Retracing the Travel Path of Marco Polo

Master’s Defence

10th May, 2021

NLP for Historical Texts

Close reading of historical texts,
take researchers a lifetime
to explore and analysis…
...in a traditional way.

Retracing travel path from historical travelogue

- 12th century travelogue of Italian explorer Marco Polo
- Narrates his own travels through Asia and exploration of China between 1271 and 1295
- It is written by Rustichello da Pisa in Franco - Italian
- English translations used in this thesis are:
    - Hugh Murray: For text and the book Index
    - Henry Yule

Retracing travel path from historical travelogue

- 12th century travelogue of Italian explorer Marco Polo
- Narrates his own travels through Asia and exploration of China between 1271 and 1295
- It is written by Rustichello da Pisa in Franco - Italian
- English translations used in this thesis are:
    - Hugh Murray: For text and the book Index
    - Henry Yule: For the book index

Visualizing the travel path

PART	Travel Route
PART I	Within China
PART II	Venice -> China
PART III	China -> Venice

Visualizing the travel path

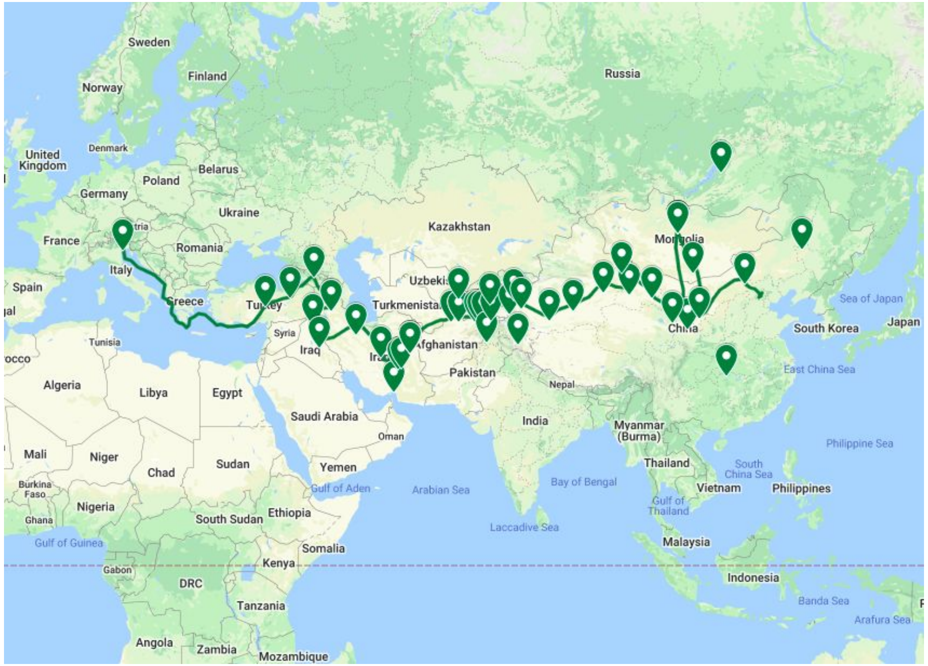
important

PART	Travel Route
PART I	Within China
PART II	Venice -> China
PART III	China -> Venice

Visualizing the travel path

Beginning of journey

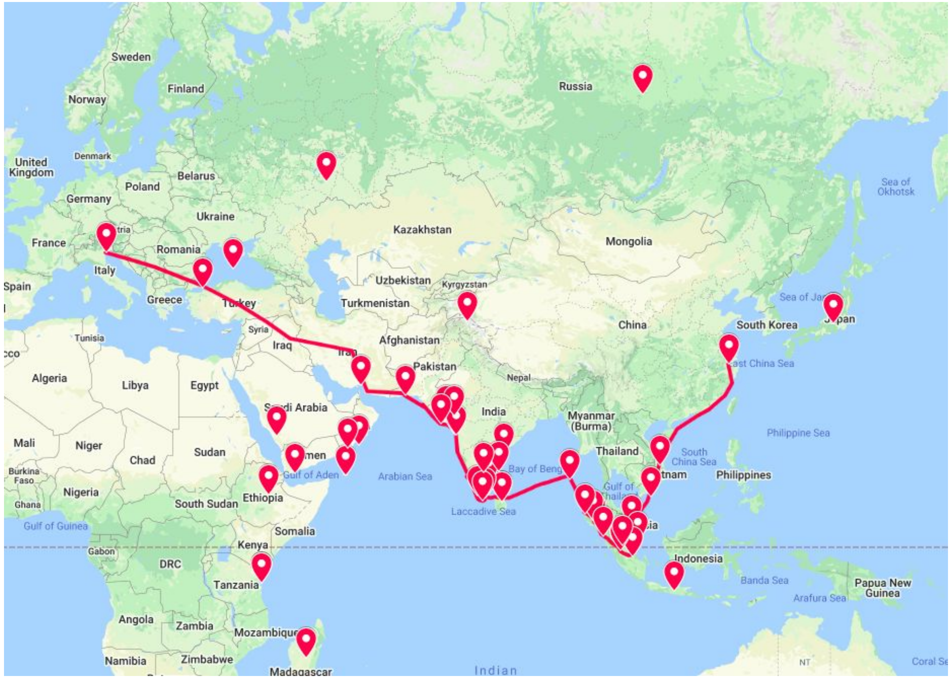
PART	Travel Route
PART I	Within China
PART II	Venice -> China
PART III	China -> Venice



Visualizing the travel path

Return journey

PART	Travel Route
PART I	Within China
PART II	Venice -> China
PART III	China -> Venice



Goal: Retracing the travel path of Marco Polo

General challenges with historical texts

- OCR errors
- Different naming conventions
- Spelling variations
- Translation errors
- Linguistic variations
- Syntax structures

KAIN-DU is a western province, $\boxed{\text{v/hich}}$ was formerly subject to its own princes; but, since it has been brought under the dominion of the grand khan, it is ruled by the governors whom he appoints. We are not to understand, however, that it is situated in the western part (of Asia), but only that it lies westward with respect to our course from the north eastern quarter. Its inhabitants are idolaters. It contains many cities and castles, and the capital city, standing at the commencement of the province, is likewise named Kain-du.
Near to it there is a large lake of salt water, in which are found abundance of pearls, of a white colour, but not round. 3

General challenges with historical texts

- OCR errors
- Different naming conventions
- Spelling variations
- Translation errors
- Linguistic variations
- Syntax structures

KAIN-DU is a western province, v/hich was formerly subject to its own princes; but, since it has been brought under the dominion of the grand khan, it is ruled by the governors whom he appoints. We are not to understand, however, that it is situated in the western part (of Asia), but only that it lies westward with respect to our course from the north eastern quarter. Its inhabitants are idolaters. It contains many cities and castles, and the capital city, standing at the commencement of the province, is likewise named Kain-du. Near to it there is a large lake of salt water, in which are found abundance of pearls, of a white colour, but not round. 3

General challenges with historical texts

- OCR errors
- Different naming conventions
- Spelling variations
- Translation errors
- Linguistic variations
- Syntax structures

KAIN-DU is a western province, v/hich was formerly subject to its own princes; but, since it has been brought under the dominion of the grand khan, it is ruled by the governors whom he appoints. We are not to understand, however, that it is situated in the western part (of Asia), but only that it lies westward with respect to our course from the north eastern quarter. Its inhabitants are idolaters. It contains many cities and castles, and the capital city, standing at the commencement of the province, is likewise named Kain-du. Near to it there is a large lake of salt water, in which are found abundance of pearls, of a white colour, but not round. 3

General challenges with historical texts

- OCR errors
- Different naming conventions
- Spelling variations
- Translation errors
- Linguistic variations
- Syntax structures

Challenges with the travelogue of Marco Polo

- Non-existing geographical entities

  e.g.: Greater India and Lesser India

- Geographical renaming

  e.g.: "Location of Barscol is very unclear but is thought to be around the eastern end of the present day Tian Shan Mountains."

- Change of boundaries

  e.g.: "Greater Khorasan includes territories that presently are part of Iran, Afghanistan, Tajikistan, Turkmenistan and Uzbekistan"

- Unnamed and descriptive place references

  E.g.: Plain of Bargu

Challenges with the travelogue of Marco Polo

- Non-existing geographical entities

  e.g.: Greater India and Lesser India

- Geographical renaming

  e.g.: "Location of Barscol is very unclear but is thought to be around the eastern end of the present day Tian Shan Mountains."

- Change of boundaries

  e.g.: "Greater Khorasan includes territories that presently are part of Iran, Afghanistan, Tajikistan, Turkmenistan and Uzbekistan"

- Unnamed and descriptive place references

  E.g.: Plain of Bargu

Challenges with the travelogue of Marco Polo

- Non-existing geographical entities

  e.g.: Greater India and Lesser India

- Geographical renaming

  e.g.: "Location of Barscol is very unclear but is thought to be around the eastern end of the present day Tian Shan Mountains."

- Change of boundaries

  e.g.: "Greater Khorasan includes territories that presently are part of Iran, Afghanistan, Tajikistan, Turkmenistan and Uzbekistan"

- Unnamed and descriptive place references

  E.g.: Plain of Bargu

Challenges with the travelogue of Marco Polo

- Non-existing geographical entities

  e.g.: Greater India and Lesser India

- Geographical renaming

  e.g.: "Location of Barscol is very unclear but is thought to be around the eastern end of the present day Tian Shan Mountains."

- Change of boundaries

  e.g.: "Greater Khorasan includes territories that presently are part of Iran, Afghanistan, Tajikistan, Turkmenistan and Uzbekistan"

- Unnamed and descriptive place references

  E.g.: Plain of Bargu

Related Work

- A study by Barbaresi[9] retraces the path from travel literature
  - Dataset :
    - Travel journals from China (1907)
    - Die Fackel (1899 - 1936)
  - Approach:
    - Combines coordinates, sequence and sense of time

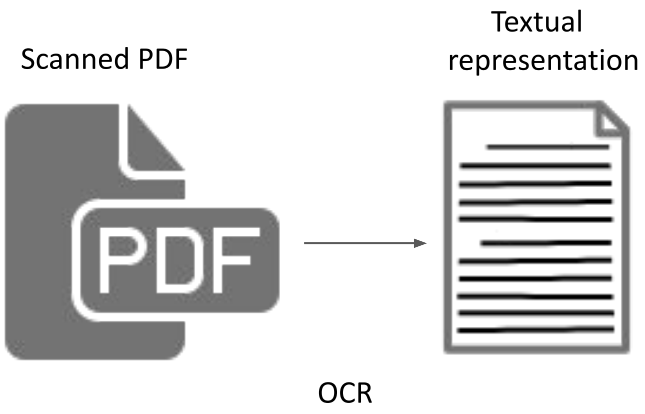
9. Barbaresi, A. (2018). A constellation and a rhizome: two studies on toponyms in literary texts. *VISUALISIERUNG*, 167

