USING LNA-NETWORKS on NEO-ASSYRIAN LOCATIONS proof-of-concept

A picture containing scatter chart

Description automatically generated

A picture containing scatter chart

Description automatically generated

Leo Törngren Bonet

2022-08-14

Front Image: Leo Törngren Bonet

# 1. iNTRODUCTION

## Aims

The present study aims to briefly research and discuss different kinds of spatial data through a quantitative analysis of textual data from the Neo-Assyrian corpus from ORACC[[1]](#footnote-1). A key aim of the study is to be a proof-of-concept of what can be achieved when using lexical network analysis (hereby abbreviated LNA), and what understanding the textual data can provide for further spatial and archaeological analyses, such as when utilizing geographical information systems (hereby abbreviated GIS). The present study does not aim to revolutionize or be overly technical and will instead refer to secondary literature regarding relevant topics.

## 1.2 Approach and Questions at Issue

While much can be achieved through using LNA, the present study will focus on a few focal points. Primarily the study will look at locations and their relevant relationship to other locations. The locations covered will be: SN (site-names), GN (geography-names) TN (temple-names) and WN (water-names). As such, a few questions at issues are required to anchor the discussion in a substant frame.

* How may one use or create and use LNA?
* Is there anything in the LNA, that is specifically interesting or unexpected?
* Are there any further historiographical or archaeological conclusions that can be drawn from the data?

## 1.3 Methodology

It is of the utmost importance to note how the programs and code used are heavily limited by the data, and ORACCS formatting of said data. For example, some locations that theoretically should be valid may be missed due to their formatting on ORACC, a pitfall that is unfixable in the scope of the present study – but still important to be mindful of. As such the data in the LNA will not be a perfect and an objective result from the corpus, and will need to be viewed dialectally, with a critical mind. The code that parses through the data works in a very specific way and may be prone to incorrect or biased results. The present study will also generalize locations and their historiographical connections and assume a basic knowledge by the reader of history, linguistics, and context. Some locations are however unknown or otherwise obscure, and we cannot pinpoint them to neither contemporary location nor specific historical context. In a more in-depth study, the method used would be to look at the specific textual context that is looked at when the data is constructed. The workflow of constructing the data and what effects certain settings and programs have on it will be outlined in 1.4.

## 1.4 Workflow

First and foremost, the Neo-Assyrian corpus from ORACC must be processed so that it is in the right format. The present study will not describe this process in detail, but it is done through code by Veldhuis, and is described very well on Veldhuis’ GitHub page.[[2]](#footnote-2) [[3]](#footnote-3) With the data in a better format, it is then run through a PMIZER, coded by Sahala. The PMIZER uses PMI (“Pointwise Mutual Information”), the likeliness that two words appear in a certain defined range of each other.[[4]](#footnote-4) [[5]](#footnote-5) The PMIZER that the present study uses is a slightly modified version.[[6]](#footnote-6)

In the present study the “words” inserted in the PMIZER are the abbreviated tags that are assigned to different kinds of locations in ORACC, i.e.: SN (site-names), GN (geography-names) TN (temple-names) and WN (water-names, i.e. rivers). These are easy to change in the code and is a simple matter of adding or removing valid initial letters of the abbreviation (see fig. 1, specifically “[TSGW]N” which indicates that any of these abbreviations may be used by the code). The string minfreq\_b = 1 (“collocate frequency”), in this case, allows for the code to use sites even if they only appear once, and the string minfreq\_ab = 4 (“bigram frequency”) decides how many times the words need to co-occur to be used by the code. In practice, the bigram frequency makes the data more concise. An increase of the bigram frequency will make the dataset smaller, but in return will show us which locations appear very often with each other. With trial and error, the present study found four to be a good middle-ground. Finally, “window-size” is set as wz = 10, and is the number of words away the code will look.[[7]](#footnote-7)

Scatter chart

Description automatically generated with low confidence

Fig 1. Sahala’s code with the configuration used for the data.

After the PMIZER has output the data in a .txt file, this will then be converted into an excel (.xlsx) file. The .xlsx file can be seen in the Appendix. The .xlsx file will in turn be imported into the Gephi, with edges set to be undirected. While Gephi allows us to study the data through a variety of lenses, the present study uses specific settings that I felt was relevant to the questions at aims. In a broader study, using more of these settings, and comparing their results would give a wider perspective on the data. The layout setting used is a tweaked “Force Atlas 2” and needs to be very precisely configured for the specific dataset, or the network breaks down and becomes unreadable. The key setting most important is normalize edge weight, which solves most issues. The specific configuration can be seen in fig. 2.

