Axel Bühler*, Gideon Yoffe, Nachum Dershowitz, Eli Piasetzky, Israel Finkelstein, Thomas Römer, Barak Sober

Exploring the Stylistic Uniqueness of the Priestly Source in Genesis and Exodus Through a Statistical/Computational Lens

https://doi.org/10.1515/zaw-2024-2001

1 Introduction

Recent developments in computational linguistics[1] have led us to discuss the contribution they can make in the field of biblical exegesis. The interest of this article is twofold: Firstly, it shows how certain statistical learning algorithms make it possible to find an editorial partition of a corpus independent of considerations from biblical scholarship and to extract linguistic features (lexemes and syntactic con-

1 Efstathios Stamatatos, »A Survey of Modern Authorship Attribution Methods,« *Journal of the American Society for Information Science and Technology* 60 (2009) 538–556.

Article note: This study was supported by a generous contribution from Mr. Jacques Chahine, delivered via the French Friends of Tel Aviv University, and by a grant from the Tel Aviv University Center for AI and Data Science (TAD). Open access was generously funded by the University of Haifa. We would like to thank John Will Rice for his careful review.

***Kontakt: Axel Bühler**, Département textes bibliques, Institut Protestant de Théologie, 83 Bd Arago, 75014 Paris, France, E-Mail: axel.buhler@ipt-edu.fr. https://orcid.org/0000-0001-5062-8806
Gideon Yoffe, Department of Statistics and Data Science, Hebrew University, Jerusalem 91905, Israel, E-Mail:gideon.yoffe@mail.huji.ac.il. https://orcid.org/0000-0002-1451-6492
Nachum Dershowitz, School of Computer Science, Tel Aviv University, Tel Aviv 6997845, Israel, E-Mail: nachum@tau.ac.il. https://orcid.org/0000-0003-0363-2735
Eli Piasetzky, School of Physics and Astronomy, Tel Aviv University, Tel Aviv 6997845, Israel, E-Mail: eip@tauphy.tau.ac.il. https://orcid.org/0000-0001-9058-2590
Israel Finkelstein, School of Archaeology and Maritime Cultures, University of Haifa, Haifa 3498838, Israel; Department of Archaeology and Ancient Near Eastern Cultures, Tel Aviv University, Tel Aviv 6997845, Israel, E-Mail: fink2@tauex.tau.ac.il. https://orcid.org/0000-0002-5643-2158
Thomas Römer, Chaire Milieux bibliques, Collège de France, 11 Pl. Marcelin Berthelot, 75005 Paris, France; and Faculty of Theology and Religions, University of Pretoria, Lynnwood Rd, Hatfield, 0002, South Africa E-Mail: thomas.romer@college-de-france.fr. https://orcid.org/0000-0002-6490-7678
Barak Sober, Department of Statistics and Data Science, Hebrew University, Mount Scopus, Jerusalem 91905, Israel, E-Mail: barak.sober@mail.huji.ac.il. https://orcid.org/0000-0001-5090-5551

structions) from it. Secondly, it seeks to underline the importance of an interdisciplinary effort to achieve an optimal understanding of the results, to define the limits of the algorithm used, and to pave the way for future work.

The aim of computational linguistics is not to replace the traditional practices of biblical scholarship but rather to introduce a new perspective for addressing the challenges thereof in a quantitative, scale-free, and (as far as possible) assumption-free manner.[2]

The test case for this article is the partition of P and non-P texts in the books of Genesis and Exodus. For a long time, the Documentary Hypothesis as formulated by J. Wellhausen[3] and A. Kuenen[4] dominated biblical exegesis.[5] According to this theory, two sources, J and E, were combined into a JE document, which was then merged with a P document. The fourth document, D, corresponds to the original core of Deuteronomy. In the wake of many scholars in the 1970s abandoning the classic Documentary Hypothesis – especially the sources J and E, or the early dates of those sources – following the work of J. Van Seters,[6] H. H. Schmid,[7] and R. Rendtorff,[8] only the distinction between P and non-P texts has remained as a possible base of consensus, at least for the texts in the first four books of the Torah. This has motivated our choice of corpus for the present study. Non-P texts are not homogeneous and were composed by different hands, and they can be earlier or later than the P texts. Hence, extracting characteristics of non-P texts reveals what is absent from P texts rather than the specifics of an editorial environment.

Our choice to apply the algorithm only to Genesis and Exodus is due to the greater coherence of the priestly framework in these books, known as P^g (the *priesterliche Grundschrift*, »priestly base text«), which runs, according to several scholars, from Gen 1 to Exod 40 (cf. § 5.1). Indeed, the book of Leviticus has its own

2 E. g., Idan Dershowitz et al., »Computerized Source Criticism of Biblical Texts,« *JBL* 134 (2015) 253–271.

3 Julius Wellhausen, *Prolegomena to the History of Israel*, trans. Allan Menzies and John Sutherland (Cambridge: Cambridge University Press, 2013).

4 Abraham Kuenen, *A Historical-Critical Inquiry into the Origin and Composition of the Hexateuch*, trans. Philip Henry Wicksteed (London: Macmillan, 1886).

5 For a history of research, see Albert de Pury and Thomas Römer, »Le Pentateuque en question: position du problème et brève histoire de la recherche,« in *Le Pentateuque en Question*, ed. Albert de Pury (Genève: Labor et Fides, ^{3}2002) 9–80.

6 John Van Seters, *Abraham in History and Tradition* (New Haven: Yale University Press, 1975).

7 Hans Heinrich Schmid, *Der sogenannte Jahwist: Beobachtungen und Fragen zur Pentateuchforschung* (Zürich: Theologischer Verlag, 1976).

8 Rolf Rendtorff, *Das überlieferungsgeschichtliche Problem des Pentateuch*, BZAW 147 (Berlin: de Gruyter, 1977); English translation: *The Problem of the Process of Transmission in the Pentateuch*, trans. John J. Scullion (Sheffield: JSOT Press, 1990).

characteristics, particularly in the Holiness Code (»H«), viz. Lev 17–26. And priestly parts of the book of Numbers are, according to recent scholarship, purportedly the most recent in the Pentateuch and have their own characteristics.[9] We have also decided to apply the algorithm to Genesis and Exodus separately, as the algorithm's partition was significantly more coherent with the one accepted by scholars when doing this than when they were taken together. Does this mean that there is a rupture within the priestly texts of Genesis and Exodus? It might. However, we could also consider that the thematic rupture between the patriarchal traditions and the exodus plays a role as well.

The priestly (P^g) texts from Gen 1 to Exod 40 are characterized by an inclusive monotheism, with the deity gradually revealing itself to humanity, the people of Israel in particular. In the first part, the primeval history (Gen 1–11), God is called »Elohim« and is presented as the creator, the one who sends the flood but nevertheless saves mankind, before concluding with the so-called Noachic Covenant, the promise not to destroy the earth again (Gen 9). The narration is in the third person and is delivered by an omniscient narrator. Genealogies connect the episodes, as do introductory titles using the word תולדות. The second part, the ancestral traditions (Gen 12–50), focuses on a more limited part of humanity: Israel and the surrounding peoples (Moab, Ammon, Edom, Arab tribes, etc.). The covenant is a promise of numerous offspring and of the land, with the rite of circumcision as a sign (Gen 17). The Isaac story is very succinct (14 verses in P^g in Gen 25:19–28:9) and cannot be separated from the Jacob narrative. In P^g, the Jacob story contains the revelation of God at Bethel, confirming the covenant made to Abraham for the giving of the land (Gen 35); then come a list of Esau's descendants in Gen 36, another list of Jacob's family descending to Egypt in Gen 46, and the account of Jacob's death and burial in Gen 49–50. The Joseph story at the end of Genesis does not contain P^g texts.[10] For P^g, name changes play a role; in particular, Abram is renamed to »Abraham«, Sarai to »Sarah« and Jacob to »Israel«. The divine name »El Shaddai« is used in the patriarchal narrative, representing a higher stage of revelation in comparison to

9 See, e. g., Reinhard Achenbach, *Die Vollendung der Tora: Studien zur Redaktionsgeschichte des Numeribuches im Kontext von Hexateuch und Pentateuch*, BZAR 3 (Wiesbaden: Harrassowitz, 2003).
10 Thomas Römer, »The Joseph Story in the Book of Genesis: Pre-P or Post-P?,« in *The Post-Priestly Pentateuch: New Perspectives on Its Redactional Development and Theological Profiles*, ed. Federico Giuntoli and Konrad Schmid (Tübingen: Mohr Siebeck, 2015) 185–201; Konrad Schmid, »Die Josephsgeschichte im Pentateuch,« in *Abschied vom Jahwisten: Die Komposition des Hexateuch in der jüngsten Diskussion*, ed. Jan C. Gertz, Konrad Schmid and Markus Witte, BZAW 315 (Berlin: de Gruyter, 2002) 83–118. Even the scholars who claim the existence of P texts in the Joseph narrative have to admit that these verses do not constitute a comprehensive text; cf., e. g., Rainer Albertz, *Die Josephsgeschichte im Pentateuch: Ein Beitrag zur Überwindung einer anhaltenden Forschungskontroverse*, FAT 153 (Tübingen: Mohr Siebeck, 2021) 134–138.

