Check for updates

RESEARCH ARTICLE

Did the language of medieval inquisition records become richer over time? A diachronic lexical richness perspective on heresy trial proceedings from Languedoc and Italy, 1231–1400

[version 1; peer review: awaiting peer review]

Gideon Kotzé, David Zbíral, Robert L.J. Shaw, Zoltán Brys

Department for the Study of Religions, Faculty of Arts, Masaryk University, Brno, South Moravian Region, 60200, Czech Republic

v1 **First published:** 06 Oct 2026, **6**:394
https://doi.org/10.12688/openreseurope.25081.1
Latest published: 06 Oct 2026, **6**:394
https://doi.org/10.12688/openreseurope.25081.1

Abstract

While medieval inquisition records have a reputation for containing standardized and formulaic administrative language, some display a thematic richness and discursive variety that extends beyond the norm, especially in records of depositions. There is also a notion that registers were becoming richer over time, moving beyond the modest beginnings of inquisition documentation in the 1230s–40s as investigative interests expanded. However, thus far there has been no formal and systematic analysis of the language of these records that would address these questions. To study the elaboration of language and assess its growth over time, we apply nine lexical richness metrics to a representative sample of over twenty inquisition registers (with 2.28 million word tokens) from Italy and Languedoc, compiled between 1231 and the end of the fourteenth century. Regression analysis indicates a significant growth in lexical richness over time in all nine metrics. We also perform a comparative analysis of the registers, report on various aspects of the behaviour of vocabulary, and offer a historical interpretation.

Keywords

heresy trial records, pre-modern legal documents, lexical diversity, regression analysis, papal inquisition, Latin

This article is included in the European Research Council (ERC) gateway.

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

H2020 This article is included in the Horizon 2020 gateway.

Corresponding author: Gideon Kotzé (gideon.kotze@mail.muni.cz)

Author roles: Kotzé G: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Zbíral D**: Conceptualization, Data Curation, Funding Acquisition, Investigation, Project Administration, Resources, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Shaw RLJ**: Conceptualization, Data Curation, Funding Acquisition, Investigation, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Brys Z**: Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000442 (project "Networks of Dissent: Computational Modelling of Dissident and Inquisitorial Cultures in Medieval Europe").
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Kotzé G *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Kotzé G, Zbíral D, Shaw RLJ and Brys Z. **Did the language of medieval inquisition records become richer over time? A diachronic lexical richness perspective on heresy trial proceedings from Languedoc and Italy, 1231–1400 [version 1; peer review: awaiting peer review]** Open Research Europe 2026, **6**:394 https://doi.org/10.12688/openreseurope.25081.1

First published: 06 Oct 2026, **6**:394 https://doi.org/10.12688/openreseurope.25081.1

Introduction

The trials conducted against those suspected of heresy in the Middle Ages have long fascinated both historians and a broader public. Among the many reasons for this is the richness in the way the records of these proceedings sometimes express the voices and lives of their subjects. For instance, the popular impact of Emmanuel Le Roy Ladurie's *Montaillou* (Le Roy Ladurie, 1975) was founded on its use of vivid testimonies produced as part of the trials led by Jacques Fournier, bishop of Pamiers, between 1318 and 1325. Through the texture of their language, these appear to open a window onto the voices, lives and ideas of ordinary medieval people in a way that other types of texts of the same period rarely do. On the other hand, it is also understood that such richness is a relative rarity in inquisition trial records. Alexander Patschovsky (1991, pp. 266–7) indeed likened testimonies offering vivid narration to "rare sugars", in contrast to the "dry bread" of administrative prose that historians typically have to chew on. Regularity of language is very often the norm in records of medieval heresy trials: for instance, a recent study of the sentence register of the Dominican inquisitor Peter Seila (1241–2) found that nearly three quarters of the crimes narrated in its 649 summaries of guilt fall into just ten lexico-semantic types (Shaw, Hampejs and Zbíral, 2026). Such limited language is typically seen as the result of the mediation of inquisitors and inquisition notaries. The ways in which standardized inquisitorial questions and notarial translation and editing constrained expression have been central to the debate over how heresy was "constructed" through the trial process and its written record, e.g. (Grundmann, 1965, pp. 522–3) (Lerner, 1972) (Arnold, 2001, pp. 7, 48–56) (Biller, 2001) (Pegg, 2001, pp. 45–51, 57–61) (Pegg, 2003) (Pegg, 2012) (Zbíral and Shaw, 2022) (Pihko, 2024).

The co-existence of richer and more standardized language within the same registers is certainly recognized. It has even inspired attempts to get closer to the voices of deponents beneath the mediation of the trial authorities: that is, by looking for so-called "excesses" or "surpluses" of language that seem to exceed the inquisitorial and notarial formulae (Arnold, 2001, p. 114) (Arnold, 2019, p. 8) (Bruschi, 2009, pp. 14–15). Despite this interpretative interest, however, our understanding of which surviving registers present richer or more regular language remains impressionistic and piecemeal. There are certainly views regarding the balance found in particular registers and particular types of documents within them; thus, while Fournier's register is noted for the vivid language of some of its depositions, the aforementioned register of Peter Seila, which presents only compact notarial summaries of fault, is known for its more standardized appearance, even if it is not without unexpected details (Taylor, 2013, pp. 333–5) (Biller, 2022, p. 364) (Shaw, Hampejs and Zbíral, 2026). There is also a sense that testimonies became more elaborately expressed over time due to growing investigative interests (e.g., in doctrinal beliefs) and accumulated experience in pushing deponents for information that might reveal further crimes and suspects. The growing ecclesiastical concern to control lay religion (Asad, 1986) (Arnold, 2001, pp. 28–29) (Biget, 2007, p. 196) is thus often believed to have found its parallel in more detailed interrogation and recording. This point has been made most clearly for the first century of the inquisition in Languedoc (1230s–1320s), beginning with a contrastive comparison of the 1270s register of the Toulouse inquisition to older ones (Wakefield, 1973, p. 25). From there, a broad historiographical consensus has emerged that, as the scope of inquisitorial investigations grew during this period, the words of deponents were afforded more space (Biget, 2007, pp. 195–196), giving later registers greater detail than earlier ones (Biget, 1991, pp. 239–240) (Arnold, 2001, pp. 52, 114–115) (Vilandrau, 2001, pp. 48–49). Nevertheless, the notion that inquisition records increase in scope and detail has not gone unchallenged: historians have also noted the growing formalization and bureaucratization of inquisitorial procedure, of the sort that might yield more regular language (Zanella, 1995, pp. 16, 31–32). Moreover, all such comparisons remain effectively untested. One might rate registers as richer on the basis of a few strikingly idiosyncratic examples, and others as more regular due to their apparent absence. Yet such impressions may be misleading in the absence of formal assessment.

In order to systematically compare how rich or regular heresy trial records are in terms of their expression, this paper applies lexical richness measures to a corpus of inquisitorial registers, and presents a formal statistical analysis of the results. While we can define the discursive variety and limitations of these sources in multiple ways, lexical richness provides a strong proxy by gauging the diversity and repetition of words. The study covers a large and representative body of trial documentation from both Languedoc and Italy, the regions from which a large portion of the surviving records of medieval heresy trials derive. Inspired by approaches to linguistic changes over corpora (Baayen, 2009) (Štajner and Mitkov, 2012) (Korkiakangas, 2024), it adopts a diachronic perspective, starting from the earliest extant inquisition trial records of the 1230s and terminating at the end of the fourteenth century. The study thus tracks the use of language from the nascence of inquisitorial procedure and recording through to a more mature state. As a result, we are able to test historical assumptions both of the richness of individual registers and of the change in richness over time.

Lexical richness as a signal of the expansiveness of inquisition records

While there is a relative lack of scholarly consensus over the definition of "lexical richness" (McCarthy and Jarvis, 2010), we use this notion as an umbrella term to refer to the complexity, variety and distinctiveness of vocabulary (the set of distinct words or "types") found in a text. In this study, we propose that such lexical patterns can be interpreted as a signal for the broader discursive richness or regularity of inquisition records. In order to produce a systematic and statistically

robust appraisal of the linguistic character of inquisition records and move beyond more subjective characterizations, we calculate scores that formally describe aspects of the vocabularies in question. These can in turn be interpreted to construct profiles of lexical richness for each register, informing our understanding of the behaviour of their respective vocabularies.

We decided to organize this selection of *metrics* into three broad classes. The first is *lexical diversity*: that is, the diversity of lexical items (words) within a document, generally in comparison to the number of distinct words (Jarvis, 2013). The second is *lexical repetition*, which is inversely related to lexical diversity: it expresses the extent to which words within the vocabulary repeat (Tweedie and Baayen, 1998). The third, which we call *lexical distribution*, is more specifically related to the shape of the lexical frequency distribution, and represents a less standard but complementary approach to lexical richness. To motivate our choice of placing our chosen metrics in the third group (see Table 3) within this class, we refer the reader to Baixeries et al. (2013), Bentz and Kiela (2014), Bentz et al. (2015), Bérubé et al. (2018), and Shi and Lei (2022).

Our approach is informed by proposals for such classifications in the literature. Tweedie and Baayen (1998), following earlier work by Linnarud (1986), proposed the lexical categories of diversity, sophistication, density, and errors. Jarvis (2013), inspired by taxonomies in ecology, identified richness, evenness, disparity, dispersion, and importance as important considerations in quantitative linguistics. Following this tradition, we used the notion of lexical richness as an umbrella term to describe various metrics that relate to the variety of words in a text (Engber, 1995) (Read, 2000) (Lissón and Ballier, 2018), noting that "lexical diversity" is sometimes interchanged with this term (Jarvis, 2013). We refer the reader to Table 3 and the subsection "Analysis metrics" for a more detailed description of these metrics.

We implement multiple metrics falling under one of these three categories in order to provide a more comprehensive view of the expansiveness of the language of inquisition records: each metric describes a specific aspect of lexical richness that can be associated with certain patterns. As the three classes of metrics can be expressed through their different relationships with lexical richness, we can represent results in a clear comparative fashion, while relating them to either the relative constraint of the discourse (more repetition, fewer new words introduced) or its greater elaboration (less repetition, with more colourful expression). In addition, the use of multiple metrics provides distinct but compatible expressions of the linguistic character of inquisition registers.

These aspects of lexical richness can provide valuable historical interpretations. The introduction of new words at increased rates would suggest a broader range of topics being addressed in greater detail and more variety in the admixture of voices in the inquisitorial process: more open inquisitorial questioning, greater room for deponent narration (both during the interrogation and its subsequent editing), and/or greater individuality in notarial approach. By contrast, if we perceive a greater reliance on a narrower, higher-frequency lexicon, this might indicate a tighter, more standardized format of interrogation and/or notarial refinement, designed to provide evidence of quite specific (and perhaps quite preconceived) categories of guilt.

