



ESTILOMETRIA DESCONTEXTUALIZADA: UMA CRÍTICA AO USO DE INTELIGÊNCIA ARTIFICIAL NA ANÁLISE DE AUTORIA BÍBLICA

Rodrigo Niskier Ferreira Barbosa¹

RESUMO: O presente artigo oferece uma resposta crítica ao estudo de Faigenbaum-Golovin et al. (2025), que emprega análise computacional de frequências lexicais e inteligência artificial para investigar a autoria de textos bíblicos do Eneateuco. Diante da crescente aplicação de métodos digitais na Alta Crítica, argumenta-se que a metodologia proposta apresenta, pelo menos, três fragilidades fundamentais que comprometem seus resultados. Primeiro, identifica-se o problema circular do *ground truth*, no qual o algoritmo foi treinado com capítulos previamente classificados segundo hipóteses documentárias do século XIX, configurando uma tautologia metodológica em que a máquina apenas confirma o viés humano preexistente. Segundo, demonstra-se a confusão entre variação de registro ou gênero literário e distinção autoral, desconsiderando-se que um único autor pode empregar vocabulários distintos em contextos discursivos diversos — legais, narrativos ou cultuais. Terceiro, apontam-se as limitações estatísticas inerentes à aplicação da estilometria a *corpora* de hebraico bíblico, tendo em vista seus limites de tamanho e sua complexidade morfológica. A partir de uma perspectiva manuscritológica reformada e da defesa da unidade

¹ O autor é Presbítero Regente da Igreja Presbiteriana de Altiplano (João Pessoa, PB) desde 2021, Seminarista do terceiro ano (2026) do Seminário Presbiteriano do Norte (SPN) da Igreja Presbiteriana do Brasil. Bacharel em Biomedicina pela Universidade Federal de Pernambuco (2003), Especialista em Biologia Molecular pela Universidade de Pernambuco (2006), Mestre em Genética e Biologia Molecular pela Universidade Federal do Rio Grande do Norte (2007), Professor Adjunto I no curso de Medicina do Centro Universitário de João Pessoa, desde 2017.

literária do Pentateuco, o artigo propõe que as distinções detectadas pelo algoritmo não evidenciam multiplicidade autoral, mas refletem primariamente diferenças funcionais de gênero e vocabulário técnico. Conclui-se que, embora as ferramentas computacionais sejam úteis, seu uso descontextualizado da teologia canônica tende a reproduzir pressupostos críticos obsoletos, reafirmando a necessidade de uma exegese que integre dados quantitativos à coerência interna do texto bíblico.

PALAVRAS-CHAVE: Estilometria Bíblica; Hipótese Documentária; Autoria do Pentateuco; Aprendizado de máquina; Crítica textual.

INTRODUÇÃO

A aplicação de métodos computacionais à crítica bíblica tem experimentado significativo desenvolvimento nas últimas décadas, com a pretensão de substituir análises subjetivas por inferências estatisticamente fundamentadas.² O estudo recente de Faigenbaum-Golovin et al. (2025), intitulado “Critical biblical studies via word frequency analysis: Unveiling text authorship”, representa um esforço ambicioso nessa direção, empregando método estatístico para distinguir entre três *corpora* supostamente autorais no Eneateuco: o Deuteronômio primitivo (D), a História Deuteronomista (DtrH) e os escritos Sacerdotais (P).³

Os autores afirmam que seu método “oferece evidências estatisticamente significativas e interpretáveis da existência de características linguísticas na escrita de autores/redatores bíblicos”.⁴ No entanto, uma análise crítica revela que tal alegação padece de problemas metodológicos fundamentais que comprometem suas conclusões. O presente artigo pretende demonstrar que o algoritmo não descobriu autores distintos; antes, detectou distinções de gênero literário e de vocabulário técnico (registro), previamente rotuladas como “autores” pelos próprios pesquisadores, configurando o clássico problema do *ground truth* enviesado.⁵

A crítica aqui desenvolvida articula-se em três frentes: (a) o vício metodológico circular na construção dos dados de treinamento; (b) a confusão entre variação de registro e distinção autoral; e (c) as limitações estatísticas da estilometria aplicada ao hebraico bíblico antigo. A análise fundamenta-se em uma

² STAMATATOS, Efstathios. A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology*, Hoboken, v. 60, n. 3, p. 538-556, mar. 2009; VAN PEURSEN, Wido Th. Computational linguistic analysis of the biblical text. In: *The Oxford Handbook of the Bible and Digital Humanities*. Oxford: Oxford University Press, 2023. cap. 5.

³ FAIGENBAUM-GOLOVIN, Shira et al. Critical biblical studies via word frequency analysis: unveiling text authorship. *PLOS One*, São Francisco, v. 20, n. 6, jun. 2025.

⁴ FAIGENBAUM-GOLOVIN et al., 2025

⁵ JATON, Florian. Assessing biases, relaxing moralism: on ground-truthing practices in machine learning. *Big Data & Society*, London, v. 8, n. 1, p. 1-15, jan./jun. 2021.

cosmovisão reformada, que reconhece a unidade literária do Pentateuco como produto da revelação divina mediada por Moisés, sem negar a possibilidade de edições posteriores sob supervisão providencial.