»Elohim«. In general, God and the nations are presented in peaceful coexistence. The book of Exodus focuses on the history of a single people – Israel. The name »YHWH«, which is revealed to Moses in Exod 6:2–8, is used to designate the deity in the priestly texts of Exodus. The confrontation with Pharaoh is presented in the P narratives as a competition of magicians (Exod 7–9*) rather than as the imposition of divine plagues in the non-P narratives. In Exod 14, the P narrative presents the parting of the sea as an act of creation in which YHWH miraculously saves and »creates« his people, whereas the non-P narrative presents a military confrontation between YHWH and Pharaoh. The P^g narrative may end with the construction of the Tabernacle (Exod 25–31*; 35–40*).[11] The priestly focus in Exodus is on cult, but there is also the motif of the progressive revelation of YHWH. These P^g texts are often dated to the early Persian period (the end of the 6th century or the beginning of the 5th century BCE), as the rites that P^g prescribes (circumcision and Sabbath) do not require a temple and because the universalistic, monotheistic, peaceful vision does not correspond to the more nationalistic and bellicose themes of monarchic times but rather to the ideology of peaceful cohabitation between peoples of the ancient Near East promoted by the Achaemenid empire. In the details, particular expressions and uses of language unite this coherent narrative framework.

As for the non-P texts, they do not form a coherent whole. These may have originally been independent narratives, such as the Joseph story and the Jacob cycle. Furthermore, many narratives exist in both P^g and non-P. Sometimes, they appear one after the other (e. g., creation, the Jacob story, and the call of Moses), while in other cases, they are intermingled (e. g., the flood, the Abraham story, the plagues of Egypt or the parting of the sea).

Our algorithm distinguishes between two sources (P and non-P) without training and extracts lexical and morphological features that show statistically significant differences between the sources. In the following section, we briefly discuss the methodology of our approach. We then discuss the results obtained in regard to biblical exegesis.

The results of the study have enabled us to distinguish between lexical and syntactic properties of P, partially in accordance with previous studies. In addition, the specificities of non-P texts indirectly define those of P. Non-P texts more often adopt the protagonists' point of view, including dialogue or descriptions of their thoughts, whereas P texts prefer third-person narration. In addition, P texts avoid certain themes such as shared meals or dreams, whereas these same themes play a major role in many non-P stories. Geography, locations and traveling are of greater

11 Thomas Pola, *Die ursprüngliche Priesterschrift: Beobachtungen zur Literarkritik und Traditionsgeschichte von P^g*, WMANT 70 (Neukirchen-Vluyn: Neukirchener Verlag, 1995).

interest in non-P texts than in P texts. Separate treatment of the P texts of Genesis on the one hand and Exodus on the other have highlighted many characteristics that are unique to each of the two corpora: Commandment formulae and cult language specific to the story of the building of the Tabernacle appear only in Exodus, while the importance of genealogies, numbers and covenant are unique to Genesis. This result allows for several hypotheses for understanding P: the possibility of two separate redactions for the P texts of Genesis or Exodus or the use of earlier textual material that would have given a different focus to the P texts in different parts of the Pentateuch.

2 Methodology

In a recently published work, we presented a pipeline for a statistical and computational exploration of partitions of texts.[12] The focal point of that work rests on the fact that, rather than attempting an attribution of authorship, we present a stylometry-based explanation and provide the statistical validation of a predefined partition. While there, we presented and scrutinized the statistical and computational aspects of our research, here we examine it from a biblical-exegetical perspective.

In this section, we provide a brief overview of our methodological framework as befits a readership in the humanities.

2.1 Text Parameterization and Mathematical Embedding

The underlying assumption of our work is that significant stylistic differences are manifested in simple observables in natural language processing (NLP), such as the distribution of words (or sequences thereof) or grammatical structures.

We consider three parameters, the combination of which results in a unique mathematical encoding of the text that enables quantitative analysis: (**1**) text representation through (Hebrew) lexemes or two variations of parts of speech (with/without morphological information), (**2**) the length of the sequence of consecutive words that are considered a single feature (i. e., *n*-gram size), and (**3**) running-window width – a parameter that determines the amount of immediate context (surrounding verses) considered for each verse.

For every unique combination of these three parameters, the text is embedded into a mathematical form using the tf-idf (term frequency divided by document

12 Yoffe, »Statistical«.

frequency) encoding scheme.[13] The resulting encoded text is a matrix in which each row represents a single unit of text (i. e., a verse, in addition to the surrounding verses added according to the desired running-window width) and whose columns are the size of all *unique* *n*-grams (i. e., features) in the corpus – where the presence (or absence) of each feature is signified by a specific entry in that column in the following manner: For every verse (row), all unique features that are not found therein receive a score of zero, whereas those that are present receive a positive score (a maximum of one), signifying their statistical significance to that verse (e. g., features found in many verses will receive a low score, whereas rare features will receive a high score).

2.2 Clustering and Optimization

Encoding the text mathematically allows for the application of a clustering algorithm to produce a partition of the corpus into a number of clusters (segments). In our work, we consider only two clusters (i. e., to reflect the P/non-P partition). We use the *k*-means algorithm[14] for the clustering task. The algorithm separates the text into the desired number of clusters by minimizing the intra-cluster variance (i. e., making each cluster most akin to itself) – equivalent to *maximizing* the inter-cluster variance (i. e., making opposing clusters as different from each other as possible). The clustering task is performed in an unsupervised manner, meaning that the decision for clustering the text into two depends solely on the statistical properties of the encoded text and is agnostic to any prior assumptions.

We use a balanced accuracy (BA) metric that signifies the measure of overlap between the scholarly and computerized partitions of the text.

Considering the combinations of the three representations of the text, a range of possible *n*-gram sizes, and running-window widths (see § 2.1), we sought the combination yielding the highest overlap between the computerized and scholarly partitions for the books of Genesis and Exodus. Table 1 lists the optimization results for both books and the statistical significance thereof, while Fig. 1 displays such an optimization round for the Book of Exodus.

13 Akiko Aizawa, »An Information-Theoretic Perspective of tf–idf Measures,« *Information Processing & Management* 39 (2003) 45–65.

14 Trevor Hastie, Robert Tibshirani and Jerome Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics (New York: Springer, 2009), 460–462, § 13.2.1.

Tab. 1: Cross-validated optimization and hypothesis testing results: For each representation, we list the optimal overlap value, its respective uncertainty, and combination of parameters (*rw* for running-window width and *n* for *n*-gram size).

Book	Opt. overlap (lexemes)	Opt. overlap (low-res POS)	Opt. overlap (high-res POS)	*p*-value
Genesis	72.95±6.45 % (*rw*: 4, *n*: 1)	65.03±5.64 % (*rw*: 14, *n*: 1)	76.96±2.91 % (*rw*: 4, *n*: 1)	0.08 (lexemes)
Exodus	89.23±2.53 % (*rw*: 8, *n*: 2)	88.63±1.96 % (*rw*: 9, *n*: 4)	86.53±2.91 % (*rw*: 6, *n*: 2)	0.06 (high-res POS)

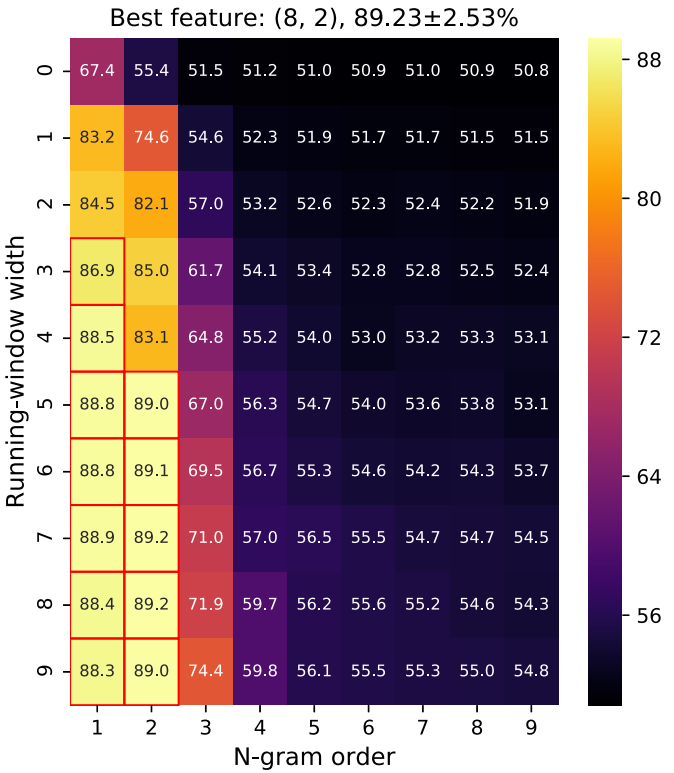

Fig. 1: Optimization results for the book of Exodus (lexeme representation): a grid search over a range of verse running-window widths and *n*-gram sizes to identify the combination yielding the optimal overlap (listed in percent). The cells outlined in red indicate the statistically-significant optimal overlap parameter combinations.

