II. Praxis

Armin Hoenen

Recurrence Analysis Function, a Dynamic Heatmap for the Visualization of Verse Text and Beyond

Abstract The Recurrence Analysis Function (ReAF) is a cross-linguistic visualization tool for (historical) verse text, especially handwritten epics. It can also provide a general visualization of various aspects of prose text. It aims to enable intuitive understanding through explorative data analysis of historical, especially bardic-oral texts.[1] The assumption behind this is that bardic/born-oral and non-bardic/born-written texts differ drastically in the way they employ repetition. The ReAF in its first implementation, as presented here, is a language-independent tool that permits the visual exploration of such structures. Firstly, general aspects and formal characteristics of oral verse text are characterized, before the main technical details and some additional applications of the ReAF are explained and illustrated.

1. Preliminaries

Plato warned through one of his characters that 'Those who acquire it [ref. to writing], will cease to exercise their memory and become forgetful' (Coe 2012). Actually, born-oral and born-written texts do have a number of different characteristics. If Plato's assumptions were true, then those differences might be primarily connected to memory, which Ong (2012) implicitly states: "In a primary oral culture, to solve effectively the problem of retaining and retrieving carefully

1 Born-oral refers to the circumstances of composition of a text. More precisely, it refers to all texts, which originated in the pre-literate ages. Born-oral encompasses such texts, which existed before their first written manifestation, that is, texts where composing them in their entirety happened exclusively in the oral medium without any use of writing. The opposite, born-written, encompasses all texts where writing was involved in the text construction processes.

Published in: Noah Bubenhofer, Marc Kupietz: Visualisierung sprachlicher Daten. Heidelberg: Heidelberg University Publishing, 2018
DOI: https://doi.org/10.17885/heiup.345.474

articulated thought, you have to do your thinking in mnemonic patterns, shaped for ready oral recurrence."[2] One striking feature of born-oral texts connected to memorization is repetition. Repetition was used to a larger extent in preliterate text (Lord 1960). This is a good reason for assuming an indication of oral origin of a text. A formal characteristic of the structure of epic texts is that they are most often composed in lines and half-lines, as verse texts. Verse text is found across cultures although rules of rhyme and rhythm are language specific. Here, patterns of verse-text such as rhyme, meter, the (re)use of particular collocates and syntactical structures or more precisely, a high repetitiveness on virtually all linguistic levels, entails a constantly high rate of priming[3]. This ensures the quick and easy retrieval of these items from memory whenever needed. That these structures involve repetition/recurrence in its various forms is an inherent requirement, as Duggan (1973) points out: "The oral poet […] has a good motive for repeating himself with stylized phrases, namely the need to sing verses before a demanding audience at the rate of ten to twenty decasyllabic lines per minute, a pace far too rapid to permit the constant generation of unique word combinations."[4] With the invention of script, the need for memorization became gradually more obsolete, because knowledge could be externalized, that is stored in books and from there be accessed any time. Verse text continued to be used, but became increasingly less repetitive. Comparing modern lyric texts to epics recorded at the beginning of the chirographic age in verse, Lord (1960) remarks about born-written verse texts that there "may be repeated phrases, but the proportion of them to the whole is small". Characteristics of oral literature are enumerated further by Lord (1960) and together with a broader analysis form the so-called Oral Formulaic Theory (OFT). This is the dominant theory for oral texts. The qualitative structural differences between texts which relate to oral literature, according to the OFT, and those which do not can be visualized and are the subject of the ReAF. Finnegan (1992) gives a detailed account of findings about oral literature or oral poetry in all its scope. She stresses the diversity of

2 The difference between recurrence and repetition is contextual; one and the same linguistic unit can be repeated verbatim but take on a different pragmatic function as the local context differs.

3 Priming is a term used in psychology to refer to pre- or coactivation which enables quicker retrieval from memory. A person who has just heard the word cat can produce the word *dog* quicker from memory than without this prior acoustic exposure since both animals belong to a semantically similar category which is activated along with *cat*. *cat* primes *dog*.

4 Reversing this statement sheds light onto the generally more thought-over character of written (especially printed) texts, where authors pause and think, reformulate and rearrange texts to form a coherent outcome. Once spoken, a word cannot be cancelled or rearranged.

this genre and remains skeptical about whether there is one definition that can be put forward explaining all diversity of oral transmission. Goody (1987) shows how even texts we perceive as born-oral might have nevertheless been deeply structurally influenced by writing and the devices, such as lists, it fosters. However, Lord (1960) gives a guideline for how to test whether a text relates to oral rather than written composition.

The ReAF intends to follow Lord (1960), and to provide an aid in classifying texts as born-oral or born-written, but does not remain neutral about Goody's findings or Finnegan's objections. Finnegan (1992) mentions that epics are typically verse: "'Pure' epics like the Iliad and Odyssey are totally in verse"; in this sense the ReAF is primarily designed for the analysis of verse texts as such, whether influenced by the emergence of writing or not.