1. O PROBLEMA DO *GROUND TRUTH*: CIRCULARIDADE METODOLÓGICA

1.1 A FALÁCIA DA TAUTOLOGIA TREINADA

O estudo de Faigenbaum-Golovin et al. (2025) fundamenta-se em uma premissa metodológica vulnerável: o algoritmo foi treinado com 50 capítulos “rotulados de acordo com considerações da exegese bíblica em três *corpora*”. Os pesquisadores admitem explicitamente que a seleção seguiu critérios da crítica bíblica tradicional, incluindo a reconstrução de Preuss para o Deuteronômio primitivo e as hipóteses de Köckert e Nihan para os textos sacerdotais.⁶

Este procedimento configura o que a literatura de aprendizado de máquina denomina viés de confirmação no *ground truth*. Conforme demonstrado por pesquisadores contemporâneos, bases de dados de *ground truth* e suas práticas concomitantes — que fundamentalmente envolvem estabelecer vieses para habilitar procedimentos de aprendizado — exercem influência decisiva sobre os resultados algorítmicos.⁷ Quando rótulos de treinamento derivam de teorias pré-existentes, o modelo aprende a reproduzir tais classificações, não a descobrir padrões independentes.⁸

⁶ No campo da crítica literária e da manuscritologia do Pentateuco, especialmente no que se refere à reconstrução do Deuteronômio primitivo e às hipóteses sobre os textos sacerdotais, a obra de Preuss (1982) tornou-se referência para os estudos sobre a reconstrução do Deuteronômio primitivo (*Urdeuteronomium*). A monografia de Köckert (1988) estabeleceu bases importantes para o debate acerca das promessas patriarcais e da formação da tradição sacerdotal. A tese doutoral de Nihan, publicada em 2007, representa um estudo considerado abrangente sobre a composição do livro de Levítico e o fim da *Grundschrift* sacerdotal, sendo amplamente citada como trabalho de referência nessa área da crítica literária do Pentateuco. Entretanto, as hipóteses de Preuss, Köckert e Nihan são contestadas por se fundamentarem em pressupostos filosóficos evolutivos que negam a autoria mosaica substancial atestada pelo testemunho interno bíblico e pela autoridade cristológica. A crítica reformada argumenta que as evidências arqueológicas e literárias do Antigo Oriente Próximo — como os tratados de suserania hititas — corroboram a origem dos textos deuteronômicos e sacerdotais no segundo milênio a.C., refutando as datações tardias, josianas ou persas, propostas pela Alta Crítica e defendendo a unidade teológica e histórica do Pentateuco. PREUSS, Horst Dietrich. *Deuteronomium*. Darmstadt: Wissenschaftliche Buchgesellschaft, 1982. (Erträge der Forschung, 164); KÖCKERT, Matthias. *Vätergott und Väterverheissungen: eine Auseinandersetzung mit Albrecht Alt und seinen Erben*. Göttingen: Vandenhoeck & Ruprecht, 1988; NIHAN, Christophe. *From Priestly Torah to Pentateuch: a study in the composition of the Book of Leviticus*. Tübingen: Mohr Siebeck, 2007.

⁷ LEOVITZ, Sarah; LEVINA, Natalia; LIFSHITZ-ASSAF, Hila. Is AI *ground truth* really true? The dangers of training and evaluating AI tools based on experts' know-what. *MIS Quarterly*, Minneapolis, v. 45, n. 3, p. 1501-1526, set. 2021.

⁸ PLANK, Barbara; HOVY, Dirk; SØGAARD, Anders. Linguistically debatable or just plain wrong? In: *Annual Meeting of the Association for Computational Linguistics*, 52., 2014, Baltimore. Proceedings [...]. Stroudsburg: Association for Computational Linguistics, 2014. p. 507-511.

O algoritmo estatístico, por mais sofisticado, apenas “aprende” a distinguir aquilo que os pesquisadores definiram *a priori* como categorias distintas. Conforme argumenta Jaton, “sem base de dados de *ground truth* validadas que forneçam os referentes do que os algoritmos de aprendizado de máquinas devem encontrar, nenhuma operação de aprendizado é concebível, pois não há indicação do que é apropriado aprender”.⁹

Se textos com vocabulário cultural (tabernáculo, sacrifício, sacerdote) foram rotulados como “P”, e textos com vocabulário legal-civil foram rotulados como “D”, o algoritmo naturalmente identificará vocabulário cultural como marcador de “P”. Isso não constitui descoberta científica; constitui tautologia computacional. O próprio estudo reconhece que palavras como זהב (ouro), קרש (tábua), משכן (tabernáculo) e אמה (côvado) são “discriminantes” para P¹⁰ — termos que aparecem inevitavelmente em descrições do Tabernáculo, independentemente de quem as tenha escrito.

