From Manuscripts to Archetypes through Iterative Clustering

Armin Hoenen

CEDIFOR

Goethe university Frankfurt

hoenen@em.uni-frankfurt.de

Abstract

Corpora of manuscripts of the same ancient text often preserve many variants. This is so because upon copying over long copy chains errors and editorial changes have been repeatedly made and reverted to the effect that most often no 2 variant texts of the same so-called textual tradition have exactly the same text. Obviously in order to save the time to read all of the versions and in order to enable discourse and unambiguous referencing, philologists have since the beginnings of the age of print embarked on providing one single textual representation of this variety. In computational terms one tries to retrieve/compose the base text which is most likely the latest common ancestor (archetype) of all observed variants using stemmata – that is trees depicting the copy history (manuscripts = nodes, Copy processes = edges). Recently, they have been computed and evaluated automatically (Roos and Heikkilä, 2009). Likewise, automatic archetype reconstruction has been introduced lately (Hoenen, 2015b; Koppel et al., 2016). A synthesis of both stemma generation and archetype reconstruction has not yet been achieved. This paper therefore presents an approach where through iterative clustering a stemma and an archetype text are being reconstructed bottom-up.

Keywords: stemmatology, clustering, archetype text reconstruction, digital textual criticism

1. Introduction

For corpora of ancient texts transmitted (and born) in the age of handwriting, collections of those documents tend to have variant texts which differ slightly due to a variety of reasons among which miscopying, editorial changes such as modernization and stylistic changes may be the most well known due to their illustrativity. In order to provide one text version to the interested scholar, generations of philologists have tried to offer single texts (even though some of those are embedded as base texts in edition types such as critical editions) which they argue best represent the most probably lost authorial original and which somehow mimick an *edicio princeps* initial print exemplar of an autograph.[1] Computationally, the task involves various subtasks. First, manuscript texts need to be digitized, for instance through manual retyping and possibly encoding. To this end, the Text Encoding Initiative[2] provides the elements and rules for an appropriate schema. Texts can be manually or computationally aligned and from this various algorithms can produce stemmata, see Figure 1 that is trees with nodes representing manuscripts and edges representing copy processes or chains thereof, which graphically display (a hypothesis about) the copy history of the variants. Although Cameron (1987) observes that the goal of a textual critic[3] or philologist is rather the reconstruction of an urtext[4] than of the copy history and that the copy history
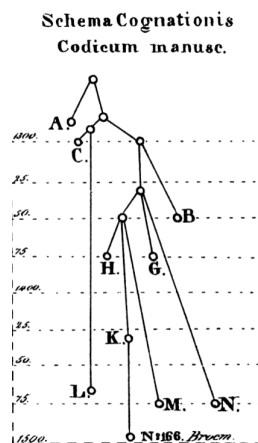

Figure 1: First modern stemma by Schlyter, 1827, from O'Hara (1996) with texts= nodes, copy processes= edges.

is only an intermediate goal on the way to such an urtext, computational attempts at reconstructing urtexts are only very recent and few (Nassourou, 2013; Hoenen, 2015b; Koppel et al., 2016). In this paper, we present an approach which combines the stemma building and archetype reproduction processes in the following way.

Starting from a number of prealigned observed variant texts, those are transformed, a) into pseudo-DNA and b) into bitvectors of so-called *leitfehler*. We derive the technical meaning of leitfehler in this context from the implemen-

---

[1] Consider Shillingsburg (2017) for discussions on (other) purposes of editions.

[2] http://www.tei-c.org/

[3] Textual criticism is the philological discipline of critically reflecting the given variation in a tradition.

[4] The term *urtext* is a loan from German which may refer to the very first (complete) version of a text. The term *autograph* refers to (any) manuscripts written by an author him or herself. This can also be a copy further down the tree and the original mustn't be an autograph in case of dictation for instance. The root of a stemmatic tree can then be an *urtext* or *autograph*, but it can and

---

most often is nothing more or less than the latest common ancestor of all surviving manuscripts and as such often a 2nd or 3rd generation copy of the original (urtext, autograph). Such a latest common ancestor, the best-we-can-do reconstruction based on evidence (further reconstruction if being done, more hypothetical) can be called *archetype* in philological jargon, compare Trovato (2014). In summary, the terms, {original, urtext, autograph, latest common ancestor, archetype} may all refer to the root of some stemmatic tree but all convey different notions about that tree.