2. Recurrence in Oral Text

In a world without script, ensuring or controlling for exact repetition of longer texts exceeds the capacity of memory and is therefore impossible. How would one verify, if not from memory, that each and every word in two performances of the 14.000 verse lines of the *Odyssey* had been the same? And why should rigid identity of a text be of importance at all? Obviously, people from oral traditions did not care much about exact text identity of longer texts such as epics. Bards produced a slightly different text each time they performed the same story. However, many factors influenced the kinds of variation that occurred. Instead of a fixed text as in later (after the onset of writing) performance-based genres (e.g. theater) the oral poet could have presumably memorized a series of events ("Hero goes to war" – "Hero gets lost" – "Hero returns"), which when performed were marked with recurrences of various repertories. Such repertories presumably encompassed, for instance, common sub-plots. When a bard had enough time, he would make extensive use of adjectives and sub-plots when retelling events, which increased repetitiveness. If he had less time, he used less *ornamentation* as termed by Lord (1960). Apart from other non-recurrent characteristics of oral poetry, such as inconsistencies, e.g. where a minor character which had already died reappears, Culley (1967), bards would use repetitive structures, so-called formulas, being non-rigid word groups expressing similar content. This is a concept described by Parry & Parry (1987) in close detail, an example being Homeric heroic epithets identified as Regular Expression: (δόλον|νέφος) ἤγαγε (δῖος Ὀδυσσεὺς|Φοῖβος Ἀπόλλων).

The author of a written verse text does have the time to think about the most appropriate word to place in a certain position, instead of being restricted to using the one which comes most quickly to mind. Thus, since bards used

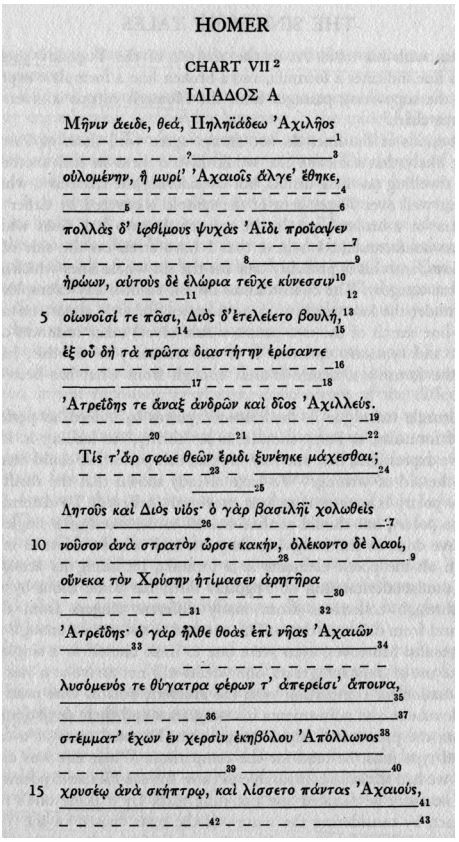
limited repertories and were furthermore forced to use those words which came most quickly to mind, written texts can be considerably less repetitive than oral verse texts.

3. The Visualization History of Verse Text

It has to be mentioned at the start that script itself is a visualization of sound. This is naturally the first visualization applied to verse texts, followed closely by the invention of the first visual prosodic boundary markers, such as dots. For verse texts, line breaks soon often supplemented or substituted the former prosodic boundary markers, which were kept in prosaic texts. The visual representation of poems using line breaks has crossed cultures and writing systems to become the dominant and most wide-spread form for representing poetry. This was stated by Culley (1967): "the unit of composition is usually the single line […] a poem is made by line being added to line. […] The line generally corresponds to a syntactic structural unit in that the end of the line coincides with a natural break such as the end of a sentence or clause."[5] Thus, apart from its other characteristics, from an aesthetic/visualization viewpoint, verse text constitutes a distinguished class of texts. This class can be further subdivided into bardic and non-bardic verse text. Although text itself is already a visualization, only abstract visual transformations allow for a holistic overview of properties of the whole text (at least if it exceeds a certain size). Parry & Parry (1987) and Lord (1960) applied some form of visualization, namely underlining the repeated and near-repeated passages of the beginning of the *Iliad* and other epics, see Figure 1.

Lord related this *extended-scriptural* visualization of verse closely to a test for the classification of oral poetry. His visualization was accepted or slightly modified by many scholars working on orality in subsequent years, for instance Culley (1967), Kailasapathy (1968), Benson (1966), Magoun (1980). A holistic extension of Lord's visualization to an entire epic may have been technically possible, but was never produced. It would have presented another problem with long epic texts: *overview*. In fact, Whallon (1969), Duggan (1973) and later Finnegan (1992) criticize the confinement of this visualization to text excerpts. Finnegan (1992): "Otherwise no overall analysis has been completed, nor any systematic sampling undertaken." Whallon (1969): "They [the visualizations] are not reliable because there is always doubt whether the specimen underlined is typical […]".

5 The invention of the line break as a visual separator interestingly seems to be a feature, which is not necessarily loaned into a new language when a writing system is taken over, but which could have been (re)invented several times. The cross linguistic spread of line breaks underscores their visual effectiveness.