2.3 Hypothesis Testing

Having identified a combination (or combinations) of parameters yielding the optimal overlap with the scholarly partition, we ensure its statistical significance (i. e., that it is not an arbitrary result) through hypothesis testing.

Essentially, the question asked here can be formulated as follows: Given some arbitrary partition of the text, can our algorithm achieve an overlap *at least as high* as it did with the scholarly partition? If the answer is »yes«, this means that the algorithm reaches equally good agreement with partitions of the text that are meaningless, which, in turn, renders the overlap with the scholarly partition coincidental, such that it does not reflect statistically significant stylistic differences between P and non-P texts. Traditionally, the *probability* (i. e., how likely it is that the optimal overlap between the scholarly and computerized partitions is coincidental) that the latter is true is quantified by the *p*-value. To compute it, one must conduct a hypothesis test, performed in the following manner: (**1**) A null hypothesis is defined. In our case, the null hypothesis is that the optimal overlap between the computerized and scholarly partitions is coincidental and does not represent statistically significant stylistic differences between P and non-P. (**2**) Many arbitrary partitions are generated, where the entire clustering and optimization routine is repeated for each (see § 2.2) and the optimal overlap value is stored. The optimal overlap values of these many arbitrary partitions thus form the null distribution (i. e., the distribution of optimal overlap values between the computerized and arbitrary partitions). (**3**) Finally, the *p*-value is derived by computing the number of optimal overlap values in the null distribution that are greater than or equal to the optimal overlap with the scholarly partition, divided by the number of arbitrary partitions generated. For example, if half of the overlap values in the null distribution are higher than the optimal overlap achieved with the scholarly partition, then the *p*-value equals 0.5, which means that the probability of the agreement between the scholarly and computerized partition being coincidental is 50 % (very high). The lower the *p*-value, the more statistically significant the result.

In our earlier article,[15] we presented a novel prohibitive hypothesis-testing routine and thoroughly discussed and demonstrated it. First, recall that a label is an integer number (in our case, either 0 or 1) assigned to each unit of text and signifies which class of text it belongs to (for the scholarly labeling: P/non-P; for unsupervised clustering: cluster 1/2). For each book, the scholarly labeling consists of a sequence of labels, assigning each verse in the book to belong to either P or non-P.

15 Yoffe, »Statistical«.

The novelty in our approach to validating the statistical significance of the optimal overlap result lies in the following caveat: Usually, hypothesis-testing routines (in the scarce instances to which they were applied in stylometric contexts) apply label-permutation techniques in order to determine whether the expert labeling can be associated with the null distribution (i. e., that the optimal overlap value can be reproduced for randomly-generated labels). We found this test to be too lenient for texts, as adjacent units of texts are strongly correlated due to either stylistic, generic, or lexical similarity. An expert scholar is very likely to probe, in some manner, these correlations and label blocks of adjacent units of text as belonging to the same class – resulting in a partition of the text that is extremely unlikely to be randomly generated through label permutation. With this in mind, we reformulate the null hypothesis as follows: The scholarly labeling reflects *some intrinsic correlations within the text* that are not stylistic and result from the semantic similarity between adjacent verses. Therefore, what we wish to test here is whether the scholarly labeling represents a partition that is based on stylistic differences or whether it relies solely on semantic correlations between adjacent verses that would be retained for any block of adjacent verses in the text. To test this null hypothesis, we *shift* the scholarly label sequence (i. e., retain the structure of the labeling but move it across the text such that the label of the last verse becomes that of the first and the rest are shifted to the following verse) and we retain the *structure* of the implicit correlations but change which verses adhere to this labeling. For example, for a sequence of six labels, a shift of one would result in the following:

$\{0, 1, 1, 1, 0, 1\} \Rightarrow \{1, 0, 1, 1, 1, 0\}$.

In this example, vv. 1 and 5 are initially assigned to class 0, whereas vv. 2, 3, 4, and 6 are initially assigned to class 1. After the shift, vv. 2 and 6 are assigned to class 0, and vv. 1, 3, 4, and 5 are assigned to class 1. If there are no stylistic differences between vv. 1, 5 and 2, 3, 4, 6, the optimal overlap reached for the shifted and unshifted cases should (in the simplified case) be similar, because in the absence of other sources of difference between verses (e. g., stylistic differences), the semantic similarity between any three consecutive verses is likely to be comparable in magnitude (and therefore yield similar optimal overlap values). However, if a significant stylistic difference between verses belonging to the original classes 0 and 1 *does exist*, then the overlap with the unshifted label sequence should be higher than that with the shifted one(s), under the assumption that the stylistic signal is stronger than that of the semantic similarity between adjacent verses.

To conclude: We assume that in the absence of a significant stylistic difference between P and non-P texts, any such shift should yield similarly high overlap as was achieved with the scholarly partition because it retains the same relationship

between blocks of verses that are labeled as belonging to the same class (0, 1), with the exception that the verses themselves change. We generate the null distribution by considering all possible shifts of the scholarly label sequence (the number of verses in each book), perform the optimization routine for each shift (see § 2.2), and save the resulting optimal overlap value. Thus, we generate the null distribution and compute the *p*-value.

2.4 Feature Importance Analysis

We leverage on mathematical properties of the *k*-means algorithm and the properties of the mathematically-encoded text, generating two sequences of all unique features (i. e., *n*-grams) – a sequence for each of the two computerized text clusters (see § 2.8 in our recent publication[16]). Each of these lists is equivalent in size to the total number of unique features found in the text, and each feature is assigned a numerical score. This numerical score indicates how effectively each feature distinguishes between the two clusters. To illustrate, if a feature has the highest numerical score within the sequence of features for cluster 1, it implies that this feature either **(1)** exhibits an uneven distribution between the two clusters, predominantly favoring cluster 1, or **(2)** is part of specific combinations of features that are highly characteristic of cluster 1. In other words, the presence of this particular combination of features in cluster 1, coupled with its absence in cluster 2, plays a crucial role in distinguishing between the two clusters.

Our analysis involves randomly selecting subsets of text units and evaluating the statistical consistency of each feature's importance. We ensure that the features associated with one cluster or the other maintain their importance consistently, regardless of the specific subset of text units chosen. Additionally, we gauge the evenness of each feature's distribution across both groups. If a feature is both highly significant and evenly distributed (meaning that it appears in roughly equal proportions in both groups), we conclude that its importance arises from its interactions with other features, particularly specific combinations of features that are unique to each group. In Figs. 2 and 3, we present the most important distinguishing features of both clusters, according to the computerized partition generated for the optimal overlap value parameterization (see Table 1), for Genesis and Exodus.