Data Preparation

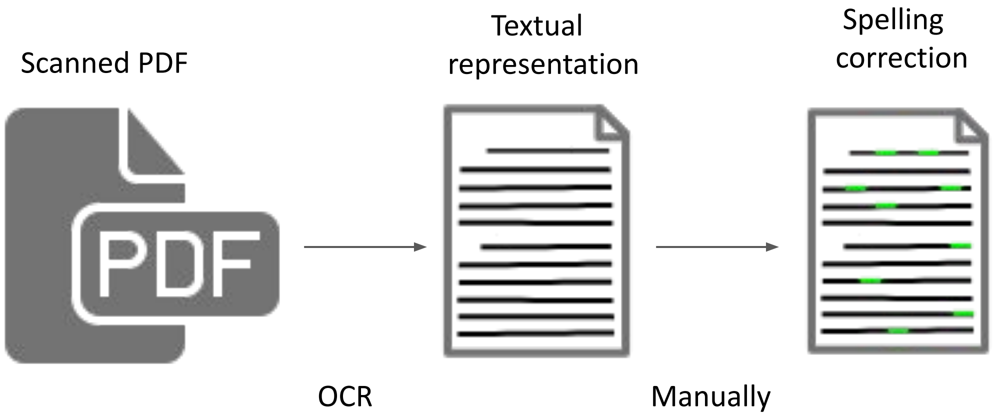
Text preparation

Scanned PDF

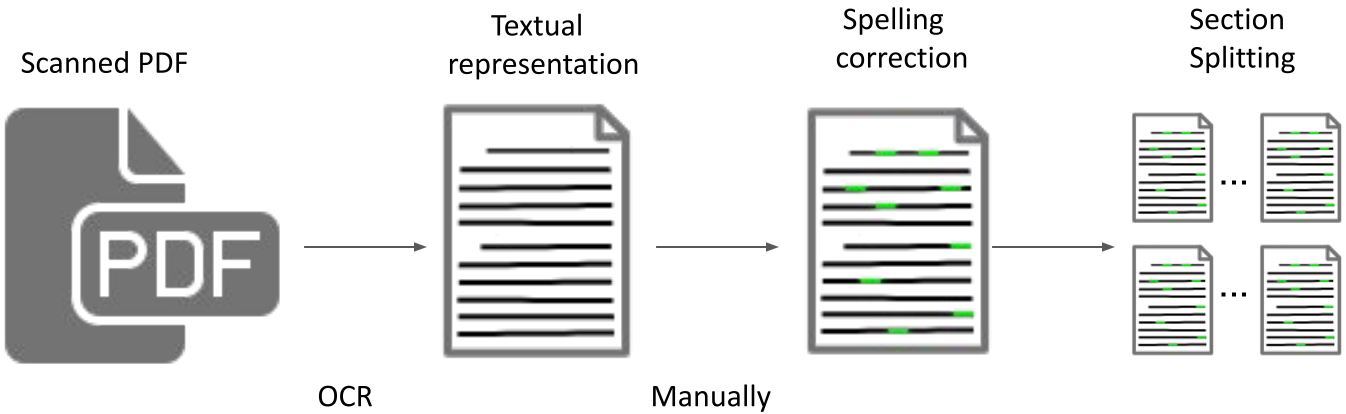
Text preparation



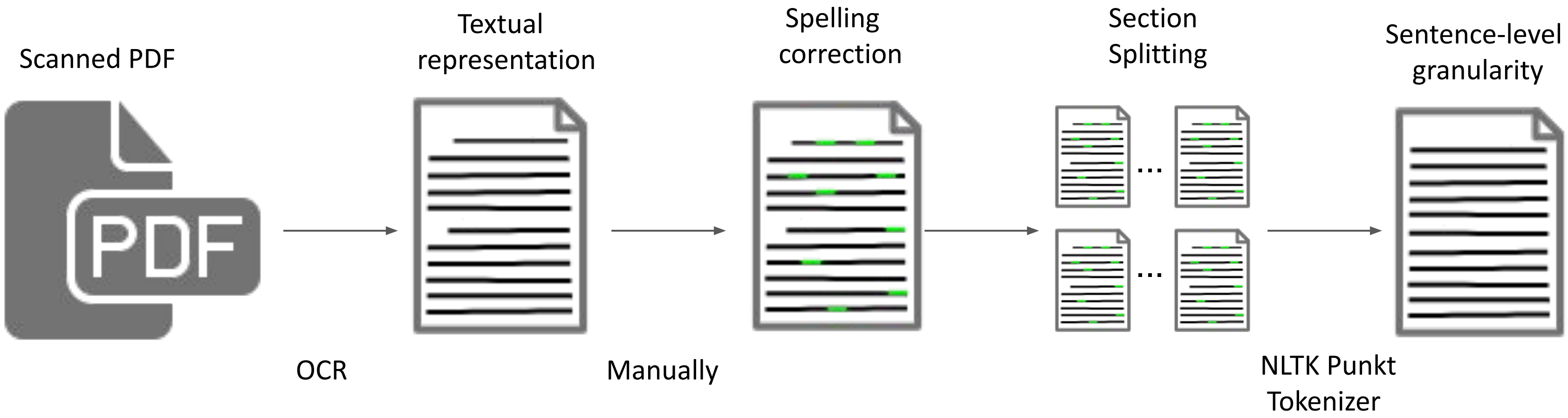
Text preparation



Text preparation



Text preparation



Index Preparation

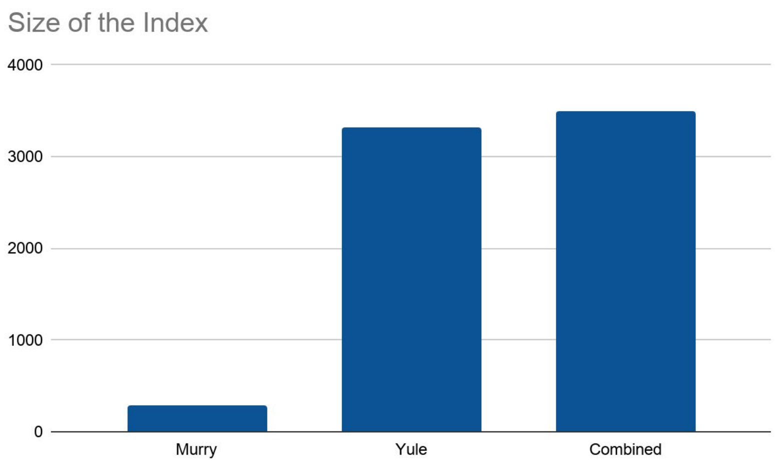
Three different versions:

- Index from the translation by Hugh Murray
- Index from the translation by Henry Yule
- Combined index from Murray and Yule translations

Index Preparation

Three different versions:

- Index from the translation by Hugh Murray
- Index from the translation by Henry Yule
- Combined index from Murray and Yule translations



Index Preparation

“ABASCIA (Abyssinia), kingdom of, 324. The inhibitants converted by St Thomas, 325. Its king defated the ruler of Adel (Aden), 326. Productions of the country, 327. Abraiamain (Bramins), order of, 293, 304-308.”

Entity	Alternative Name	References	Page No.
ABASCIA	Abyssinia	kingdom of	324
ABASCIA	Abyssinia	The inhibitants converted by St Thomas	325
ABASCIA	Abyssinia	Its king defated the ruler of Adel (Aden)	326
ABASCIA	Abyssinia	Productions of the country	327
ABASCIA	Abyssinia	Abyssinia & Abraiamain (Bramins), order of	293,304-308

What is a Gazetteer?

"Gazetteers are reference list...

….that are labeled,

relevant to the task"

Gazetteer Preparation

Index + Tag (Entity type) -> Gazetteer

Entity	Alternative Name	References	Page No.	Tag
ABASCIA	Abyssinia	kingdom of	324	Location
ABASCIA	Abyssinia	The inhibitants converted by St Thomas	325	
ABASCIA	Abyssinia	Its king defated the ruler of Adel (Aden)	326	
ABASCIA	Abyssinia	Productions of the country	327	
ABASCIA	Abyssinia	Abyssinia & Abraiamain (Bramins), order of	293,304-308	

Need for gazetteers

- Tasks where entity extraction is challenging
- Scarcity of proper resources for particular domain specific knowledge
- To encode additional background knowledge

In case of Marco Polo narrative….

- Absence of universal tooling for historical texts
- Lack of common standards for gazetteers

Need for gazetteers

- Tasks where entity extraction is challenging
- Scarcity of proper resources for particular domain specific knowledge
- To encode additional background knowledge

In case of Marco Polo narrative….

- Absence of universal tooling for historical texts
- Lack of common standards for gazetteers

Ambiguity Resolution

Entity Name	Alternative Name	Entity Tag
Alau	Hookalu	Person

Gazetteer from Murray translation

Ambiguity Resolution

Entity Name	Alternative Name	Entity Tag
Alau	Hookalu	Person

Gazetteer from Murray translation

Entity Name	Alternative Name	Entity Tag
Hukalu Khan	<u>Alau</u>, Hukalu	Person

Gazetteer from Yule translation

Ambiguity Resolution

Entity Name	Alternative Name	Entity Tag
Alau	Hookalu	Person

Gazetteer from Murray translation

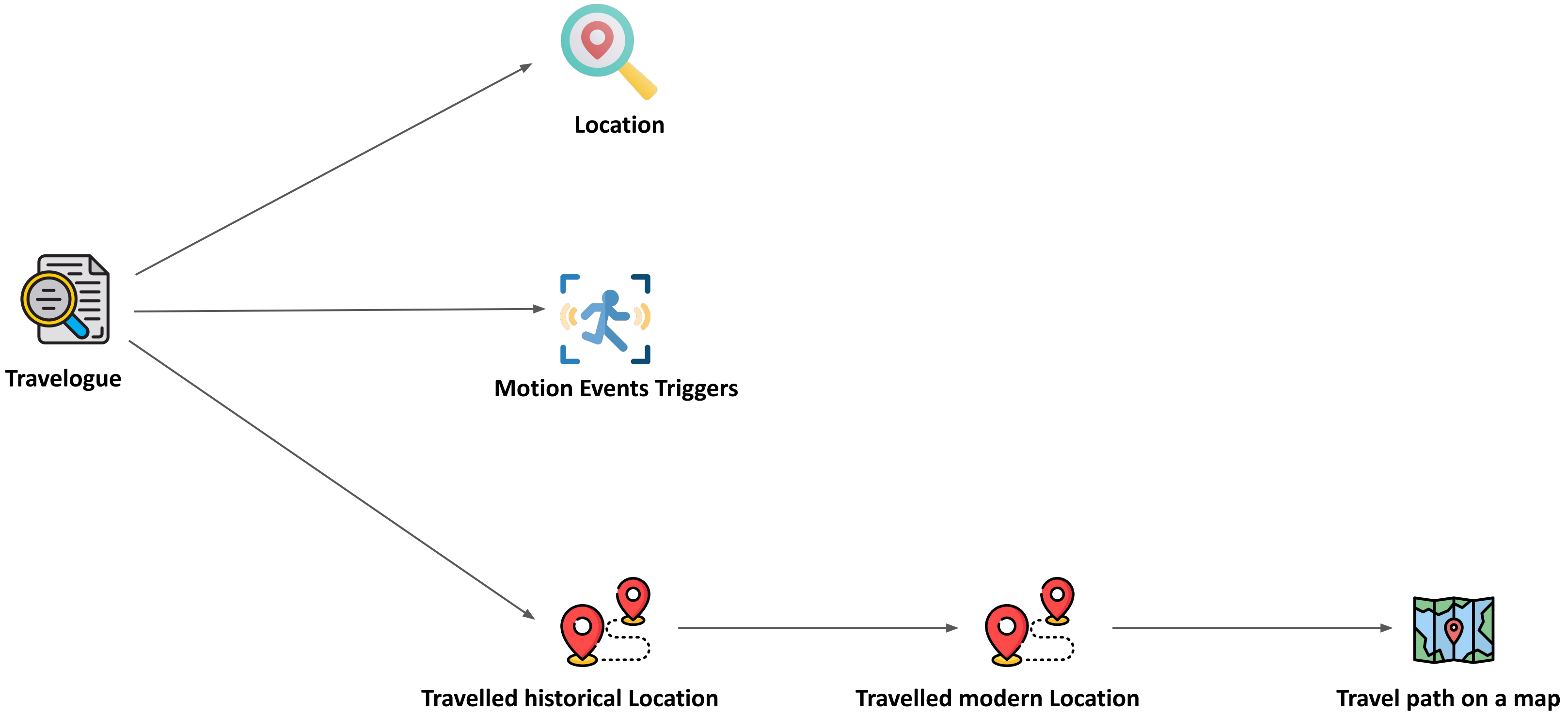
Entity Name	Alternative Name	Entity Tag
Hukalu Khan	Alau, Hukalu	Person