Position	m_1	m_2	m_3	m_4	m_5	m_6	X	ζ	π	**Variants**
1	that	this	ðis	ðis	t'	this	ðis	this	this	A-D
2	iz	iz	is	is	is	'	is	is	is	A-C
3	one	one	a	a	an	an	an	one	an	A-D
4	text	text	tekst	text	text	tekst	text	text	text	A,B
pseudo-DNA	DCBA	ACDA	BACB	BACA	CAAA	ABAB	BAAA	AABA	AAAA	**pseudo-DNA**
Leitfehler bitvector	01010	01000	10001	10000	00100	00101	10100	10010	00100	Leitfehler bitvector

Table 1: An example of how pseudo-DNA can be generated which reflects a binary distance between words (same/different). Bitvectors of leitfehler, if the list of good Leitfehler were (ðis, iz, an, one, tekst).

tation by Roelli and Bachmann (2010) which is explained in more detail in the next section. For now it may suffice to understand that this refers to a pruned list (Roelli & Bachmann do not necessitate pruning, but for the application intended here it is necessarily a pruned list) of genealogically indicative variants. That is, by this method, for each variant it is determined if it can serve as a leitfehler (1) or not (0). For an example of the transformations see Table 1.[5] Then several steps are taken to produce both a stemma and an archetype text bottom-up.

1. Vectors or texts are non-hierarchically clustered by some algorithm (Farthest First, Simple K-means, Threshold dependent similarity based clustering) into groups.

2. Incompatible groups (in case one manuscript is detected as belonging to more than one cluster) are resolved into one single cluster.

3. For each cluster, an ancestral text is constructed which then replaces all its descendents in the corpus.[6] New reconstructed nodes are saved as nodes and the ancestral relations as edges of a stemmatic tree which "grows" bottom-up.

4. Iteration of 1.-3. until only one cluster or only clusters with 1 member each exist. The last text reconstruction step assembles the assumed archetypical text (root of the stemma, latest common ancestor to all extant texts).

After obtaining both stemma and archetype this way, we evaluate both of them separately.

2. Artificial Data Sets

For experimentation and evaluation, we use so called artificial datasets or benchmark datasets. Such a dataset is one for which the true copy relationships and thus the true stemmata are known. This is so, because the texts had been given to volunteers (students, friends, etc.) who had copied them by hand recording who had copied which version from whom. The artificial datasets have thus been produced by recent scribes, not medieval ones. Furthermore, their time and generation depth are not representative of

historical cases. Thus, those datasets are but an extremely limited testbed for automatic stemma generation which is by no means representative of the variety of processes and topologies historical stemmata can display (especially cross fertilization or contamination). Furthermore, their small size makes them challenging in terms of the application of data hungry statistical approaches (many machine learning approaches). However, they are the only datasets for which the *copy truth* is known and not a matter of scholarly debate. An alternative is to use stemmata created by the philological community on certain historical works, but this way of conduct has, in the authors view some crucial disadvantages. First, if a stemma, which the community agrees upon already exists, an automatic stemma generation can tendentially at maximum serve to rearrange or corroborate details. Secondly, the evaluation of two different algorithmic methods with different outcomes may be obscured/skewed in favour of the less accurate algorithm by a wrong assumption in the data with only massive amounts of data and massive interpretational intervention providing some remedy, a point which is currently not yet reached in computational stemmatology in the authors view. Thirdly, if the philological community heavily debates the question when (for instance for which tradition) to use stemmatology or even if to deem it useful at all (compare the best-text-editing debate (Bédier, 1928; Maas, 1937), consider also (Haugen, 2015; Hoenen et al., 2017)), demonstrable objectivity in stemma evaluation may be one of the best arguments in trying to win philological users. Fourthly, the method can also be used in non-philological contexts, for instance reconstructing Wikipedia versioning histories (Marmerola et al., 2016) for which no historical uncertainty must be coped with. In sum, the situation is very challenging and some sort of dilemma, on the one hand the benchmark datasets do not represent even a tiny fraction of historical variation. On the other hand, they are the only datasets where we do know the ground truth for sure. The author decided to pursue the path of working primarily on them based under more on the aforementioned reasons, but does not assume that this is the only reasonable or the only objective procedure.

