

Cuneiform Fragment Matching

Fabian Simonjetz, Jussi Laasonen, Yunus Cobanoglu, Alexander Fraser, Enrique Jiménez
Ludwig-Maximilians-Universität München, Germany

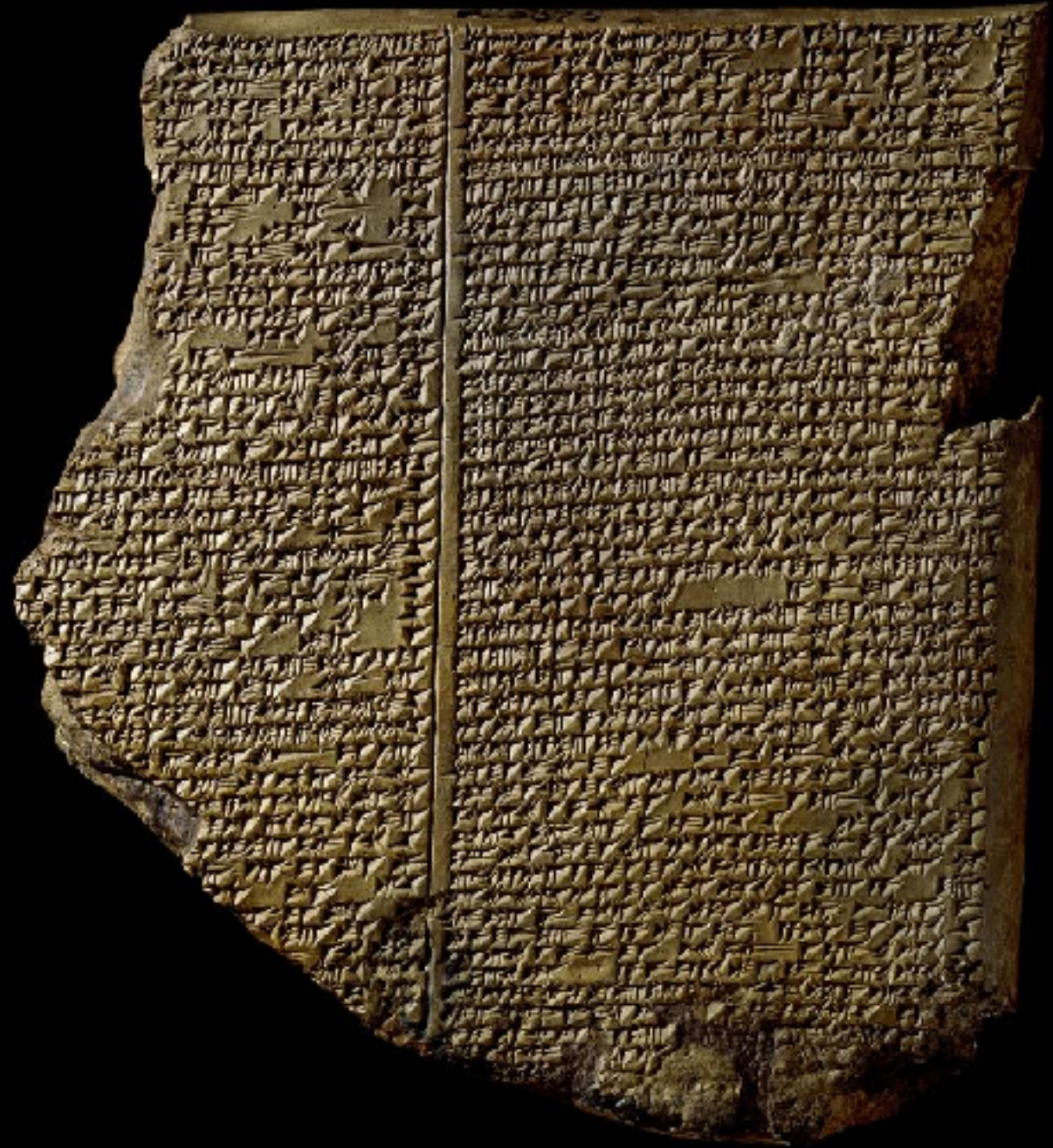


Overview

1. Background
2. Task description
3. Corpus and Experiments
4. Conclusions

What is Cuneiform?