Gazetteer from Yule translation

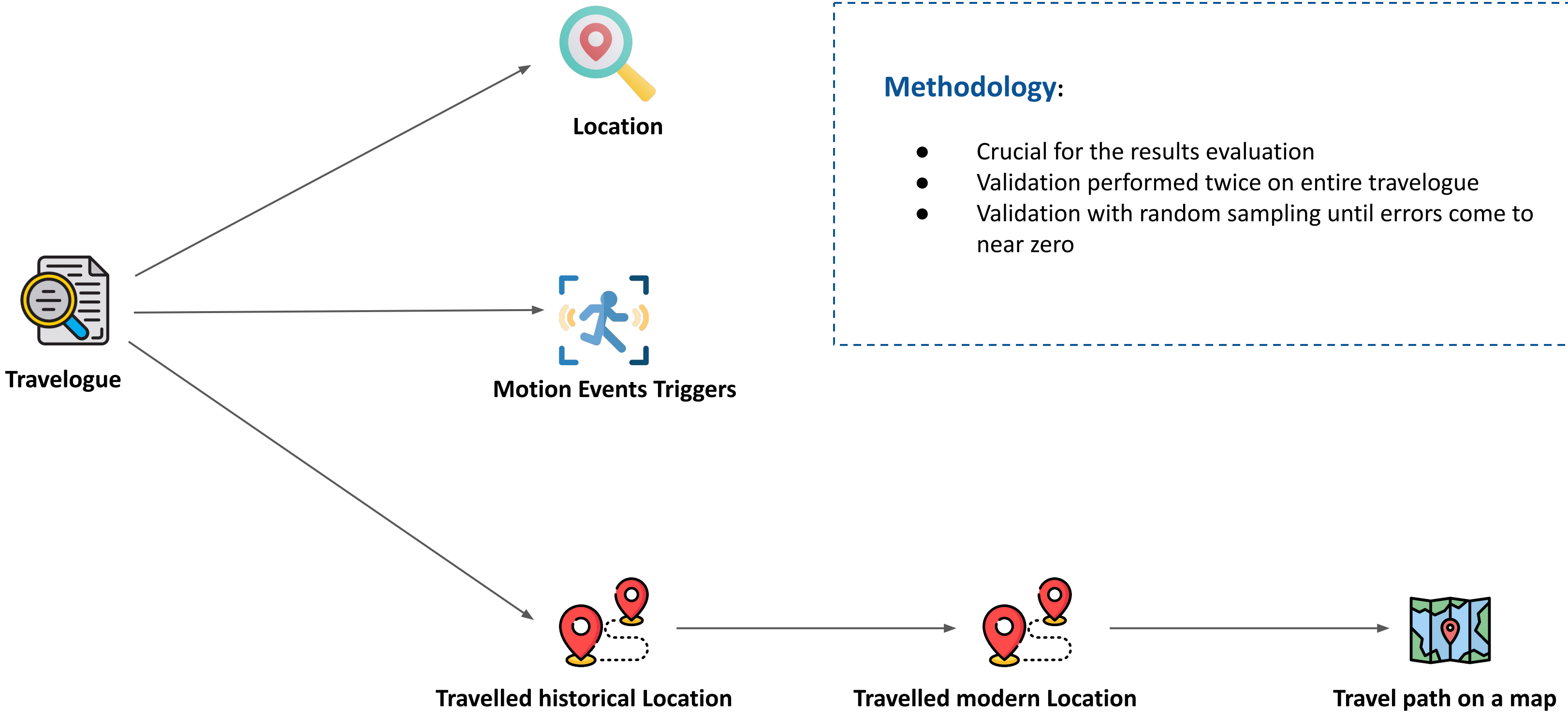
Entity Name	Alternative Name	Entity Tag
Alau	Hookalu, Hukalu, Hukalu Khan	Person

Combined gazetteer after ambiguity resolution

Gold Standard Setup



Gold Standard Setup



Methodology:

- Crucial for the results evaluation
- Validation performed twice on entire travelogue
- Validation with random sampling until errors come to near zero

Methodology

Raw Text

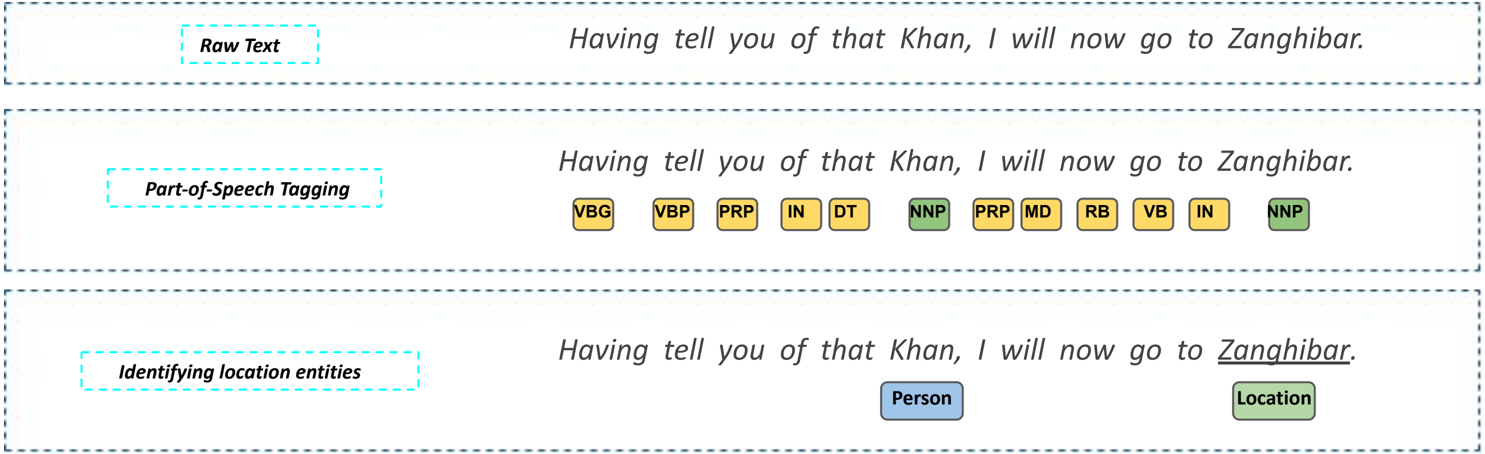
Having tell you of that Khan, I will now go to Zanghibar.

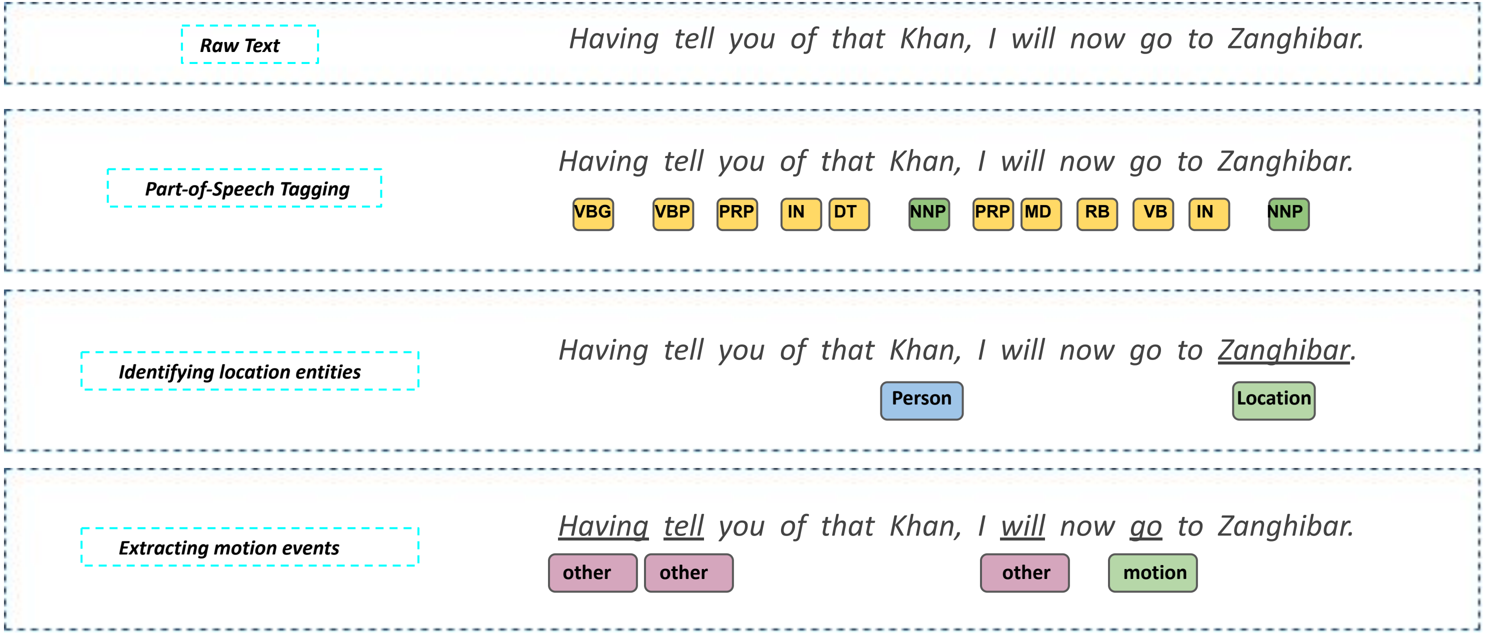
Raw Text

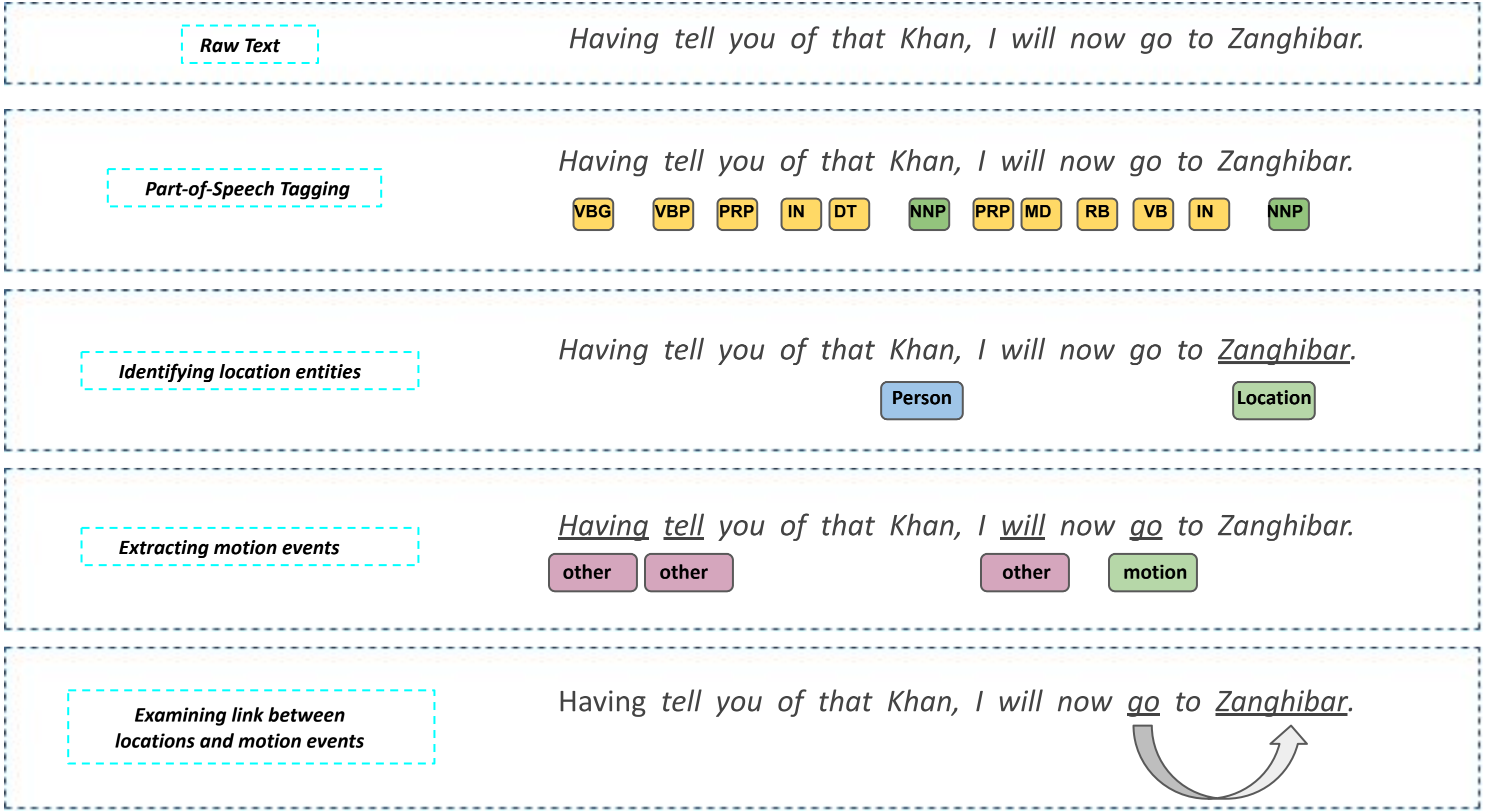
Having tell you of that Khan, I will now go to Zanghibar.

