εἶτα Δαμάστης ὁ Κητιεὺς τὰ πλείστα ἐκ τῶν Ἑκαταίου μεταγράψας περίπλουν ἔγραψεν·
- εἶτα δαμάστης ὁ κητιεὺς τὰ πλείστα ἐκ τῶν ἑκαταίου μεταγράψας, περίπλουν ἔγραψεν·

εἶτα	Δαμάσσης	ὁ	Κιτιεύς	τὰ	πλείστα	ἐκ	τῶν	Ἑκαταίου	μεταγράψας	περίπλουν	ἔγραψεν	.	
εἶτα	Δαμάσσης	ὁ	Σιγίειός	τὰ	πλείστα	ἐκ	τῶν	Ἑκαταίου	μεταγράψας	περίπλουν	ἔγραψεν	·	
εἶτα	Δαμάσσης	ὁ	Κητιεύς	τὰ	πλείστα	ἐκ	τῶν	Ἑκαταίου	μεταγράψας	περίπλουν	ἔγραψεν	·	
εἶτα	δαμάσσης	ὁ	κητιεύς	τὰ	πλείστα	ἐκ	τῶν	ἐκαταίου	μεταγράψας	,	περίπλουν	ἔγραψεν	·

Four texts aligned with *iAligner*

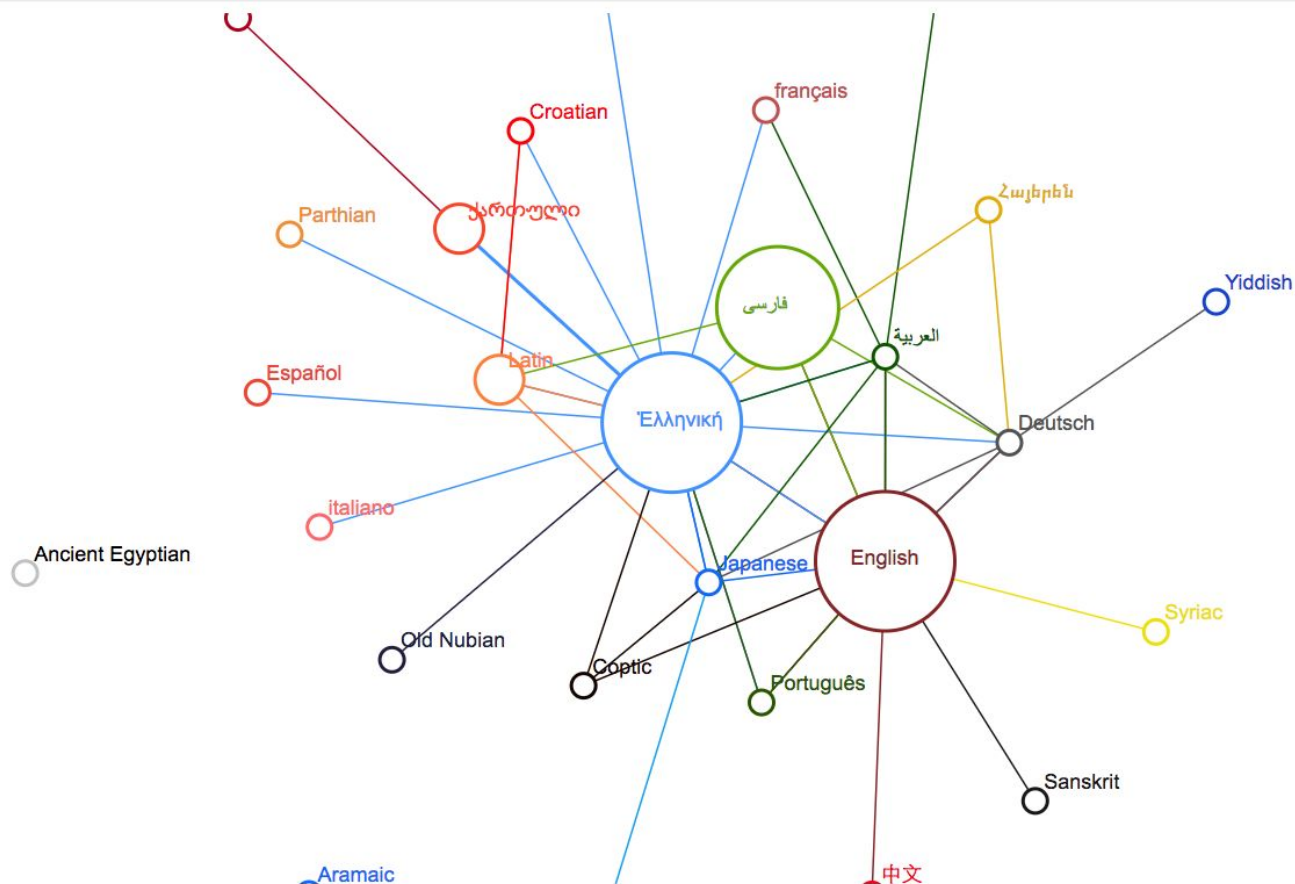
- | | | | | | | | | | | | | | |
|------|----------|---|----------|----|---------|----|-----|----------|------------|-----------|-----------|---------|---|
| εἶτα | Δαμάστης | ὁ | Κιτιεύς | τὰ | πλείστα | ἐκ | τῶν | Ἑκαταίου | μεταγράψας | περίπλουι | ἔγραψεν | . | |
| εἶτα | Δαμάστης | ὁ | Σιγίειος | τὰ | πλείστα | ἐκ | τῶν | Ἑκαταίου | μεταγράψας | περίπλουι | ἔγραψεν | . | |
| εἶτα | Δαμάστης | ὁ | Κητιεύς | τὰ | πλείστα | ἐκ | τῶν | Ἑκαταίου | μεταγράψας | περίπλουι | ἔγραψεν | . | |
| εἶτα | δαμάστης | ὁ | κητιεύς | τὰ | πλείστα | ἐκ | τῶν | ἐκαταίου | μεταγράψας | , | περίπλουι | ἔγραψεν | . |