HOMER

CHART VII 2

ΙΛΙΑΔΟΣ Α

Μῆνιν ἄειδε, θεά, Πηληϊάδεω Ἀχιλῆος
οὐλομένην, ἡ μυρί' Ἀχαιοῖς ἄλγε' ἔθηκε,
πολλὰς δ' ἰφθίμους ψυχὰς Ἄϊδι προΐαψεν
ἡρώων, αὐτοὺς δὲ ἑλώρια τεῦχε κύνεσσιν
οἰωνοῖσί τε πᾶσι, Διὸς δ' ἐτελείετο βουλή,
ἐξ οὗ δὴ τὰ πρῶτα διαστήτην ἐρίσαντε
Τίς τ' ἄρ σφωε θεῶν ἔριδι ξυνέηκε μάχεσθαι;
Λητοῦς καὶ Διὸς υἱός· ὁ γὰρ βασιλῆϊ χολωθείς
νοῦσον ἀνὰ στρατὸν ὦρσε κακήν, ὀλέκοντο δὲ λαοί,
οὕνεκα τὸν Χρύσην ἠτίμασεν ἀρητῆρα
Ἀτρεΐδης· ὁ γὰρ ἦλθε θοὰς ἐπὶ νῆας Ἀχαιῶν
λυσόμενός τε θύγατρα φέρων τ' ἀπερείσι' ἄποινα,
στέμματ' ἔχων ἐν χερσὶν ἑκηβόλου Ἀπόλλωνος
χρυσέῳ ἀνὰ σκήπτρῳ, καὶ λίσσετο πάντας Ἀχαιούς,

Figure 1: A. B. Lord, *The Singer of Tales,* 1960, p.143, Scan, (Copyright Harvard University Press).

He enumerates instead all repetitions in long sequences consisting of elements such as "1:13–16 = 1:372–375". A complete list of a work's repetitions in this way is again not easily exploitable in terms of a visual overview. The completeness of it is however superior to that of text excerpts.

To summarize, while a manual visualization line for line, like the one Lord proposed is feasible for smaller sections of text, it would be cumbersome for a complete text. Such a visualization of roughly 14.000 lines of the *Odyssey* would be completely unsatisfactory in terms of an overview.[6]

6 Foley (2002) mentions another individual visual rendering of oral poetry, yet not focussing on repetitions, which is not suitable to generate an overview over many lines.

In the digital age, especially since programmable visualization techniques have become widespread, a holistic analysis/overview is no longer difficult to achieve. ReAF aims at providing consecutive proof of the concept of visualizing repetition, as attempted by Lord (1960). The problem that ReAF will have to overcome is the need to display an overview on the one hand, while on the other hand allowing for examination of the text itself, which was solved by Lord through a confinement to text excerpts.

4. Verbatim Repetition and Bag of Words

In ReAF, verbatim repetitions, that is verbatim repeated verses, are the main feature. They are colored in the same non graded-color. For non-verbatim repeated verses, the words of each verse are used in a bag-of-words representation, that is, an unordered set of the words, where any positioning information is discarded. Depending on the absolute maximum number of shared words with any other verse,[7] each verse's cell is marked with color grading. Consider two lines in two different performances of the same song by the same bard in Lord (1960): "Nit' mu porez ni vergiju daje" and "Nit' mu porez daje ni vergiju" are different in terms of verbatim repetition but the same in terms of bag-of-words repetition. The same scenario is used for repetition within one song. The choice of color for verbatim and bag-of-words should be sufficiently different so as to not confuse the reader.

The number of highly frequent (function) words such as articles, leads to inter-verse repetition even in modern text sentences, especially between verses which are very different. Language itself is redundant when robustly conveying a message. It is repetitive in using the same words for the same references. In other words, there is a base rate of repetition. Since the ReAF is designed to highlight repetition induced by oral formulaic principles, the natural repetition of language can be seen as noise in this context. This noise can be limited by restricting bag-of-words coloring by means of a threshold. Trigrams and other ngrams have been shown to be highly indicative of language (see for instance (Cavnar & Trenkle 1994)). Excluding those verses which share less than three words from a visualization is an ambiguous reduction; it excludes some oral structures such as epithets, but excludes also many arbitrary repetitions. Comparing

7 The more formulaic a verse, the larger this number could get through the same items and protagonists being combined in another verse. This number is also printed on the cell. Another approach would be to color according to the overall number of repeated words or using some measure such as the Jaccard similarity, while keeping the number. Many more approaches thinkable.

the visualizations resulting from a non-restricted bag-of-words repetitiveness visualization with a restricted one, where the boundary is set to above 3, this setting clearly reduces visual overcrowding while higher thresholds result in sparse coloring. Further research employing a usability study will be needed to determine the best measures and thresholds. For the time being, the threshold is heuristically set to three, but the user can modify this. Likewise, the user can choose a color theme which accords to the principles of effective presentation of graded color schemes, as used, for example, in web design.

5. How to Quantify the Oral

We compared various different measures for how indicative they were of an oral composition. One major problem with this is the uncertainty concerning the classification of an epic as born-oral. A gold standard data set for this purpose would ideally contain not only relatively uncontroversial specimens, it would also be balanced in language, text length and other factors. Since it is outside the scope of this article to create such a set, a few well-known examples which were available online (some as base texts of critical editions) have been considered. The main contribution of this article is to demonstrate how overview and details about recurrence can be combined holistically into a visualization. This section is a small side-note on possible measurements for orality (and repetitivity, see for instance Altmann (1988)) which would have to be tested for significance as soon as a suitable data set is compiled. One of the tested measures ranked the texts roughly in accordance with their historically assumed orality; the *verse type/verse ratio* (VVR). This bares a superficial similarity to the Type Token Ratio (TTR). It measures the number of verse types divided by the number of verses. If no verse is repeated the value becomes 1, if there is only one verse type repeated n times, the measure becomes 1/n, which converges against 0, and the value range is thus]0,1]. While the TTR is text length dependent, verses do have different dynamics and distributions. Another measure we tested appears in (Bennet et al. 2003): pairwise comparison of chain-letter texts is achieved by setting the compression size of a text using a compression algorithm. While this measure is readily applicable within one language, the different UTF-8 block sizes and memory sizes lead to them not being readily comparable between languages using different writing systems.