Using formal metrics as proxies of lexical richness also allows us to deploy inferential statistics. Most centrally, our focus on the change of richness over time can be operationalized through the application of a series of regression models that report whether any given measure shows a statistically significant change over time on the corpus selected for this study.

We can thus restate our research questions more formally: Firstly, can we observe an increase in lexical richness in a selection of inquisition registers from Italy and Languedoc from 1231 to 1400? Secondly, where do individual registers stand in lexical richness measures in comparison to others, and what are the implications for how we characterize them?

Corpus composition, processing, and metric analysis

Corpus preparation

The corpus selected for this study is well-adapted to a diachronic analysis of the lexical richness of heresy trial records in the first century and a half in which inquisitions occurred, comprising documentation from Languedoc and Italy. These regions witnessed sustained inquisitorial activity from the second third of the thirteenth century onwards, and a far larger and more continuous body of trial documentation survives from there than from anywhere else for the delimited period. They were also closely connected regions: although, as Kelly (1989) has emphasized, there was no single, centralized "Inquisition" and procedures remained locally directed, the inquisitorial cultures of Languedoc and Italy did not develop in isolation. In both, heresy trials were typically conducted by members of the Dominican Order, and there is good evidence for the circulation of inquisitorial handbooks between the two regions (Dondaine, 1947) (Parmeggiani, 2002) (Sackville, 2019, p. 373). Bringing these neighbouring traditions together therefore provides a sufficiently extensive and

continuous corpus by which to trace developments from the earliest documented inquisitions of heresy to the end of the fourteenth century.

The corpus used for this study (Table 1) comprises twenty-six inquisition registers of different lengths, all of which are in Latin for the vast majority of the text. Figure 1 presents a timeline of these registers. As our research question is on the development of the lexical richness of heresy trial records, the corpus only covers those, thus excluding other types of material related to the inquisition of heresy or circulated among inquisitors (e.g., papal legislation or inquisitors' manuals). While the corpus does not contain the totality of extant, or even edited, heresy trial records from Languedoc and Italy in the given period, it can be considered quite representative.

The corpus texts derive from modern scholarly editions (referenced in Table 1). A smaller part is drawn from digital editions. The majority was digitized from printed editions via professional robotic scanning and ABBYY FineReader OCR, producing PDF and DOCX files. After digitization, a team of historians competent in Latin used Microsoft Word — chosen for its ability to search by formatting and regular expressions (including in combination) — to remove editorial matter and correct OCR errors through a combination of string replacement and manual edits. Finally, we produced plain text files (UTF-8) for analysis.

In order to produce a more accurate account of the number of words and word types, we separated words from surrounding punctuation marks by applying a Latin tokenizer (Burns, 2023) to the text. (In this article, we use the term

Table 1. List of heresy inquisition registers from Languedoc and Italy constituting the corpus for this study.

Full name	Short name	Region	Date	Token count	Citation
Process against Bernard Niort and his family	Niort	Languedoc	1234–1235	6,145	(Grau Torras, 2015, pp. 214–229)
Sentences of William Arnold and Stephen of Saint-Thibéry	Arnold Sentences	Languedoc	1235–1242	8,092	(Wallnerová and Zbíral, 2025c)
Peter Seila's Register of Penances	Seila	Languedoc	1241–1242	32,362	(Duvernoy, 2001)
Sentences of Ferrer and William Raymond	Ferrer Sentences	Languedoc	1244	2,482	(Wallnerová and Zbíral, 2025b)
Register of Bernard of Caux and John of Saint-Pierre	Caux Doat 22	Languedoc	1243–1247	23,923	(Douais, 1900, pp. 90–114) (Duvernoy, 2002)
Register of Bernard of Caux and John of Saint-Pierre	Caux ms. 609	Languedoc	1245–1246	411,889	(Rehr, 2019)
Sentences of Bernard of Caux and John of Saint-Pierre	Caux Sentences	Languedoc	1244–1248	17,162	(Douais, 1900, pp. 1–89)
Register HHH of the Carcassonne inquisition	Carcassonne HHH	Languedoc	1246–1247	9,426	(Duvernoy, 1990)
Register of the notary (greffier) of the Carcassonne inquisition	Carcassonne Greffier	Languedoc	1250–1267	39,159	(Douais, 1900, pp. 115–301)
Book of sentences from Orvieto	Orvieto	Italy	1268	38,147	(D'Alatri, 1996, pp. 209–338)
Libellus of Armanno Pungilupo	Pungilupo	Italy	1208–1301	10,676	(Zanella, 1986)
Register of Pons of Parnac, Ranulph of Plaissac et al., Toulouse 1273–1282	Parnac	Languedoc	1273–1282	102,287	(Biller, Bruschi and Sneddon, 2011)

Table 1. *Continued*

Full name	Short name	Region	Date	Token count	Citation
Register of Nicolas of Abbeville et al., 1299–1300	Abbeville	Languedoc	1299–1300	40,152	(Davis, 1948)
Quaternus imbriviaturarum Beltrami Salvagni	Guglielmites	Italy	1300–1302	41,376	(Benedetti, 1999)
Acta Sancti Officii Bononie ab anno 1291 usque ad annum 1310	Bologna	Italy	1291–1310	207,834	(Paolini and Orioli, 1982)
Toulouse inquisition register fragment preserved by F. Graverol	Graverol	Languedoc	1282–1319	5,965	(Toti, 2020)
Four depositions (1301–1305) on beliefs and the way of life of the Autier group in Doat 34, 93v–103r	AutierBeliefsDoat34	Languedoc	1301–1305	2,495	(Wallnerová and Zbíral, 2025a)
Register of John Galand and William of Saint-Seine, 1283–1302	Galand 1	Languedoc	1283–1316	55,692	(Wallnerová and Zbíral, 2025d)
Register of Geoffroy of Ablis	Ablis	Languedoc	1308–1309	69,796	(Palès-Gobilliard, 1984)
Bernard Gui's Book of Sentences	Gui Sentences	Languedoc	1308–1323	277,676	(Palès-Gobilliard, 2002)
Register of Jacques Fournier	Fournier	Languedoc	1318–1325	680,040	(Duvernoy, 1965b) (Duvernoy, 1972)
Register DDD of the Carcassonne inquisition	Carcassonne DDD	Languedoc	1323–1327	48,874	(Duvernoy, 2003a)
Register GGG of the Carcassonne inquisition	Carcassonne GGG	Languedoc	1328–1329	57,821	(Duvernoy, 2003b)
Process against heretics in Giaveno by Alberto de Castellario	Castellario	Italy	1335	35,096	(Merlo, 1977, pp. 161–255)
Process against heretics in Piedmont by Tomasso of Casasco	Casasco	Italy	1373–1388	10,228	(Merlo, 1977, pp. 257–283)
Process against heretics in Piedmont by Antonio of Settimo	Settimo	Italy	1387–1388	44,312	(Amati, 1865a) (Amati, 1865b) (Pazé, 2016a) (Pazé, 2016b)
TOTAL				2,279,107	

Legend: List of heresy inquisition registers from Languedoc and Italy constituting the corpus for this study, sorted by mean date. The "Citation" column indicates the edition of the text used. In two cases, Caux Doat 22 and Settimo, we merged two incomplete editions, using the newer one as the basis and supplying missing material from the older one.

"token" to denote words and punctuation marks produced by the tokenization process.) We then applied a series of string transformations to standardize the texts (Table 2).

Our question relates to word choice, and thus could be compromised by including named entities. Their presence might result in unduly ascribing high lexical richness to registers which simply happen to record names of people and places

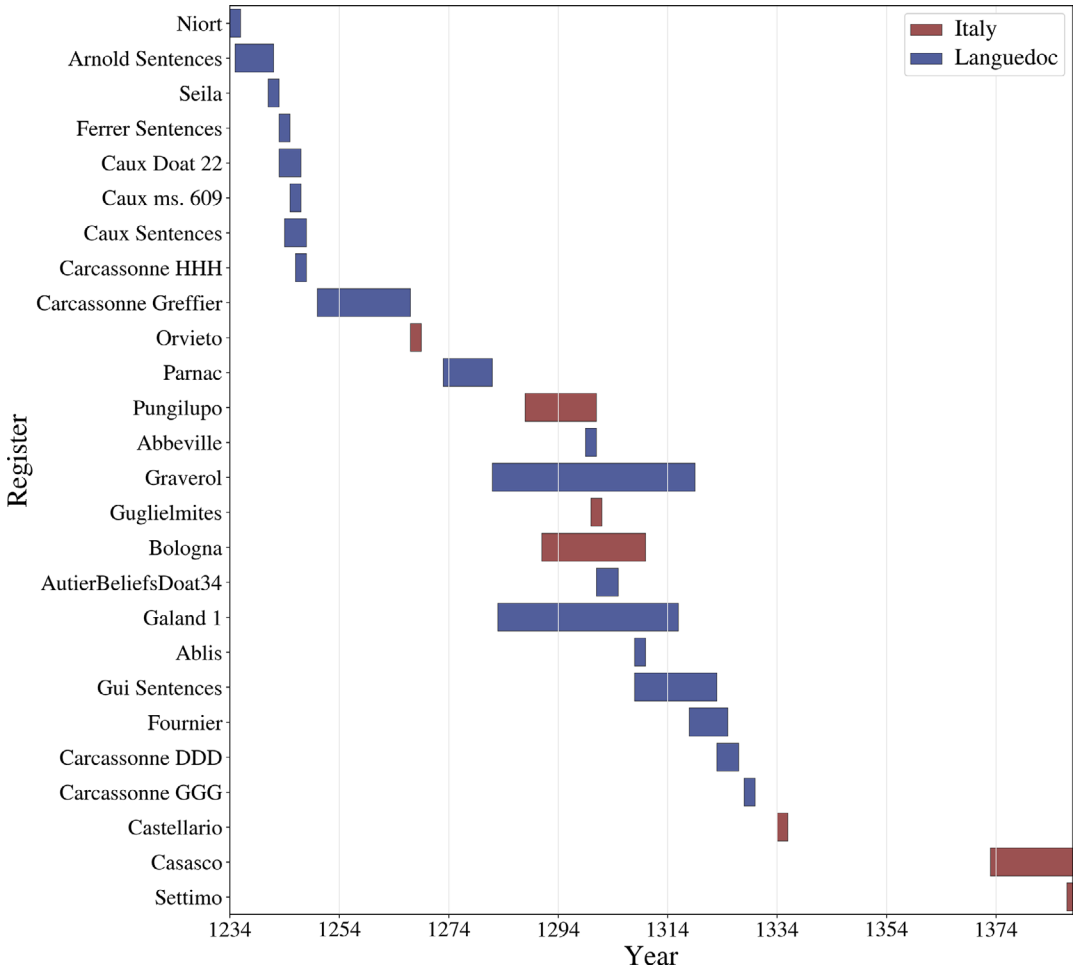


Figure 1. Timeline of inquisition registers from Languedoc and Italy, 1231–1400, in the corpus.
Legend: Timeline of inquisition registers from Languedoc and Italy, 1231–1400, in the corpus. The time span covers the earliest to latest possible date of register redaction, which sometimes results in a date range exceeding the dates of depositions covered (Pungilupo, Galand 1).