In the context of this study, we use the three previously most used artificial datasets, called Parzival PRZ (English), Notre Besoin NB (French) and Heinrichi HR (Finnish), (Baret et al., 2004; Spencer et al., 2004; Roos and Heikkilä, 2009).[7] PRZ has 21 manuscripts and the alignment has 855 lines, NB features 13 manuscripts of 1035 lines and HR 64

---

[5]The unit of all subsequent comparisons are tokens mostly coinciding with words or punctuation, more precisely alignment positions. Some phenomena, which may be common in historical data such as transpositions are not captured directly but produce several positions aligning with gaps.

[6]This step is similar to other matrix updating procedures for instance the one in the Neighbor Joining algorithm (Saitou and Nei, 1987).

[7]Reduced sets (as a simulation of historical loss) available online. We obtained the complete sets from the authors.

manuscripts of 1208 words.[8] We use the complete datasets and various reduced sets representing loss (among which the reduced sets from the above mentioned authors on PRZ and HR are not present/not analysed separately since they represent only one of many possible loss scenarios).

3. Method

Philologists often operate with tables of numbers of shared variants (agreements and disagreements) and postulate a group of manuscript texts which go back to a common ancestor where the amount of shared features makes this plausible, compare (West, 1973; Timpanaro, 2005). Numerical considerations do play a role in this type of human guided manual clustering, but in cases of doubt, hermeneutical argument can always over-rule numerical considerations. Since we aim at automatical methods, since the computer is not capable of complex hermeneutical reasoning, at least not without explicit knowledge bases and sophisticated algorithmic architectures which to date are not available for this specfic task, we cluster computationally in two flavours and 4 variants.

First, we mimick simplistically philological clustering and detect groups which share a certain percentage of variants (threshold set to 0.95). That is we simply define a threshold of for instance 95% and then for each manuscripts detect all other manuscripts which share at least that amount of variants (proportion of positions in the alignment, where both have exactly the same text). Then we resolve this information into groups (clusters) for the whole set. In case of ambiguity, ambiguous groups are merged.

Secondly, we compile a bitvector of so-called *leitfehler*. Leitfehler are variants, which witness certain distributional patterns on the (true) stemma. Philologically, a leitfehler according to Roelli and Macé (2015) is understood as a genealogically significant error. The authors of the only algorithm implementing a method based on the leitfehler (Roelli and Bachmann, 2010; Roelli, 2014) which according to Roelli and Macé (2015), p.129, still needs development, try to measure the capacity of any variant to serve as such a 'guiding error' for the build-up of a stemma. Roelli and Macé (2015, p.129) state that their method is:

> [...] a subcategory of distance-based methods. Thus the traditional scholarly concept of Leitfehler is taken to be a quantitative one: a variant's usefulness as Leitfehler may be assigned a number or weight. In classical stemmatology the Leitfehler is the most important tool to arrive at a filiation of witnesses that is believed to be most correct representation. [..., evaluating how good a leitfehler a variant is] for every pair of them [...] If one of the four combinations of absence / presence of any of these two candidates [intending a truth table for the variants A

and B: AB ,$A\neg B$,$\neg AB$,$\neg A\neg B$] is not represented in any witness, this is taken to be a hint that both variants suffered their change from absence to presence (or vice versa) exactly once in the tradition, which is characteristic for good traditional Leitfehler (Maas 1937). Such a comparison can be made for all combinations of potential Leitfehler while both Leitfehler in pairs with only three combinations get their score increased.