VVR values: **Odyssey** 0.92, **Iliad** 0.94, **Beowulf** 0.98, **Kalevala**: 0.95, **Chanson de Roland** 0.99, **Psalms** 0.97, **Rg Veda** (beginning) 0.85, *Shahname* 0.96, *Parzival* 0.99, *Heliand* 0.97, *Knight With The Tiger Skin* 1, Divina Comedia 1, Faust 0.98, MacBeth 0.99, Luthien 1. For the majority of the most likely born-oral texts (bold), the VVR measure tends to be y low, for born-written texts (underlined)

it is on average higher. There are counter examples, which must be explored visually and qualitatively for the probability of the written medium interfering with the mode of composition. For the time being, measure and visualization are thus meant to be tools for exploration – not classifiers – designed to help the researcher of a specific text to qualitatively evaluate his/her text.

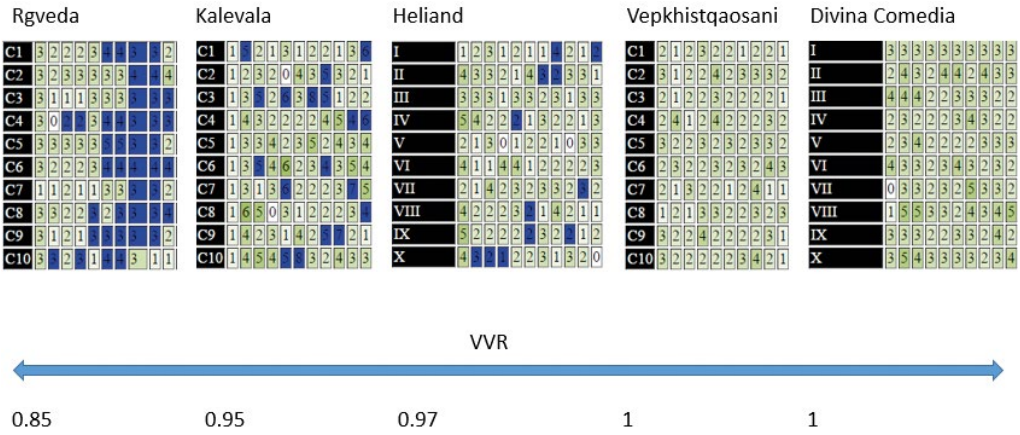


Figure 2: Some selected texts in ReAF overview with VVR value (Copyright A. Hoenen).

6. Technical Details of ReAF

ReAF is a dynamic heatmap with extended features such as text display on demand. While there is software such as R[8] generating heatmaps, for the testing of hypotheses connected with text genesis or text structure, using them without subtle programming intervention is problematic. It requires movement back and forth between the textual and the visual representation, meaning the user must locate certain positions himself each time, jumping from one to the other representation. This is a process prone to errors, especially considering verbatim repetitions which occur close to each other, or thinking about the prevalence of line skips in manuscript copying, or even regressions in eye-movements. The problem is furthermore one of unequal dimension or scaling, where the text must be scrolled in order to be read, whereas the heatmap can only be of use once it presents an overview and therefore displays the whole text or larger portions of it in a small display area (figure 3).