- | | | | | | | | | | | | | |
|-------|------------|-----|---------|-----|-------|-------|-----|-----------|-----|-----------|-----------------|---|
| έξηξη | Δημόκριτος | καί | Εὐδοξος | καί | ἄλλοι | τινὲς | γῆς | περιόδους | καί | περίπλους | ἐπραγματεύσαντο | . |
| έξηξη | Δημόκριτος | καί | Εὐδοξος | καί | ἄλλοι | τινὲς | γῆς | περιόδους | καί | περίπλους | ἐπραγματεύσαντο | . |
| έξηξη | Δημόκριτος | καί | Εὐδοξος | καί | ἄλλοι | τινὲς | γῆς | περιόδους | καί | περίπλους | ἐπραγματεύσαντο | . |
| έξηξη | δημόκριτος | καί | εὐδοξος | καί | ἄλλοι | τινὲς | γῆς | περιόδους | καί | περίπλους | ἐπραγματεύσαντο | . |



What is **UGARIT** ?

- Web-based tool for manual translations alignment.
- The first aim was to generate accurate training dataset for our automatic word alignment system.
- Training datasets will be open and downloadable in different formats.
- Crowdsourcing project (cumulative results).
- Ugarit serves also as reading environment and learning tool.



Languages

28

Texts

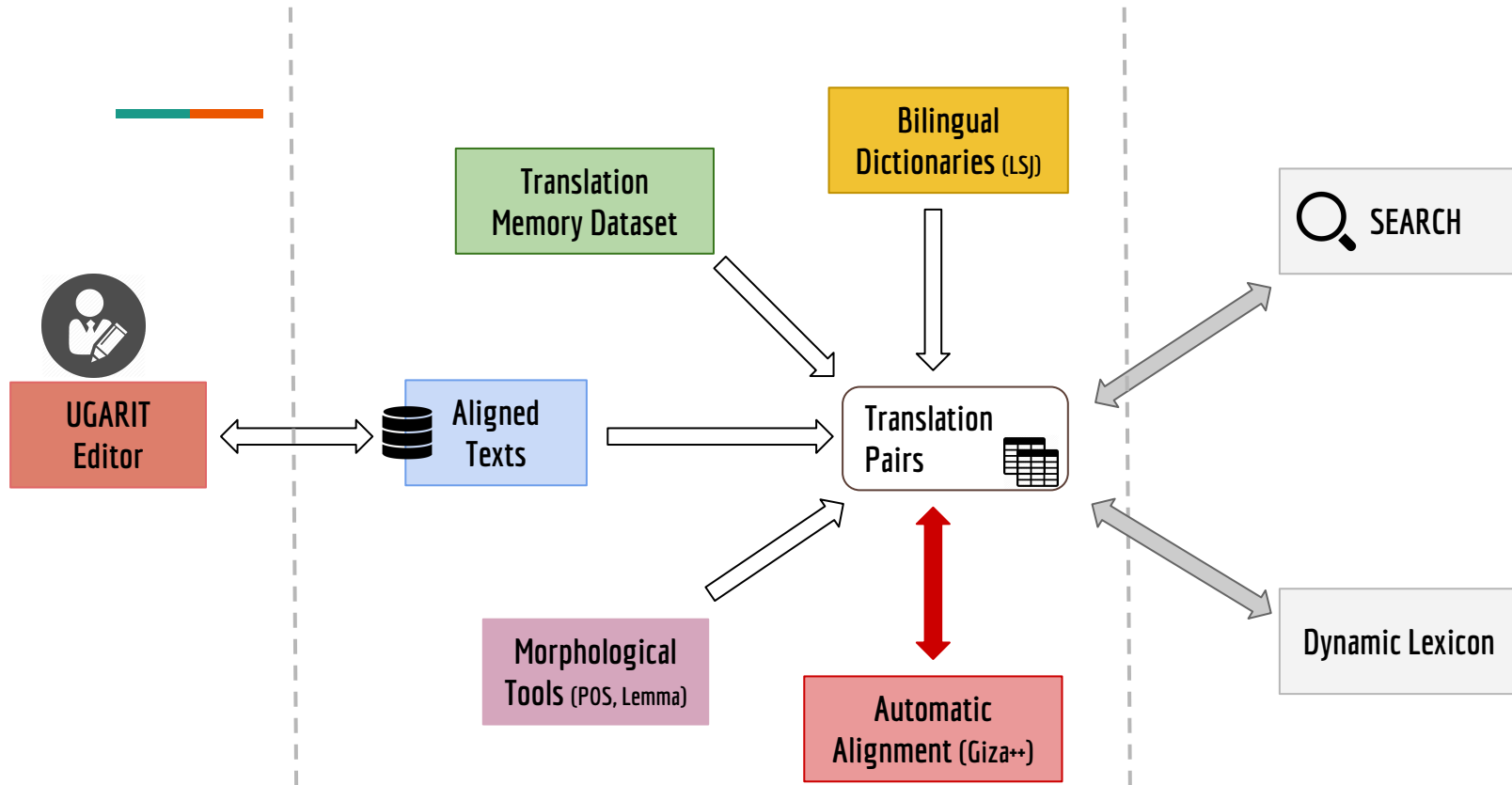
22065

Translation
Pairs

358850

Users

61





Application of translation alignment

One type of research done by translators

How we can use translation alignment for this type of research

How we can create manual aligned data



Sufi terms fall into the category of CSIs

“Cultural words” or “culture specific items (CSIs)” refer to terms or expressions not existing in the target language

They are problematic in the process of translation and consist of objects, proper nouns, institutions, expressions, etc.

Understanding mystical terms is central to the appreciation of Hafiz’s poems



Sufism