16 Yoffe, »Statistical«.

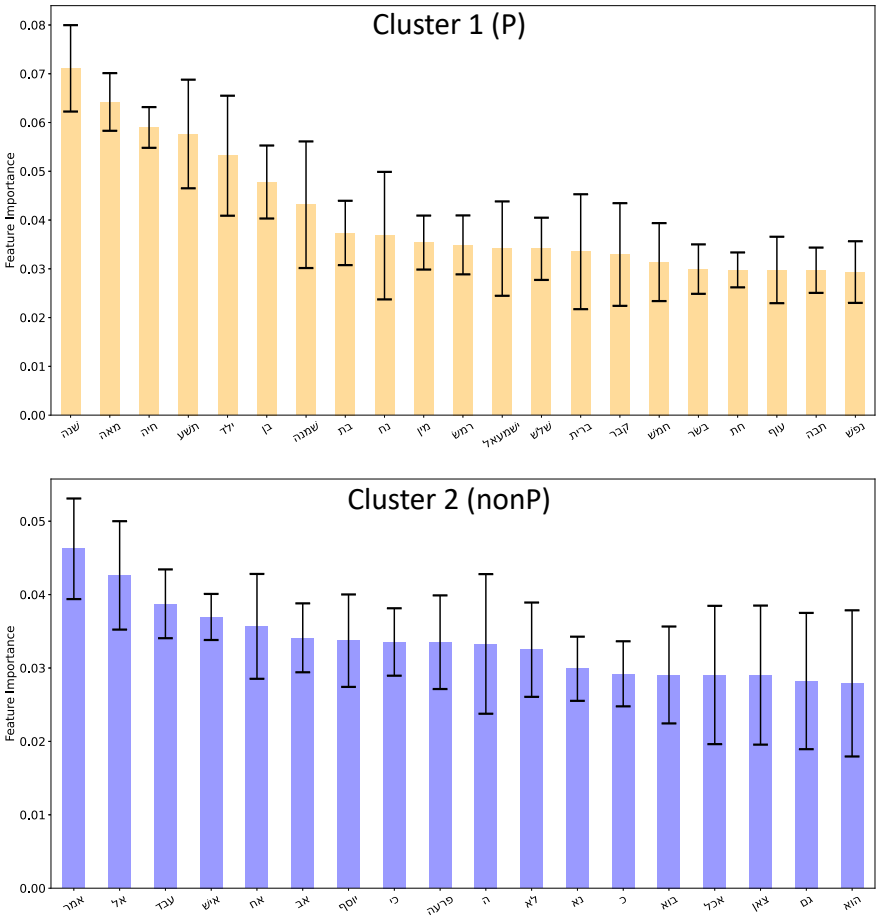

Fig. 2: Feature importance analysis results for the book of Genesis: **lexemes** (running-window width of 4; *n*-gram size of 1).
Note: Only features carrying 75 % of the explained variance are displayed.

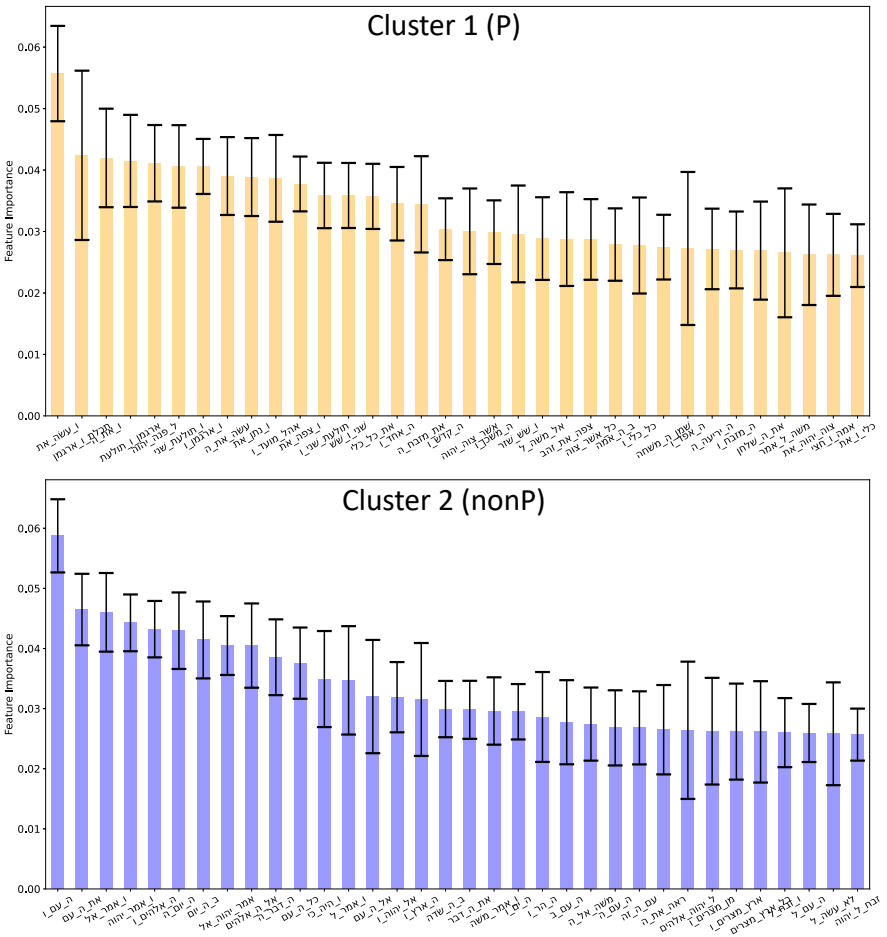

Fig. 3: Feature importance analysis results for the book of Exodus: **lexemes** (running-window width of 6; *n*-gram size of 3).

Note: Only features carrying 75 % of the explained variance are displayed.

3 Incorrectly Attributed Verses

The reconstruction of P texts has achieved a certain consensus among scholars. The various reconstructions for Genesis-Exodus are 97 % similar.[17] In order to measure the effectiveness of the algorithm, we compare the number of correct attributions against the corpus considered as P, defined as follows:[18] Gen 1:1–2:3; 5:1–28,30–32; 6:9–22; 7:6,11,13–15,18–21,24; 8:1,4–5,14–19; 9:1–17,28–29; 10:1–7,20,22–23,31–32; 11:10–27,31–32; 12:5; 13:6; 16:3,15–16; 17*; 19:29; 21:2–5; 23*; 25:7–10,12–17,19–20; 26:34–35; 27:46; 28:1–9; 35:9–13,15,23–29; 36:1–14; 37:1; 46:6–27; 47:28; 48:3–6; 49:29–33; 50:12–13; Exod 1:1–5,7,13–14; 2:24–25; 6:2–30; 7:1–13,19,22; 8:1–3,12–15; 9:8–12; 11:9–10; 12:1–14,28,40–51; 14:1–4,15–18,22–23,26,28–29; 15:27; 16:1–3,6–26,32–34; 19:1; 24:16–17; 25:1–31:18; 35–40*. The unsupervised computerized reconstruction of P by our algorithm deviates more significantly from the consensus. Therefore, the majority of attributions that disagree with the scholarly P/non-P partition can be attributed to the statistical nature of our approach. The algorithm performs efficiently when the texts are thematically different and when P and non-P texts are not mixed (e. g., the Joseph story and the construction of the Tabernacle). On the other hand, when P and non-P texts deal with the same theme (e. g., the two creation stories or the two stories about the call of Moses), or when P and non-P verses are mixed, the algorithm fails to scrutinize the subtle differences that are obscured by larger statistical trends (such as the use of theme-specific words in both accounts). Thus, the stories of the flood (Gen 6–9), the table of nations (Gen 10), and Terah's genealogy (Gen 11:10–32) characterize the cluster that is predominantly associated with P – the dominant stratum, whereas the stories of the plagues of Egypt (Exod 7–9) or the crossing of the sea (Exod 14) are attributed to the opposite cluster.

Compared with the analyses of biblical scholars, the algorithm does not distinguish between thematic words (e. g., the words »Joseph« or »ark«) and words presenting a particular use of an editorial context. Scribes can redefine the meaning of a word or give it a particular meaning (e. g., the word »covenant« is used differently in P texts vs. in Deuteronomistic (Dtr) texts, which are the result of redactors of the history of Israel in the books of Deuteronomy, Joshua, Judges, Samuel, and Kings),

17 E. g., Philip P. Jenson, *Graded Holiness: A Key to the Priestly Conception of the World* (Sheffield: JSOT Press, 1992); Israel Knohl, *The Sanctuary of Silence: The Priestly Torah and the Holiness School* (Winona Lake, IN: Eisenbrauns, 2007); Thomas Römer, »From the Call of Moses to the Parting of the Sea: Reflections on the Priestly Version of the Exodus Narrative,« in *The Book of Exodus*, ed. Thomas B. Dozeman, Craig A. Evans and Joel N. Lohr, VT.S 164 (Leiden: Brill, 2014) 121–150; Avraham Faust, »The World of P: The Material Realm of Priestly Writings,« *VT* 69 (2019) 173–218.

18 There is some discussion of whether all of the chapters Gen 17; 23; Exod 25–31; 35–40 belong entirely to P^g or whether parts of these chapters are due to later (priestly) revisions.

but the algorithm does not distinguish between different meanings of the same word. The algorithm does, however, take into account the lexical environment of a word. In narrative texts, specialists use arguments of narrative logic to distinguish between several strata, which the algorithm does not do.