8 http://r-project.org

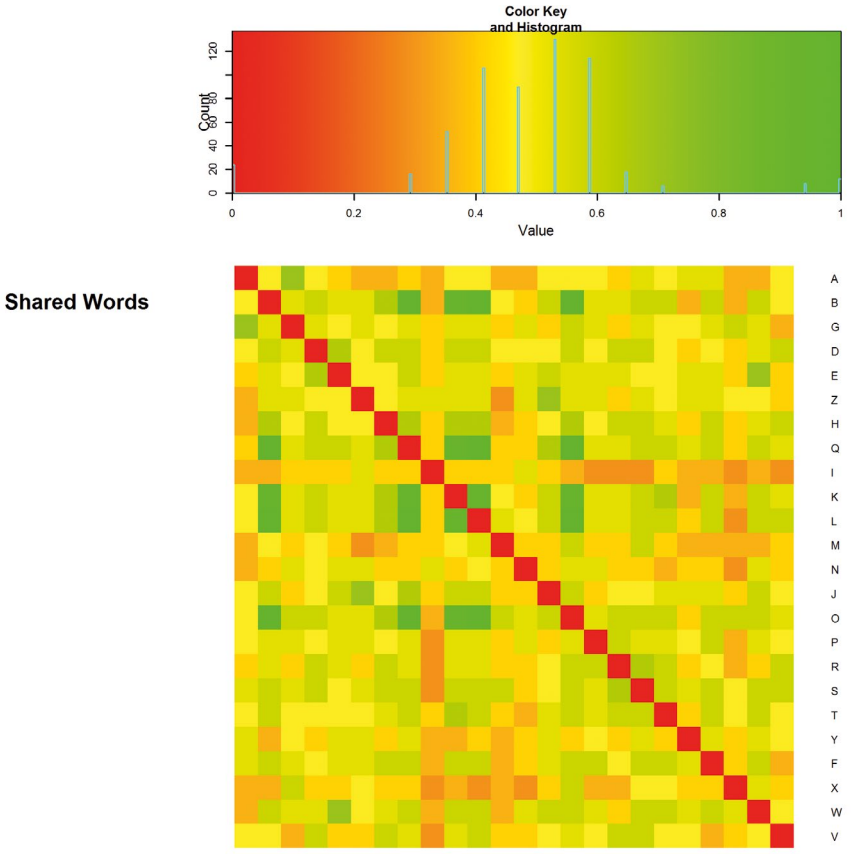


Figure 3: Hoenen ibd. Heatmap generated with R (Copyright A. Hoenen).

Each row and column in Figure 3 represent a chapter of the "Odyssee" and the coloring refers to the maximum proportion of shared words observed in any verse pair of the two chapters (with the diagonal set to 0).

ReAF tries to resolve both the need for such an overview and the need for close reading. It uses HTML, more precisely the rendering of tables in HTML. If in an HTML table, two rows do not have the same number of cells or the cell content differs in length, there is no automatic adjustment of breadth of all rows to one length, although adjusted table cells can easily be generated. This lack of specification allows a presentation of structures with unequal numbers of subunits, such as chapters with unequal numbers of verses. The advantage of this rendering is that differences in length are immediately discernible. ReAF visualizes verse text by representing one verse as one table cell, with a row representing a chapter or paragraph (or in case of plays the speech of one character).

An extension in JavaScript allows the user to read the text by expanding the cell content by double clicking. A second function dynamically colors the sister verses in one color for verbatim repetitions, and in another for bag-of-words-repetitions.[9]In this way not the whole text, but only the passages of interest are readable, allowing the size of the display to be small enough to still exhibit an overview. On mouseover, the sister verse numbers are displayed to facilitate searching. When exploring single cells, the user can choose a color scheme and decide if successive cells should be colored additionally, see Figures 4 and 5.

From the perspective of automated visualization resolving the conflict between text reading or gaining an overview, the ReAF fulfils the requirement formulated by Mazza (2009): "It is necessary for a picture to give the reader as much data as can be processed quickly, using as little space as possible." Furthermore, Shneiderman (1996) postulates a sequence as the Visual Information-Seeking Mantra: "overview first, zoom and filter, then details on demand." This sequence is one that the ReAF complies with, as will be seen in the following example. Furthermore, this is a way to combine distant reading and close reading in one interactive, digital visualization. (Mis)using HTML in this way is however not ideal and more suitable implementations are obviously possible. Text collation in digital scholarly editing, and established visualizations in this field represent possible repositories for more sophisticated technical backends.

7. Application Scenario – Text Exploration of Born-written Text

In the case of a born-written text, it is obvious that a repeated verse line can in principle be copied from a pre-existing instance on paper or from memory. Sometimes, the copied verse text precedes the first text genealogical appearance of the same verse text due to rearrangements made by the author. This is why in philology the term *urstelle* has emerged, referring not to the first *sequential* occurrence of a verse or an ensemble of verses, but to the one place in the text where the *oldest/first authorial version* of the respective verse is located. One example points to an investigation of two similar verses seen through visual inspection using ReAF's rendering of Goethe's *Faust*. This example is simple and does not give new insights into the text genesis of Goethe's *Faust* or any of his

9 To avoid long "loading" intervals, the repetition information is previously computed in the Java programming language and stored directly into the HTML code, which may produce larger files. In this way, the user only has to wait for loading the document upon opening, not at each computational step.

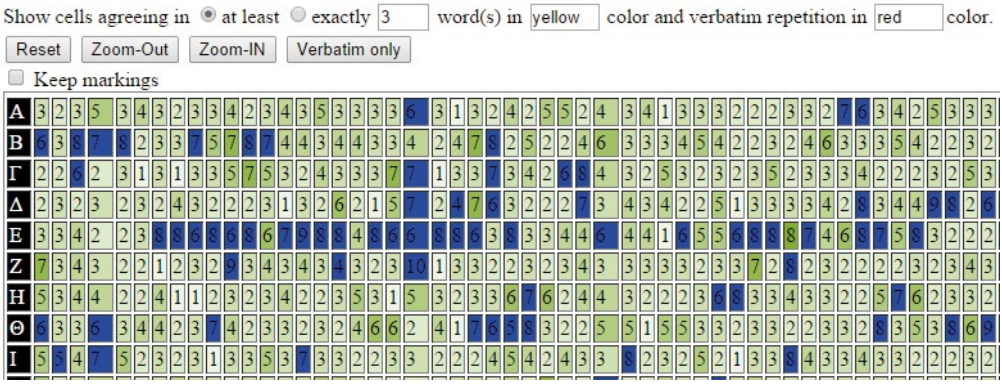


Figure 4: Hoenen ibd. ReAF *Odyssey* overview (Copyright A. Hoenen).