Part-of-Speech Tagging

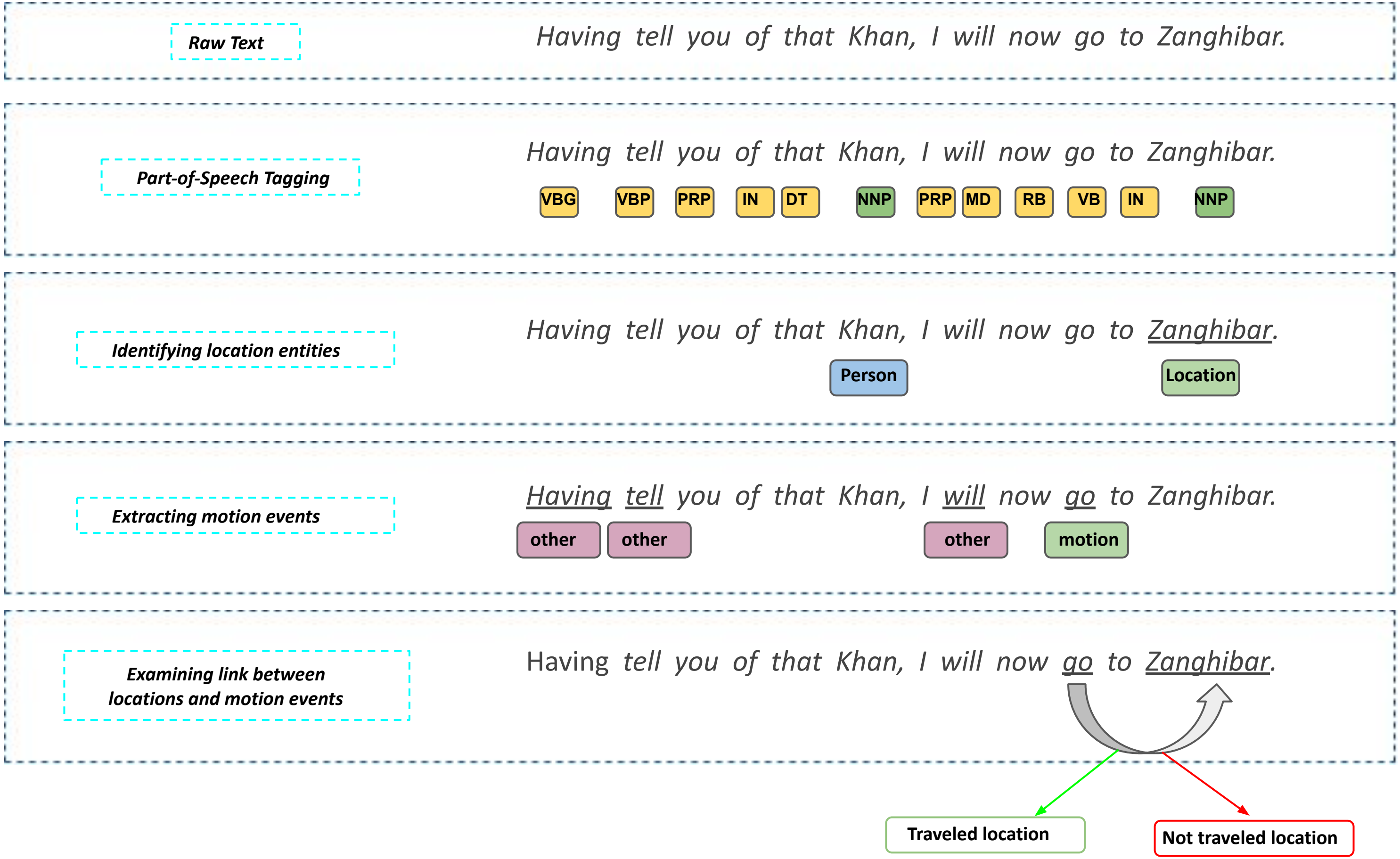
Having	*tell*	*you*	*of*	*that*	*Khan,*	*I*	*will*	*now*	*go*	*to*	*Zanghibar.*
VBG	VBP	PRP	IN	DT	NNP	PRP	MD	RB	VB	IN	NNP

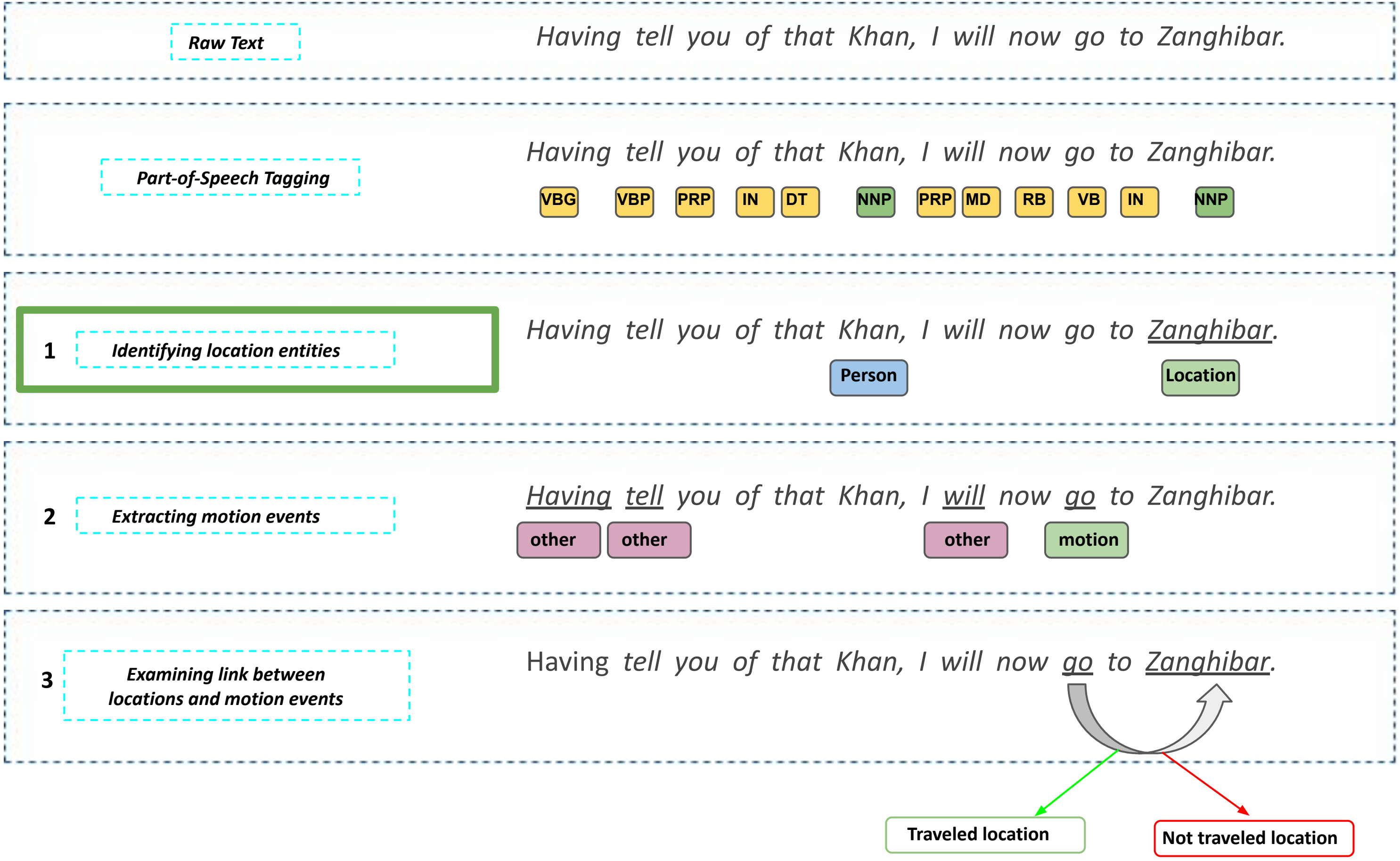

Raw Text
Having tell you of that Khan, I will now go to Zanghibar.
Part-of-Speech Tagging
Having tell you of that Khan, I will now go to Zanghibar.
VBG VBP PRP IN DT NNP PRP MD RB VB IN NNP
Identifying location entities
Having tell you of that Khan, I will now go to Zanghibar.
Person
Location


Raw Text
Having tell you of that Khan, I will now go to Zanghibar.
Part-of-Speech Tagging
Having tell you of that Khan, I will now go to Zanghibar.
VBG VBP PRP IN DT NNP PRP MD RB VB IN NNP
Identifying location entities
Having tell you of that Khan, I will now go to Zanghibar.
Person
Location
Extracting motion events
Having tell you of that Khan, I will now go to Zanghibar.
other
other
other
motion



Raw Text
Having tell you of that Khan, I will now go to Zanghibar.
Part-of-Speech Tagging
Having tell you of that Khan, I will now go to Zanghibar.
VBG VBP PRP IN DT NNP PRP MD RB VB IN NNP
Identifying location entities
Having tell you of that Khan, I will now go to Zanghibar.
Person
Location
Extracting motion events
Having tell you of that Khan, I will now go to Zanghibar.
other
other
other
motion
Examining link between locations and motion events
Having tell you of that Khan, I will now go to Zanghibar.


Raw Text
Having tell you of that Khan, I will now go to Zanghibar.
Part-of-Speech Tagging
Having tell you of that Khan, I will now go to Zanghibar.
VBG VBP PRP IN DT NNP PRP MD RB VB IN NNP
Identifying location entities
Having tell you of that Khan, I will now go to Zanghibar.
Person
Location
Extracting motion events
Having tell you of that Khan, I will now go to Zanghibar.
other
other
other
motion
Examining link between locations and motion events
Having tell you of that Khan, I will now go to Zanghibar.
Traveled location
Not traveled location


Raw Text
Having tell you of that Khan, I will now go to Zanghibar.
Part-of-Speech Tagging
Having tell you of that Khan, I will now go to Zanghibar.
VBG VBP PRP IN DT NNP PRP MD RB VB IN NNP
1 Identifying location entities
Having tell you of that Khan, I will now go to Zanghibar.
Person
Location
2 Extracting motion events
Having tell you of that Khan, I will now go to Zanghibar.
other other other motion
3 Examining link between locations and motion events
Having tell you of that Khan, I will now go to Zanghibar.
Traveled location
Not traveled location

What is Named Entity Recognition?

“ Identifying words,

classifying them into predefined categories,

such as person, location, organization, etc. ”

Identifying location entities

1. Application of pre-trained NER models
    - that have shown promising performance for standard corpora
    - that are widely and commonly used
    - that can be generalized to any data
    - that are representative of different approaches

Identifying location entities

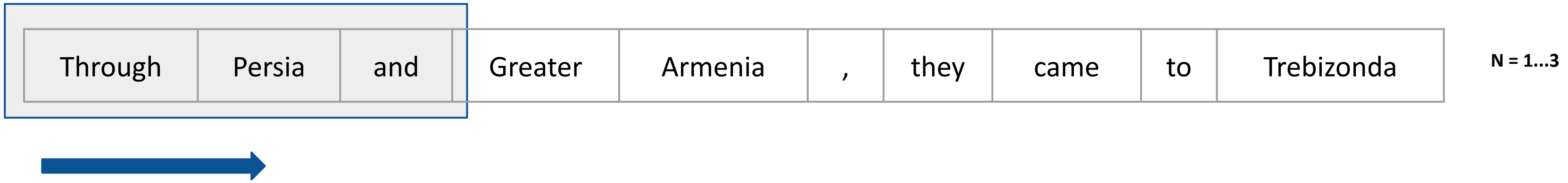
1. Application of pre-trained NER models
   - that have shown promising performance for standard corpora
   - that are widely and commonly used
   - that can be generalized to any data
   - that are representative of different approaches

NER model	Data	Model	F1-score
NLTK NER[1]	ACE 2004	MaxEnt classifier	0.89 ± 0.11
Stanford NER[2]	CoNLL, MUC-6, MUC-7 and ACE	CRF Classifier	87.94 %
spaCy NER[3]	OntoNotes	Multi-task CNN	85.85 %
AllenNLP NER[4]	CoNLL	ELMo	90.87 ± 0.13

1. https://nlp.stanford.edu/software/CRF-NER.shtml 2. https://spacy.io/ 3. https://www.nltk.org/book/ 4. https://allennlp.org/