Thus, the algorithm produces a list, where every variant in the textual tradition gets a value that indicates how good a leitfehler it is (obtained through the comparison of the distributions of this variant in conjunction with each other variant). This list can be pruned at a certain point. We do so using roughly the top third of the list (initial parameter setting of Roelli and Bachmann (2010), numerically 400) and by that token can achieve a binary classification for each variant: good leitfehler (1), bad leitfehler (0). We encode the manuscripts as bitvectors where each dimension corresponds to one leitfehler. The number of positions equals thus the number of position bound variants in the corpus (the sum of the number of variants per aligment position where variation occurrs).

For the clustering, we again detect for each manuscript those manuscripts, which do share a certain percentage of leitfehler (95%). For comparison, we cluster the bitvectors also with Simple K-means (SKM) and Farthest First (FF) as implemented in Weka (Hall et al., 2009).

3.1. Resolving Incompatible Groups

There is one problem emerging from detecting threshold dependent clusters: The outcome of this step can be an *incompatible* grouping where at least one manuscript belongs to more than one cluster. The reason for this is, that any manuscript m can share t or more % with another manuscript a but at the same time t or more % with yet another manuscript b, where the variants both share with m are not the same ones. Consequently, a and b can share less than $t\%$. The same problem is known in biology, where DNA sequences can similarly share nucleic acids of their multiple sequence alignment. Huson et al. (2010) describe how to transform an incompatibility graph into a phylogenetic network by a *divide and conquer* strategy using decomposition. This can be applied here similarly: incompatible clusters (clusters which share at least one manuscript) are all merged into a single cluster.

3.2. Inferring an Archetype for a Cluster

West (1973, p.32 ff) gives some guidelines how to determine if in a group of several (clustered) manuscripts the group internal archetype (root) is to be considered contained. This is the case if there is one manuscript among them, which has no idiosyncratic variants – it will have a high probability of being the group internal root. If there is such a group internal root we take it as the groups ancestor. If this is not the case, Hoenen (2015b) describes how along the lines of a known stemma, the archetype text can be reconstructed. The algorithm is rule-based and simply takes the majority variant at each position of the alignment for the archetype text and in case there are more than one

---

equally most probable variants just writes them all. We use the same algorithm in order to compute an archetype text for each cluster where no group internal archetype[9] exists. Now, we have an archetype text for each group and can build a stemma bottom-up by connecting the archetype of each group with the members by an edge.

3.3. Updating the Corpus

This procedure is repeated. Instead of all manuscript texts as in the first round, in the second round all members of all clusters are replaced by their archetype text. The vectors and variants for this text are updated/generated if necessary. Conflicting information for any one position as to whether or not a variant is a good leitfehler (numerical tie) leads to that position being ignored. Then, the new text corpus is clustered again until either all clusters have only one member or until there is only one cluster. In this way a stemma is built bottom-up. For an illustration of the whole process, see Figure 2.

In case a cluster has for instance 2 manuscripts, there is no easy way to decide which variant to assign to a reconstructed ancestral text. Yet, it is important to reconstruct an archetype with some text at all positions of the alignment. The algorithm pragmatically generates an archetype, which in such cases accumulates all equiprobable variants (if they are most probable). That means, that each reconstructed text may carry more than one variant at each position. In the final root of the stemma, the philologist may then start to decide which variant is the correct one. For future research more sophisticated computational aids are thinkable. For instance one can use some language model and then let some algorithm such as Viterbi find the most probable final text. However, this will only produce variants already extant in the manuscripts, but not such, which have been ancestral, but irretrievably altered. If for instance the word <bath> had been changed to <path> in an early copy event above the archetype and all earlier manuscripts having the correct variant had been lost, the only way to reestablish this variant would be through more sophisticated reasoning involving analyses of individual writers language and orthographic systems, sound shifts, orthographic conventions and the like. In philological literature such cases are abundant. Additionally, meta data such as the age of a manuscripts material (although this doesn't have to coincide with the age of the manuscript text) can and should be taken into account, especially since many manuscripts have been encoded in TEI schemata, which is not the case for the artificial traditions.[10]

---

[9]Here, archetype refers to the root of the cluster not the whole tree.