Table

Description automatically generatedThe specific statistics that are used for the edges in this study are: edge weight ranking that is shown by a black-red colour scale (in practice this shows the total number of connections), and with the same effect weight is also applied on edge width. As for the nodes, the node size is configured as “eigenvector centrality” ranking (which shows node importance based on a its connections), and the node colour is its modularity class ranking (which cluster of nodes it is part of).[[8]](#footnote-8) Finally, to be able to see the labels, one needs to turn on and make labels visible, and then run “Label Adjust” layout once. Fig 2. Shows the Force Atlas 2 layout   
 settings that were used to make fig. 3

# 2. rESULTS

## 2.1 Results

A picture containing scatter chart

Description automatically generated

Fig. 3. The LNA configured as outlined in 1.4. For the full Gephi file and 4k screenshot, see the Appendix.

In fig. 3 a few patterns and trends are clear: colours are coloured according to clusters, and clusters are roughly grouped according to geography. Since it would be a herculean task to cover every single nuance of the LNA, we will instead briefly look at a few case studies. The red group is a good example of the LNA working, and all the sites presently connected are located in Northern Mesopotamia, around the Assyrian heartland. In addition, many of the orange cities are widely regarded as important in the Neo-Assyrian period, also reflected in the node size. There are a few noteworthy things in this cluster, firstly is the “Libbi-ali[Inner-city-(Assur)]”, which probably represents the city of Assur itself in some capacity, given the close relation to the temple of Esarra – an important temple in Assur.[[9]](#footnote-9) There are a few more cases were Assur appears: once connected to only Urartu, and another as “Mat-Assur”, under which we also can see “Al-Assur” (which is the location’s only connection). “Mat-Assur” is also connected to most big orange locations, but not Arrapha or Kilizi[[10]](#footnote-10) (though obviously we can not know whenever this stems from regarding these as outside the sphere of “Mat-Assur”, or from a lack of data). This leaves a stark lack of Assur in the LNA, and may further explain the “Libbi-ali”. There are a few outliers, notably Ashkelon, located in the levant, and Que[[11]](#footnote-11), likely located in the central south of Turkey, but these are not very connected in the LNA and may as such stem from the settings used in the PMIZER.

Another interesting aspect is that the blue region, which is centred around “Madaya” (the kingdom of Medes), features a lot of cities located in the Northwest of Syria. A few examples are: Naṣibina[[12]](#footnote-12), Tille[[13]](#footnote-13), Raṣappa[[14]](#footnote-14) and Guzana[[15]](#footnote-15). Some of the locations in the Madaya cluster are hard to find reference about however, and some are openly stated to be unknown. An example is Kurbail, probably in the vicinity of Kalhu, but an exact geographical location is unknown.[[16]](#footnote-16) In fig. 3 Kurbail is connected to both Arbail and Isana[[17]](#footnote-17), which means this proposed location could be plausible. Further, the Madaya cluster kind of mirrors the spread that is commonly attributed to the kingdom of Medes[[18]](#footnote-18). This spread is shown in fig. 4 where some of the locations (that either has a location or theorized location in the “Assyrian empire builders” project) are put into GIS. This is understandably a very large area, but the only real oddity is Damascus (which, again, may have been caused by a variety of issues), otherwise all the locations are located in the North. Lastly, we have a pink cluster representing most of the large southern locations such as Babylon, Uruk, Nippur and Sippar. In contrast to the North, despite being fewer in number of locations and smaller in terms of modularity, a lot of the southern locations has their respective temple connected, such as the Eanna in Uruk and the Esagila in Babylon.

Map

Description automatically generated

Fig. 4. Shows some of the locations in the “Madaya Cluster”. Made in GIS with data edited from the Assyrian empire builders project. URL: <https://www.ucl.ac.uk/sargon/peoplegodsplaces/>

The current dataset does however include a lot of isolate nodes which are a biproduct from how the PMIZER works with structuring the data. A common denominator is that all of these has self-loops, or are connected with one or two other nodes, but it does not tell us much. Further, there are the purely geographic names, such as Mat-Assur. One should note that most of the locations starting with “Mat” appear in the green cluster, which in addition mostly consists of “Mat”-locations. Noteworthy is that Akkad is attested here in the green cluster, only being connected to Mat-Sumer and Mat-Akkad. Why the green cluster exists is interesting, but not something that I can explain – perhaps it may stem from people counting countries in certain contexts?

# 3. Discussion and conclusion

## 3.1 discussion and Conclusion

To conclude, while the data yields interesting results, some parts of it are clearly flawed. For example, the facts that rivers are completely missing in the results, the green cluster dilemma, and the isolates are problematic. The data does, surprisingly, however allow us to draw rudimentary conclusions about certain aspects of geography and interconnectivity between locations. While there are oddities in the data, the nodes do broadly conform to respective geographic zones and unsurprisingly these zones may in turn conform to historiographic or archaeological data. What is perhaps more interesting is the locations that we have not yet found archaeologically but appear in the LNA. While not a substant way to determine geography on, studies such as this may aid in giving a broader context and in some cases I can see it able to disprove wild guesses or dated research regarding spatial data. In addition, I can also see a potential use in landscape archaeology, in the future, if the usable corpus grows and programs get more advanced. In such a case the LNA could act as a proxy to spatialize textual data and make it more applicable or usable on archaeological data.