1.2 A HERANÇA CRÍTICA DO SÉCULO XIX

A classificação dos *corpora* de treinamento reproduz, em essência, a Hipótese Documentária de Wellhausen, embora seus proponentes contemporâneos tenham abandonado alguns de seus pressupostos mais controversos.¹¹ O estudo assume como dado que D, DtrH e P representam “autores ou autores/redatores com contexto cultural/ideológico distintivo”, sem submeter essa premissa à verificação independente.¹²

Umberto Cassuto, em suas célebres oito conferências sobre a Hipótese Documentária, demonstrou que os cinco “pilares” dessa teoria — uso dos nomes divinos, variações de linguagem e estilo, contradições internas, repetições narrativas e evidências de estrutura composta — admitem explicações alternativas igualmente plausíveis.¹³ As variações no uso de YHWH e Elohim, por exemplo, refletem contextos teológicos distintos (transcendência *versus* imanência divina), não necessariamente fontes documentárias diferentes.¹⁴

⁹ JATON, 2021.

¹⁰ FAIGENBAUM-GOLOVIN et al., 2025.

¹¹ A literatura crítica recente, embora ainda trabalhe com distinções como Deuteronômico (D), Deuteronomista (DtrH) e Sacerdotal (P), já não repete alguns dos pressupostos mais radicais de Wellhausen: a reconstrução de quatro “documentos” rigorosamente separados (J, E, D, P), com fronteiras claras, cronologia rígida e perfis teológicos mutuamente excludentes, tem sido amplamente relativizada; fala-se hoje em camadas, escolas redacionais e processos contínuos de transmissão, não em arquivos estanques combinados tardiamente. Nesse quadro, permanece, porém, a matriz básica da Hipótese Documentária (a segmentação do Pentateuco em blocos atribuídos a círculos deuteronomistas e sacerdotais), que o artigo em exame assume como dado ao escolher D, DtrH e P como *Ground truth*. WELLHAUSEN, Julius. *Prolegomena to the History of Israel*. Trad. de J. Sutherland Black e Allan Menzies. Edinburgh: Adam & Charles Black, 1885.

¹² FAIGENBAUM-GOLOVIN et al., 2025

¹³ CASSUTO, Umberto. *The Documentary Hypothesis and the composition of the Pentateuch: eight lectures*. Trad. de Israel Abrahams. Jerusalem: Magnes Press, 1961.

¹⁴ CASSUTO, 1961, p. 18-41.

Conforme observado por Kikawada e Quinn, muitos dos argumentos da Hipótese Documentária baseiam-se em pressupostos literários anacrônicos, projetando convenções modernas de autoria sobre textos antigos que operavam com padrões composicionais distintos.¹⁵ Estudos recentes sobre a *Neo-Documentary Hypothesis*, embora refinando metodologicamente a teoria clássica, ainda não resolvem o problema fundamental da circularidade epistêmica.¹⁶

2. VARIAÇÃO DE REGISTRO *VERSUS* DISTINÇÃO AUTORAL

2.1 O FENÔMENO DA VARIAÇÃO LINGUÍSTICA INTRATEXTUAL

A sociolinguística contemporânea estabelece firmemente que um mesmo autor emprega registros linguísticos distintos em contextos comunicativos diversos.¹⁷ A variação de registro refere-se às mudanças sistemáticas no vocabulário, na sintaxe e no estilo em função do tema tratado, do gênero discursivo e da audiência presumida. Essa variação é particularmente pronunciada em textos que abrangem múltiplos gêneros literários — precisamente o caso do Pentateuco.

Pesquisas recentes confirmam a universalidade desse fenômeno: Li, Dunn e Nini demonstraram que a variação de registro permanece estável em 60 línguas, indicando que as características linguísticas que compõem um registro são motivadas pelas necessidades e restrições da situação comunicativa, e não por diferenças autorais.¹⁸ Henriksson desenvolveu métodos de segmentação automática de registro em textos *web*, demonstrando que unidades segmentadas exibem características linguísticas mais consistentes que documentos completos — evidenciando que a variação interna de registro é um fenômeno comum e esperado.¹⁹

O estudo de Faigenbaum-Golovin et al. reconhece o problema apenas marginalmente: “Uma vez que um único autor pode escrever textos sobre tópicos divergentes usando gêneros diferentes, um grande desafio em estudos de autoria é a identificação de características do autor que sejam, ao mesmo tempo, não relacionadas a tópico ou gênero”. No entanto, os autores não demonstram que

¹⁵ KIKAWADA, Isaac M.; QUINN, Arthur. *Before Abraham was: the unity of Genesis 1-11*. Nashville: Abingdon Press, 1985.

¹⁶ BADEN, Joel S. *The composition of the Pentateuch: renewing the Documentary Hypothesis*. New Haven: Yale University Press, 2012.

¹⁷ BIBER, Douglas; CONRAD, Susan. *Register, genre, and style*. Cambridge: Cambridge University Press, 2009.

¹⁸ LI, Haipeng; DUNN, Jonathan; NINI, Andrea. *Register variation remains stable across 60 languages*. arXiv preprint, arXiv:2209.09813, 2022.

¹⁹ HENRIKSSON, Erik; HELLSTRÖM, Saara; LAIPPALA, Veronika. Analyzing register variation in web texts through automatic segmentation. In: *International Conference on Natural Language Processing for Digital Humanities*, 5., 2025, Albuquerque. Proceedings [...]. Albuquerque: Association for Computational Linguistics, 2025. p. 7-19.

seu método supera esse desafio; limitam-se a pressupor que a estatística empregada é capaz de distinguir estilo autoral de variação temática.