Table 2. String transformation patterns applied to the register texts.

String to match	String to replace with	Condition
ui	vi	word-initial
ue	ve	word-initial
æ	ae	none
œ	oe	none

Legend: String transformation patterns applied to the register texts.

more frequently, or in which those names are more diverse. A similar issue could arise from tokens related to dates, which represent noise rather than signal for the question at hand, since they simply identify times when proceedings and events occurred. Thus, we applied a masking approach where we ignored named entities and numerals in the calculation of metrics. We masked the corpus by converting each affected token to the arbitrary '@' character, which otherwise appears nowhere in the corpus, thereby enabling our scripts to recognize and ignore it. While the masking operationalization employed here is our own, the omission of named entities or proper nouns in the context of lexical richness and vocabulary studies has some backing in the literature (Laufer and Nation, 1995) (Dawson et al., 2021). The documentation of the LancsLex vocabulary profiling tool (Brezina, 2018) (Brezina and Gablasova, no date) also notes that "inclusion of proper nouns can lead to over-estimation of lexical diversity and sophistication of text".

Our masking approach was applied in the same way across all registers according to the following rules:

1. To remove named entities, we masked all tokens starting with uppercase letters under these conditions: (a) they do not begin with a lowercase letter anywhere in the corpus and (b) they do not appear exclusively in the start-of-sentence position in the corpus. Since the tokenizer failed to detect the boundaries of some sentences ending in abbreviations, we checked whether the token following the abbreviation started with a capital letter. If so, it was handled as a sentence boundary, triggering the start-of-sentence rules. We designed this general approach to avoid the imbalanced performance of pretrained named entity recognition models, which might better capture named entities in some registers than in others.

2. We masked all capitalized abbreviations, since they generally referred to names (e.g., "B." for "Bernardus"), but left lowercased abbreviations such as "etc." unmasked. Abbreviations are defined as tokens that contain one or more alphabetic characters and end in a full stop.

3. We masked all Roman and Arabic numerals.

After performing the masking across our twenty-six registers, 2,066,884 unmasked tokens remained (90.68% of the original size).

Analysis metrics

In order to examine the lexical richness of these registers, we applied a comprehensive selection of metrics (Table 3). They can be subdivided into three general types:

- Lexical diversity, which measures the vocabulary diversity.

- Lexical repetition, which measures the repetition of tokens, and is inversely related to diversity (as more repetition reduces the amount of unique tokens).

- Lexical distribution, which reflects aspects of the frequency distribution of words in the text. We apply this classification as an umbrella term for metrics that measure the evenness of the token distribution and those that measure the shape of the distribution curve.

Regarding lexical diversity, we did not consider the well-known type-token ratio (TTR), since it is particularly sensitive to text length (Tweedie and Baayen, 1998); as the number of words increases, the TTR value decreases monotonically, as dictated by Heaps' Law (Heaps, 1978): the rate of introduction of new words slows down relative to the accumulation of words. TTR will therefore be biased towards longer registers, making them appear more lexically diverse than would be merited by the actual range of vocabulary in the text. The same problem also affects metrics that are mathematical transformations of TTR to varying degrees (Hess, Sefton and Landry, 1986) (Tweedie and Baayen, 1998): specifically, the root TTR (Guiraud, 1960), the corrected TTR (Carroll, 1964), Somers' Index (Somers, 1966) and Herdan's *C* (Herdan, 1960). However, we have included several other lexical diversity metrics that apply various alternative approaches (Maas's a^2, Dugast's *U*, MATTR, MSTTR, MTLD, vocd-*D*, and HD-*D*; see Table 3 for citations).

Another set of metrics measures lexical repetition. This is generally inversely related to lexical diversity (Tweedie and Baayen, 1998), where higher scores indicate a more constrained diversity. Within this category, we included Yule's *K,* Herdan's V_m, and Simpson's Dominance Index (or *D*).

The third class of lexical richness metrics, concerning lexical distribution, is more diverse. They describe different aspects of token distribution. In this study, we included the Gini coefficient, Shannon's entropy, and Zipf's α from this category of metrics.

For most metrics, we used the implementation as available in the *lexicalrichness* package (Shen, 2022). For Shannon's entropy, we used the *sciPy* package (Virtanen et al., 2020). We implemented the formulas for the Gini coefficient (Kleiber and Kotz, 2003), ensuring that a score of 0 indicates perfect equality and 1 maximal inequality, and for Zipf's α, estimating the exponent of Zipf's law using Maximum Likelihood Estimation, using the *sciPy* Python package for optimization. We cross-validated these custom implementations against published Python code (Guest, 2016) (Rey, 2018) (Dovgalecs, 2025) and a base-R reimplementation (*optimize()*). The specific versions of the code are bundled in the supplementary data (Kotzé et al., 2026) — see Supplementary Data S2. For the implementation of Zipf's α in the context

Table 3. List of lexical richness metrics applied in the study. Higher values indicate a higher degree of lexical richness, if not indicated otherwise. "LR" refers to "lexical richness".

Metric	Type	Citation	Interpretation
MSTTR (Mean Segmental Type-Token Ratio)	lexical diversity (segmental)	(Johnson, 1944)	Splits text into segments of 100 tokens and calculates the type-token ratio (TTR) for each window, then averages them. The choice of a 100-token segment size is based on a suggestion in the original paper. It is also the default in the *lexicalrichness* package. Since the remaining tokens after the last segment are discarded, we thought it best to keep the size of the segment relatively low, as opposed to MATTR, where no tokens are discarded.
MATTR (Moving Average Type-Token Ratio)	lexical diversity (segmental)	(Covington and McFall, 2010)	Similar to MSTTR, but the window slides to the right, one token at a time. We base our choice of a window size of 500 tokens on a recommendation in the original paper. A larger window size would provide a more accurate TTR score for the window, while we justify our choice based on the fact that even our smallest registers used in analysis consist of over 6,000 tokens (see "Bootstrap resampling").
MTLD (Measure of Textual Lexical Diversity)	lexical diversity (segmental)	(McCarthy and Jarvis, 2010)	MTLD calculates the mean length of sequential text segments that maintain a TTR above a specific threshold. A higher score means higher LR over longer continuous stretches. We applied the default TTR threshold of 0.72 as suggested in the original paper.
HD-*D*	lexical diversity (hypergeometric)	(McCarthy and Jarvis, 2010)	Calculated by drawing a random sample of words and estimating the probability of encountering any type (unique word) in the sample. This is done for all types and then the probabilities summed. We used the standard number of random draws, 42, as implemented in the *lexicalrichness* package (Shen, 2022).
vocd-*D*	lexical diversity (hypergeometric)	(McKee, Malvern and Richards, 2000) (MacWhinney, 2014)	Repeatedly samples sets of tokens and compares their mean TTR to an "ideal" TTR curve. 100 samples of 35 tokens are taken, the mean TTR computed, then again from 36 random tokens, and so on, up until and including 50 random tokens. A curve is fitted from the TTR score to the token size to obtain a score. All previous steps are applied 3 times, then averaged. We applied the standard implementation in the *lexicalrichness* package (Shen, 2022).
Maas's a^2	lexical diversity (logarithmic)	(Maas, 1972)	Uses squared logarithms of the token count that represent the rate of type/token change irrespective of text length. A lower score indicates a higher LR.
Dugast's U^2	lexical diversity (logarithmic)	(Dugast, 1978) (Dugast, 1979)	The inverse of Maas's constant. A higher U^2 indicates a higher LR.
Yule's *K*	lexical repetition	(Yule, 1944)	Calculates the probability that any two words sampled from the text are the same. A higher score means lower LR.
Simpson's *D* (Dominance Index)	lexical repetition	(Simpson, 1949)	A simplification of Yule's *K*. A higher score indicates lower LR.
Herdan's V_m	lexical repetition	(Herdan, 1955)	Based on Yule's *K*, it disregards the scaling factor. A higher score contributes to lower LR.
Gini coefficient	lexical distribution (evenness)	(Gini, 1912) (Kleiber and Kotz, 2003)	A score closer to zero indicates perfect equality, where every word is used with the same frequency (higher LR) while a score closer to one indicates maximum inequality, where a few words dominate the text, which we interpret as contributing to lower LR. We applied the bias-corrected formula (Kleiber and Kotz, 2003) in order to ensure correct scaling of the coefficient.

Table 3. *Continued*

Metric	Type	Citation	Interpretation
Zipf's α	lexical distribution (shape)	(Zipf, 1935) (Clauset, Shalizi and Newman, 2009)	Models the shape of the word frequency rank distribution. α is the exponent parameter, where lower values indicate flatter distributions, interpreted as contributing to higher LR. We estimated the exponent of Zipf's law using Maximum Likelihood Estimation (MLE) over rank-ordered word frequencies, which is the standard choice for fitting power-law distributions (Clauset, Shalizi and Newman, 2009). We optimized the parameter using Brent's method, falling back to log-log linear regression if MLE failed.
Shannon's entropy	lexical distribution (evenness)	(Shannon, 1948)	Measures the average level of uncertainty and evenness of words. A more uneven spread of words with high rates of repetition has lower entropy, while a higher degree of evenness and lower predictability is associated with a higher rate of LR.

Legend: List of lexical richness metrics applied in the study. Higher values indicate a higher degree of lexical richness, if not indicated otherwise. "LR" refers to "lexical richness".

of lexical richness, we refer the reader to Baixeries et al. (2013), Koplenig (2018), and Fang et al. (2025). Similarly, for the application of the Gini coefficient to word frequency distributions, see Popescu and Altmann (2006).

The metrics, listed in Table 3 alongside key references, are best applied to coherent textual units. We analysed whole registers as units, rather than individual documents, as many of them are too short for metric stability (Tweedie and Baayen, 1998), and would have to be dropped. This is not without limitations, as we cannot highlight individual depositions in which lexical richness significantly exceeds the register average — a feature that might relate to the aforementioned "excesses" or "surpluses" of speech. Nevertheless, this less granular approach achieves our main analytical goal: comparing in a valid manner the overall lexical richness of Italian and Languedocian registers.

Bootstrap resampling

It is well known that lexical richness metrics are generally sensitive to text length (albeit to various degrees), making them unreliable when applied to shorter texts (McCarthy and Jarvis, 2010). This informed our choice of metrics and unit of analysis as explained above. Nevertheless, even taking the whole-register approach, some of our registers (a few of which constitute only fragments of the original) are still quite short. To determine whether we can reliably apply the chosen metrics, we ran a bootstrap resampling analysis. Its main goal was to determine the minimum number of tokens at which the chosen metrics become stable enough for confident use in our study (McCarthy and Jarvis, 2010).