One thing I discovered that can be fixed with coding, and increase the understanding of the data, is a way to more clearly differentiate KUR and URU in the data. Perhaps even being able to configure Gephi in such a way that it shows weighted KUR-URU – which sites only are attested as KUR, URU or both. In the present study I made a brief connection to geography but expanding this to be broader could yield many interesting results. While the LNA made in this study can roughly be described as an upside-down map of the Near East, it would be interesting to further correlate Gephi data to GIS data. Perhaps by comparing distance in Gephi to real world distance, or otherwise using mathematical functions and relationships to explain and contextualize the parts of the data that may be harder to grasp without graphs.

# 4. Bibliography

Blondel, D. Vincent, Guillaume, Jean-Loup, Lambiotte, Renaud & Lefebvre, Etienne. Fast unfolding of communities in large networks, in Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P1000

Pleiades. URL: <https://pleiades.stoa.org/> (2022-08-14)

Porter, B. N., 1993. *Images, power, and politics: Figurative aspects of Esarhaddon's Babylonian policy*. Philadelphia: American Philosophical Society.

Sahala, Aleksi. PMIZER. URL: <https://github.com/asahala/Pmizer> (2022-08-12)

Sahala, Aleksi, Svärd, Sanna. 2021. Language technology approach to “seeing” in Akkadian. In The Routledge Handbook of the Senses in the Ancient Near East. pp. 561-575

The Assyrian Empire builders project. People, Gods and Places. URL: <https://www.ucl.ac.uk/sargon/peoplegodsplaces/> (2022-08-14)

Veldhuis, Niek. The Computational Assyriology project (Compass). URL: <https://github.com/niekveldhuis/compass> (2022-08-12)

Wikipedia. URL: <https://commons.wikimedia.org/wiki/File:Median_empire_map.png> (2022-08-14)

# 5. Appendix

Spreadsheet URL: [https://github.com/Leo-TorngrenBonet/DAA/blob/main/Appendix.zip](https://github.com/Leo-TorngrenBonet/DAA/blob/main/Appendix.zip%20) (2022-08-14)

1. Abbreviation of “the Open Richly Annotated Cuneiform Corpus”. [↑](#footnote-ref-1)
2. Veldhuis, N., Compass. (<https://github.com/niekveldhuis/compass>) [↑](#footnote-ref-2)
3. For the sake of simplicity, this data was in this case provided by Aleksi Sahala from the University of Helsinki. [↑](#footnote-ref-3)
4. Sahala, A., PMIZER (<https://github.com/asahala/Pmizer>) [↑](#footnote-ref-4)
5. Sahala, A., Svärd, S., 2021. [↑](#footnote-ref-5)
6. Questions regarding the modifications done to the PMIZER are directed to Aleksi Sahala. [↑](#footnote-ref-6)
7. For further description of the code, see Sahala, A., PMIZER (<https://github.com/asahala/Pmizer>) [↑](#footnote-ref-7)
8. For modularity class, cf. Blondel, V. et al. 2008. [↑](#footnote-ref-8)
9. Porter, B. N., 1993. p. 60 [↑](#footnote-ref-9)
10. Kilizu in Pleiades <https://pleiades.stoa.org/places/894027> [↑](#footnote-ref-10)
11. Que in Pleiades. <https://pleiades.stoa.org/places/652445664> [↑](#footnote-ref-11)
12. Also known as Nisibis. <https://pleiades.stoa.org/places/874623> [↑](#footnote-ref-12)
13. Tille. <https://pleiades.stoa.org/places/658623> [↑](#footnote-ref-13)
14. Raṣappa. <https://pleiades.stoa.org/places/427568541> [↑](#footnote-ref-14)
15. Guzana. <https://pleiades.stoa.org/places/874739> [↑](#footnote-ref-15)
16. Kurbail is attested as in the vicinity of Kalhu. <https://www.ucl.ac.uk/sargon/peoplegodsplaces/> [↑](#footnote-ref-16)
17. Isana is stated as located on the bank of where the Tigris meets the Great Zab. <https://www.ucl.ac.uk/sargon/peoplegodsplaces/> [↑](#footnote-ref-17)
18. I.e., Wikipedia. URL: <https://commons.wikimedia.org/wiki/File:Median_empire_map.png> (2022-08-14) [↑](#footnote-ref-18)