- Ca. 3,200 BCE - 100;
Mesopotamia (Iraq)
- Clay tablets used for
bookkeeping
- Sumerian
- Akkadian (Assyrian,
Babylonian)



Reconstruction



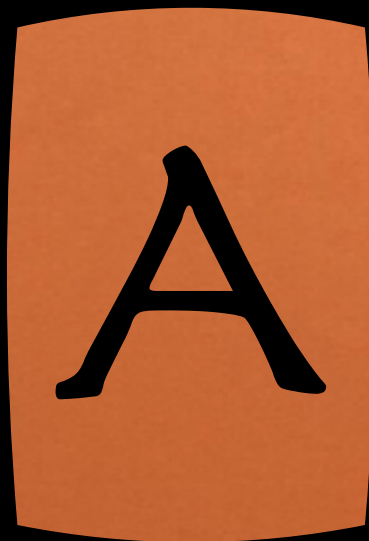
Reconstruction



Reconstruction



Reconstruction



Reconstruction



Data Properties

- Orthographic variations
- Diachronic change
- Different versions
- Structured linebreaks
- Partially solved

附錄全

a-par-ra-as

下半部有紅線

a-pa-ar-ra-as

解法全

a-pa-ra-as

4

KUD

'I will divide'

給與不給與

le-mut-ta

第一回

le-mut-tu

各一册半紅

 $le-mut-tu_4$

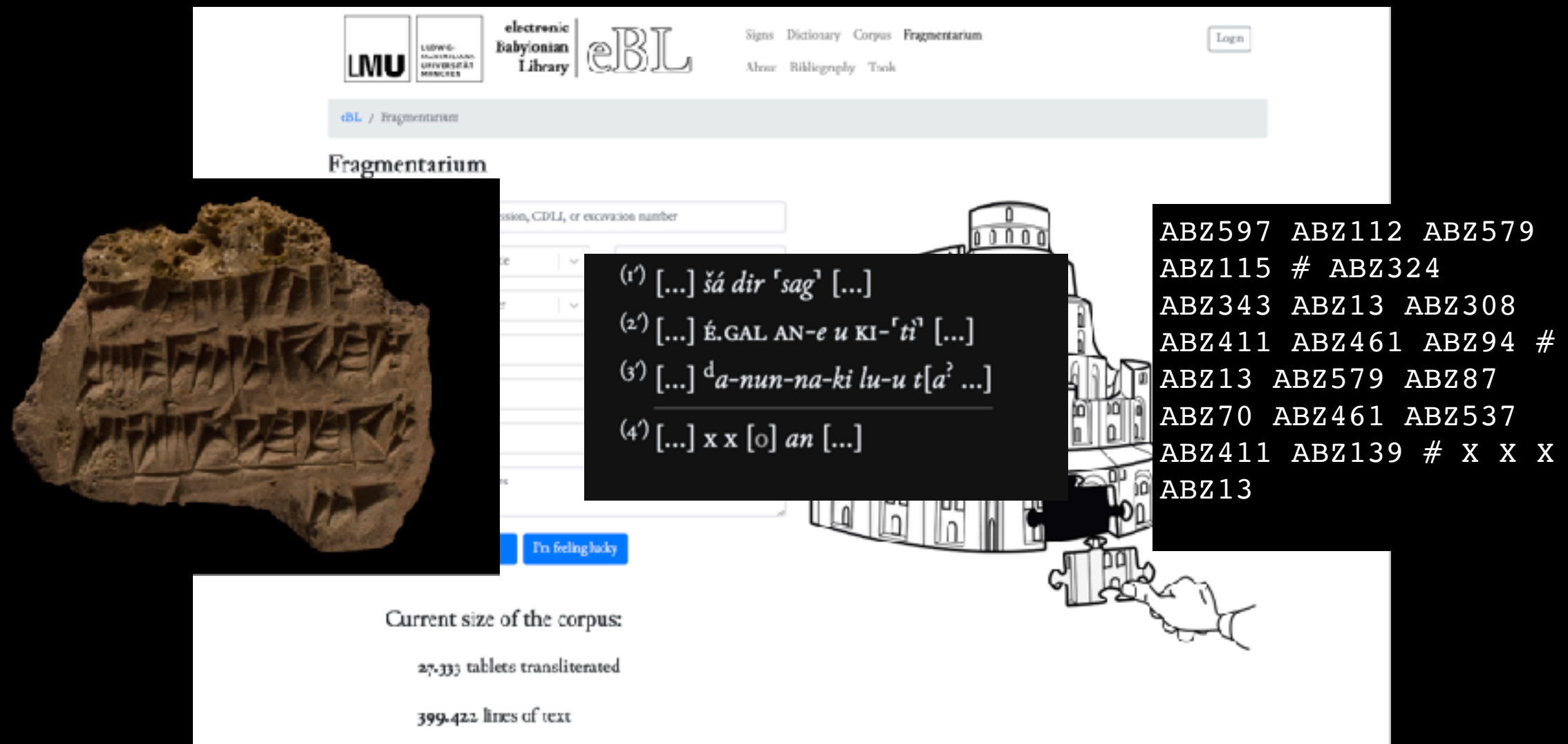
一不女不

le-mut-ti

'evil'

Data

- electronic Babylonian Library (eBL; www.ebl.lmu.de)
- *ASCII Transliteration Format (ATF)*



The screenshot displays the eBL (electronic Babylonian Library) website interface. On the left, there is a photograph of a fragment of a clay tablet