Our strategy involved applying the metrics in Table 3 to expanding selections of unmasked tokens from the registers, taking note of the relative change of the metric value between steps.

Before analysis, we preprocessed all texts to ensure compatibility with the *lexicalrichness* package (Shen, 2022), which implements all lexical diversity and lexical repetition metrics except for repeat rate. This entailed:

- the removal of punctuation marks, digits, as well as variations of dashes and hyphens;
- the lowercasing of tokens.

These steps further reduced the number of tokens from 2,066,884 to 1,794,300.

Each selection of tokens was randomly drawn from each register with replacement, allowing reuse of the same tokens. We then calculated metrics starting at 2,000 tokens, adding 2,000 more for each new sampling step. Since we required a minimum of two consecutive steps, and thus three sampling points, to measure stability (see below), we excluded four registers that fell below the 6,000 token threshold: Ferrer Sentences (2,333), AutierBeliefsDoat34 (2,469), Niort (5,563),

and Graverol (5,590), leaving us with 22 registers to analyse. (Throughout the text, we refer to registers through their custom short name; a conventional full name and reference to the edition used are listed in Table 1.)[1]

Across a total of fifty iterations, we calculated all metrics on these selections, as well as the mean values for the metrics across all samples. We determined the stability threshold for any metric by calculating the first token count where the average relative change across all registers fell below 2.5% over two consecutive steps. This identifies the "elbow" of the stabilization curve (McCarthy and Jarvis, 2010), beyond which the metric can be regarded as sufficiently stable.

Most metrics reached this point by 6,000 tokens, with a few at higher thresholds: Zipf's α (8,000), Dugast's U^2 (10,000), Maas's a^2 (10,000) and Gini coefficient (12,000). Including all metrics initially considered would require opting for a higher token threshold, and thus dropping additional registers. We prioritized retaining registers over retaining metrics, and thus decided to apply the 6,000 tokens threshold to select metrics and registers for our subsequent analysis, retaining all 22 registers after the omission of the four smaller works as mentioned, and omitting the aforementioned four metrics but retaining, out of the original thirteen, HD-*D*, vocd-*D*, Herdan's V_m, MATTR, MSTTR, MTLD, Shannon's entropy, Simpson's *D*, and Yule's *K*.

Regression analysis

To address our research question whether lexical richness increased in inquisition registers over the period in our focus, we performed a regression analysis of the nine retained metrics on the remaining twenty-two registers (seven from Italy and fifteen from Languedoc). This analysis estimates nine ordinary least squares (OLS) regressions (Montgomery, Peck and Vining, 2021), one for each lexical diversity measure, using register mean year (average between the earliest and latest possible date of its original redaction) and preprocessed unmasked token count as predictors. While our primary focus was the association with time, we included the count of preprocessed unmasked tokens as a control variable to adjust for variation in register length.

Document type, or 'genre', is usually included as a control variable in similar studies (Heng, Pu and Liu, 2023) (Zheng, 2025b), as genre differences can influence lexical choice, vocabulary, and word distribution levels. However, an additional control variable would require more observations in order to draw robust statistical conclusions (Brooks and Barcikowski, 2012). Moreover, a considerable proportion of the registers consist of multiple document types in differing proportions. We also observed notable differences even within genres, such as records of sentencing, within and across our inquisition registers. We therefore proceeded without considering genre, although with the caveat that genre is a potential explanatory variable that we did not test.

Each model used all twenty-two registers. The descriptive characteristics of the untransformed variables are shown in Table 4.

Data transformations

The unmasked preprocessed token count was transformed using the natural logarithm of one plus the value $log(1 + x)$ to reduce skew related to register length and to stabilize residual variance. All variables were then standardized using z-scores. We inverted Herdan's V_m, Simpson's *D*, and Yule's *K* to ensure that higher values indicate greater lexical richness, for consistency with the other measures.

Data screening and pre-model evaluation

Pearson correlation matrices showed low correlations between predictors ($r = 0.16$). Variance inflation factors (VIF) were also below critical thresholds (VIF = 1.03), indicating no collinearity concerns (Hahs-Vaughn, 2024). An a priori sensitivity analysis for regression ($\alpha = .05$, $n = 22$, two predictors) indicated that the minimum detectable effect size with 80% power corresponded to approximately $R^2 = 0.34$, which means that effects below this level may remain undetected. The Bonferroni outlier test (Fox and Weisberg, 2019) identified Orvieto as an outlier in MSTTR ($t = 3.54$, $p = 0.048$), MATTR ($t = 4.23$, $p = 0.01$) and in MTLD ($t = 16.54$, $p < 3E\text{-}11$). Consequently, winsorization was applied to Orvieto, replacing the three extreme values with the closest values within the non-outlying ones. After winsorization, Cook's distance values for all models were well below the conventional cutoff of one, indicating no other multidimensional outlier in the dataset (Cook and Weisberg, 1986).

[1]Note that we distinguish between the unmasked token count as the appropriate reflection of register size, and the further removal of tokens as a prerequisite for the metric calculations. This distinction is however inconsequential in our case, as all registers with an unmasked token count exceeding 6,000 also retain more than 6,000 tokens after the aforementioned preprocessing.

Table 4. Characteristics of the variables before transformations.

	Minimum	Maximum	Mean	SD	n
Predictor variables					
Mean year	1239	1388	1294	43	22
Preprocessed token count	6648	558343	80940	130203	22
Outcome variables					
MSTTR	0.56	0.84	0.70	0.06	22
MATTR	0.32	0.68	0.48	0.08	22
MTLD	37.83	364.97	91.44	64.18	22
HD-D	0.75	0.90	0.85	0.04	22
vocd-*D*	47.85	170.03	107.28	31.82	22
Yule's *K*	82.01	218.77	128.45	39.82	22
Simpson's *D*	0.008	0.022	0.013	0.004	22
Herdan's V_m	0.09	0.15	0.11	0.02	22
Shannon's entropy	7.32	10.18	8.78	0.77	22

Legend: Characteristics of the variables before transformations.

Correlation among lexical diversity measures

Because the sample size (n = 22) was insufficient for stable multiple-outcome regression models, we estimated nine parallel OLS models instead (Stevens, 2012). However, for context, correlations (signed r^2) among the nine lexical diversity measures are presented in Figure 2.

Results

In Table 5, we present results from the nine OLS models applied to the lexical richness measures. Despite the modest sample size (n = 22), F-tests indicated that seven of the nine models were significant at the $p < .05$ level and the remaining two (MSTTR and MTLD) at the $0.05 \leq p \leq .08$ marginally significant level. Across these models, the regression assumptions were largely satisfied. Durbin–Watson statistics ranged from 1.09 to 1.75 (Durbin and Watson, 1951) suggesting possible residual dependence; however, this deviation reached statistical significance only in two models (MTLD and vocd-*D*; p = 0.02, p = 0.04, respectively). Residuals showed no meaningful deviation from a normal distribution, as indicated by Kolmogorov–Smirnov tests ($\forall p \geq 0.59$) and by Shapiro–Wilk tests ($\forall p > 0.34$). Post-hoc power estimates of the fitted models indicated that five models exceeded the 0.80 threshold, while four models (MSTTR, MATTR, MTLD and vocd-*D*) fell below this level (0.56–0.79). However, the Breusch–Pagan test (Breusch and Pagan, 1979) indicated acceptable residual variance only in five of the nine models ($\forall p > 0.05$). In four models (HD-*D*, Yule's *K*, Simpson's *D* and Herdan's V_m) the test yielded $0.03 \leq p \leq 0.05$, indicating departure from constant residual variance. Accordingly, all coefficient tests (standard errors, t-values, and p-values) were conducted using heteroscedasticity-consistent (HC3) covariance estimators (Long and Ervin, 2000). This estimator is also generally recommended in small samples (MacKinnon and White, 1985).

Using HC3 robust standard errors, seven of the nine models showed a statistically significant positive association between mean register year and lexical richness ($p < .05$), with the remaining two (MTLD, p = .052; MSTTR, p = .074) showing effects in the same direction at marginal significance levels. Taken together, the results converge across all nine measures. Our results indicate a diachronic rise in lexical diversity in the inquisitorial registers from Languedoc and Italy between 1231 and 1400.

Comparative analysis

Global analysis

Besides examining the relation between the lexical richness of inquisition registers and the time of their creation, we conducted a systematic comparison. In the first step, we measured the deviation of each register from the predicted temporal trend. In the second step, we compared each register against a precomputed corpus average of lexical richness measures.

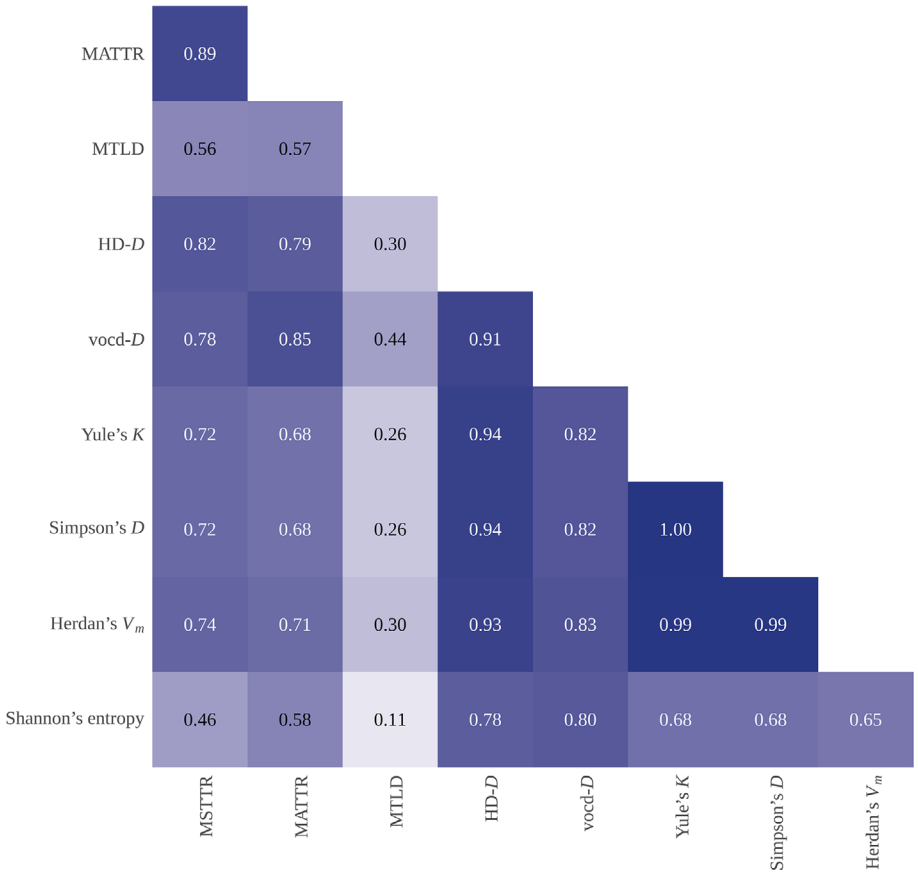


Figure 2. Signed r^2 values among lexical diversity measures. Legend: Signed r^2 values among lexical diversity measures.