When P verses are isolated within a non-P text, the algorithm has erroneously assigned its surroundings as P as well (thus Gen 7:6,11; 10:20; 12:5; 37:1). When non-P verses are isolated in a P text, they have been attributed as P (thus Gen 13:6; 16:3; 19:29; 35:15; 47:28; Exod 7:22; 12:28; 19:1). Similarly, when verses are partially P, the algorithm has considered them to be P or non-P according to the verse's direct environment. Thus, the algorithm considered verses such as Gen 2:4a; 7:16a,17a; 8:2a,3b,13a; 12:4b; 25:11a; Exod 8:11a; 24,18a to be P but Gen 13:11b–12a; 16:1a; 21:1a; 25:26b; 31:18a; 35:6a,22b; 37:2a; 41:46a; 47:27b; 49:1a; Exod 2:23a; 7:20a,21b; 14:8a, 10a,21a,27a; 15:22a; 16:35a; 17:1a; 19:2a; 24:15b were considered to be non-P.

In its current state, the exact seams between P and non-P texts are out of the resolution scope of the algorithm and therefore cannot be precisely identified here. However, since the analysis is based on unsupervised learning, it demonstrates that the division into P and non-P can be found independently of the exegetical results and that the algorithm can extract known distinct features as well as a more subtle combination of features, or features whose stylistic signal is less prominent and can be overlooked by humans.

4 Important Features of P

In the following, we offer an exegetical analysis of our results for each of the two books. All data to which this analysis was applied are available online.[19] P texts correspond to about 20 % of the text in Genesis (292/1533 verses), 50 % of that in Exodus (596/1213 verses), and 33 % of the total (888/2746 verses).

4.1 Genesis

For Genesis (unlike Exodus), any partition according to the syntactic features of P and non-P did not achieve high statistical significance and will therefore not be discussed.

19 https://github.com/YoffeG/PnonP.

4.1.1 Lexical Features

We extract significant features from the two clusters of Genesis, generated according to the optimal-overlap partition (i. e., 1-grams of lexemes, running-window width of 4; see Table 1), and list some of the most important ones (see § 2.4) in Fig. 2. Many significant features (and combinations thereof) of the cluster that is most associated with P are agreed to be characteristics of the priestly stratum.[20] In what follows, we discuss the features extracted from the optimal-overlap partition of Genesis (see § 2.4).

The characteristic use of numbers in P (here, in descending order of importance, the algorithm considered the numbers 100, 9, 8, 3, 5, 6, 4 as characteristic of P) appears mainly in the genealogies, such as Gen 5 and 11, but also in the use of ordinal numbers to indicate the months and in the definition of the dimensions of the Tabernacle. The term שׁנה »year« is used in both dates and P genealogies. Furthermore, some names of patriarchs are (almost) exclusively in P texts (נח, ישׁמעאל, חת, עפרון, חנוך, שׂרה). The name אברהם »Abraham« points towards P according to the algorithm, perhaps because of differentiated use, but it also appears frequently in non-P texts ($43^{P}/90^{non\text{-}P}$). Some place names are also typical of P, such as פדן ארם »Paddan-Aram«. The algorithm considers the terms קבר »grave« and מערה »cave« in the account of the cave of the patriarchs in Gen 23 to be features of P. The term »Canaan« is often found in P in the expressions ארץ כנען »land of Canaan« and בנות כנען »daughters of Canaan«. The word ארץ »land« is also found in the P texts of creation, the flood, and as a prefix before the name of a region, e. g., Canaan, Egypt, etc. The term בן »son« appears not only in genealogies but also in typical P expressions such as בן X שׁנה, lit. »son of X year« and idiom. »X years old«, בני ישׂראל »sons of Israel«, etc. The root ילד »to beget/to give birth« is found not only in the genealogies in Gen 5; 11; Exod 6; but also in other P narratives of the patriarchs (Gen 16–17; 21; 25; 35; 36; 46; 48) that focus on filiation. The term דור »generation« is also recognized by the computational analysis as typically P (Gen 6:9; 9:12; 17:7,9,12; etc.), as is the term תולדות »offspring, generations«, which serves to introduce a narrative section or a genealogy. This term structures the narrative and genealogical sections in the book of Genesis (Gen 2:4; 5:1; 6:9; 10:1,32; 11:10,27; 25:12,19; 36:1,9; 37:2). It is often introduced by the determiner אלה »these«. The terms עוף »fowl«, בהמה »living being«, נפשׁ חיה »swarming«, שׁרץ »creeping«, רמשׂ »beast/flesh«, בשׂר

20 Avi Hurvitz, »Once Again: The Linguistic Profile of the Priestly Material in the Pentateuch and Its Historical Age: A Response to J. Blenkinsopp,« *ZAW* 112 (2000) 180–191; Eckart Otto, »Forschungen zur Priesterschrift,« *TRu* 62 (1997) 1–50; Erich Zenger, »Priesterschrift,« *TRE* 27 (1997) 435–446; Joseph Blenkinsopp, »An Assessment of the Alleged Pre-Exilic Date of the Priestly Material in the Pentateuch,« *ZAW* 108 (1996) 495–518; Pola, *Ursprüngliche Priesterschrift*; Heinrich Holzinger, *Einleitung in den Hexateuch*, vol. 1 (Freiburg i. Br.: Mohr Siebeck, 1893).

»cattle«, and מין »kind« are found in the typically P expression »living creatures of every kind: cattle and creeping things and wild animals of the earth of every kind« (Gen 1:24; cf. Gen 1:25–26; 6:7,20; 7:14,23; etc.). These expressions are often associated with the roots רבה »to be many« and פרה »to bear fruit«, an essential theme for P that also appears in the blessings of P accounts as in Gen 17; 48; etc. The term נפשׁ »being« is also used in P texts to refer to a person, e. g., in Gen 12; 17; 36; 46. As for the term כל »all«, it is used overwhelmingly in both P and D texts. The ברית »covenant« plays an important role in the structuring of P (the covenant, though used differently, also appears in non-P texts like Gen 15). In Gen 9, after the flood, God promises not to destroy the world again, establishes the sign of the covenant, the rainbow, and gives the prohibition not to consume animal blood. In Gen 17, the sign of the covenant is the circumcision (מול) of the foreskin (ערלה). This covenant states that Abraham will be the ancestor of many nations (גוי; Gen 17:4–6,16,20; 35:11; the term is also found in P in the table of nations in Gen 10). These Hebrew terms are correctly characterized as P. According to P, God's covenants are linked to a promise of offspring (זרע; cf. Gen 17; etc.) and are valid forever (עולם; Gen 9; 17; 48:4; Exod 12:14; etc.). The term זרע »seed/descendant« is also used by P in the creation narrative in Gen 1. The term בין »between« is used several times to indicate the parties concerned in the covenant in Gen 9 and Gen 17. The term is also found frequently in the creation story of Gen 1, where creation is the result of separation בין »between« different elements – presenting God as the creator is not typical of a national god whose role primarily guarantees protection, military success, and fertility. The transformation of the God of Israel into a creator God appears only in exilic or postexilic texts. Thus, the root ברא »to create« is rightly associated with P (Gen 1:1–2:4; 5:1–2). The use of divine names is particular to the priestly narratives. אלהים »God« is the term used in the primeval history (Gen 1–11), »El Shaddai« for the patriarchs (Gen 12–Exod 6), and »YHWH« from Exod 6:2–3 onward. Here, the algorithm did understand that P uses the term אלהים »God« in a particular way. One of the differences with Holzinger's list of P texts[21] is the fact that the algorithm considers the terms נח »Noah«, מבול »flood« and תבה »ark« as typical of P. This is probably because the flood narrative is much more developed in P than in non-P or because the lexical environment is attached to other P expressions. Nevertheless, all three terms appear in non-P texts as well. The term בת »daughter« is probably considered P not because of its frequency, which is admittedly somewhat higher in the P narratives of Genesis, but probably because of its lexical environment. Thus, the term appears in the expression בנים ובנות »sons and daughters«, which is very frequently used in Gen 5; 11. The preposition אחר »after« appears in the expression

21 Holzinger, *Einleitung*.

אחריך »after you« in the promise to Abraham in Gen 17 and the expression אחרי הולידו »after his begetting« in the genealogies in Gen 5; 11. The appearance of the term מות »to die« as characteristic of P is explained not only by its presence in the genealogies of Gen 5; 11 but also in the succession of each patriarchal generation. Finally, the terms מים »water« and שמים »heaven« play a major role in the creation narrative in P (Gen 1:1–2:4) and in the flood narrative (Gen 6–9*). These two terms also appear in Exodus, where water is mentioned in the account of the duel with the magicians (Exod 7–9*), in the crossing of the Sea of Reeds (Exod 14), which parallels creation in Gen 1, and as a means of purification during the building of the Tabernacle (Exod 29–30; 40). This latter function of water is probably the root of its symbolism in the other narratives. The term רקיע »firmament« and the root אור »to shine« appear in the creation story of Gen 1 (P) but are of little significance elsewhere.