Identifying location entities

2. Gazetteer for identifying location entities



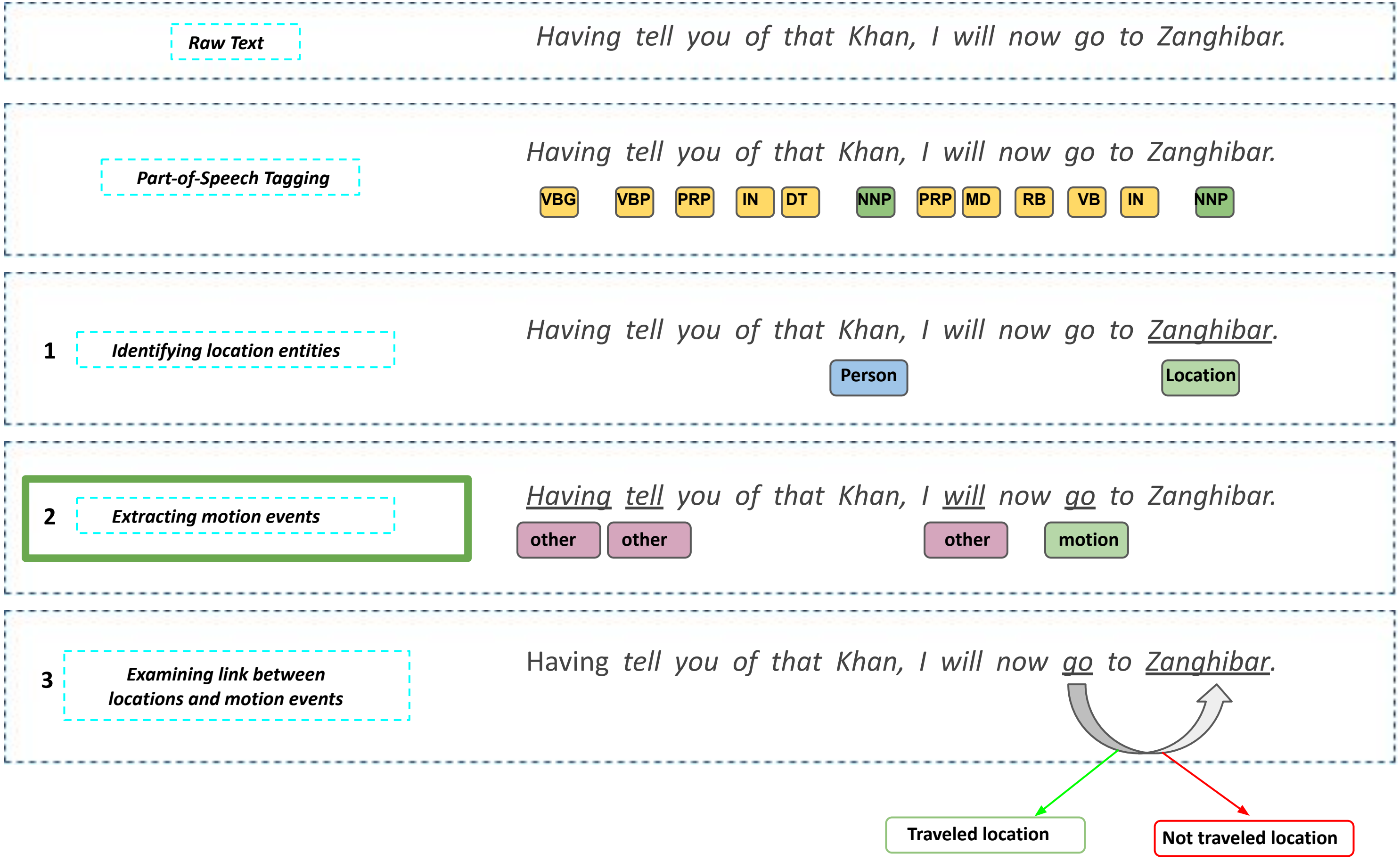
Identifying location entities

2. Gazetteer for identifying location entities

Through	Persia	and	Greater	Armenia	,	they	came	to	Trebizonda

N = 1…3

If extracted N-gram exists in Gazetteer, check the entity type from gazetteer.


Raw Text
Having tell you of that Khan, I will now go to Zanghibar.
Part-of-Speech Tagging
Having tell you of that Khan, I will now go to Zanghibar.
VBG VBP PRP IN DT NNP PRP MD RB VB IN NNP
1
Identifying location entities
Having tell you of that Khan, I will now go to Zanghibar.
Person
Location
2
Extracting motion events
Having tell you of that Khan, I will now go to Zanghibar.
other
other
other
motion
3
Examining link between locations and motion events
Having tell you of that Khan, I will now go to Zanghibar.
Traveled location
Not traveled location

What is Event Extraction?

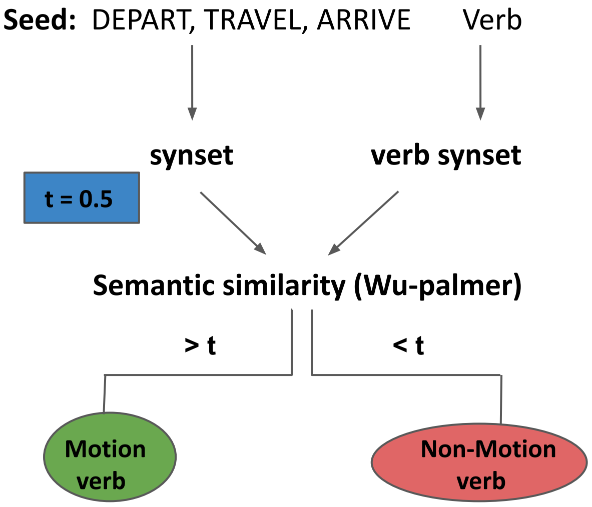
“ Information extraction task,

to extract specific knowledge,

related to certain incidents / events ”

Extracting motion event triggers

WordNet[5]



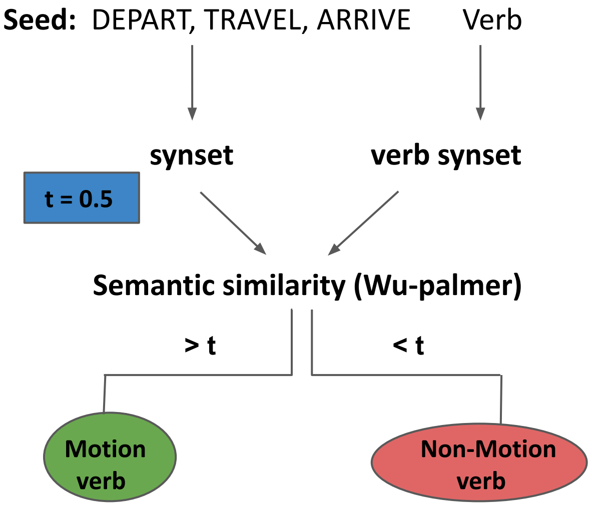
VerbNet[6]

FrameNet[7]

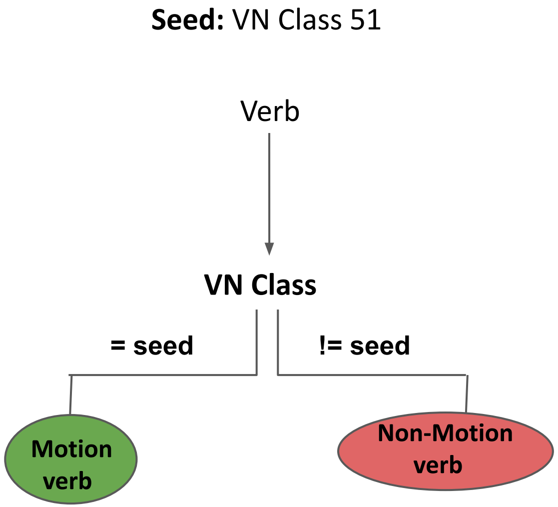
5. https://wordnet.princeton.edu/
6. https://verbs.colorado.edu/verbnet/
7. https://framenet.icsi.berkeley.edu/fndrupal/

Extracting motion event triggers

WordNet[5]



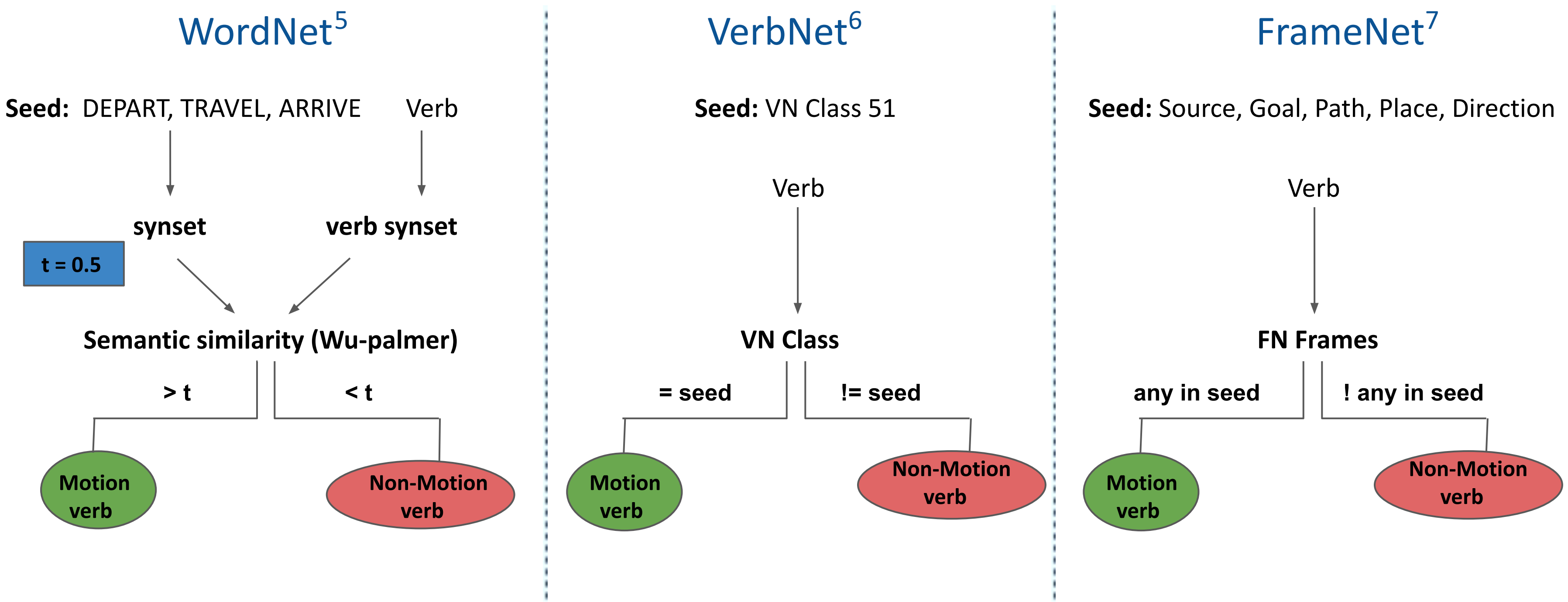
VerbNet[6]



FrameNet[7]

5.https://wordnet.princeton.edu/ 6. https://verbs.colorado.edu/verbnet/ 7. https://framenet.icsi.berkeley.edu/fndrupal/

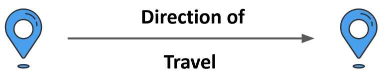
Extracting motion event triggers



5. https://wordnet.princeton.edu/ 6. https://verbs.colorado.edu/verbnet/ 7. https://framenet.icsi.berkeley.edu/fndrupal/

Same verb - different context

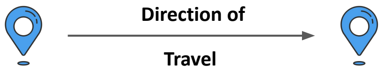
After climbing down a mountain, they descended on a plain area.