2.2 EVIDÊNCIAS DE VARIAÇÃO DE REGISTRO NOS DADOS

A lista de lemas discriminantes apresentada no estudo mostra que o algoritmo detectou, *a priori*, vocabulário temático, não estilístico. Os termos identificados como característicos de P — זהב (ouro), קנה (cana), אמה (côvado), טבעת (argola), ארון (arca) — constituem vocabulário técnico de construção do Tabernáculo, não impressões digitais autorais.²⁰

Da mesma forma, os lemas discriminantes para D — לא (não), אלהים (Deus), מלך (rei) — representam linguagem típica de exortação legal e narrativa, enquanto os marcadores de DtrH — אשר (que), יום (dia), דבר (palavra) — são elementos comuns de narrativa histórica.²¹ Quando Moisés prescreve as medidas do altar, naturalmente emprega vocabulário de mensuração; quando narra a conquista de povos, emprega vocabulário militar; quando exorta à obediência, emprega vocabulário parenético. A variação lexical reflete a diversidade funcional do texto, não multiplicidade de autores.

2.3 A PERSISTÊNCIA DO PROBLEMA TÓPICO-GÊNERO NA ESTILOMETRIA

Pesquisas recentes em atribuição de autoria demonstram que a confusão entre estilo e tópico permanece um desafio central não resolvido.²² A proposta de Altakrori, Cheung e Fung evidencia sistematicamente a incapacidade de modelos estilométricos de distinguir erros decorrentes de mudança de tópico daqueles produzidos pela falha em capturar o estilo autoral.²³

Wegmann, Schraagen e Nguyen investigaram representações de estilo independentes de conteúdo, demonstrando que abordagens baseadas em

²⁰ FAIGENBAUM-GOLOVIN et al., 2025

²¹ FAIGENBAUM-GOLOVIN et al., 2025

²² MIKROS, George K. *Authorship attribution and gender identification in Greek blogs*. Methods and Applications of Quantitative Linguistics, Tübingen, p. 21-32, 2018; ZHENG, Wei et al. A review on authorship attribution in text mining. *WIRES Data Mining and Knowledge Discovery*, Hoboken, v. 13, n. 2, e1584, 2023.

²³ Vários experimentos mostram que os classificadores “aprendem” mais o assunto do que o estilo: se um mesmo autor escreve, por exemplo, um artigo técnico de medicina e um sermão devocional, muitos modelos de atribuição passam a tratá-los como se fossem de autores distintos, porque o léxico temático muda radicalmente; por outro lado, textos de autores diferentes, mas sobre o mesmo tema e no mesmo gênero (por exemplo, artigos de pesquisa seguindo o mesmo modelo de revista) tendem a ser confundidos, pois partilham o mesmo vocabulário e estrutura retórica. Em cenários controlados em que se obriga um algoritmo a decidir autoria quando “autor” e “tópico” são combinados de maneiras inesperadas, a taxa de erro sobe justamente quando o autor escreve sobre um tema que o modelo só tinha visto associado a outro autor, evidenciando que a classificação estava ancorada em pistas de conteúdo, não em um estilo autoral estável. ALTAKRORI, Malik; CHEUNG, Jackie Chi Kit; FUNG, Benjamin C. M. The topic confusion task: a novel evaluation scenario for authorship attribution. In: *Findings of the Association for Computational Linguistics: EMNLP 2021*, 2021, Punta Cana. Proceedings [...]. Punta Cana: Association for Computational Linguistics, 2021. p. 4242-4256.

aprendizado contrastivo ainda confundem estilo com tópico.²⁴ O estudo de Sawatphol, Udomcharoenchaikit e Nutanong sobre verificação de autoria demonstrou que pode haver vazamento de tópico nos dados de teste, causando desempenho enganoso do modelo e rankings instáveis. Em resposta a esse problema, os autores propõem o método HITS (*Heterogeneity-Informed Topic Sampling*) para criar bases de dados com conjuntos de tópicos heterogeneamente distribuídos.²⁵

Um estudo de Fincke e Boschee demonstra que mesmo modelos sofisticados baseados em BERT exibem desempenho deficiente em cenários de transferência de gêneros, sublinhando “a necessidade de modelos de autoria robustos a efeitos específicos de gênero”.²⁶ Pesquisas mais recentes com *Large Language Models* (LLMs) confirmam este desafio: embora LLMs demonstrem capacidades promissoras em análise estilística, abordagens supervisionadas e contrastivas frequentemente confundem estilo com tópico.²⁷ Faigenbaum-Golovin et al. não apresentam controles experimentais para lidar com esse confundimento metodológico.

²⁴ WEGMANN, Anna; SCHRAAGEN, Marijn; NGUYEN, Dong. Same author or just same topic? Towards content-independent style representations. In: *Workshop on Representation Learning for NLP*, 7., 2022, Dublin. Proceedings [...]. Stroudsburg: Association for Computational Linguistics, 2022. p. 249-268.