[10]According to Barabucci et al. (2014): "Non sono mai state sperimentate, tuttavia, soluzioni per generare stemma a partire da documenti codificati in XML." translated by the author as: "Anyway, there have never been experiments of stemma generation starting from XML encoded documents." Especially meta data should not be dismissed by automatic stemmatology as an important source of for instance conditions on th tree topology.
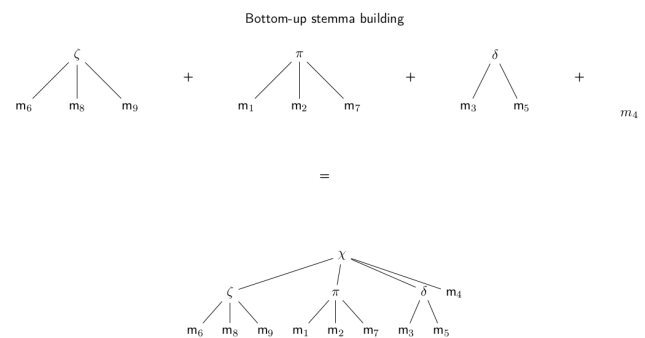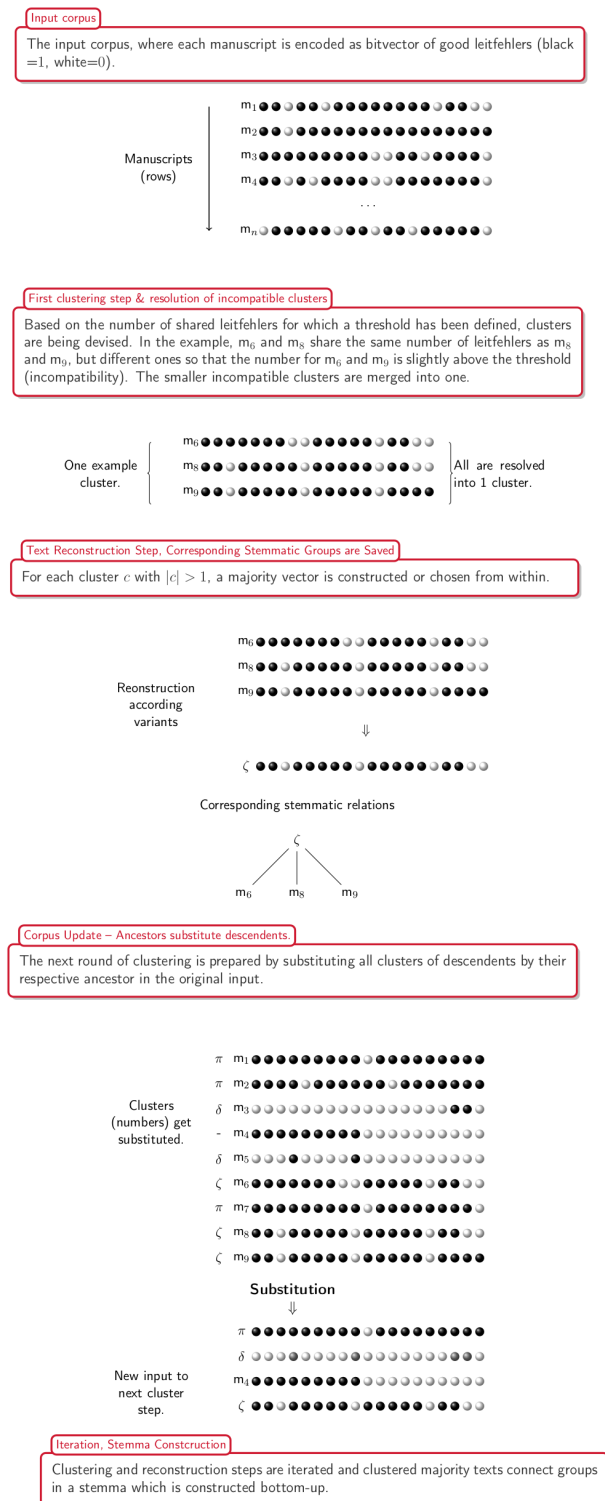
Figure 2: Iterative archetype clustering illustrated.