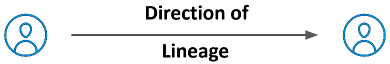


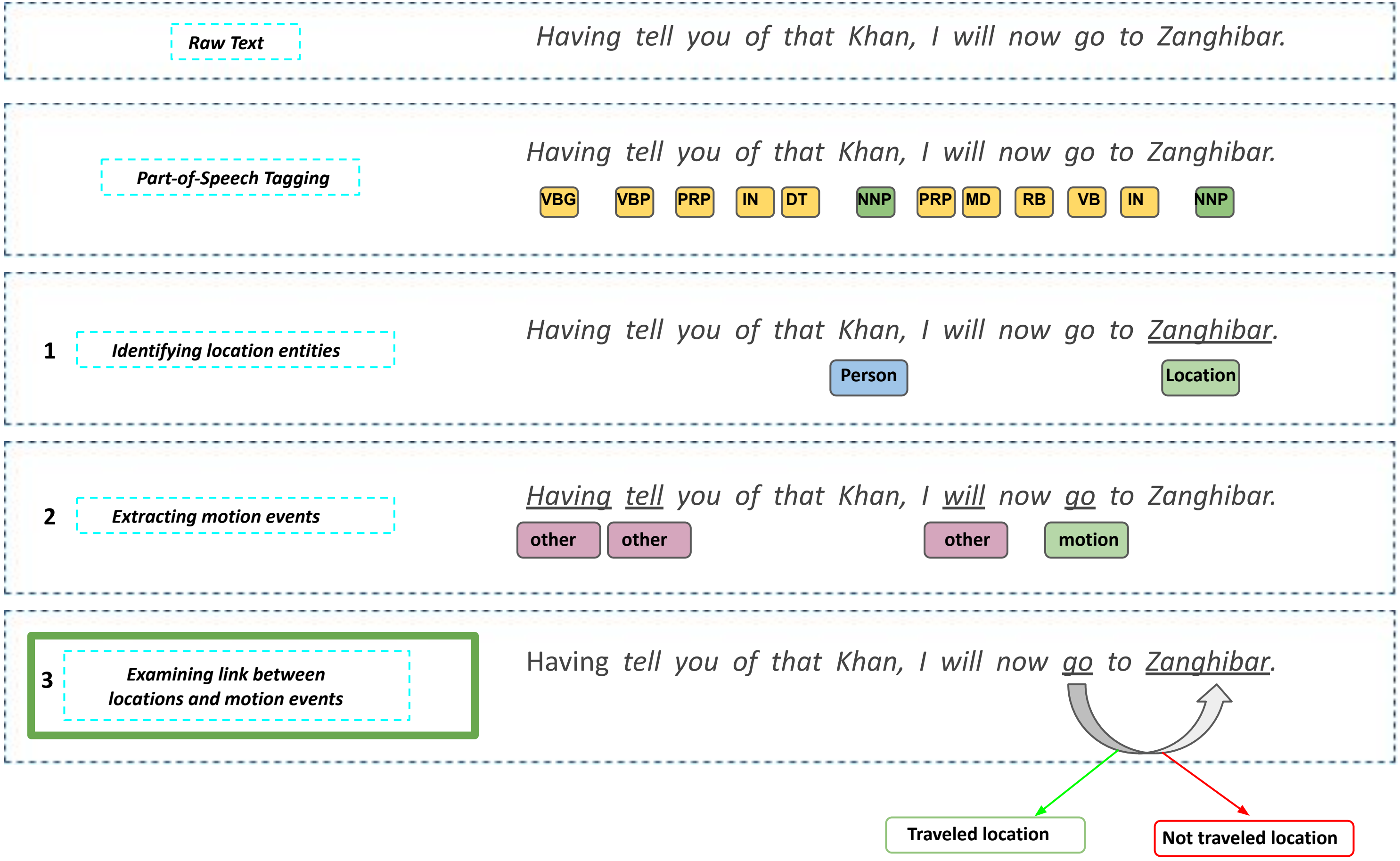
Same verb - different context

After climbing down a mountain, they descended on a plain area.

In this province there is a king named George, descended from that prince, and who indeed enjoy his power.






Raw Text
Having tell you of that Khan, I will now go to Zanghibar.
Part-of-Speech Tagging
Having tell you of that Khan, I will now go to Zanghibar.
VBG VBP PRP IN DT NNP PRP MD RB VB IN NNP
1 Identifying location entities
Having tell you of that Khan, I will now go to Zanghibar.
Person
Location
2 Extracting motion events
Having tell you of that Khan, I will now go to Zanghibar.
other other other motion
3 Examining link between locations and motion events
Having tell you of that Khan, I will now go to Zanghibar.
Traveled location
Not traveled location

Location and Motion Event Linking

Now I will leave this topic here and go to further describe the country of Cianba.

motion motion location

Location and Motion Event Linking

Now I will leave this topic here and go to further describe the country of Cianba.

motion **motion** **location**

VP -> VBD NP* PP* *OR* VP -> VBD PP* NP*

Find all noun phrases

Find associated prepositional phrase

Find all noun phrases

any

any

NP has location

Traveled location by Marco Polo

Results & Discussion

Evaluation Criteria for Named Entity Recognition

$$\textbf{Precision} = \frac{\text{True Positives}}{\text{True Positives + False Positives}} = \frac{\text{True Positives}}{\text{Total Predicted Positives}}$$

$$\textbf{Recall} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}} = \frac{\text{True Positives}}{\text{Total Actual Positives}}$$

$$\textbf{F1 score} = \frac{2 * (\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}}$$

Evaluation Criteria for Named Entity Recognition

Ground truth:

[Marco Polo] noted that, [Armenia the Greater] is a large country and at the entrance of it is a city called [Arzinga]

Person Location Location

Evaluation Criteria for Named Entity Recognition

Ground truth:

[Marco Polo] noted that, [Armenia the Greater] is a large country and at the entrance of it is a city called [Arzinga]

Person Location Location

NER Model:

[Marco Polo] noted that, [Armenia] the Greater is a large country and at the entrance of it is a city called Arzinga

Person Location

Evaluation Criteria for Named Entity Recognition

[Marco Polo] noted that, [Armenia the Greater] is a large country and at the entrance of it is a city called [Arzinga]

[Marco Polo] noted that, [Armenia] the Greater is a large country and at the entrance of it is a city called Arzinga

Ground Truth Entity	Predicted Entity	Match
Marco Polo	Marco Polo	Exact
Armenia the Greater	Armenia	Partial
Arzinga	-	No match

Evaluation Criteria for Named Entity Recognition

[Marco Polo] noted that, [Armenia the Greater] is a large country and at the entrance of it is a city called [Arzinga]

[Marco Polo] noted that, [Armenia] the Greater is a large country and at the entrance of it is a city called Arzinga

Ground Truth Entity	Predicted Entity	Match
Marco Polo	Marco Polo	Exact
Armenia the Greater	Armenia	Partial
Arzinga	-	No match

Matching Criteria:

Exact Match
Or
Partial Match

Evaluation Criteria for Named Entity Recognition

[Marco Polo] noted that, [Armenia the Greater] is a large country and at the entrance of it is a city called [Arzinga]

[Marco Polo] noted that, [Armenia] the Greater is a large country and at the entrance of it is a city called Arzinga

Ground Truth Entity	Predicted Entity	Match
Marco Polo	Marco Polo	Exact
Armenia the Greater	Armenia	Partial
Arzinga	-	No match

Matching Criteria:

Exact Match
Or
Partial Match

Armenia -> correct Greater -> Incorrect

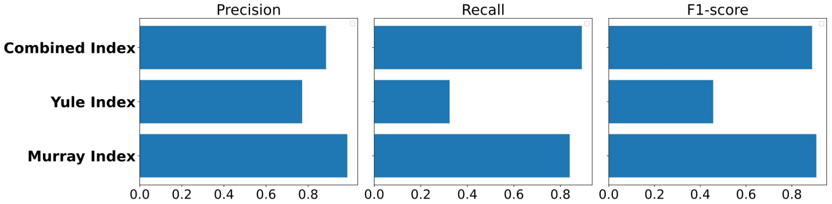
Evaluation Criteria for Named Entity Recognition

[Marco Polo] noted that, [Armenia the Greater] is a large country and at the entrance of it is a city called [Arzinga]

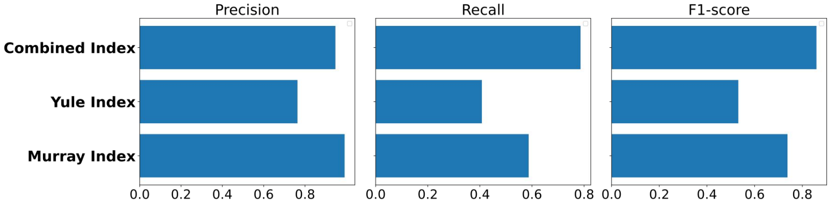
[Marco Polo] noted that, [Armenia] the Greater is a large country and at the entrance of it is a city called Arzinga

Ground Truth Entity	Predicted Entity	Match	True Positive	False Positive	False Negative	True Negative
Marco Polo	Marco Polo	Exact	1	0	0	0
Armenia the Greater	Armenia	Partial	0	1	1	0
Arzinga	-	No match	0	0	1	0

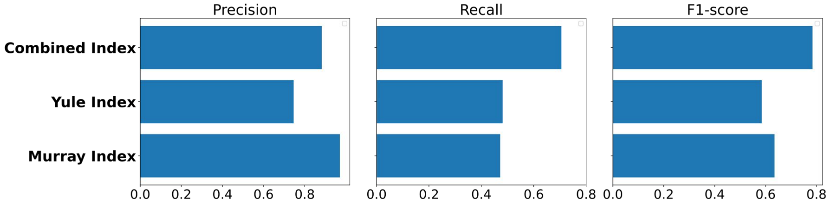
Evaluating Gazetteers for location extraction



Gazetteer comparison for PART 1



Gazetteer comparison for PART 2

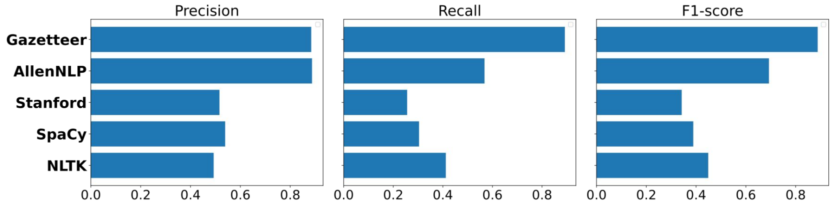


Gazetteer comparison for PART 3

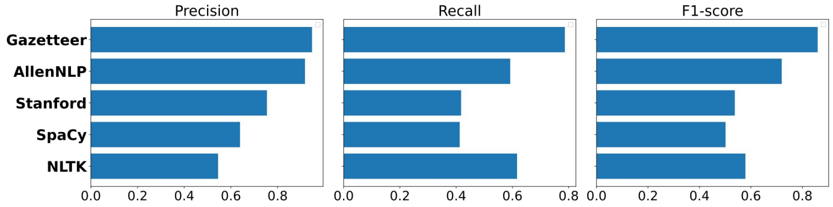
Gazetteer Analysis:

- Combined Index **outperformed** each of them individually
- Murray Gazetteer - high precision, low recall
- Yule Gazetteer - high precision, low recall

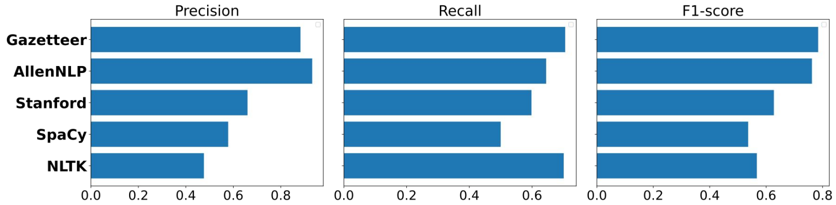
Evaluating NER for location extraction



NER comparison for PART 1



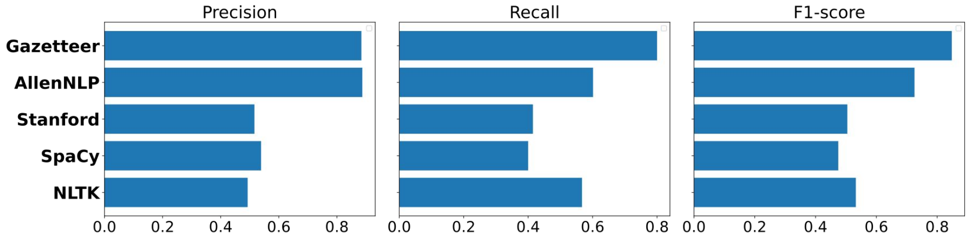
NER comparison for PART 2



NER comparison for PART 3

NER Analysis:

- Gazetteer has best F1-score
- Stanford, spaCy, NLTK performs poorly: low precision and low recall



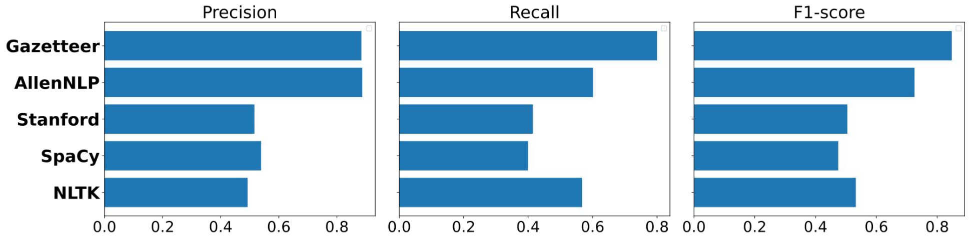
NER comparison for entire travelogue

Entire Book NER Analysis:

- Gazetteer has best F1-score
- Stanford, spaCy, NLTK performs poorly: low precision and low recall

Results:

- Gazetteer **outperforms** all the other pre-trained NER
- AllenNLP performs well compare to other pre-trained NERs
- AllenNLP has almost identical precision as gazetteer, but low recall and hence, low F1-score



NER comparison for entire travelogue

Entire Book NER Analysis:

- Gazetteer has best F1-score
- Stanford, spaCy, NLTK performs poorly: low precision and low recall

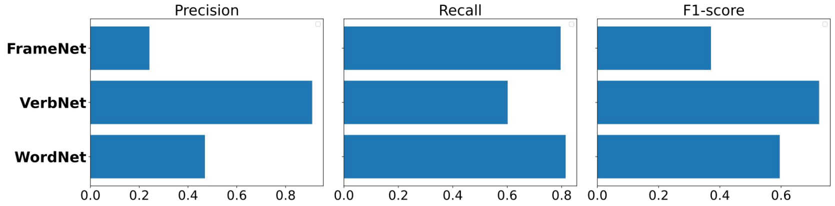
Results:

- Gazetteer **outperforms** all the other pre-trained NER
- AllenNLP performs well compare to other pre-trained NERs
- AllenNLP has almost identical precision as gazetteer, but low recall and hence, low F1-score

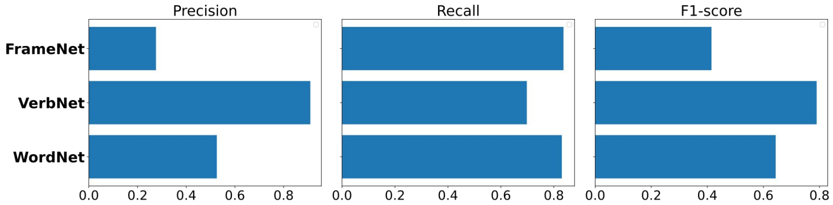
Observations:

- Pre-trained NERs do not perform well on entities with different naming convention
- Incorrectly identified entity type
  - Location -> Person
  - False negative for location
  - False positive for Person

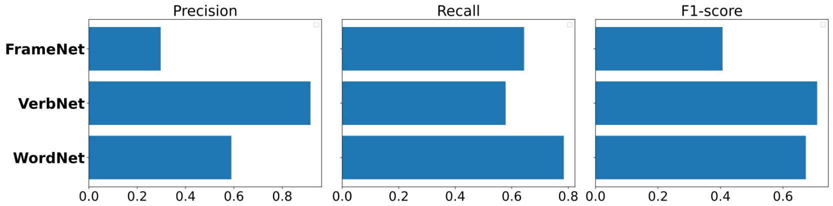
Evaluating Motion Events Extraction



Motion Event Extraction for PART 1



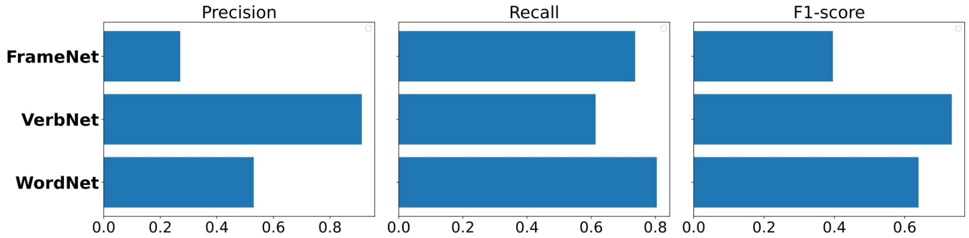
Motion Event Extraction for PART 2



Motion Event Extraction for PART 3

Event Extraction Analysis:

- FrameNet and WordNet has poor precision but high recall
- VerbNet has high precision but low recall
- Overall, VerbNet has highest f1 score



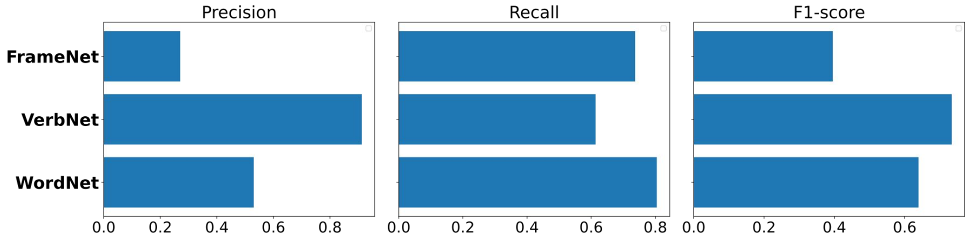
Motion Event Extraction for entire travelogue

Event Extraction Analysis:

- FrameNet and WordNet has poor precision but high recall
- VerbNet has high precision but low recall
- Overall, VerbNet has highest f1 score

Results:

- Framenet and Wordnet has high precision but low recall
- VerbNet has high precision but low recall
- High recall represents that all 3 are able to identify correct motion event triggers
- But low precision shows that they also falsely identify other events as motion events



Motion Event Extraction for entire travelogue

Event Extraction Analysis:

- FrameNet and WordNet has poor precision but high recall
- WordNet has highest recall
- VerbNet has high precision but low recall

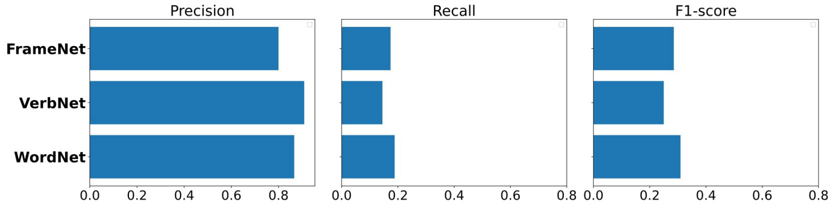
Results:

- Framenet and Wordnet has high precision but low recall
- VerbNet has high precision but low recall
- High recall represents that all 3 are able to identify correct motion event triggers
- But low precision shows that they also falsely identify other events as motion events

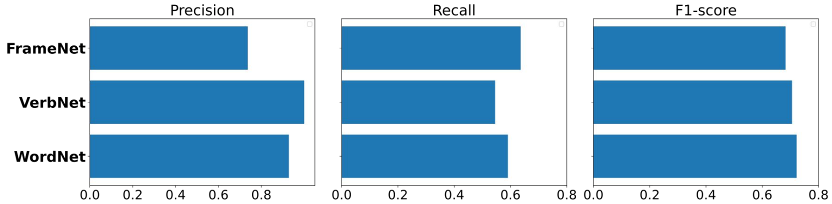
Observations:

- Choice of seeds affects the results
  - Choice of seeds can be narrowed down
  - However, it needs in-depth domain knowledge
  - Goal is to automate path retracing with minimal domain knowledge

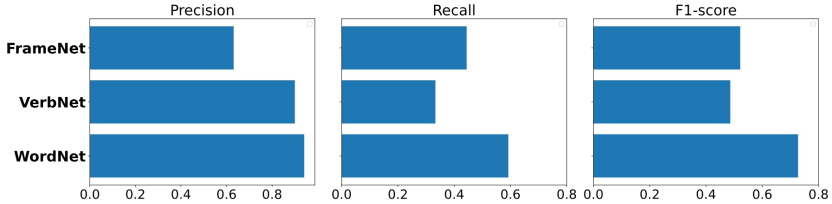
Motion Event and Location Linking



Location Identification for PART 1



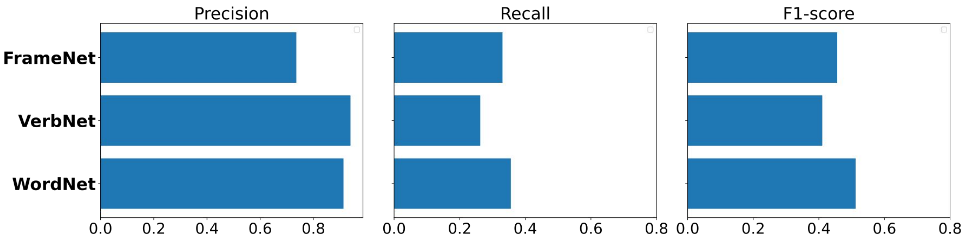
Location Identification for PART 2



Location Identification for PART 3

Traveled Location Identification Analysis:

- Poor recall
- PART 1 has a very low recall and PART 3 has highest recall
- High precision for all three parts
- WordNet has highest F1-score



Location Identification for entire travelogue

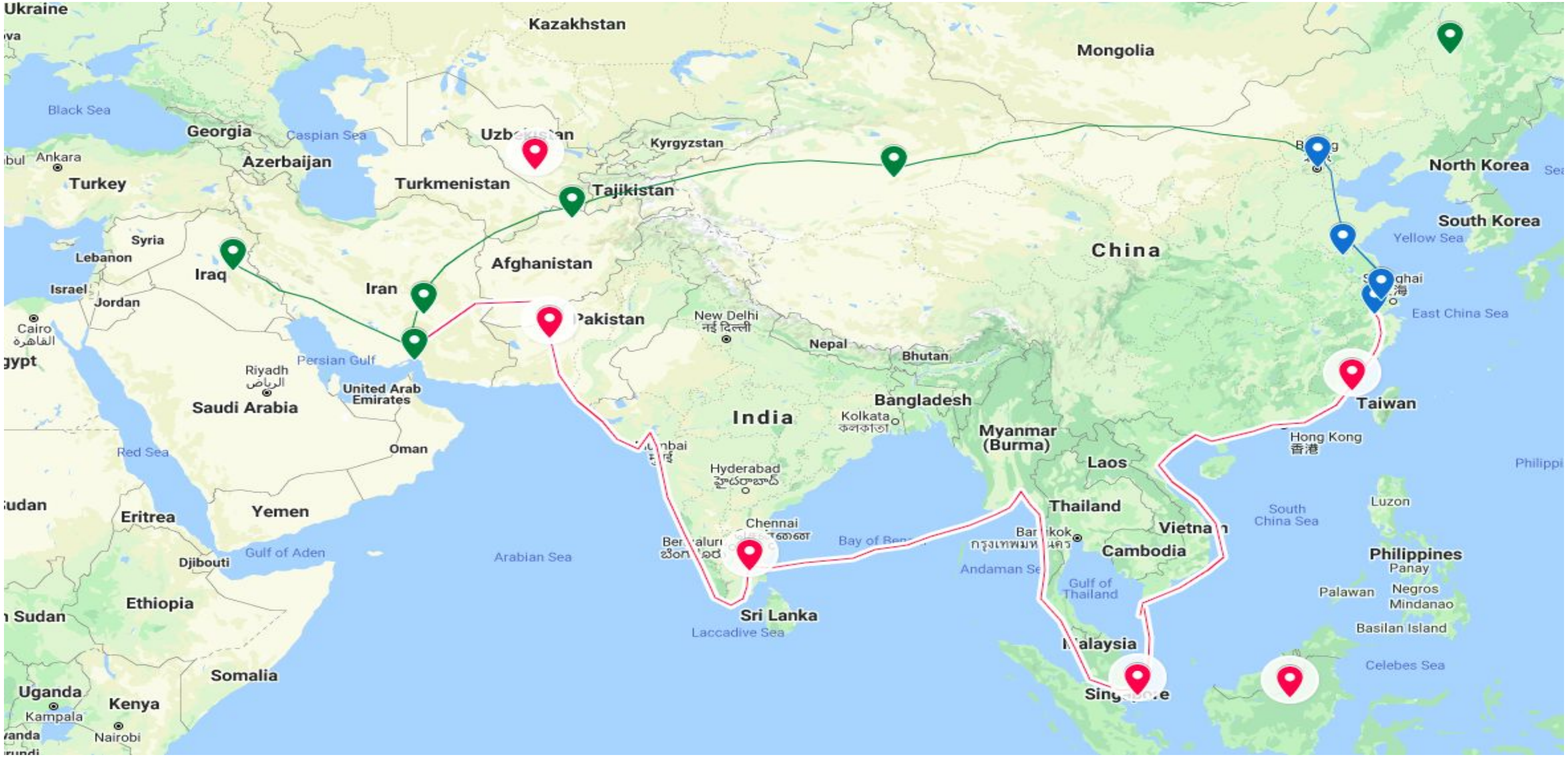
Traveled Location Identification Analysis:

- Poor recall
- High precision
- WordNet has highest F1-score

Results:

- Recall shows the identification of traveled locations which is low
- Precision shows correct identification of traveled location which is important as well and it is quite high here
- All three lexical resources have high precision but low recall
- WordNet has highest F1-score

Result: Extracted locations and interpolated travel path



Limitations

- **Verbs linking**

Location

" Leaving Ta-in-fu, and riding westward full seven days through very fine districts, amid numerous merchants, you find a large town, named Pi-an-fu, supported by commerce and the silk manufacture.