²⁵ O método HITS (Heterogeneity-Informed Topic Sampling) foi proposto exatamente para construir corpora em que os tópicos não ficam “amarrados” a autores específicos, mas são distribuídos de forma heterogênea entre vários autores e textos. Em vez de aceitar um conjunto de treino em que, por exemplo, Autor A só escreve sobre Tópico 1 e o Autor B só escreve sobre Tópico 2 (situação em que o classificador tende a confundir tópico com autoria), o HITS seleciona amostras de modo a garantir que cada autor apareça em múltiplos tópicos e que cada tópico seja representado por múltiplos autores. Com isso, a base de dados força o modelo de atribuição de autoria a depender menos de pistas de conteúdo e mais de traços estilísticos estáveis, reduzindo artificialidades do *ground truth* e tornando os testes de estilometria mais rigorosos e informativos. SAWATPHOL, Jitkapat; UDOMCHAROENCHAIKIT, Can; NUTANONG, Sarana. Addressing topic leakage in cross-topic evaluation for authorship verification. *Transactions of the Association for Computational Linguistics*, Cambridge, v. 12, p. 1363-1377, 2024.

²⁶ BERT (*Bidirectional Encoder Representations from Transformers*) é um modelo de linguagem baseado em Transformers que aprende representações contextualizadas de palavras e frases, lendo o texto de forma bidirecional (considerando o contexto anterior e posterior). Após um treinamento genérico em grandes corpora, ele é ajustado para tarefas específicas, como atribuição de autoria, funcionando como extrator de características de alto nível sobre as quais um classificador decide o autor. Quando Fincke e Boschee (2024) testam esse tipo de modelo em cenários de transferência de gênero e observam queda acentuada de desempenho, o resultado indica que mesmo essas representações avançadas ainda capturam muito das marcas de gênero e de formato textual, e menos de um estilo autoral realmente estável. FINCKE, Steven; BOSCHEE, Elizabeth. Separating style from substance: enhancing cross-genre authorship attribution through data selection and presentation. 2024. E-print arXiv:2408.05192.

²⁷ HUERTAS-TATO, Javier et al. LLM one-shot style transfer for authorship attribution and verification. arXiv preprint, arXiv:2510.13302, 2025.

3. LIMITAÇÕES DA ESTILOMETRIA EM HEBRAICO BÍBLICO

3.1 O PROBLEMA DO *CORPUS* RESTRITO

A estilometria computacional requer *corpora* substanciais para inferências estatisticamente robustas. O estudo analisado trabalha com apenas 50 capítulos de treinamento, com o *corpus* D contendo apenas 9 textos. Os próprios autores reconhecem que “determinar textos bíblicos suficientemente longos que pudessem ser atribuídos ao *corpus* D foi uma tarefa desafiadora”.²⁸

A literatura especializada demonstra que a acurácia de atribuição de autoria aumenta significativamente com o tamanho do *corpus* de referência.²⁹ Estudos subsequentes de Eder refinaram estes achados, demonstrando que em condições ideais o mínimo pode ser reduzido para aproximadamente 2.000 palavras, embora textos menores apresentam acurácia substancialmente comprometida.³⁰ Com textos tão curtos quanto 300 palavras (o mínimo reportado no estudo), a confiabilidade estatística é intrinsecamente limitada. O estudo reporta que capítulos com menos de 10 versículos apresentam “probabilidade muito baixa de atribuição correta”, e mesmo textos de 10-30 versículos alcançam apenas “cerca de 80% de chance de atribuição correta”.³¹

Van Peursen observa que pesquisadores de linguística computacional do hebraico bíblico têm demonstrado crescente consciência das limitações desses modelos em razão do tamanho restrito do *corpus*, e que técnicas de enriquecimento de dados e ampliação de *corpus* — como a construção de bases de dados com línguas semíticas relacionadas — têm sido empregadas para mitigar essas limitações.³²

3.2 COMPLEXIDADES MORFOLÓGICAS DO HEBRAICO

O hebraico bíblico apresenta características morfológicas que confundem análises estilométricas convencionais. Sua rica morfologia flexional — com prefixos, sufixos e infixos — faz com que uma única raiz verbal possa gerar dezenas de formas distintas, diluindo a frequência de cada forma específica. O estudo emprega a lematização para mitigar esse problema, mas reconhece a limitação do procedimento: “ao contar lemas, podemos perder um sinal potencial de autoria, isto é, as diferentes maneiras como vários autores usam o mesmo lema”.³³

²⁸ FAIGENBAUM-GOLOVIN et al., 2025.

²⁹ EDER, Maciej. Does size matter? Authorship attribution, small samples, big problem. *Digital Scholarship in the Humanities*, Oxford, v. 30, n. 2, p. 167-182, jun. 2015.

³⁰ EDER, Maciej. Short samples in authorship attribution: a new approach. In: *Digital Humanities Conference*, 2017, Montreal. Abstracts [...]. Montreal: McGill University, 2017. p. 1-3.

³¹ FAIGENBAUM-GOLOVIN et al., 2025.

³² VAN PEURSEN, 2023.