On the non-P side, terms like »Joseph« and »Pharaoh« in Gen 37–50 are non-P features, since the story of Joseph is non-P. The algorithm considers the name »YHWH« to be non-P, since P uses only »Elohim« or »El Shaddai« in Genesis to refer to God. The terms »brother« ($15^{P}/180^{non\text{-}P}$), »father« ($13^{P}/219^{non\text{-}P}$), and »mother« ($4^{P}/33^{non\text{-}P}$) as features of non-P can be understood through a greater emphasis on family in the non-priestly patriarchal accounts, whereas P emphasizes genealogy. The terms אדון »master«, and עבד »slave/servant« reflect the hierarchical structures of the household (בית) of the wealthy landowners in the narratives of the patriarchs but are of no interest to the priestly editors. Similarly, non-P texts show more interest in livestock, with terms such as חמור »donkey« or צאן »flock«. Dialogue is more present in the non-P stories than in P. Thus, the analysis considers the terms that open direct discourse, דבר »to speak«, אמר »to say« and נגד »to tell«, to be typical non-P terms, as well as the set of Hebrew particles in direct discourse (אל, כי, ה, לא, נא, גם, הנה, אם, מה, עתה, זה). The prepositions ל and כ are also considered non-P, although the reason for this is less clear. The use of personal pronouns such as הוא »he/him« ($27^{P}/185^{non\text{-}P}$) and אנכי »I/me« ($1^{P}/77^{non\text{-}P}$) is considered more typical of non-P texts, probably because of the narrative style and direct speech. In the second case, two spellings are possible for the possessive pronoun and P systematically uses the short form (אני) with the exception of Gen 23:4, a verse that may belong to a later revision.

The term איש »man« can be used in many ways: »man«, »husband«, »human«; »someone«. Its use alone or in broader expressions is significantly more frequent in non-P texts of Genesis ($7^{P}/152^{non\text{-}P}$) but the difference is insignificant in the Exodus texts ($30^{P}/66^{non\text{-}P}$). This may be the result of language evolution rather than a deliberate or theological change on the part of P. The term יד »hand« ($4^{P}/91^{non\text{-}P}$) hardly ever appears in the P texts of Genesis. In the priestly texts of Exodus and Numbers, the expression »by the hand of Moses/Aaron« designates the human intermediary

of divine action. »The hand« is often also mentioned in the miraculous deeds that Moses and Aaron perform. The terms בוא »to enter« ($20^{P}/197^{non\text{-}P}$), שׁוב »to return« ($0^{P}/68^{non\text{-}P}$), הלך »to go« ($13^{P}/108^{non\text{-}P}$), and שׁכם »to rise« ($0^{P}/8^{non\text{-}P}$) are features of non-P; this may reflect a stronger interest in places and traveling in the original texts, probably composed to legitimize sanctuaries or as etiological narratives, whereas these aspects are less marked in the P texts. For the same reasons, the word עיר »city« ($4^{P}/51^{non\text{-}P}$) is considered to be non-P by the algorithm. For P, access to God is guaranteed by rituals and the cult led by priests. Oneiromancy only plays a role in non-P stories in which the word חלום »dream« ($0^{P}/34^{non\text{-}P}$) is found, whether in the story of Abimelech (Gen 20), Jacob (Gen 31) or Joseph (Gen 37–41). The motif of the common meal is very present in the ancestral narratives and is found in many non-P stories, but not in P, which has little interest in this matter: אכל »to eat« ($2^{P}/63^{non\text{-}P}$; the expression »for food« לאכלה is nevertheless typical of P: Gen 1:29–30; 6:21; 9:3). Another motif is that of »(water) springs« (עין), reflecting the traditional tension between nomadic animal breeders and sedentary farmers (cf. Gen 24).

This motif was not of interest to the P editors. עין in the sense of »eye« is used by non-P texts to show the subjective point of view of one or another protagonist, unlike P, for whom the narration is omniscient and made by the narrator. The expression »in the eyes of YHWH« is anthropomorphic and avoided by P.

Certain roots are particularly present in non-P narratives but are rarely, if ever, used in P narratives in Genesis: ידע »to know« ($0^{P}/57^{non\text{-}P}$), מצא »to find« ($0^{P}/56^{non\text{-}P}$), לקח »to take« ($18^{P}/124^{non\text{-}P}$), שׂים »to put/to set« ($1^{P}/47^{non\text{-}P}$), עשׂה »to do/make« ($21^{P}/132^{non\text{-}P}$), עמד »to stand« ($0^{P}/16^{non\text{-}P}$). It is difficult to define whether these are simply thematic differences or genuine redactional variations since in biblical texts, a narrative is often built around one or more thematic roots appearing many times in a few verses. For some roots, the difference, though less marked, can also be seen in Exodus (ידע; $12^{P}/33^{non\text{-}P}$; מצא; $3^{P}/19^{non\text{-}P}$; עמד; $4^{P}/14^{non\text{-}P}$), while for others this is no longer the case: (עשׂה; $41^{P}/39^{non\text{-}P}$; שׂים; $237^{P}/86^{non\text{-}P}$; לקח; $20^{P}/30^{non\text{-}P}$).

4.2 Exodus

4.2.1 Lexical Features

For the P texts of Exodus, we comment on 3-grams (Fig. 3). Certain words appear in both the 1-gram and the 3-grams because they are typical of the description of the construction of the Tabernacle. Among the 3-grams characteristic of P, we find the typical P command formulae: אשׁר צוה יהוה »as YHWH commanded« (32^{P} and $2^{non\text{-}P}$ in Exod 7:20; 34:4), כל אשׁר צוה »all as X commanded« ($12^{P}/0^{non\text{-}P}$), צוה יהוה את »YHWH commanded« (21^{P} and $1^{non\text{-}P}$ in Exod 34:4). We find also the expression לפנה יהוה »before YHWH« (20^{P} and $1^{non\text{-}P}$ in Exod 34:4) and the call formula אל משׁה לאמר

»to Moses, saying:« (12^{P} and $1^{non\text{-}P}$ in Exod 13:1). Furthermore, the algorithm has extracted the lexemes used to describe the construction of the Tabernacle in Exod 25–31; 35–40 but does not give features of the P-texts that would be found elsewhere. We find the following features: the different names of the Tabernacle, »the holy one«, »the dwelling«, »the tent of meeting«; the materials used for the construction, »acacia wood«, »pure gold«, »bronze«, »linen«, »blue, purple, crimson yarns«, etc.; the spatialization, »around«, »outside«; the dimensions, »length«, »cubit«, »five«; the components, »altar«, »curtain«, »ark«, »utensils«, »table«, and YHWH's orders to Moses, »You shall make…«. Thus, the algorithm has a good understanding of the terms specific to the construction of the Tabernacle.

The 3-grams' non-P features contain expressions such as the construction והיה כי to express »when« or »because« ($0^{P}/10^{non\text{-}P}$), זבח ליהוה »to sacrifice to YHWH« ($0^{P}/9^{non\text{-}P}$) or לא עשׂה ל »not to make« ($1^{P}/5^{non\text{-}P}$). In the latter case, negation is responsible for this attribution, as P rarely uses it. Some expressions are used by both P and non-P, so it is not so easy to understand how the algorithm came to the conclusion that they are features of non-P, such as expressions using היום »today« or אמר »to say«. Finally, some 3-grams do not form expressions but are sequences of lexemes that actually appear more in non-P texts, often because they are used several times in the same pericope: ההר ו »the mountain and« (cf. Exod 19; 24; 32), בשׂדה »in the field« (cf. Exod 9; 22–23), ראה את ה »to see the«. For the 1-gram, the almost exclusive use of the name YHWH by P in Exodus leads the algorithm to consider divine appellations such as יהוה אלהים »YHWH Elohim« ($4^{P}/35^{non\text{-}P}$) or האלהים »HaElohim« as typical of non-P ($0^{P}/29^{non\text{-}P}$).