To calculate the temporal deviations, we fitted separate OLS models for each metric using the regression parameters described in the previous section. However, we omitted winsorization, as extreme values are especially interesting for this analysis. To mitigate the overrepresentation of highly-correlated metrics (Figure 2) and simplify the results, we grouped metrics into functionally coherent categories (Table 6): due to their focus on technical mechanism, these categories sit at a level below the threefold division of lexical richness measures defined earlier. The limited sample size meant that we could not pursue formal dimensionality reduction or reliability testing (e.g., factor analysis or internal consistency measures); therefore, we treated these category scores as purely descriptive. Shannon's entropy is a single-category member, and thus has the same name as the metric.

We averaged the residuals — the differences between the expected values according to the fitted line produced by the models and the observed values — within each category for each register. We then averaged these category means to produce a single residual for each register representing its deviation from the temporal trend. Using these values, Figure 3 ranks registers from more to less lexically diverse than predicted, where each is categorized by geographical area (colour) and year mean (shade).

There are no readily observable patterns in deviation from the trend with respect to region or period. For instance, the most extreme outliers — Orvieto (positive) and Seila (negative) — are both early and from different areas, while we also observe a mixed representation of Italian registers in both halves of Figure 3. Nevertheless, the observed deviations highlight interesting perspectives on individual registers, discussed in the next subsection.

We also measured the relative difference in lexical richness for each register by converting each metric into a z-score, relative to the corpus mean for that metric. The z-scores were then averaged to determine each register's composite deviation from the corpus mean across all four categories, which we present as a heat map in Figure 4. The registers are sorted by date, from older to later; as a result, the general trend from lexically poorer (red shades) to richer (blue shades) is apparent.

Table 5. **Results from nine ordinary least squares models predicting lexical richness measures.**

Standardized MSTTR (window = 100)				
	b	**SE**	**t**	**p**
Year	0.50	0.27	1.89	0.07
Token count	−0.23	0.25	−0.92	0.37
R^2 = .27, adj. R^2 = .19, F(2, 19) = 3.45, p = .05; DW = 1.47, p = .19; KS p = .92; BP p = .07; pw = .64				
Standardized MATTR (window = 500)				
	b	**SE**	**t**	**p**
Year	0.58	0.21	2.79	0.01
Token count	−0.15	0.19	−0.82	0.42
R^2 = .34, adj. R^2 = .27, F(2, 19) = 4.82, p = .02; DW = 1.48, p = .19; KS p = .64; BP p = .47; pw = .79				
Standardized MTLD				
	b	**SE**	**t**	**p**
Year	0.49	0.24	2.07	0.05
Token count	−0.12	0.19	−0.60	0.56
R^2 = .23, adj. R^2 = .15, F(2, 19) = 2.91, p = .08; DW = 1.09, p = .02; KS p = .85; BP p = .29; pw = .56				
Standardized HD-*D*				
	b	**SE**	**t**	**p**
Year	0.60	0.26	2.32	0.03
Token count	0.01	0.24	0.03	0.98
R^2 = .36, adj. R^2 = .29, F(2, 19) = 5.28, p = .02; DW = 1.27, p = .05; KS p = .96; BP p = .04; pw = .83				
Standardized vocd-*D*				
	b	**SE**	**t**	**p**
Year	0.54	0.22	2.43	0.03
Token count	0.08	0.21	0.37	0.71
R^2 = .31, adj. R^2 = .24, F(2, 19) = 4.35, p = .03; DW = 1.10, p = .04; KS p = .59; BP p = .26; pw = .75				
Standardized and inverted Yule's *K*				
	b	**SE**	**t**	**p**
Year	0.67	0.25	2.68	0.01
Token count	−0.01	0.25	−0.03	0.97
R^2 = .44, adj. R^2 = .38, F(2, 19) = 7.54, p < .01; DW = 1.38, p = .12; KS p = .78; BP p = .03; pw = .94				
Standardized and inverted Simpson's *D*				
	b	**SE**	**t**	**p**
Year	0.67	0.25	2.68	0.01
Token count	−0.01	0.25	−0.03	0.97
R^2 = .44, adj. R^2 = .38, F(2, 19) = 7.54, p < .01; DW = 1.38, p = .11; KS p = .78; BP p = .03; pw = .94				
Standardized and inverted Herdan's V_m				
	b	**SE**	**t**	**p**
Year	0.66	0.24	2.71	0.01
Token count	−0.04	0.24	−0.16	0.87
R^2 = .43, adj. R^2 = .37, F(2, 19) = 7.12, p < .01; DW = 1.24, p = .05; KS p = .83; BP p = .04; pw = .93				
Standardized Shannon's entropy				
	b	**SE**	**t**	**p**
Year	0.62	0.21	2.95	<0.01

Table 5. *Continued*

Standardized Shannon's entropy				
	b	**SE**	**t**	**p**
Token count	0.28	0.21	1.30	0.21
R^2 = .51, adj. R^2 = .46, F(2, 19) = 10.05, p < .01; DW = 1.75, p = .47; KS p = .94; BP p = .87; pw = .98				

Legend: Results from nine ordinary least squares models predicting lexical richness measures. Year is the mean register year, z-standardized. Token count is the preprocessed token count, log(1 + *x*)-transformed and *z*-standardized. Coefficient standard errors, t-values, and p-values are based on heteroscedasticity-consistent covariance (HC3) estimators. DW = Durbin–Watson statistic assessing residual autocorrelation; KS = Kolmogorov–Smirnov test for residual normality; BP = Breusch–Pagan test for heteroscedasticity; pw = post-hoc power of the fitted model.

Table 6. Categorization of lexical richness metrics.

Category	**Metrics**
Segmental lexical diversity	MATTR, MSTTR, MTLD
Hypergeometric lexical diversity	vocd-*D*, HD-*D*
Lexical repetition	Yule's *K*, Herdan's V_m, Simpson's *D*
Shannon's entropy	Shannon's entropy

Legend: Categorization of lexical richness metrics. Segmental lexical diversity metrics rate local word introduction, hypergeometric lexical diversity metrics estimate the probability of drawing unique words from a given finite population, lexical repetition metrics quantify the extent to which words in a text are repeated, while Shannon's entropy measures the evenness of word usage across the vocabulary — here, a higher value means that more words are used often, meaning a consistently broader set of words from which the writer is drawing, corresponding to higher lexical richness.

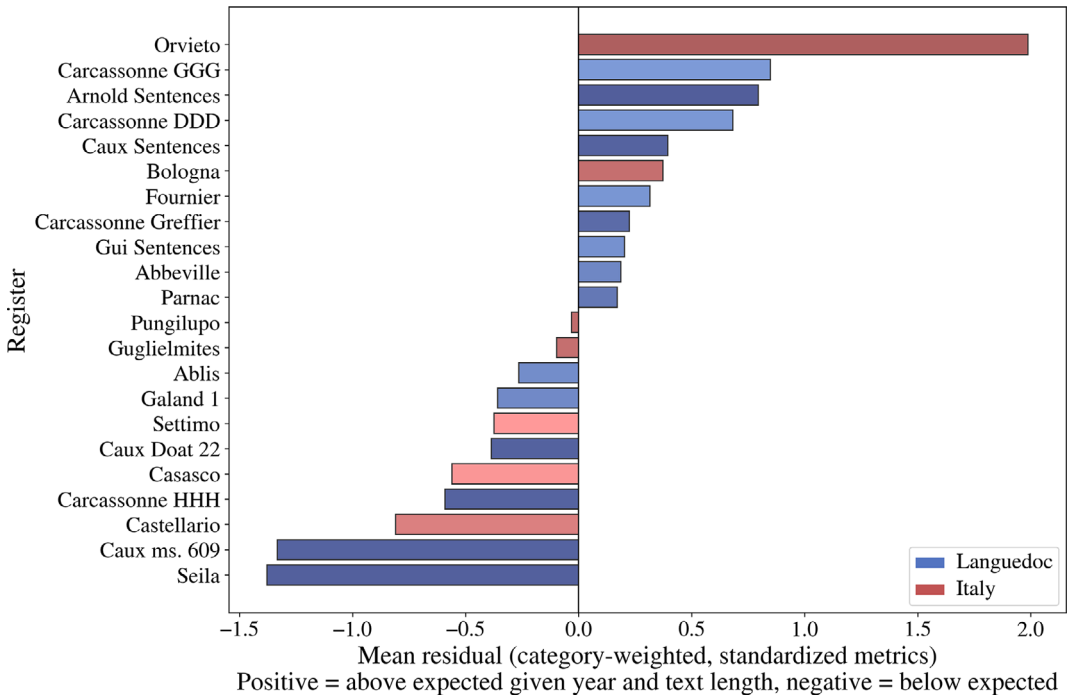


Figure 3. Inquisition registers from Languedoc and Italy, 1231–1400, ranked by their mean residual. Legend: Inquisition registers from Languedoc and Italy, 1231–1400, ranked by their mean residual, after controlling for year mean and unmasked token count. Different shades of blue (Languedoc) and red (Italy) indicate the date of the register, where older registers are darker, and more recent ones lighter.