³³ Lematização é o processo de reduzir cada forma flexionada hebraica (com prefixos, sufixos, estados e conjugações diversas) à sua forma de dicionário, o chamado lema (por exemplo, todas as formas verbais derivadas da raiz כתב são contabilizadas como um único item). Isso diminui o

Pesquisas recentes em Processamento de Linguagem Natural (PLN) para hebraico confirmam estas dificuldades. Seker et al. desenvolveram o *AlephBERT* especificamente para abordar os desafios morfológicos do hebraico moderno, reconhecendo que modelos PLN padrão são inadequados para línguas morfológicamente ricas.³⁴ Shmidman et al. demonstraram que homógrafos desbalanceados no hebraico representam um grande desafio para analisadores morfológicos automatizados.³⁵

Chriqui e Yahav desenvolveram o HeBERT reconhecendo que “o hebraico é uma língua morfológicamente rica, definida como uma língua na qual informação significativa concernente a unidades e relações sintáticas é expressa no nível da palavra”, o que exige escolhas arquiteturais cuidadosas para garantir que as características da língua sejam adequadamente representadas.³⁶

Além disso, o hebraico bíblico tardio apresenta variações diacrônicas significativas.³⁷ Distinções atribuídas a diferentes “autores” podem,

número de “palavras distintas” no *corpus* e aumenta a frequência relativa de cada lema, facilitando a aplicação de testes estatísticos sobre tabelas de frequência. Em contrapartida, essa normalização apaga diferenças sutis de uso entre formas concorrentes de uma mesma raiz, o que pode misturar sinal de estilo (escolhas preferenciais entre variantes) com mera necessidade morfológica da construção. FAIGENBAUM-GOLOVIN et al., 2025.

³⁴ AlephBERT é um modelo de linguagem do tipo BERT treinado especificamente em grandes corpora de hebraico, de modo a aprender representações contextualizadas sensíveis às particularidades morfológicas e sintáticas dessa língua. Em vez de depender de modelos genéricos, ele utiliza um vocabulário e pesos ajustados ao sistema consonantal hebraico, o que melhora tarefas como análise morfossintática e classificação de textos em hebraico. Esse tipo de ferramenta pode ser recebido como um recurso técnico útil para explorar regularidades linguísticas do corpus bíblico, sem, contudo, substituir a exegese canônica nem autorizar reconstruções de autoria que entrem em conflito com o testemunho explícito das Escrituras acerca da origem mosaica e da unidade teológica do Pentateuco. SEKER, Amit et al. AlephBERT: language model pre-training and evaluation from sub-word to sentence level. In: *Annual Meeting of the Association for Computational Linguistics*, 60., 2022, Dublin. Proceedings [...]. Stroudsburg: Association for Computational Linguistics, 2022. p. 46-56.

³⁵ SHMIDMAN, Avi et al. A novel challenge set for Hebrew morphological disambiguation and diacritics restoration. In: *Conference on Empirical Methods in Natural Language Processing*, 2020. Findings of the Association for Computational Linguistics: EMNLP 2020. Stroudsburg: Association for Computational Linguistics, 2020. p. 3545-3552.

³⁶ HeBERT é um modelo de linguagem do tipo BERT treinado especificamente em grandes corpora de hebraico contemporâneo (e, em algumas variantes, também em textos bíblicos e rabínicos), de modo a produzir representações vetoriais contextualizadas adaptadas às estruturas morfológicas e sintáticas do hebraico. Tal como outros modelos Transformer, ele é pré-treinado em tarefas genéricas de compreensão de texto e depois ajustado para aplicações específicas (classificação de textos, análise de sentimento, rotulagem morfossintática etc.), oferecendo um perfil estatístico detalhado do uso real da língua. Esse tipo de recurso pode ser valorizado como instrumento auxiliar para estudar padrões linguísticos e estilísticos em hebraico, mas não confere, por si só, critério teológico ou canônico para redefinir a autoria ou a unidade dos livros bíblicos, que permanecem ancoradas no testemunho inspirado da própria Escritura e na fé da Igreja ao longo dos séculos. CHRIQUI, Avihay; YAHAV, Inbal. HeBERT and HebEMO: a Hebrew BERT model and a tool for polarity analysis and emotion recognition. *INFORMS Journal on Data Science*, Catonsville, v. 1, n. 1, p. 81-95, 2022.

³⁷ YOUNG, Ian; REZETKO, Robert; EHRENSVÄRD, Martin. Linguistic dating of biblical texts. London: Equinox, 2008. v. 1.

alternativamente, refletir diferentes estágios linguísticos de transmissão ou atualização por atividade escribal ao longo dos séculos de cópia manuscrita.

4. PERSPECTIVA REFORMADA SOBRE A UNIDADE DO PENTATEUCO

4.1 AUTORIA MOSAICA E SUPERVISÃO PROVIDENCIAL

A tradição reformada, fundamentada no testemunho interno das Escrituras e na confissão de fé histórica da Igreja, sustenta a autoria mosaica substancial do Pentateuco.³⁸ Isso não implica negação de que Moisés possa ter empregado fontes anteriores, nem de que redatores posteriores (como Josué ou Samuel) tenham feito acréscimos menores sob inspiração divina. O que se afirma é que a obra possui unidade literária e teológica coerente com um autor principal sob direção providencial.