with cuneiform inscriptions. The main content area shows a search bar and a list of tablet references and their corresponding transcriptions in ASCII Transliteration Format (ATF). The references are listed on the right, and the transcriptions are shown in a central box. The website header includes the LMU logo and navigation links. The footer provides statistics on the corpus size.

LMU electronic Babylonian Library eBL

Signs Dictionary Corpus Fragmentarium

Home Bibliography Tools

Log in

eBL / Fragmentarium

Fragmentarium

Search by tablet number, CDLI, or excavation number

On feeling lucky

Current size of the corpus:

27.333 tablets transliterated

399.422 lines of text

(1') [...] šá dir 'sag' [...]

(2') [...] É.GAL AN-e u KI-'ti' [...]

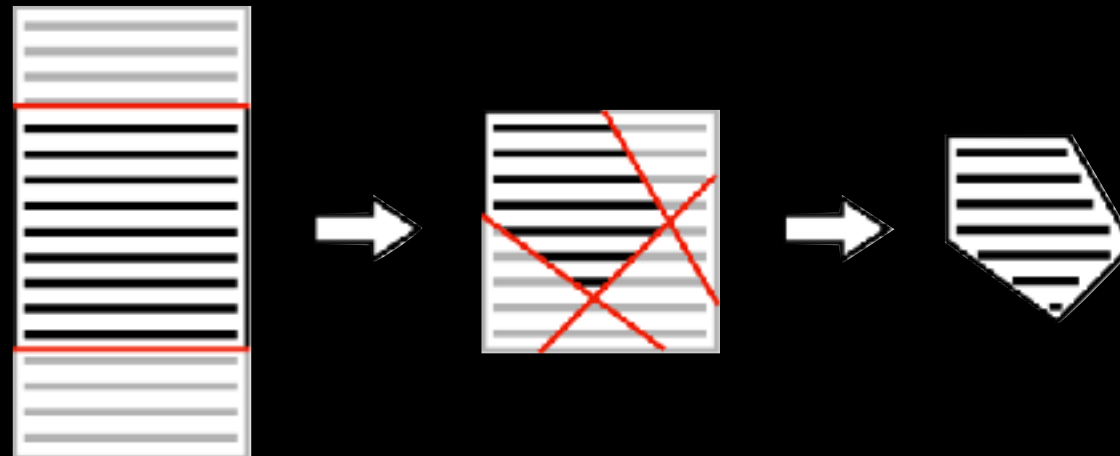
(3') [...] ^da-nun-na-ki lu-u t[a² ...]