Figure 5: Hoenen ibd. ReAF *The Knight With the Tiger Skin*, close reading (Copyright A. Hoenen).

preferences. Furthermore, the text section can be seen whilst reading without the need for a visualization. *Faust*, Chapter 3, Prologue in the heavens:

Ihr Anblick gibt den Engeln Stärke	*Der Anblick gibt den Engeln Stärke*
Wenn keiner Sie ergründen mag;	*Da keiner dich ergründen mag*
die unbegreiflich hohen Werke	*Und alle deine hohen Werke*
Sind herrlich wie am ersten Tag	*Sind herrlich wie am ersten Tag*

These two passages could have been separately and spontaneously created. However, since Goethe is not an oral poet and the amount of verbatim repetition in his works is fairly low when compared to Homer, an alternative hypothesis would be that the verse has been created only once, and that the other occurrence of the near same sequence has resulted from a copy of the urstelle and an abridgement. Thematically, the subject of the verse is the sun and/or the Lord, the first sequential occurrence being one of three angels singing a praise hymn, and the second occurrence being a chorus of these angels repeating - not verbatim - the last four lines of the first of the three angels. It remains a question for German studies to determine the genesis of the work and which of the two occurrences is the more probable urstelle, drawing from resources such as study notes of the author.

Using the ReAF for born-written verse text such as *Faust*, see Figure 6, 7 and 8, one does not immediately explore places such as the one discussed but spots the obvious preference of Goethe's in Faust for verbatim repetition in close vicinity which works of other composers of written verse text such as Dante's *Divina Comedia* do not show, serving as an entry point to such closer exploration as we have discussed with a simple example.

8. ReAF and Preprocessing

Avestan, an extinct Indo-Iranian language maintained as liturgical language by the Zoroastrians, formed the basis for some individualized renderings for the ReAF, whose production provided the initial impulse for its development. Skjærvoe (2012) connects aspects of the Avestan written witnesses to the OFT. Renderings were based on manuscripts rather than on editions. Lost verses, the extent of which have been determined through numbering, and similar texts, were marked in black. In one rendering, prayers were visualized in order to understand the text structuring function of this special class of recurring elements.

The Avestan manuscripts are mostly written in more than one language, where e.g. Middle Persian commentaries were inserted into the original text. Visualizing the positions of the Middle Persian text in another rendering allowed

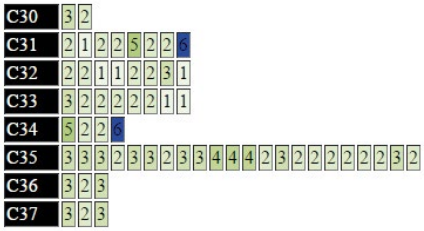


Figure 6: Hoenen ibd. ReAF *Faust* overview (Copyright A. Hoenen).

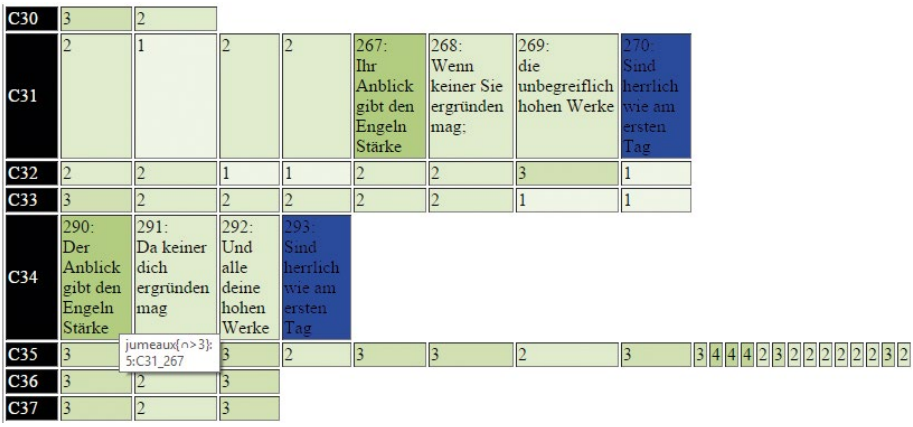


Figure 7: Hoenen ibd. ReAF *Faust*, close reading (Copyright A. Hoenen).

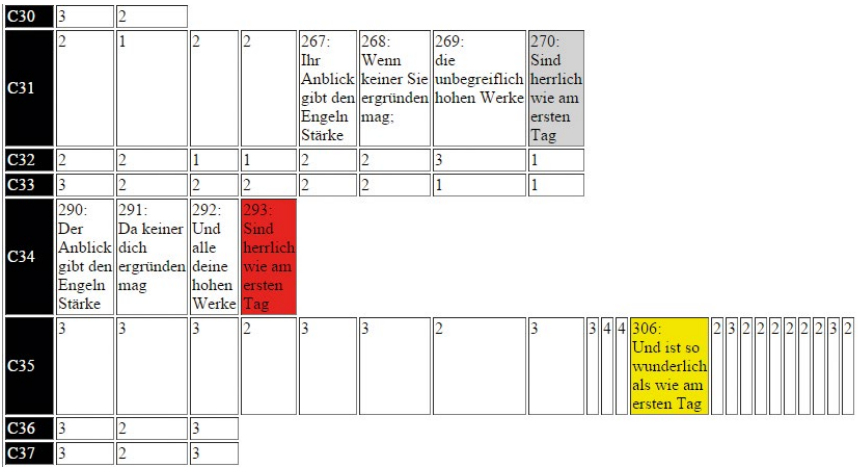


Figure 8: Hoenen ibd. ReAF Faust details, verbatim repeated in red, near verbatim in yellow (Copyright A. Hoenen).

for an intuitive understanding of the text structure, induced or perceived through them, see Figure 9.