3.4. Evaluation

For evaluation, we evaluate both stemmata and archetypes. This is necessary because one stemma can be consistent with several archetypes and one archetype can be explicable through several stemmata. For stemmata, we use the Average Sign Distance or ASD, a measure introduced by Roos and Heikkilä (2009).

$$u(A, B, C) = 1 - \frac{1}{2}|sign(d(A, B) - d(A, C)) -$$
$$sign(d'(A, B) - d'(A, C))|$$

A, B and C are nodes present in both the true and the estimated stemma, $d(A, B)$ is the distance of the two nodes in the true stemma defined as the number of edges on the shortest path between them, $d'(A, B)$ the same distance for the estimated tree. $sign(d(A, B) - d(A, C))$ returns so to speak only the sign, discarding length, thus -1 if $d(A, B) < d(A, C)$, 1 in the opposite case and 0 if both are equal. The index equals 1 if both stemmata agree and 0 if they differ ($\frac{1}{2}$ in case of partial agreement) and is computed and turned into a proportion for all such triples. Roos and Heikkilä (2009) provide a python script for ASD evaluation.[11]

For evaluation of the archetype text on the other hand, we use the evaluation introduced by Hoenen (2015b). It gives the percentage of correct variants at all variant alignment positions (ignoring those where all manuscripts agree). We term this the variant hit rate VHR. If a reconstructed archetype has p equiprobable variants at a position, where one is the correct one, $\frac{1}{p}$ is assigned.

Additionally, to assess historical circumstances, for each number from 1 to $n - 1$ (n, the number of ms in the corpus) that many randomly chosen manuscripts were deleted to simulate historical loss. This process was repeated 100 times for each tradition and evaluation results (over all random samples and all loss numbers) averaged.

4. Results

Average ASD results of all evaluation (loss) scenarios for the single traditions are listed in Tables 2, 3, and 4.

Farthest First and Simple-K-Means were initiated with an estimated generation depth from the clustering steps of the former two algorithms (for comparability) and did a little worse than leitfehler and word based percentages. Practically, the archetype texts coincide in almost every instance for both. However, the stemmas of both differ and on the whole on our data sets and with our randomization scheme, the Farthest First algorithm produced a slightly better average result.

The best ASDs were reasonable with NB: 70.98, PRZ: 73.53, HNR: 64.53 and due to a different approach to simulating loss not directly comparable to previous results. For NB the result was obtained at a loss of 2 manuscripts with the threshold clustered leitfehler vectors; for the Parzival and the Heinrichi traditions by Simple K-means at a loss of

Algorithm	VHR	ASD
Leitfehler	0,91	58.86
Word perc	0.91	56.80
FF	0.42	56.95
SKM	0.42	57.78

Table 2: Notre Besoin.

Algorithm	VHR	ASD
Leitfehler	0.85	59.50
Word perc	0.77	59.68
FF	0.85	59.22
SKM	0.85	59.16

Table 3: Parzival.

2 manuscripts. Overall, the best performing algorithm for archetype text production was the one based on threshold clustered leitfehler. As for the stemmas, the word based token difference percentage produced stable results.[12] The overall average over all traditions (including TASCFE) can be seen in Table 5 but is skewed towards TASCFE, which contributed 4 traditions.