The use of the word עם »people« appears primarily in the non-P texts because the priestly redactors usually preferred to use the terms עדה »assembly« or בני ישׂראל »Sons of Israel«. The word »I« in the long form אנכי is considered non-P because the short form אני appears in P texts. The expression »to YHWH« appears 24 times in non-P texts, e. g., »to cry out to YHWH«/»to speak to YHWH«/»to turn to YHWH«, whereas P avoids this expression. This is easily understandable by a desire to give YHWH the initiative in all interactions. In P, it is he who demands, commands, and speaks. There is little dialogue. As for the terms מצרים »Egypt« ($53^{P}/119^{non\text{-}P}$) and פרעה »Pharaoh« ($33^{P}/82^{non\text{-}P}$), they are indeed quantitatively more frequent in the non-P texts of Exodus as in Genesis.

4.2.2 Syntactic Features

As we have already seen, non-P texts more often adopt the protagonists' point of view by including dialogue or thoughts, whereas P texts prefer a third-person narration. One of the consequences thereof is the privileged use of third-person singular or plural suffixes, unlike non-P texts, where first-person singular or sec-

ond-person singular suffixes are more often used. Moreover, the extreme use of the third person in P texts can also be explained by the presence of pleonasms that use a form with this suffix: עמו, אתו, etc.[22] Concerning verbs, the qal and piel stems as well as the *qatal* conjugation in the second-person masculine singular are prevalent in P texts. This is understandable because of P's theology, according to which God orders using the second person and then the protagonists act according to YHWH's orders. On the side of the non-P texts, the *wayyiqtol* narrative form, mostly in the qal stem in the third-person masculine singular, is significant, although it is also present in P texts. According to our algorithm, another peculiarity is the use in non-P of »name in construct state + toponym«. P seems to have avoided this type of syntactic construction because of a lesser interest in localizations. The remaining terms are persistent elements. Further analysis would be needed to understand the relevance of the distinction made by the algorithm.

4.3 Summary

Our method performed separate partitions of the books of Genesis and Exodus into two clusters, which under a specific parametric setting was able to attribute the majority of verses (or sequences of verses) associated with P to the same cluster, in both books. Out of this cluster, we extracted features that account for most of the distinction between the two clusters and showed that many of them are typical features that biblical scholars associate with priestly texts. In addition, other P features have also been found that may be specific to a single narrative; they correspond to repeated use of an expression or a significant theological theme (such as water). On the other hand, the features of non-P texts do not indicate a coherent editorial milieu or style but rather allow us to better distinguish between P texts and non-P texts by pointing out particular theological or linguistic features. On this point, certain observations such as the absence of shared meals or dreams define P by what its editors had avoided. Such features are not often found in lists that attempt to define P by listing what it contains rather than what it does not. The rationale behind our approach allows for the detection of particularities that require explanation. Certain lexical features are found only in the priestly texts of Genesis or Exodus, but not in both. This is an interesting point in regard to the question of the homogeneity of the P texts, but it is difficult to draw conclusions at this stage. For the texts of Exodus, the excessive importance of the chapters devoted to the construction of the Tabernacle (Exod 25–31; 35–40) was the only significantly-distinguished P-associated

22 Holzinger, *Einleitung*.

block. Moreover, by manually removing the Tabernacle-related chapters,[23] we found that the remainder of the P-associated texts cannot be distinguished from the texts in which it is embedded, despite the fact that the distribution of the lengths of its sequences is closer to that of P-associated texts in Genesis – where the algorithm made a significant (albeit incomplete) distinction between P and non-P texts. Nevertheless, the characterization of non-P texts is relevant, as are the results concerning grammar.

5 Overview

Generally speaking, the prospects for research using algorithms on priestly texts are significant. Algorithms can be used to separate what is statistically significant from what is coincidental, offering a different perspective on textual data. In what follows, we list a few examples in which further computational work could shed new light on the priestly texts.

5.1 Subdivisions of P

Several questions concerning P have not been addressed in this article but could be the subject of further research. Firstly, biblical scholars have noted that P texts are not homogeneous; even when the Documentary Hypothesis was formulated under Wellhausen and Kuenen, there was still the separation between an original document P^g (*priesterliche Grundschrift* »priestly base text«) and secondary additions P^s (*priesterliche sekundäre Texte* »secondary priestly texts«).[24] When the Documentary Hypothesis was called into question in the 1970s and 1980s, many new proposals emerged.[25] The J and E sources were often abandoned in favor of the designation »non-P«, as in the present article. As for the priestly document, E. Blum (and followed by some scholars) suggested that some parts seem closer to a composition presupposing a number of traditions than an independent docu-

23 Yoffe, »Statistical«: § 3.1.

24 Wellhausen, *Prolegomena*; Kuenen, *Inquiry*.

25 David M. Carr, »Changes in Pentateuchal Criticism,« in *Hebrew Bible/Old Testament: The History of Its Interpretation*, vol. 3.2, *The Twentieth Century – From Modernism to Post-Modernism*, ed. Magne Sæbø (Göttingen: Vandenhoeck & Ruprecht, 2015) Chap. 40: 433–466; Thomas Römer and Albert de Pury, eds., *Le Pentateuque en question: les origines et la composition des cinq premiers livres de la Bible à la lumière des recherches récentes* (Genève: Labor et Fides, ^{3}2002).

ment. This idea is known as the »KP« in German (or »PC« in English).[26] In addition, some scholars are discussing the existence of proto-P texts that were incorporated into the P texts.[27] The increasing complexity of the origins of the priestly writings can also be observed regarding the secondary priestly texts, P^s. By using new acronyms, scholars seek to distinguish more precise realities behind these texts, such as »ThB«[28] to designate late priestly redactions of the Book of Numbers, »H« (for »Holiness School«) to designate the redaction of Lev 17–26 as well as scattered texts in the Pentateuch presenting linguistic and thematic affinities with the Holiness Code (e. g., Gen 1*; 2:2–3; 17:9–14,23–27; 21:4; 23*; 25:9–10; 49:29–32; 50:13; Exod 27:20–21; 29:38–46; 31:12–17; 35:1–3; Lev 3:16–17; 6:12–18; 7:22–29; 9:17; 10:10–11; 11:43–45; Num 3:11–13; 8:1–4,14–19).[29] The use of the acronym »H« goes back to the end of the 19th century,[30] but the tendency to attribute P texts outside Lev 17–26 to this same editorial group is more recent. These late priestly traditions are also characterized by the influence of Deuteronomistic texts, making them post-P and post-Dtr texts.[31]

With so many new theories emerging without consensus, the question arises as to the subjective value of each observation. By isolating the P texts, it would be interesting to see whether an algorithm would observe linguistic and stylistic breaks similar to the research proposals of biblical scholars.

Another unresolved question is the end of P^g. While in 19th century scholarship, the Priestly Writing ended – according to most scholars – with the death of Moses in Deut 34 or, in the context of the idea of an Hexateuch, at the end of the book of Joshua, other earlier endings have been proposed over the last forty years,

26 »Priesterliche Komposition« is used by Erhard Blum, »Once Again: The Literary-Historical Profile of the P Tradition,« in *Farewell to the Priestly Writing?*, ed. Friedhelm Hartenstein and Konrad Schmid, Ancient Israel and Its Literature 38, (Atlanta, GA: SBL Press, 2022) 27–62; for the expression »Priestly Composition,« see also Jürg Hutzli, *The Origins of P: Literary Profiles and Strata of the Priestly Texts in Genesis 1–Exodus 40*, FAT 164 (Tübingen: Mohr Siebeck, 2023).

27 E. g., Hutzli, *Origins*.

28 »ThB« stands for »Theokratische Bearbeitungen« (»theocratic redactions«); cf. Achenbach, *Vollendung*.

29 List drawn from Hutzli, *Origins*, 205 and Jacob Milgrom, *Leviticus 17–22: A New Translation with Introduction and Commentary*, AB 3A (New York: Doubleday, 2000) 1337–1344; see also Knohl, *Sanctuary* and Jakob Wöhrle, »The Integrative Function of the Law of Circumcision,« in *The Foreigner and the Law: Perspectives from the Hebrew Bible and the Ancient Near East*, ed. Reinhard Achenbach, Rainer Albertz and Jakob Wöhrle, BZAR 16 (Wiesbaden: Harrassowitz Verlag, 2011) 71–87.

30 As far back as 1877: August Klostermann, »Ezechiel und das Heiligkeitsgesetz,« *ZLThK* 38 (1877) 401–445.

31 Cf. Norbert Lohfink, »Die Abänderung der Theologie des priesterlichen Geschichtswerks im Segen des Heiligkeitsgesetzes: zu Lev 26,9.11–13,« in *Studien zum Pentateuch*, SBAB 4 (Stuttgart: Verlag Kath. Bibelwerk, 1988) 157–168.

such as Exod 6,[32] Exod 29,[33] Exod 40,[34] Lev 9,[35] Lev 16,[36] and Num 27.[37] Here again, an algorithmic study could possibly observe a statistically significant break in the P texts.