Several registers are associated with a mix of poorer and richer categories, underscoring the multidimensionality of lexical richness. Pungilupo, for example, represents a combination of below average entropy (associated with a lower amount of evenness, contributing to lexical richness), but a lower rate of lexical repetition (inversely related to lexical

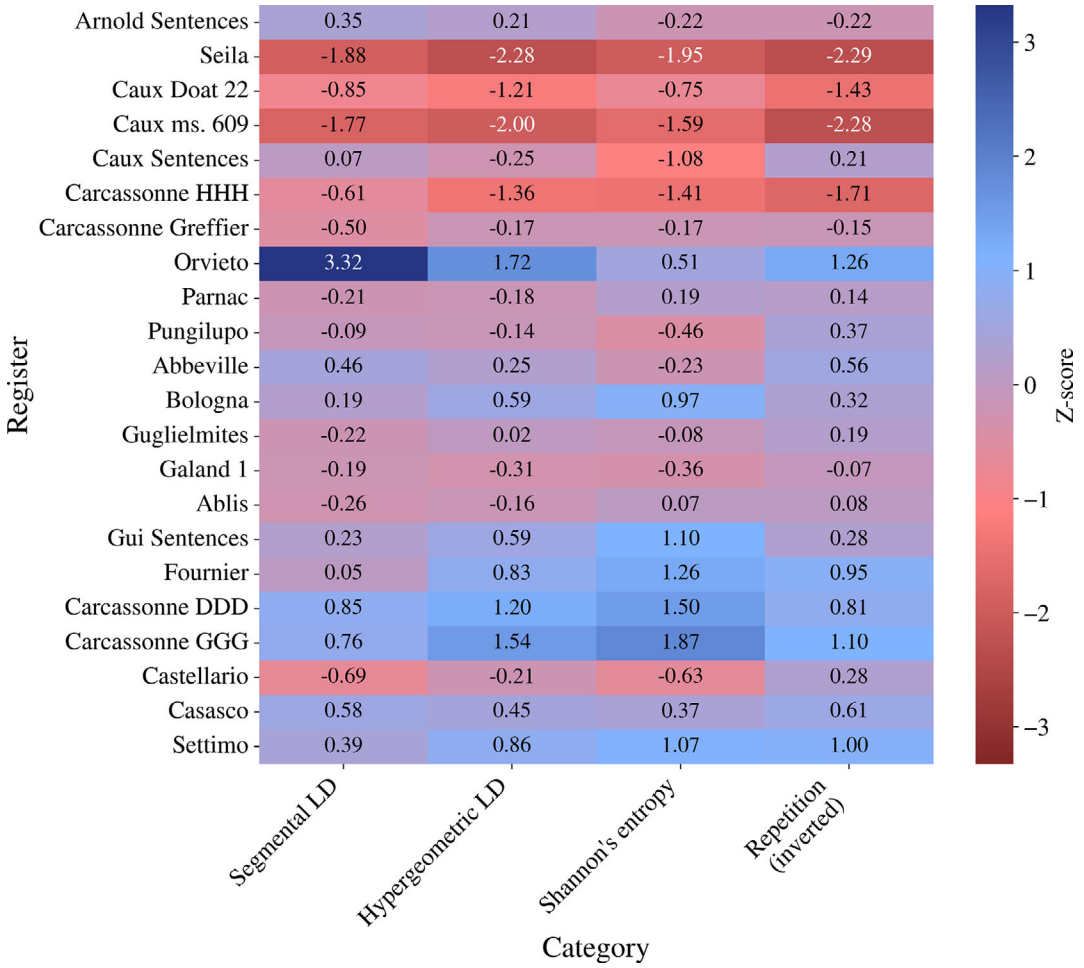


Figure 4. Heat map comparing registers against the corpus average for each category of lexical richness metrics. Legend: Heat map showing how individual inquisition registers from Languedoc and Italy, 1231–1400, compare against the corpus average for each category of lexical richness metrics. The z-scores express, in standard deviations, the extent of the deviation of the category value from the corpus average. The z-scores for repetition are inverted, as in this category, higher scores suggest lower lexical richness. The registers are sorted from earliest (top) to latest (bottom).

richness), at 0.83 SDs higher. Conversely, several registers show a relatively consistent tendency towards the corpus average across all metric categories (Ablis, Galand 1, Parnac, Guglielmites).

Complementing this output, Figure 5 presents parallel coordinates visualizing lexical richness profiles using the category-based z-scores against the corpus mean (the horizontal line at zero). It illustrates the trend discovered by the regression models towards higher lexical richness in later registers across multiple categories, despite some variation. Notably, larger registers from around 1300 and some smaller, older registers display similar levels of diversity around the corpus average, while the consistently non-diverse, large Caux ms. 609 and the smaller Seila appear at the bottom.

To facilitate comparison, we calculated a composite lexical richness z-score by averaging the z-scores from all four categories, generating a list sorted by descending composite z-score (Table 7). While this composite score is an analytical construct that assumes equal weighting from all categories, it nevertheless provides a useful indication of general trends in category metrics. For each row, we highlighted in bold the category that has the highest absolute value, while we italicized the highest category absolute value across all registers. A Pearson correlation of the composite z-score against the year means is moderately positive ($r = 0.59$), aligning with our findings that lexical richness increases over time.

Focusing on regional composite z-score means, Languedoc is slightly below (−0.21) while Italy is higher (+0.45). Moreover, each of the four metric categories follows the trend (i.e., all negative for Languedoc, all positive for Italy).

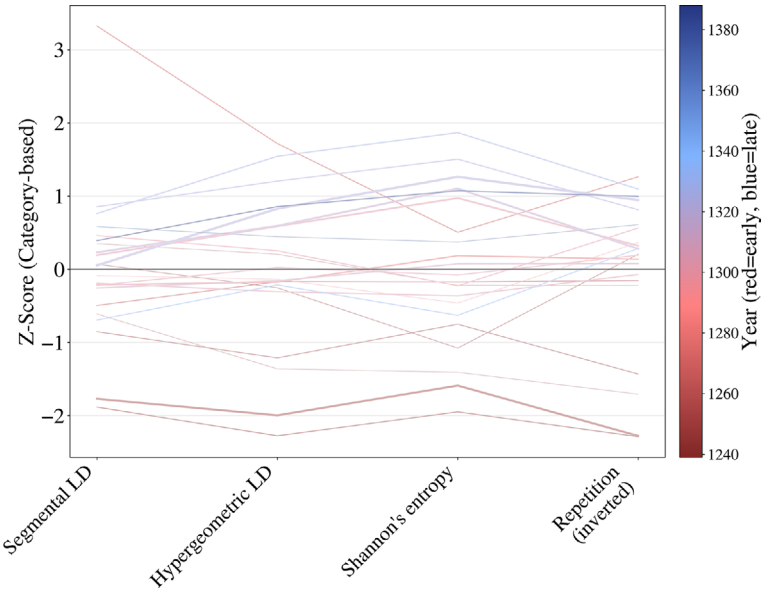


Figure 5. Parallel coordinates chart visualizing lexical richness profiles of inquisition registers using the category-based z-scores. Legend: Parallel coordinates chart visualizing lexical richness profiles of inquisition registers from Italy and Languedoc, 1231–1400, using the category-based z-scores. Each register is represented by a line that connects z-score values for each metric category. The y-axis reflects the extent to which the z-score deviates from the corpus mean. Thicker lines indicate larger registers, based on log-transformed values of the unmasked token counts. Z-scores for repetition are inverted, as higher scores in this category contribute to lower lexical diversity.

Although Italy's small sample precludes formal significance testing, the means suggest regional variation. Nevertheless, we see a clear temporal progression also among registers from Languedoc, suggesting that increasing richness is not merely a regional artefact. For instance, two earlier Carcassonne registers (HHH, 1246–1247; Greffier, 1250–1267) score − 1.27 and − 0.25, respectively, while two from later years (DDD, 1323–1327; GGG, 1328–1329) reach 1.09 and 1.32, respectively. Standing chronologically in between, Parnac (1273–1282; −0.02 SD) and Ablis (1308–1309; −0.07 SD) remain near the mean, thereby following the general trend.

Some registers perform consistently across all categories. Guglielmites, Parnac, Ablis, and Galand 1 display well-balanced levels of lexical diversity, repetition, and distribution, with Casasco (range: 0.24 SD) and Galand 1 (range: 0.29 SD) having the most consistent z-scores across categories, being quite close to each other. Several registers (Seila, Caux ms. 609, Caux Doat 22, Carcassonne HHH, Carcassonne Greffier, Galand 1) are consistently lexically poor, while others (Orvieto, Gui Sentences, Fournier, Carcassonne DDD, Carcassonne GGG, Casasco, Settimo, Bologna) are consistently lexically rich.

A few registers exhibit a higher degree of imbalance, with higher variance across category z-scores, most notably Orvieto (range = 2.82 SD), Caux Sentences (1.28 SD), and Fournier (1.21 SD). These contrasting category values demonstrate that metric categories capture different aspects of lexical richness, where the performance in one category by a register does not necessarily determine its performance in another. For example, the extremely high segmental lexical diversity z-score of Orvieto (3.32 SD), which represents higher rates of local word introduction, is counterbalanced by a mildly positive Shannon's entropy z-score (0.51 SD). Similarly, in Caux Sentences, we observe a relatively uneven lexical distribution (Shannon's entropy SD: −1.08), but a slightly above-average inverted lexical repetition z-score (0.21 SD).

Given these observed variances, the presence of specific passages characterized either by idiosyncratic diction or (on the other hand) by extensive boilerplate is thus far from defining, in itself, for the overall lexical richness of registers. This underlines the usefulness of our comprehensive use of different metrics, as they indeed proved to express different aspects of lexical richness.

Specific registers

Beyond these general comparisons, our analysis provides a lexical richness perspective on specific registers. The highest-ranking text (Table 7), Orvieto, is a clear outlier, exceeding both the corpus average and the temporal trend (Figure 3).

Table 7. List of registers ranked by the average of their category z-scores.

Register	Area	Rank	Z-score	Seg.	Hyp.	Ent.	Rep.
Orvieto	Italy	1	+1.70	***+3.32***	+1.72	+0.51	+1.26
Carcassonne GGG	Languedoc	2	+1.32	+0.76	+1.54	*+1.87*	+1.10
Carcassonne DDD	Languedoc	3	+1.09	+0.85	+1.20	*+1.50*	+0.81
Settimo	Italy	4	+0.83	+0.39	+0.86	*+1.07*	+1.00
Fournier	Languedoc	5	+0.77	+0.05	+0.83	*+1.26*	+0.95
Gui Sentences	Languedoc	6	+0.55	+0.23	+0.59	*+1.10*	+0.28
Bologna	Italy	7	+0.52	+0.19	+0.59	*+0.97*	+0.32
Casasco	Italy	8	+0.50	+0.58	+0.45	+0.37	*+0.61*
Abbeville	Languedoc	9	+0.26	+0.46	+0.25	−0.23	*+0.56*
Arnold Sentences	Languedoc	10	+0.03	*+0.35*	+0.21	−0.22	−0.22
Parnac	Languedoc	11	−0.02	*−0.21*	−0.18	+0.19	+0.14
Guglielmites	Italy	12	−0.02	*−0.22*	+0.02	−0.08	+0.19
Ablis	Languedoc	13	−0.07	*−0.26*	−0.16	+0.07	+0.08
Pungilupo	Italy	14	−0.08	−0.09	−0.14	*−0.46*	+0.37
Galand 1	Languedoc	15	−0.23	−0.19	−0.31	*−0.36*	−0.07
Carcassonne Greffier	Languedoc	16	−0.25	*−0.50*	−0.17	−0.17	−0.15
Caux Sentences	Languedoc	17	*−0.26*	+0.07	−0.25	−1.08	+0.21
Castellario	Italy	18	−0.31	*−0.69*	−0.21	−0.63	+0.28
Caux Doat 22	Languedoc	19	−1.06	−0.85	−1.21	−0.75	*−1.43*
Carcassonne HHH	Languedoc	20	−1.27	−0.61	−1.36	−1.41	*−1.71*
Caux ms. 609	Languedoc	21	−1.91	−1.77	−2.00	−1.59	*−2.28*
Seila	Languedoc	22	−2.10	−1.88	**−2.28**	**−1.95**	***−2.29***

Legend: List of registers ranked by the average of their category z-scores (standard deviations above/below the corpus mean). In order to preserve all values in the given amount of space, we abbreviated the names of the categories: "Seg." is "segmental lexical diversity", "Hyp." is "hypergeometric lexical diversity", "Ent." is "Shannon's entropy", and "Rep." is "lexical repetition". Values that show the highest magnitude for each row, irrespective of direction, are in italics. The highest absolute values for categories overall are in bold.