5. Discussion

The results, given that many scenarios of loss have been conducted need some brief interpretational intervention. A single specific scenario where manuscript loss is simulated may result in remaining texts which are all quite bad that is to say which differ considerably in their texts from the true original. If this is the case, reconstruction can only result in a bad archetype even in the best case. Naturally, averaging over loss scenarios with partly heavy loss will thus deteriorate average values but may be more faithful to historical circumstances. However, Trovato (2014) conducting a case study on historical loss finds that for many scenarios an amount of loss above 73% is expectable, p. 108. The data sparsity and the small number of copy generations in the artificial datasets indicate that much more research may be necessary in order to reach more conlusive results. However, this is rather secondary to the main aim of this paper, which is to present a method, which produces both stemma and archetype in conjunction.

The algorithm, apart from the fact that it does not disambiguate between variants which are equiprobable or inserts new ones based on the observed ones, is oriented at philological practice and is thus hopefully more straightfor-

---

[11]While on the level of path comparison operating on distance, in terms of the overall manuscript comparison, the ASD is rather a similarity and referred to as Average Sign Similarity by other authors.

[12]For the TASCFE artificial tradition, which is both quite small (137 alignment positions) and unusual (written in Persian, containing 4 subversions), we found similar values in ASD between 63 and 73% and VHRs between 0.68 and 0.88.

Algorithm	VHR	ASD
Leitfehler	0.82	55.47
Word perc	0.78	56.59
FF	0.61	55.04
SKM	0.61	55.63

Table 4: Heinrichi.

Algorithm	VHR	ASD
Leitfehler	0.82	61.03
Word perc	0.77	62.99
FF	0.74	60.98
SKM	0.74	60.39

Table 5: Average over all traditions including TASCFE.

wardly interpretable and adjustable in its results and inner workings. Along the lines of this it produces both a stemma and an (undisambiguated) archetype. Previous methods have produced either separately (independently), or automatic archetype reconstruction followed automatic stemma generation. The production in conjunction is not only oriented at philological practice, it is an essentially stronger hypothesis than either only an archetype text (which can be consistent with several stemmata) or a stemma (which can be consistent with different archetypical texts). By this token it can claim to automatically solve both tasks leaving for manual postprocessing much fewer work than previous attempts if one assumes the philological users want to have both: stemma and archetype. The herein presented first implementation has shown to produce reasonable results, which can and should be improved in future research (especially concerning the stemmatic fit). As such this may make it plausible that future developments can achieve a high degree of automatization in the task albeit with the to be affronted danger of using a blackbox and overseeing crucial givens in the data. Such a position has been challenged in the philological literature repeatedly, see for instance (West, 1973; Griffith, 1984, p. 72, p. 83). If such a high degree of automatization is desirable is another question. The same method of iterative clustering has been applied to produce multilingual stemmata (Hoenen, 2017).

6. Conclusion

Cameron (1987) stated: "Reconstruction is a serious business, and the only point in studying manuscripts at all." A simple algorithm has been presented, which reconstructs archetypical texts automatically. The reconstuctions have been compared to the true archetype for artificial traditions, finding that they performed well.

Bottom-up text reconstruction and stemma building are modelled as alternating and interdependent processes, much like in philological practice. A clustering step clusters together groups of similar texts for which a common ancestral text is being reconstructed. Substituting the children of this text by it, subsequent clustering steps are performed until the program arrives at a stemma and an archetype text. As a criterium for clustering either the percentage of different variants or the number of shared leitfehler was taken. Alternatively, well-known clustering algorithms performed the cluster steps. Averaging over all possible conditions of loss, the produced archetypes had a fair hit rate but the stemmas evaluated not extremely well on average (reasonably in the best scenarios). The main aim of the paper has been to show that automatic archetype production and stemma building can be combined into an interdependent single tasks.

8. Bibliographical References

Barabucci, G., Di Iorio, A., and Vitali, F. (2014). Stemma codicum: analisi e generazione semi-automatica. *Quaderni DigiLab*, 3(1):129–145.

Baret, P., Macé, C., and P. Robinson, e. (2004). Testing methods on an artificially created textual tradition. In *Linguistica Computazionale XXIV-XXV*, volume XXIV-XXV, pages 255–281, Pisa-Roma. Instituti Editoriali e Poligrafici Internationali.