Location

Riding -> find -> named -> **Pi-an-fu**

Limitations

- **Verbs linking**

Location

" Leaving Ta-in-fu, and riding westward full seven days through very fine districts, amid numerous merchants, you find a large town, named Pi-an-fu,supported by commerce and the silk manufacture.

Location

Riding -> find -> named -> **Pi-an-fu**

- **Extracting traveled locations described using non-motion verbs**

*"Having told you all about these Tartars of East, I might go on to **treat** of **Great Turkey**."*

Limitations

- **Ambiguity in the narrative**

  *"Having nothing more to tell of this island, I will **go** to **Zanghibar**."* **Not visited**

  *"Now I will **go** to another city named **Balk**."* **visited**

Limitations

- **Ambiguity in the narrative**

  *"Having nothing more to tell of this island, I will **go** to **Zanghibar**."* Not visited

  *"Now I will **go** to another city named **Balk**."* visited

- **There has also been a difference of opinion between various historians about the exact travel path of Marco Polo and there has been not single agreed and verified path.**

Contributions

- Corpus creation
- Generating a mapping between historical and contemporary locations in the context of Marco Polo travelogue
- Evaluating state-of-the-art NER tools for 12th century historical travelogue
- Approach for extracting traveled locations from 12th century historical texts

Future Work

- Improving state-of-the-art NER systems to avoid problems with out-of-vocabulary words
- Implementing multiple relation extractions and verbs linking
- Identification of meaning from the text[8]
- A way to determine the order of the travel
- Finding individual route segments and connect route segments to generate a travel path

8. https://arxiv.org/pdf/2005.09099.pdf

“Deo Gratias. Amen, Amen.”

- Marco Polo -

Icons made by Freepik, Those Icons from Flaticon, Findicons

XLII. The Province and City of Sin-din-fu.

When a man has **left this country** and **travelled** twenty days westward, he **approaches** a province **on the borders of Manji named Sin-din-fu**. **The capital, bearing the same name**, was anciently very great and noble, governed by a mighty and wealthy sovereign. He died, leaving three sons, who divided the city into three parts, and each enclosed his portion with a wall, which was within the great wall of twenty miles in circuit. They ranked still as kings, and had ample possessions ; but the great khan overcame them, and took full possession of their territory. Through the city, a large river of fresh water, abounding with fish, passes and flows on to the ocean, distant eighty or a hundred days' journey ; **it is called Quian-su**. On that current is a very great number of cities and castles, and such a multitude of ships as no one who has not seen could possibly believe. Equally wonderful is the quantity of merchandise conveyed ; indeed it is so broad as to appear a sea and not a river. Within the city, it is crossed by a bridge wholly of marble, half a mile long and eight paces broad ; the upper part is supported by marble columns, and richly painted ; and upon it are many houses where merchants expose goods for sale ; but these are set up in the morning and taken down in the evening. At one of them, larger than the others, stands the chamberlain of the khan, who receives the duty on the merchandise sold, which is worth annually a thousand golden bezants.The inhabitants are all idolaters ; and **from that city a man goes five days' journey through castles, villages, and scattered houses.** The people subsist by agriculture, and the tract abounds with wild beasts. There are also large manufactures of gauzes and cloth of gold. After **travelling** these five days, he **comes** to **Thibet**.

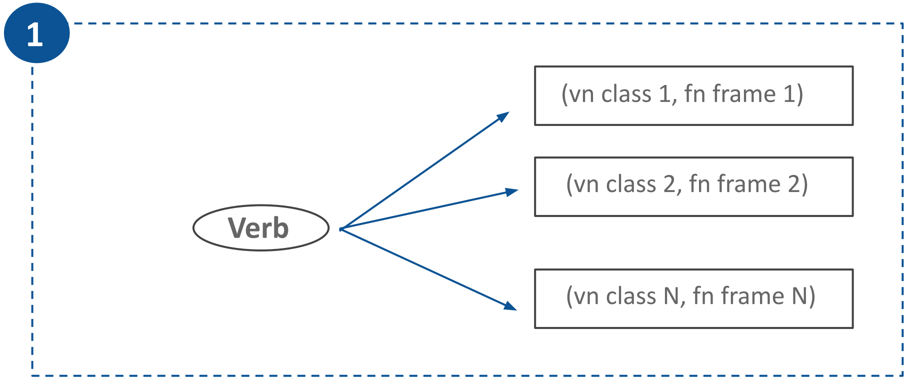
XLII. The Province and City of Sin-din-fu.

When a man has **left this country** and **travelled** twenty days westward, he **approaches** a province **on the borders of Manji named Sin-din-fu**. **The capital, bearing the same name**, was anciently very great and noble, governed by a mighty and wealthy sovereign. He died, leaving three sons, who divided the city into three parts, and each enclosed his portion with a wall, which was within the great wall of twenty miles in circuit. They ranked still as kings, and had ample possessions ; but the great khan overcame them, and took full possession of their territory. Through the city, a large river of fresh water, abounding with fish, passes and flows on to the ocean, distant eighty or a hundred days' journey ; **it is called Quian-su**. On that current is a very great number of cities and castles, and such a multitude of ships as no one who has not seen could possibly believe. Equally wonderful is the quantity of merchandise conveyed ; indeed it is so broad as to appear a sea and not a river. Within the city, it is crossed by a bridge wholly of marble, half a mile long and eight paces broad ; the upper part is supported by marble columns, and richly painted ; and upon it are many houses where merchants expose goods for sale ; but these are set up in the morning and taken down in the evening. At one of them, larger than the others, stands the chamberlain of the khan, who receives the duty on the merchandise sold, which is worth annually a thousand golden bezants.The inhabitants are all idolaters ; and **from that city a man goes five days' journey through castles, villages, and scattered houses.** The people subsist by agriculture, and the tract abounds with wild beasts. There are also large manufactures of gauzes and cloth of gold. After **travelling** these five days, he **comes** to **Thibet**.

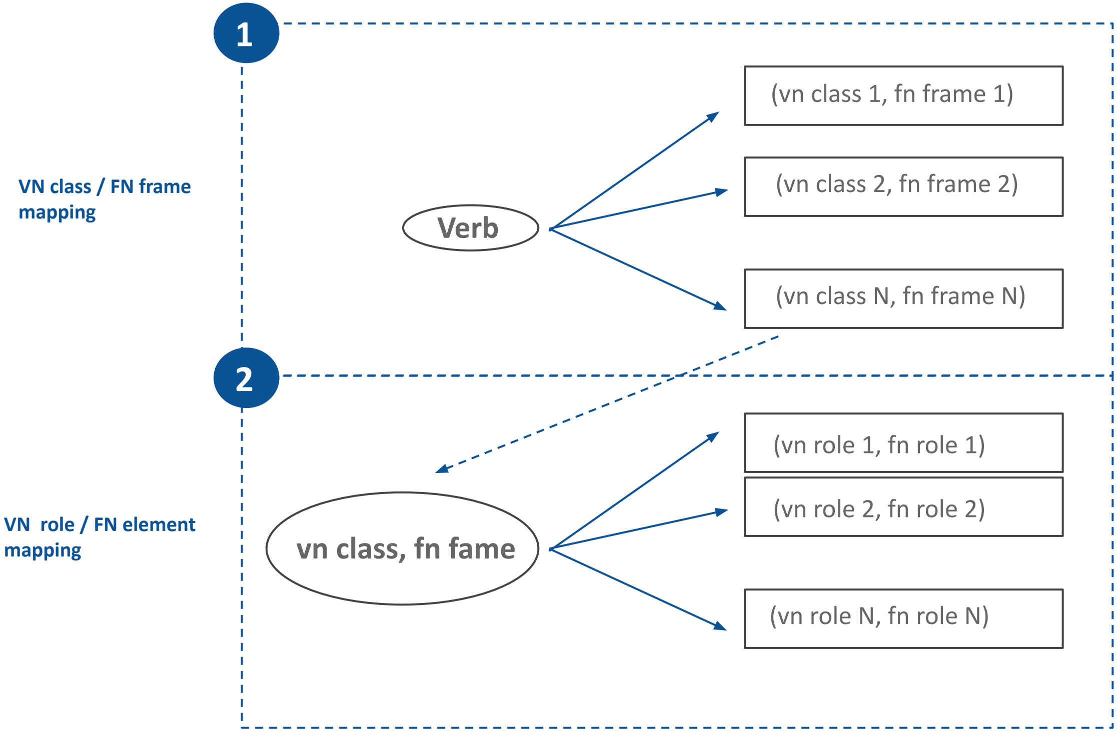
- Coreference resolution
- Verbs linking
- Ordering and establishing a connection through distance or sense of time or direction

Identifying context using SemLink Mapping

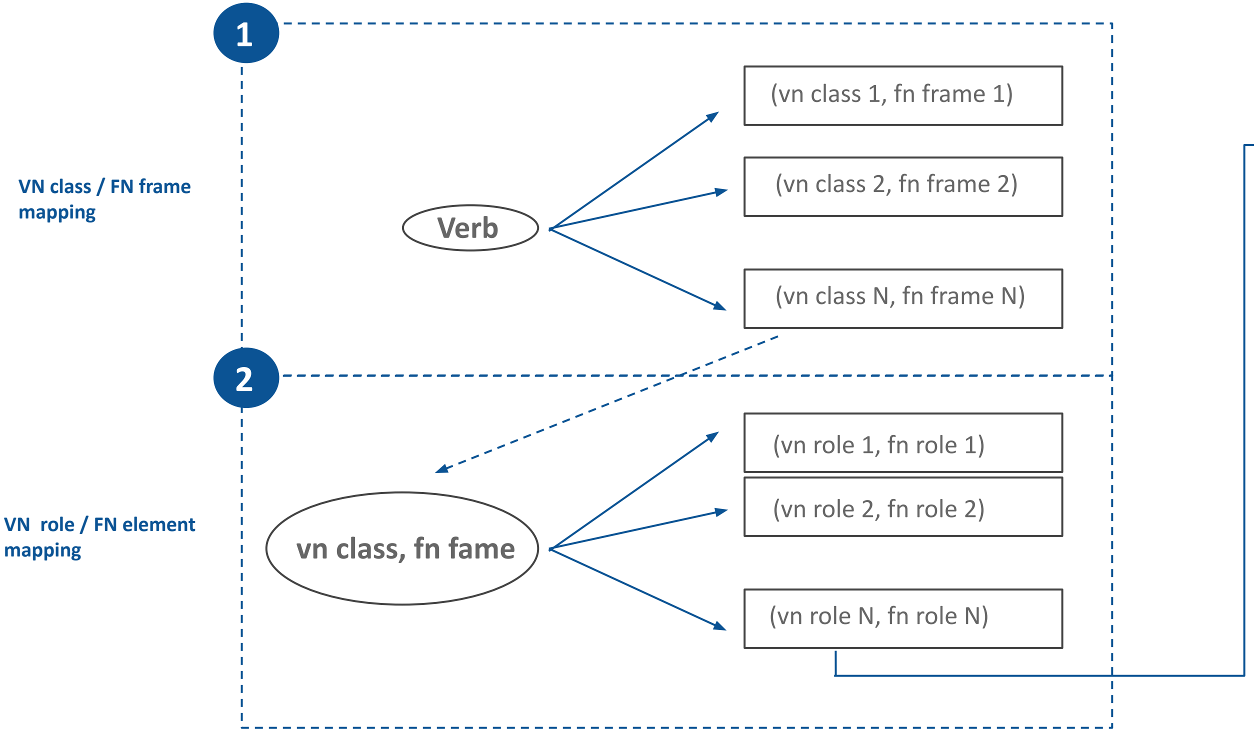
VN class / FN frame mapping



Identifying context using SemLink Mapping



Identifying context using SemLink Mapping



VerbNet Thematic Roles
Source
Destination
Location

Identifying context using SemLink Mapping