It ranks first in three of the four categories: segmental lexical diversity (driven by MTLD, showing sustained diversity over long stretches), hypergeometric lexical diversity, and inverted (lesser) repetition. Such a high ranking is far from expected; sentence registers tend to narrate crimes in a more summary fashion, editing down the contextual detail of underlying depositions to focus on specific guilt categories, and as such, we would expect a tighter, more formulaic vocabulary. Orvieto distinguishes itself by highly rhetorical language, including elaborated words and word forms usual in literature but less so in heresy trial records (*oblitus*, *dampnabilius*, *decernentes* etc.). Such florid language may well have been used to enhance the performative impact of sentencing: it was indeed common for sentences to be pronounced publicly. Orvieto stands here perhaps as an extreme example of a more general finding that goes against expectations: other registers of sentences (Arnold Sentences, Gui Sentences, Caux Sentences) also outperform several registers of depositions.

The second highest-ranking text, Carcassonne GGG — the latest Languedocian register in the corpus — scores consistently high in multiple categories, most notably Shannon's entropy and hypergeometric lexical diversity, suggesting the relatively even growth of new lexical types with less repetition. This indicates a very unformulaic register, in which the lexical categories imposed by the inquisitors and/or notaries were less rigid than was generally the case in the period under study, leaving greater room for the natural variety of deponent diction.

Conversely, we observe two early (1240s) registers from Languedoc, Caux ms. 609 and Seila, ranking lowest. Both registers rank consistently low in all categories of metric. With more than 5,500 deponents, the largest number among all medieval inquisition registers (Rehr, 2019), the Caux ms. 609 collection is composed of typically very short and thematically limited depositions. Notably, this register has been central to arguments that the inquisition in Languedoc "constructed" heresy through rigid questioning and strict notarial editing (Pegg, 2001, pp. 45–51, 57–61) (Pegg, 2003).

Our findings on its limited lexical richness lend further, if indirect, support to this judgement. At the same time, however, the exceptional nature of these results suggests caution in extending the same degree of "construction" to other registers, whose language appears less consistently constrained.

Seila's register of penances exhibits both the lowest composite z-score and the largest negative residuals, ranking as lexically much poorer than predicted even given its early date (Figure 3). Moreover, all four of its category z-scores are lower than those of Caux ms. 609. These results align with other examinations of the register that have highlighted the highly formulaic language that runs throughout the register despite the presence of subtle irregularities at the narrative level (Taylor, 2013, pp. 333–335) (Biller, 2022, p. 364) (Shaw, Hampejs and Zbíral, 2026).

Fournier, the longest extant medieval inquisition register, is well-known for its high degree of detail and thematic richness (Duvernoy, 1965a, p. 8) (Le Roy Ladurie, 1975) (Paul, 1991, p. 41) (Arnold, 2001, pp. 21, 164–165) (Théry-Astruc, 2003, p. 478) (Kras, 2020, p. 303). In our analysis, it indeed ranks among the lexically richest registers, coming fifth in the composite score. Its high Shannon's entropy z-score (at third place) indicates a less predictable vocabulary compared to most other registers, with a less rigid textual style and/or larger and more varied blocks of testimony. However, it is still striking that Jacques Fournier's register does not surface as the richest one in either this composite score or indeed any of the four categories.

Settimo is known for the manipulatory nature of some of its documents, which ascribe orgiastic misbehaviour to heretics and posit a heretical anti-pope (Merlo, 1977, pp. 42–43, 71–72). While such testimonies must have been closely guided by the inquisitorial hand, the register nevertheless presents high lexical richness scores, especially in Shannon's entropy and inverted repetition, and ranks fourth in the composite z-score. This serves as a useful reminder that narrative richness, signalled here by the richness of vocabulary, is not straightforwardly related to the authenticity and reliability of deposition records. Stories confected under duress were also prone to florid narration.

Key findings

- All nine regression models corroborate that the lexical richness of inquisition registers from Italy and Languedoc increased between 1231 and 1400.
- Different metric types occasionally produce divergent results for the same registers, underlining the value of drawing on multiple metrics to capture different aspects of lexical richness.
- Contrary to expectations, sentence registers outperform several deposition registers in lexical richness metrics. The Orvieto book of sentences in fact achieves the highest average z-score across all metric categories.
- The register of Jacques Fournier confirms its reputation for richness even from the perspective of vocabulary, but it does not rank first in any of the lexical richness categories.

Limitations

The results of this study need to be read in the context of its limitations.

While considerable efforts were expended in correcting the optically recognized printed editions, some errors inevitably persist. Furthermore, some texts are not critical editions but working transcriptions (e.g., Galand 1, Arnold Sentences, Carcassonne DDD, Carcassonne GGG): some errors of transcription can be expected in them. Such OCR or transcription errors could affect our results, even if we expect that their prevalence is too low to change them.

The second limitation is our unit of analysis, which was, for practical reasons related to the stability of the metrics, a register rather than an individual document. A more granular, document-level analysis, perhaps by application of a different approach, such as functional diversity (Karsdorp, Manjavacas and Fonteyn, 2022), might show more nuanced patterns that our design could not capture.

The third limitation is interpretative. Our metrics studied the place of words and their distribution in the context of overall vocabulary. We thus do not directly measure the presence of repetitive multi-word constructions generally associated with formulaic language. While we expect that lexical richness will have a negative association with the use of fixed multi-word formulas and a positive association with thematic richness, we have not examined either of them.

Conclusion

To the best of our knowledge, our analysis represents the first quantitative examination of the lexical richness of a corpus of medieval heresy trial records. Applying a set of diverse metrics to a quite representative corpus of twenty-two registers (2.28 million tokens) from Languedoc and Italy, we employed regression modelling to measure the growth of lexical richness between the establishment of the papal inquisition in 1231 and the late fourteenth century. The results from all metrics confirm that the vocabulary from inquisition records did become more diverse and elaborate over time. As there is some evidence that lexical richness constitutes a reasonable proxy for thematic richness (Baayen, 2001) (Van Gijsel, Speelman and Geeraerts, 2005), we have also provided the first quantitative argument to further pursue the idea of gradual thematic expansion of inquisition records over time in the regions and period examined.

Our analysis also offers new perspectives on individual registers and their groups. Notably, registers of sentences are sometimes richer than registers of depositions, despite their narratives of guilt being typically more condensed by notaries and accompanied by boilerplate concerning sentencing itself. This suggests a more complicated relationship between the formula and richness of content than previously assumed: the textual interaction between them is far more complex than that implied by the metaphor of a struggle between inquisitorial questionnaires and notarial framing on the one hand and the deponent's voice occasionally breaking through on the other (Patschovsky, 1991, p. 266) (Arnold, 2001, pp. 114–115) (Arnold, 2019, p. 8) (Bruschi, 2009, pp. 14–15).

On a more general methodological level, we hope to have made a case for the use of the quantitative methods of corpus linguistics (Zheng, 2025a) and computational literary studies (Müngen, 2023) in the study of premodern trial sources. The perspective these fields offer is especially pertinent in the case of inquisition records, whose multivocal construction has been intensely debated, often on the basis of anecdotal evidence offered by exceptional testimonies. By gaining a more principled and comparative understanding of the different aspects of their language, we bring their overall character and composition into sharper definition.

Author roles

Gideon Kotzé: conceptualization, methodology, software, data curation, validation, formal analysis, investigation, writing – original draft, writing – review & editing, visualization, supervision. **David Zbíral:** conceptualization, investigation, resources, data curation, writing – original draft, writing – review & editing, supervision, project administration. **Robert L. J. Shaw:** conceptualization, investigation, data curation, writing – original draft, writing – review & editing, project administration. **Zoltán Brys:** methodology, software, validation, formal analysis, investigation, data curation, writing – original draft, writing – review & editing, visualization.

Ethics and consent

Ethical approval and consent were not required.

Declaration of AI use

During the preparation of this manuscript, large language models (OpenAI GPT, Anthropic Claude, and Google Gemini) were used in the development and optimization of the analytical code and to refine the clarity and flow of the prose. These tools were also utilized for brainstorming and exploratory discussions regarding research concepts. The final ideas, interpretations, and the specific expression of the writing remain the work of the authors, who have reviewed and edited all AI-assisted output, and who take full responsibility for the integrity and accuracy of this publication.

Data and software availability

Source data

The corpus texts used in this study derive from modern scholarly editions protected by copyright, which we are not licensed to redistribute. We obtained the editions under standard institutional library access and by purchase, with each cited in full in the reference list, which will enable the interested reader to assemble the corpus by the same means as the authors.

Underlying data

The datasets and software supporting the findings of this study have been deposited in Zenodo.

Data supplement for: Did the language of medieval inquisition records become richer over time? A diachronic lexical richness perspective on heresy trial proceedings from Languedoc and Italy, 1231–1400. https://doi.org/10.5281/zenodo.19593453 (Kotzé et al., 2026).

The project contains the following underlying data:

- **heresy_inquisition_registers_with_metadata.csv**: Table with all registers used in the study (Table 1), including their names, token counts, relevant metadata, and lexical richness scores.
- **OLS_output.txt**: Full regression analysis output.
- **OLS_sessioninfo.txt**: Data on imported libraries used in the regression analysis.
- **register_richness_ranking.csv**: Table with all registers after filtering, ranked by lexical richness composite z-score (Table 7).
- **lexical_richness_stability_results.csv**: Data created by bootstrap resampling.

Data is available under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) license.

Software availability

In the same repository as mentioned under "Data availability", we also include the following software scripts, which should enable the reader to reproduce the findings listed below:

- **register_level_regression_lexical_richness_time_token_count.R**: R script used to perform the regression analysis (Tables 4 and 5).
- **supplementary_code.py**: Python script used to cross-validate the custom lexical richness implementations (Gini coefficient and Zipf's α) against reference implementations in the **external/**directory. Note that the other metrics were implemented using third-party libraries available online: We therefore did not redistribute them. For most metrics, we used the implementation as available in the *lexicalrichness* package (Shen, 2022). For Shannon's entropy, we used the *sciPy* package (Virtanen et al., 2020).
- **external/**: A directory containing three third-party implementations of the aforementioned metrics, including their respective license files and the sources from where they were bundled:
  - **gini_oliviaguest.py**: Calculates the Gini coefficient (CC0 1.0 Universal)
  - **inequality_pysal.py**: Calculates the Gini coefficient (BSD 3-Clause)
  - **zipf_usptact.py**: Calculates the log-likelihood of observed frequencies under the Zipf distribution (Apache License 2.0).The code requires the download of supporting packages, but is otherwise self-contained.