Bédier, J. (1928). La tradition manuscrite du 'Lai de l'Ombre': Réflexions sur l'Art d'Éditer les Anciens Textes. *Romania*, 394:161–196, 321–356.

Cameron, H. D. (1987). The upside-down cladogram: problems in manuscript affiliation. In *Biological Metaphor and Cladistic Classification: an Interdiscplinary Approach*, pages 227–242. University of Pennsylvania.

Griffith, J. (1984). A three-dimensional model for classifying arrays of manuscripts by cluster analysis. *Studia Patristica*, XV(I).

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD Explorations*, 11(1):10–18.

Haugen, O. E. (2015). The silva portentosa of stemmatology Bifurcation in the recension of Old Norse manuscripts. *Digital Scholarship in the Humanities*, 30(2).

Hoenen, A., Eger, S., and Gehrke, R. (2017). How many stemmata with root degree k? In *Proceedings of the 15th Meeting on the Mathematics of Language*, pages 11–21.

Hoenen, A. (2015a). Das artifizielle Manuskriptkorpus TASCFE. In *DHd 2015 - Von Daten zu Erkenntnissen - Book of abstracts*. DHd.

Hoenen, A. (2015b). Lachmannian archetype reconstruction for ancient manuscript corpora. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT)*.

Hoenen, A. (2017). Beyond the tree - a theoretical model of contamination and a software to generate multilingual stemmata. In *Book of Abstracts of the annual conference of the AIUCD 2017, Sapienza, Rome*. AIUCD.

Huson, D. H., Rupp, R., and Scornavacca, C. (2010). *Phylogenetic networks*. Cambridge University Press.

Koppel, M., Michaely, M., and Tal, A. (2016). Reconstructing ancient literary texts from noisy manuscripts. In *CLfL@NAACL-HLT*.

Maas, P. (1937). Leitfehler und Stemmatische Typen. *Byzantinische Zeitschrift*, 37(2):289–294.

Marmerola, G. D., Oikawa, M. A., Dias, Z., Goldenstein, S., and Rocha, A. (2016). On the reconstruction of text

phylogeny trees: Evaluation and analysis of textual relationships. *PloS one*, 11(12):e0167822.

Nassourou, M. (2013). *Computer-Supported Textual Criticism: Theory, Automatic Reconstruction of an Archetype*. BoD–Books on Demand.

Robinson, P. (2015). Four rules for the application of phylogenetics in the analysis of textual traditions. *Digital Scholarship in the Humanities*.

Roelli, P. and Bachmann, D. (2010). Towards Generating a Stemma of Complicated Manuscript Traditions: Petrus Alfonsi's Dialogus. *Revue d'histoire des textes*, 5(4):307–321.

Roelli, P. and Macé, C. (2015). Parvum lexicon stemmatologicum. a brief lexicon of stemmatology.

Roelli, P. (2014). Petrus alfonsi, or on the mutual benefit of traditional and computerised stemmatology. In T. L Andrews et al., editors, *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches*, pages 43–68. Brepols, Turnhout.

Roos, T. and Heikkilä, T. (2009). Evaluating methods for computer-assisted stemmatology using artificial benchmark data sets. *Literary and Linguistic Computing*, 24:417–433.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, 4(4):406–425.

Shillingsburg, P. (2017). *Textuality and Knowledge: Essays*, volume 27. Penn State Press.

Spencer, M., Davidson, E. A., Barbrook, A. C., and Howe, C. J. (2004). Phylogenetics of artificial manuscripts. *Journal of Theoretical Biology*, 227:503–511.

Timpanaro, S. (2005). *The Genesis of Lachmann's Method*. University of Chicago, Chicago.

Trovato, P. (2014). *Everything You Always Wanted to Know about Lachmann's Method, A Non-Standard Handbook of Genealogical Textual Criticism in the Age of Post-Structuralism, Cladistics, and Copy-Text*. libreriauniversitaria.it.

West, M. L. (1973). *Textual Criticism and Editorial Technique: Applicable to Greek and Latin texts*. Teubner, Stuttgart.