The term “Sufi” roots in Arabic meaning “Wool”

Sufism is an attempt to interiorize Quranic revelation

Sufism tries to bring the hidden meaning of Islam out of legalistic and juridical religion

Sufis believe Sufism is acting according to God’s orders



Sufi terms and translation of Hafez

Hafez (14th century) is known as the poet of mysteries; he is called as “tongue of secrets”

Some expressions bear a meaning totally paradoxical and opposite to what people or scholars are familiar with

The meaning of the terms has been changed in the past centuries

There is no unified approach for practicing mystical tradition

Most of Sufi terms are derived from Quran



Two Sufi terms

Saki:

literal meaning: the one who bears wine

mystical meaning: the perfect mankind

love:

literal meaning: going beyond borders of friendship

mystical meaning: includes the perfection



Same work with dynamic lexicon

Let's take a look in www.diwan-hafez.com

We try to search for both discussed words in the previous part



Data generation through citizen scientists

win-win situation:

- a) data is generated
- b) users learn the language

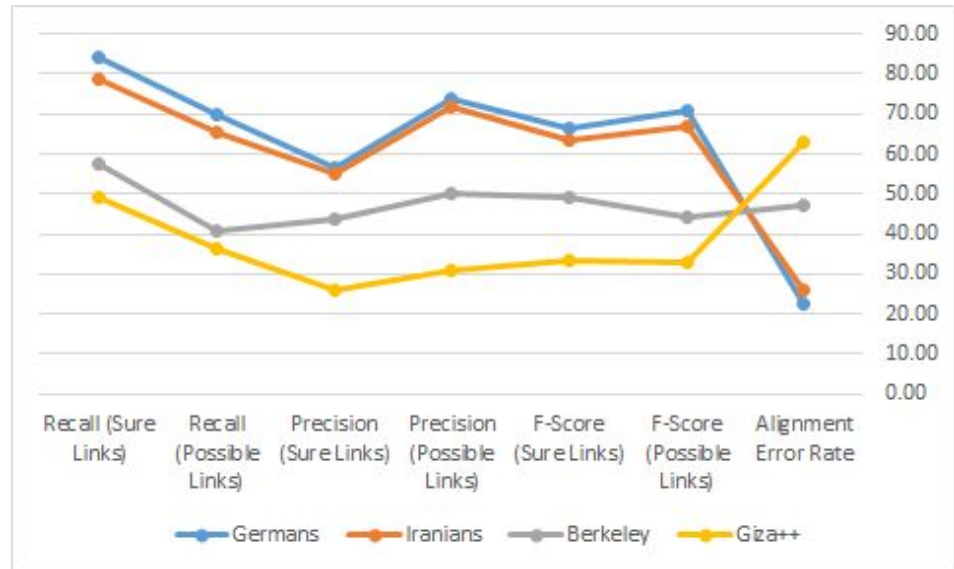
result:

- a) data is as accurate as speakers of both languages
- b) user engagement is high

Evaluation

Parameters are described in the suggested literature

The performance of both groups of human annotators is very close to each other





User engagement

Average time spent on translation alignment: 58 minutes

Average time spent on digital flashcards: 42 minutes



Thank you

tariq@informatik.uni-leipzig.de

gabriel.bodard@sas.ac.uk

saman.rezzaie@gmail.com

maryam.foradi@uni-leipzig.de