One characteristic of the Avestan texts is that the actual performances are so repetitive, that priestly scribes abbreviated multiple repetitions in statements of a meta comment in another language, such as "from here to verse X 3 times"; one rendering of the ReAF spelled out these implied repetitions and colored the relevant cells in light grey, so what the ReAF actually represented was not the manuscript text but the intended oral ritual.

A lemmatic repetition ReAF, where instead of the words, their lemmas have formed the basis for the visualization that was produced. From this, a second visualization highlighted only those verses which were repeated on a lemma but not on a verse basis, which brought forth previously undiscovered principles of formulaic and text genetic alternations, see (Jügel 2015).

These individualized ReAF renderings may serve as examples of what the ReAF can be used for when investigating individual texts, and when preprocessing other data. Other renderings, such as ones displaying syntactic similarity, are among possible additional preprocessing steps.

9. The ReAF principle as a Generalized Visualization Technique for Texts and Language

Instead of visualizing only verse text, the ReAF can be used to visualize text in general. The question arising immediately after breaking free of the "ready-made" chunks called verse, is how to define basic units of visual representation. Unlike verse, where the content is regulated very strictly, as in case of Greek meter which has a very rigid, almost fixed number of syllables, in prose text, sentence length can differ drastically and it is usually only coincidence when two sentences have the same number of syllables. In other words, when the ReAF was applied to verse it did not transform textual information on verse length very much, allocating the same amount of visual space to each cell in the overview, but for prose text this may differ. To summarize, using the ReAF for non-verse text may mean having to compensate for such idiosyncrasies of visualization. In the following its application to non-verse text is outlined.

10. Reference Terms

As a fictional example, the visualization of reference chains is presented. This example is similar to a visualization for reference terms based on trigrams invented by Stede (2007). While Stede's visualization does not supply the close

Figure 9: Hoenen ibd. ReAF for Avestan, showing where the Middle Persian is inserted (Copyright A. Hoenen).

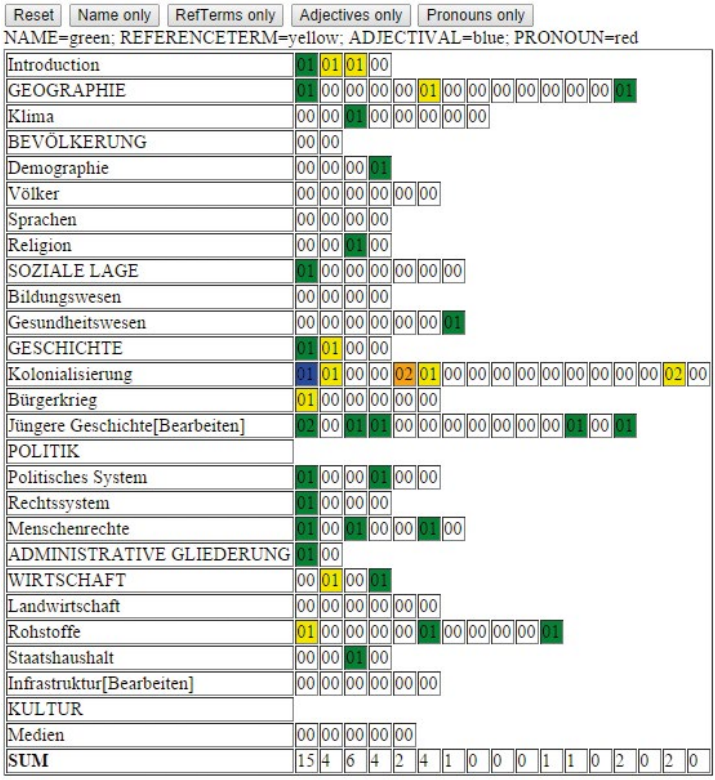


Figure 10: Hoenen ibd. ReAF as general visualization, reference chains (Copyright A. Hoenen).

reading facility, the overview in ReAF looks quite similar. It has, however, been created absolutely independently without prior knowledge whatsoever and any similarity is purely coincidental. Such independent creation of a similar visualization may be a good indicator for its visual effectiveness.

When visualizing reference chains, it is assumed, that a sentence is a basic unit and that sentence length is negligible. A Wikipedia article about Sierra Leone in German from the 3rd of October 2014, is taken and annotated for the occurrence of the title and its (anaphoric) references. Then a visualization is constructed, where the type of reference (noun, pronoun, other) is denoted by color and the number of references within the sentence printed onto the cell, see Figure 10.

With an additional table row summing-up the numeric values at the end, it can immediately be seen that the first sentence of a larger textual unit, in our case in particular and most probably in general, has a higher occurrence of direct reference, which becomes more intuitive via this visualization. Thus, the summary row is one additional customization that can be employed when adapting to a specific research question. An elaboration would be to add a cluster analysis connecting the columns, as is often supplemented in heatmaps (but not in HTML tables).