Jesus Cristo e os apóstolos consistentemente atribuíram o Pentateuco a Moisés (Mt 19.7-8; Mc 12.26; Jo 5.46-47; Rm 10.5), conferindo autoridade canônica a essa atribuição. A hermenêutica reformada reconhece que o testemunho de Cristo sobre a autoria das Escrituras constitui dado teológico que não pode ser descartado por reconstruções hipotéticas.³⁹

4.2 EXPLICAÇÕES ALTERNATIVAS PARA AS VARIAÇÕES OBSERVADAS

As variações estilísticas detectadas pelo algoritmo admitem explicações compatíveis com a unidade autoral:

- Variação funcional: Moisés, como líder, legislador e historiador, naturalmente empregava registros distintos para narrativa, legislação cultural e exortação.⁴⁰
- Desenvolvimento composicional: A redação do Pentateuco ao longo de 40 anos de peregrinação no deserto implicaria variações naturais no estilo do próprio Moisés.
- Edições redacionais: Atualizações linguísticas e geográficas por escribas posteriores (explicando, por exemplo, o uso de “Dã” para uma cidade que só recebeu esse nome após a morte de Moisés) não comprometem a unidade autoral substancial.⁴¹
- Fontes incorporadas: Moisés pode ter incorporado tradições patriarcais orais ou escritas, preservando suas características estilísticas originais,

³⁸ HARRISON, Roland Kenneth. *Introduction to the Old Testament: with a comprehensive review of Old Testament studies and a special supplement on the Apocrypha*. Grand Rapids: Eerdmans, 1969.

³⁹ ARCHER, Gleason L. *Merece confiança o Antigo Testamento?* 4. ed. São Paulo: Vida Nova, 1994.

⁴⁰ GARRETT, Duane A. *Rethinking Genesis: the sources and authorship of the first book of the Pentateuch*. Grand Rapids: Baker Book House, 1991.

⁴¹ KITCHEN, Kenneth A. *On the reliability of the Old Testament*. Grand Rapids: Eerdmans, 2003.

semelhante ao uso que Lucas faz de fontes distintas em seu Evangelho.⁴²

5. AVALIAÇÃO CRÍTICA DA METODOLOGIA ESTATÍSTICA

5.1 O MÉTODO HIGHER CRITICISM DE DONOHO E JIN

O estudo emprega o método estatístico *Higher Criticism* (HC), desenvolvido por Donoho e Jin, originalmente projetado para detectar sinais esparsos em misturas heterogêneas.⁴³ O método calcula um índice de discrepância entre tabelas de frequência lexical, identificando palavras cuja distribuição desvia significativamente do esperado sob alocação aleatória.

Embora estatisticamente robusto para seus propósitos originais, a aplicação do HC a textos bíblicos enfrenta desafios específicos que não foram adequadamente tratados:

- Pressuposto de independência: O HC assume que as observações são independentes, mas palavras em texto são altamente correlacionadas por coocorrência temática.⁴⁴
- Sensibilidade a distribuições de cauda: O método pode ser subótimo quando a distribuição dos sinais apresenta caudas polinomiais,⁴⁵ comum em dados linguísticos que seguem a Lei de Zipf.⁴⁶ Wallkulle demonstrou em estudo de simulação que HC não é bem-sucedido próximo às fronteiras teóricas onde detecção de sinal e seleção de características são assintoticamente possíveis, ou sob condições muito esparsas.⁴⁷
- Calibração do *ground truth*: A validação do método depende inteiramente da qualidade dos rótulos de treinamento — precisamente o ponto mais

⁴² WALTKE, Bruce K.; YU, Charles. *An Old Testament theology: an exegetical, canonical, and thematic approach*. Grand Rapids: Zondervan, 2007.

⁴³ DONOHO, David; JIN, Jiashun. Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics*, Washington, v. 32, n. 3, p. 962-994, jun. 2004.

⁴⁴ KIPNIS, Alon. Higher criticism for discriminating word-frequency tables and authorship attribution. *The Annals of Applied Statistics*, v. 16, n. 2, jun. 2022.

⁴⁵ Diz-se que uma distribuição tem “caudas polinomiais” quando a probabilidade de observar valores muito extremos diminui devagar, seguindo aproximadamente uma lei de potência (algo da forma $1/x$), em vez de cair muito rapidamente como na curva normal (cauda exponencial). Isso significa que observações grandes são bem mais frequentes do que se esperaria de uma distribuição gaussiana, o que torna alguns testes estatísticos — entre eles certas versões do Higher Criticism — menos eficientes ou mal calibrados quando assumem, implicitamente, caudas leves. DONOHO, David; JIN, Jiashun. Higher criticism for large-scale inference, especially for rare and weak effects. *Statistical Science*, v. 30, n. 1, p. 1-25, 2015.

⁴⁶ Lei de Zipf é uma regularidade estatística observada em muitos corpora linguísticos: se as palavras de um texto (ou de uma língua) forem ordenadas da mais frequente para a menos frequente, a frequência f de uma palavra é aproximadamente inversamente proporcional à sua posição no ranking r (isto é, $f \sim 1/r$). Em termos intuitivos, isso significa que poucas palavras muito comuns concentram grande parte das ocorrências (artigos, preposições, pronomes), enquanto a imensa maioria do vocabulário aparece raramente, gerando justamente uma distribuição com “cauda pesada”, na qual itens pouco frequentes são muito numerosos.