5.2 P in Other Biblical Books

The algorithm presented in this article was used for the books of Genesis and Exodus. Of course, the texts of Leviticus, Numbers, and to a lesser extent Deuteronomy and the Deuteronomistic History also contain P, P-like, or post-P texts. Using the characterization of P, could the algorithm find the set of texts inserted into these corpora?

Biblical scholars have also noted that the book of Ezekiel[38] and Deutero-Isaiah[39] have a significant number of affinities with the P texts of the Pentateuch. Is there a literary dependency between these different texts? Do they stem from the same group of authors? Were there late insertions seeking to make the link with priestly texts?

32 Rendtorff, *Überlieferungsgeschichtliche Studien.*

33 Otto, »Forschungen«.

34 Pola, *Ursprüngliche Priesterschrift.*

35 Zenger, »Priesterschrift«.

36 Christophe Nihan, *From Priestly Torah to Pentateuch: A Study in the Composition of the Book of Leviticus*, FAT II/25 (Tübingen: Mohr Siebeck, 2007).

37 Lothar Perlitt, »Priesterschrift im Deuteronomium?,« in *Lebendige Forschung im Alten Testament*, ed. Otto Kaiser (Berlin: de Gruyter, 1988) 65–88; Jean-Louis Ska, »Le récit sacerdotal: Une ›histoire sans fin‹?,« in *The Books of Leviticus and Numbers*, ed. Thomas Römer, BETL (Leuven: Peeters, 2008) 631–653; Suzanne Boorer, *The Vision of the Priestly Narrative: Its Genre and Hermeneutics of Time*, Ancient Israel and Its Literature 27 (Atlanta, GA: SBL Press, 2016).

38 Menahem Haran, »Ezekiel, P, and the Priestly School,« *VT* 58 (2008) 211–218; Jaeyoung Jeon, »A Source of P? The Priestly Exodus Account and the Book of Ezekiel,« *Semitica* 58 (2016) 77–92; Michael A. Lyons, »How Have We Changed? Older and Newer Arguments about the Relationship between Ezekiel and the Holiness Code,« in *The Formation of the Pentateuch*, ed. Jan C. Gertz et al., FAT 111 (Tübingen: Mohr Siebeck, 2016) 1055–1074; Walter Bührer, »Ezechiel und die Priesterschrift,« in *Das Buch Ezechiel*, ed. Jan C. Gertz, Corinna Körting and Markus Witte, BZAW 516 (Berlin: de Gruyter, 2020) 175–206.

39 Arvid S. Kapelrud, »The Date of the Priestly Code (P),« *ASTI* 3 (1964) 58–64; Philip B. Harner, »Creation Faith in Deutero-Isaiah,« *VT* 17 (1967) 298–306; Carroll Stuhlmueller, »The Theology of Creation in Second Isaias,« *CBQ* (1959) 429–467.

5.3 Linguistic Dating of P

Linguistic dating is a field that has attracted considerable research attention in recent decades.[40] It is a complex field, since language usage varies not only according to the date of the text but also according to its social milieu, geographical origin, theological intentions, and sources. Still, certain grammatical and syntactic evolutionary processes (probably more than lexical ones) bear witness to the evolution of the language. A systematic and quantitative assessment of such features could help classify the various corpora according to relative dating.

6 Conclusion

Implementing our computational and statistical analysis to study biblical texts provides a fresh perspective on the texts. The characteristics of P could be defined both by what is found in P frequently or through a specific lexical environment, as well as by what is absent from this corpus in relation to the other accounts of creation, the patriarchs, and the exodus. While most of the observations are similar to those previously listed by Holzinger,[41] some, especially in what is absent, are not listed in his work. Such results suggest that this type of algorithm could be used routinely in many corpora to provide results that can easily be used by exegetes.

The continuity of the P texts between the books of Genesis and Exodus is apparent in certain specific expressions, but in general, many usages are specific to a given book. The long text of the Tabernacle account in Exodus contrasts with

40 Erhard Blum, »The Linguistic Dating of the Biblical Texts: An Approach with Methodological Limitations,« in *The Formation of the Pentateuch*, ed. Jan C. Gertz et al., FAT 111 (Tübingen: Mohr Siebeck, 2016) 303–325; Ohad Cohen, »Linguistics and the Dating of Biblical Literature,« in *The Wiley Blackwell Companion to Ancient Israel*, ed. Susan Niditch (Wiley-Blackwell: Chichester, 2016) 118–130; Jan Joosten, »Diachronic Linguistics and the Date of the Pentateuch,« in *The Formation of the Pentateuch*, ed. Jan C. Gertz et al., FAT 111 (Tübingen: Mohr Siebeck, 2016) 327–344; Robert Rezetko and Ian Young, »Currents in the Historical Linguistics and Linguistic Dating of the Hebrew Bible: Report on the State of Research as Reflected in Recent Major Publications,« HIPHIL Novum 5 (2019) 1–93; Martin Ehrensvärd, »The Contemporary Debate over Linguistic Dating of Biblical Texts,« in *History, Archaeology and the Bible Forty Years After »Historicity,«* ed. Ingrid Hjelm and Thomas L. Thompson, Changing Perspectives 6 (London: Routledge, 2016) 60–67; Shimon Gesundheit, »Introduction: The Strengths and Weaknesses of Linguistic Dating,« in *The Formation of the Pentateuch*, ed. Jan C. Gertz et al., FAT 111 (Tübingen: Mohr Siebeck, 2016) 295–302; William M. Schniedewind, »Linguistic Dating, Writing Systems, and the Pentateuchal Sources,« in *The Formation of the Pentateuch*, ed. Jan C. Gertz et al., FAT 111 (Tübingen: Mohr Siebeck, 2016) 345–356.

41 Holzinger, *Einleitung*.

the shorter P texts of Exodus. These findings deserve to be discussed and taken into account in scholarship. Perhaps P^g should be separated from P^s by adopting a shorter stratum of the original narrative of the construction of the Tabernacle, as suggested, for example, by T. Pola[42] and E. Otto,[43] among others.

The algorithmic approach presented here is a first step toward the development and integration of a robust computational framework for biblical exegesis. While the distinction between P and non-P strata was found to be significant, independent of previous considerations and based on objective literary features, much future effort is required to improve its precision and sensitivity in more ambiguous units of text.

Abstract: We have recently introduced a computational framework that, given a partition of a text into two literary constituents, finds the best parameters for successfully distinguishing linguistic features to support that partition and evaluates the statistical significance thereof. We applied our algorithm to assess the literary uniqueness of the Priestly source in the books of Genesis and Exodus, focusing on the mathematical and statistical underpinning of our approach. Here we take a close philological look at the linguistic features found to characterize the two distinct categories of texts.

Keywords: Priestly Source, Computational Humanities, Biblical Exegesis, Hebrew Bible, Genesis, Exodus

Zusammenfassung: Vor kurzem haben wir einen Algorithmus entwickelt und vorgestellt, der die besten Parameter zur erfolgreichen Unterscheidung der sprachlichen Merkmale für eine Aufteilung eines Textes in zwei literarische Gruppen findet und die statistische Signifikanz dieser Parameter bewertet. Wir haben unseren Algorithmus auf die Bücher Genesis und Exodus angewendet, um die Priesterschrift von den nicht-priesterlichen Texten zu unterscheiden, wobei wir uns auf die mathematischen und statistischen Grundlagen unseres Ansatzes stützten. Hierbei gehen wir näher auf die sprachlichen Besonderheiten ein, die die beiden literarischen Textgruppen kennzeichnen.

Schlagwörter: Priesterschrift, Digital Humanities, biblische Exegese, Hebräische Bibel, Genesis, Exodus

42 Pola, *Ursprüngliche Priesterschrift*.
43 Otto, »Forschungen«.

Résumé: Récemment, nous avons développé un algorithme qui trouve les meilleurs paramètres pour distinguer avec succès les caractéristiques linguistiques d'une partition d'un texte en deux ensembles littéraires et en évalué la significativité statistique. Pour cet article, nous avons appliqué notre algorithme aux livres de la Genèse et de l'Exode pour différencier le document sacerdotal du reste des textes, en nous appuyant sur les fondements mathématiques et statistiques de notre approche. Nous examinons ici de près les spécificités linguistiques qui caractérisent les deux ensembles littéraires.

Mots-clés: document sacerdotal, humanités numériques, exégèse biblique, Bible hébraïque, Genèse, Exode