11. Summary

An initial version of an interactive visualization has been presented (ReAF 1.0). The visualization combines distant and close reading, is platform independent and can be extended through the use of web technology. It can be used to explore, analyze or compare both born-oral and born-written texts. It can also be used for prose text. The ReAF is an interactive visualization which is only feasible in the digital medium and can thus be termed a second generation digital humanities text representation system, not merely imitating print, but extending its capabilities.

12. References

Altmann, Gabriel. 1988. *Wiederholung in Texten*. Bochum: Studienverlag Bockmeyer.

Bennett, Charles, Li, M., and B. Ma. 2003. "*Chain letters and evolutionary histories*." *Scientific American* 32: 76–81.

Benson, Larry D. 1966. "*The literary character of anglo-saxon formulaic poetry.*" *PMLA,* 81 (5): 334–341.

Cavnar, William B., and J. M. Trenkle, J. M. 1994. "*N-gram-based text categorization.*" In *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*, 161–175.

Coe, Michael D. 2012. *Breaking the Maya Code*. New York: Thames and Hudson Ltd.

Culley, Robert. 1967. *Oral formulaic language in the Biblical psalms*. Toronto: University of Toronto Press (Near and Middle East Series).

Duggan, Joseph J. 1973. *The Song of Roland – Formulaic Style and Poetic Craft*. Berkeley: University of California Press.

Finnegan, Ruth. 1992. *Oral Poetry: Its Nature, Significance and Social Context*. Cambridge: Cambridge University Press.

Foley, John. 2002. *How to Read an Oral Poem*. Urbana, Ill.: University of Illinois Press.

Gonda, Jan. 1959. *Stylistic Repetition in the Veda*. Amsterdam: Noord-Holland (Volume 65(3) of Verhandelingen der koninklijke nederlandse Akademie van Wetenschappen, Afd. Letterkunde, N.R. Noord-Hollandsche Uitgevers Maatschappij).

Goody, Jack. 1987. *The interface between the written and the oral*. Cambridge: Cambridge University Press.

Jügel, Thomas. 2015. "*Repetition analysis function (ReAF I)*: Identifying textual units in Avestan." *Indogermanische Forschungen* 120: 177–208.

Kailasapathy, Kanagasabapathy.1968. *Tamil heroic poetry*. Oxford: Clarendon Press.

Lord, Albert B. 1960. *The Singer of Tales*. Cambridge, Mass.: Harvard University Press.

Magoun, Francis P. 1980. *The oral-formulaic character of anglo-saxon narrative poetry. Speculum* 28 (1953): 446–467.

Mazza, Ricardo. 2009. *Introduction to Information Visualization*. London: Springer.

Ong, Walter J. 2012. *Orality and Literacy. The technology of the word*. London: Routledge.

Parry, Millman, and A. Parry. 1987. *The Making of Homeric Verse: The Collected Papers of Milman Parry*. Oxford: Oxford University Press.

Rubanovich, Julia. 2011. "Orality in Medieval Persian Literature." In *Medieval Oral Literature*, edited by Karl Reichl. Berlin: de Gruyter, 653–680.

Shneiderman, Ben. 1996. "*The eyes have it: A task by data type taxonomy for information visualizations.*" In *Proceedings of the 1996 IEEE Symposium on Visual Languages, VL '96*, 336–343.

Skjærvoe, P. O. 2012. "*The zoroastrian oral tradition as reflected in the texts.*" In *The Transmission of the Avesta*, edited by Alberto Cantera. Wiesbaden: Harrassowitz, 3–48 (Iranica 20).

Stede, Manfred. 2007. *Korpusgestützte Textanalyse. Grundzüge der Ebenen-orientierten Textlinguistik*. Tübingen: Narr.

Whallon, William. 1969. *Formula, Character and Context – Studies in Homeric, Old English and Old Testament Poetry.* Cambridge, Mass.: Harvard University Press.

Yamamoto, Kumiko. 2003. *The Oral Background of Persian Epics*. Leiden: Brill.

Online sources of the verse texts, date of last access and traditional classification:

Odyssey: http://titus.uni-frankfurt.de/ 08.09.2014, oral acc. to Lord (1960)
Iliad: http://gutenberg.spiegel.de 08.09.2014 oral acc. to Lord (1960)
Beowulf (halflines): http://www.humanities.mcmaster.ca 26.09.2014 oral acc. to Lord (1960)
Kalevala: http://sacred-texts.com/ 08.09.2014 mainly oral acc. to Foley (2012)
Chanson de Roland: http://www.hs-augsburg.de/ 08.09.2014 oral acc. to Lord (1960)
Psalms: http://www.mechon-mamre.org/ 08.09.2014 oral acc. to Culley (1967)
Rg Veda: http://titus.uni-frankfurt.de/ 08.09.2014 oral, see e.g. Gonda (1959)
Shahname: http://fa.wikisource.org 08.09.2014 acc. to Foley (2002) oral, Yamamoto (2003), Rubanovich (2011) more complex assessment
Parzival: http://titus.uni-frankfurt.de/ 08.09.2014
Heliand: http://titus.uni-frankfurt.de/ 08.09.2014
Knight With Tiger Skin: http://titus.uni-frankfurt.de/ 10.09.2014
Divina Comedia: http://www.filosofico.net 08.09.2014 written
Faust: http://www.gutenberg.org/ 08.09.2014 written
MacBeth: http://www.gutenberg.org/ 08.09.2014 written
Luthien: http://allpoetry.com 08.09.2014 written