⁴⁷ WALLKULLE, Therese. *Detection and identification of rare and weak signals using Higher Criticism*. 2025. Dissertação (Mestrado) - Uppsala University, Uppsala, 2025.

vulnerável do estudo.

5.2 RESULTADOS QUE CONTRADIZEM AS CONCLUSÕES

Os próprios resultados do estudo sugerem que a distinção D/DtrH é tênue. Os autores reportam que “cinco dos oito (62%) erros de classificação ocorreram entre D e DtrH”,⁴⁸ e reconhecem que “os escribas das camadas mais antigas de Deuterônômio e da DtrH provavelmente pertenciam à mesma ‘escola’”. Essa admissão contradiz a tese de autores distintos: se D e DtrH são estatisticamente indistinguíveis, a distinção entre ambos pode ser artificial.

As oito classificações incorretas reportadas na Tabela 2 do referido artigo revelam que o algoritmo frequentemente atribui capítulos deuteronomísticos ao *corpus* errado — precisamente o que se esperaria caso as categorias D e DtrH fossem construções acadêmicas sobrepostas, e não realidades autorais.

CONSIDERAÇÕES FINAIS

O estudo de Faigenbaum-Golovin et al. (2025), embora tecnicamente sofisticado, não demonstra aquilo que afirma demonstrar. O algoritmo de *Higher Criticism* não “descobriu” autores distintos no Pentateuco; antes, classificou textos em categorias definidas *a priori* pela crítica bíblica tradicional, reproduzindo computacionalmente os pressupostos da Hipótese Documentária. As principais fragilidades identificadas são as seguintes:

1. Circularidade do *ground truth*: O treinamento com categorias derivadas da crítica liberal do século XIX garante que o algoritmo confirme essas categorias. Conforme demonstrado por estudos recentes em práticas de *ground-truthing*, vieses são cruciais para o *Machine Learning*: qualquer tarefa de classificação necessita de um referente que jaz fora da tarefa para fundamentar seu princípio classificatório.⁴⁹
2. Confusão registro-autoria: As distinções lexicais detectadas refletem variação de gênero literário e vocabulário técnico, não impressões digitais autorais. A literatura recente confirma que este permanece um problema não resolvido na estilometria computacional.⁵⁰
3. Limitações do *corpus*: O hebraico bíblico oferece um *corpus* estatisticamente restrito e morfologicamente complexo, inadequado para inferências robustas de autoria múltipla. Modelos de linguagem desenvolvidos para hebraico confirmam os desafios únicos desta língua morfologicamente rica.⁵¹

⁴⁸ FAIGENBAUM-GOLOVIN et al., 2025.

⁴⁹ JATON, 2021.

⁵⁰ ALTAKRORI; CHEUNG; FUNG, 2021; SAWATPHOL; UDOMCHAROENCHAIRIT; NUTANONG, 2024.

⁵¹ SEKER et al., 2022; CHRIQUI; YAHAV, 2022.

A abordagem reformada não nega a utilidade de métodos computacionais para a análise textual; questiona, porém, o emprego de tais métodos para validar pressupostos que deveriam ser testados, e não assumidos. Uma metodologia genuinamente científica testaria a hipótese nula de autoria única antes de pressupor a multiplicidade autoral.

O Pentateuco, conforme recebido pela Igreja ao longo dos milênios, apresenta uma coerência teológica e literária que transcende as supostas costuras editoriais. A fé reformada, firmada no testemunho de Cristo e dos apóstolos, reconhece em Moisés o autor substancial dessa obra fundacional, sem desconsiderar o processo providencial de transmissão e preservação textual que assegurou sua integridade até os nossos dias.

ABSTRACT: This article offers a critical response to the study by Faigenbaum-Golovin et al. (2025), which employs computational analysis of lexical frequencies and artificial intelligence to investigate the authorship of biblical texts within the Enneateuch. Given the growing application of digital methods in Higher Criticism, this article argues that the proposed methodology exhibits at least three fundamental weaknesses that compromise its results. First, it identifies the circular problem of *ground truth*, insofar as the algorithm was trained on chapters previously classified according to nineteenth-century documentary hypotheses, thereby creating a methodological tautology in which the machine merely confirms preexisting human bias. Second, it demonstrates a confusion between variation in register or literary genre and authorial distinction, overlooking the fact that a single author may employ distinct vocabularies across diverse discursive contexts (legal, narrative, cultic). Third, it highlights the inherent statistical limitations of applying stylometry to corpora of Biblical Hebrew, given the restricted size and morphological complexity of such corpora. From a Reformed manuscriptological perspective, and in defense of the literary unity of the Pentateuch, the article argues that the distinctions detected by the algorithm do not constitute evidence of authorial multiplicity but instead reflect primarily functional differences of genre and technical vocabulary. It concludes that, although computational tools are useful, their use when detached from canonical theology tends to reproduce obsolete critical assumptions, thereby reaffirming the need for an exegesis that integrates quantitative data with the internal coherence of the biblical text.

KEYWORDS: Biblical stylometry; Documentary Hypothesis; Authorship of the Pentateuch; Machine learning; Textual criticism.