(4') [...] x x [o] an [...]

ABZ597 ABZ112 ABZ579
ABZ115 # ABZ324
ABZ343 ABZ13 ABZ308
ABZ411 ABZ461 ABZ94 #
ABZ13 ABZ579 ABZ87
ABZ70 ABZ461 ABZ537
ABZ411 ABZ139 # X X X
ABZ13

Test Data

- “Break” already identified fragments
- Aim: Associate test fragments with their original text



Matching Approaches

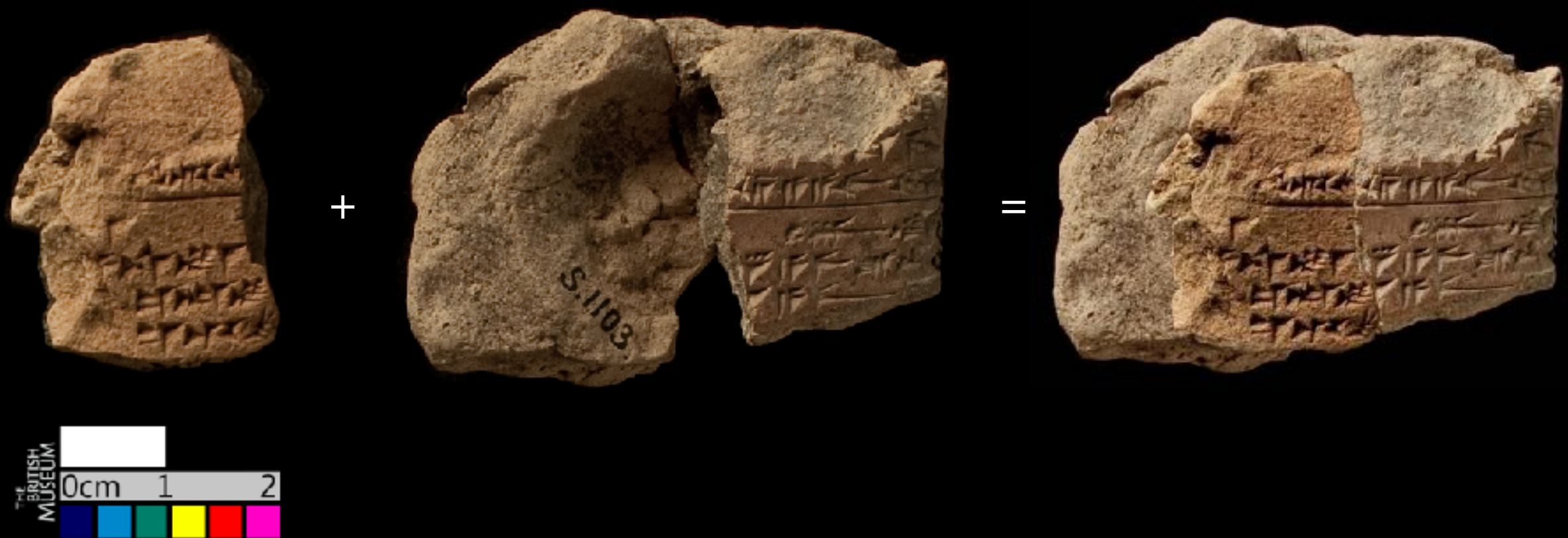
1. Bag of signs + Jaccard
2. Longest common substring
3. Needleman-Wunsch¹ alignment
4. N-gram overlap (different combinations of n + weighting)

¹Needleman, S.B., & Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, 48 3, 443-53 .

Results

Approach	Precision@3
Bag of signs + Jaccard	0.14
Longest common substring	0.90
Needleman-Wunsch alignment	0.79
n-grams ¹	0.91
n-grams (length weighting) ¹	0.92
n-grams (TF-IDF weighting) ¹	0.92
n-grams (length + TF-IDF weighting) ¹	0.94

¹All n-gram results based on $n \in [1, 2, 3, 4]$





Thank you!