Acknowledgments

We thank Stanislaw Banach, Soňa Hudíková, Katia Riccardo, Davor Salihović, and Katalin Suba for their contributions to curating and cleaning corpus data. We also thank Matthew Nicholls for linguistic advice.

References

Amati G, (ed.). **Processus contra Valdenses in Lombardia Superiori, anno 1387 I.** *Archivio storico italiano.* 1865a; **3**(1): 3–52.

Amati G, (ed.). **Processus contra Valdenses in Lombardia Superiori, anno 1387 II.** *Archivio storico italiano.* 1865b; **3**(2): 3–61.

Arnold JH: *Inquisition and power: Catharism and the confessing subject in medieval Languedoc.* Philadelphia: University of Pennsylvania Press (The Middle Ages); 2001.
Publisher Full Text

Arnold JH: **Voicing dissent: heresy trials in later medieval England.** *Past & Present.* 2019; **245**: 3–37.
Publisher Full Text

Asad T: **Medieval heresy: an anthropological view.** *Soc. Hist.* 1986; **11**(3): 345–362.
Publisher Full Text

Baayen RH: *Word frequency distributions.* Springer Science & Business Media; 2001.
Publisher Full Text

Baayen RH: **Corpus linguistics in morphology: morphological productivity.** *Corpus linguistics: An international handbook.* 2009; pp. 900–919.
Publisher Full Text

Baixeries J, Elvevåg B, Ferrer-i-Cancho R: **The Evolution of the Exponent of Zipf's Law in Language Ontogeny.** *PLoS ONE.* 2013; **8**(3): e53227.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Benedetti M, (ed.). *Milano 1300: I processi inquisitoriali contro le devote e i devoti di santa Guglielma.* Milano: Libri Scheiwiller (Milano medievale; 1999; 2).

Bentz C, Verkerk A, Kiela D, *et al.*: **Adaptive Communication: Languages with More Non-Native Speakers Tend to Have Fewer Word Forms.** *PLoS ONE.* 2015; **10**(6): e0128254.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Bentz C, Kiela D: **Zipf's law across languages of the world: Towards a quantitative measure of lexical diversity.** *Evolution Of Language, The Proceedings Of The 10th International Conference (Evolang X). Evolang X.* Vienna, Austria: World Scientific; 2014.

Bérubé N, Sainte-Marie M, Mongeon P, *et al.*: **Words by the tail: Assessing lexical diversity in scholarly titles using frequency-rank distribution tail fits.** *PLoS ONE.* 2018; **13**(7): e0197775.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Biget J-L: **I catari di fronte agli inquisitori in Languedoc 1230-1310.** Biget J-L, Maire Vigueur J-C, Paravicini Bagliani A, (eds.). *La parola all'accusato.* Palermo: Sellerio; 1991; pp. 235–251.

Biget J-L: **L'action inquisitoriale: Origines, formes, portée.** Biget J-L, (ed.). *Hérésie et inquisition dans le Midi de la France.* Paris: Picard; 2007; pp. 172–205.

Biller P: **Through a glass darkly: seeing medieval heresy.** Linehan P, Nelson JL, (eds.). *The medieval world.* London; New York: Routledge; 2001; pp. 308–326.
Publisher Full Text

Biller P: **Inquisitors' interrogations of Waldenses.** Benedetti M, Cameron E, (eds.). *A companion to the Waldenses in the Middle Ages.* Leiden: Brill; 2022; pp. 354–393.
Publisher Full Text

Biller P, Bruschi C, Sneddon S, (eds.). *Inquisitors and heretics in thirteenth-century Languedoc: edition and translation of Toulouse inquisition depositions, 1273–1282.* Leiden; Boston: Brill; 2011.
Publisher Full Text

Breusch TS, Pagan AR: **A simple test for heteroscedasticity and random coefficient variation.** *Econometrica.* 1979; **47**(5): 1287.
Publisher Full Text

Brezina V: *Statistics in corpus linguistics: a practical guide.* Cambridge; New York, NY: Cambridge University Press; 2018.
Publisher Full Text

Brezina V, Gablasova D: *LancsLex: Lancaster Vocab Analysis Tool.* Lancaster University; no date.
Reference Source

Brooks GP, Barcikowski RS: **The PEAR method for sample sizes in multiple linear regression.** *General Linear Model Journal.* 2012; **38**(2): 1–16.
Reference Source

Bruschi C: *The wandering heretics of Languedoc.* Cambridge; New York: Cambridge University Press (Cambridge studies in medieval life and thought); 2009.
Publisher Full Text

Burns PJ: **LatinCy: Synthetic Trained Pipelines for Latin NLP.** *arXiv.* 2023.
Publisher Full Text

Carroll JB: *Language and Thought.* Englewood Cliffs, NJ: Prentice-Hall; 1964.

Clauset A, Shalizi CR, Newman MEJ: **Power-Law Distributions in Empirical Data.** *SIAM Rev.* 2009; **51**: 661–703.
Publisher Full Text

Cook RD, Weisberg S: *Residuals and influence in regression.* Chapman and Hall; 1986.

Covington MA, McFall JD: **Cutting the Gordian knot: The Moving-Average Type–Token Ratio (MATTR).** *Journal of Quantitative Linguistics.* 2010; **17**(2): 94–100.
Publisher Full Text

D'Alatri M: *L'inquisizione francescana nell'Italia centrale del Duecento: con il testo del "Liber inquisitionis" di Orvieto trascritto da Egidio Bonanno.* Roma: Istituto Storico dei Cappuccini; 1996.

Davis GW, (ed.). *The inquisition at Albi, 1299–1300: Text of register and analysis.* New York: Columbia University Press; 1948.

Dawson N, *et al.*: **Features of lexical richness in children's books: Comparisons with child-directed speech.** *Language Development Research.* 2021; **1**(1).
Publisher Full Text

Dondaine A: **Le manuel de l'inquisiteur (1230-1330).** *Archivum Fratrum Praedicatorum.* 1947; **17**: 85–194.

Douais C: *Documents pour servir à l'histoire de l'Inquisition dans le Languedoc.* Paris: Renouard; 1900.

Dovgalecs V: **usptact/zipf-pymc3: Fitting Zipf's Law using PyMC3.** 2025. (Accessed: April 15, 2026).
Reference Source

Dugast D: **Sur quoi se fonde la notion d'étendue théorique du vocabulaire?.** *Le francais moderne.* 1978; **46**(1): 25–32.

Dugast D: *Vocabulaire et stylistique: Théâtre et dialogue.* Paris: Slatkine-Champion (Travaux de Linguistique Quantitative); 1979.

Durbin J, Watson GS: **Testing for serial correlation in least squares regression II.** *Biometrika.* 1951; **38**(1–2): 159–178.
Publisher Full Text

Duvernoy J: **Introduction.** Duvernoy J, (ed.). *Le registre d'inquisition de Jacques Fournier, évêque de Pamiers (1318–1325).* Toulouse: Privat; 1965a; pp. 7–33.

Duvernoy J, (ed.) *Le registre d'inquisition de Jacques Fournier, évêque de Pamiers (1318–1325).* Toulouse: Privat; 1965b.

Duvernoy J, (ed.). *Le registre d'inquisition de Jacques Fournier, évêque de Pamiers (1318–1325): Corrections.* Toulouse: Privat; 1972.

Duvernoy J, (ed.). **Registre de Bernard de Caux, Pamiers, 1246-1247.** *Foix: Société Ariégeoise des Sciences, Lettres et Arts (Bulletin de la Société Ariégeoise des Sciences, Lettres et Arts).* 1990.

Duvernoy J, (ed.). *L'inquisition en Quercy: Le registre des pénitences de Pierre Cellan, 1241–1242.* Castelnaud-la-Chapelle: L'Hydre; 2001.

Duvernoy J, (ed.). **Cahiers de Bernard de Caux: Ms Doat XXII B.n. Paris: Agen, Cahors, Toulouse, 1243–1247.** 2002. (Accessed: May 20, 2019).
Reference Source

Duvernoy J, (ed.). **Le registre DDD de l'inquisition de Carcassonne.** 2003a.
Reference Source

Duvernoy J, (ed.). **Registre GGG de l'inquisition de Carcassonne.** 2003b.
Reference Source

Engber CA: **The relationship of lexical proficiency to the quality of ESL compositions.** *J. Second. Lang. Writ.* 1995; **4**(2): 139–155.
Publisher Full Text

Fang Y, Wang F, Dai Z: **A Zipfian Approach to Exploring Lexical Complexity Changes in Mandarin Chinese: A Corpus-Based Analysis of Word N-Grams.** *Journal of Quantitative Linguistics.* 2025; **32**(3): 251–278.
Publisher Full Text

Fox J, Weisberg S: *An R Companion to Applied Regression.* Melbourne: Los Angeles London New Delhi Singapore Washington DC; 2019.

Gini C: *Variabilità e mutabilità.* Rome: Libreria Eredi Virgilio Veschi; 1912.

Grau Torras S: *L'herètica pravitat a la Corona d'Aragó: Documents sobre càtars, valdesos i altres heretges (1155–1324).* Barcelona: Fundació Noguera (Diplomataris, 70); 2015.

Grundmann H: **Ketzerverhöre des Spätmittelalters als quellenkritisches Problem.** *Deutsches Archiv für Erforschung des Mittelalters.* 1965; **21**: 519–575.
Publisher Full Text

Guest O: **oliviaguest/gini.** 2016. (Accessed: April 15, 2026).
Reference Source

Guiraud: *Problèmes et méthodes de la statistique linguistique.* Dordrecht: D. Reidel; 1960.

Hahs-Vaughn DL: *Applied multivariate statistical concepts.* New York: Routledge; 2nd ed 2024.
Publisher Full Text

Heaps HS: *Information Retrieval: Computational and Theoretical Aspects.* USA: Academic Press, Inc; 1978.

Heng R, Pu L, Liu X: **The effects of genre on the lexical richness of argumentative and expository writing by Chinese EFL learners.** *Front. Psychol.* 2023; **13**: 1082228.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Herdan G: **A new derivation and interpretation of Yule's 'Characteristic' K.** *Zeitschrift für angewandte Mathematik und Physik ZAMP.* 1955; **6**(4): 332–339.
Publisher Full Text

Herdan G: *Type-token mathematics: A textbook of mathematical linguistics.* The Hague: Mouton; 1960.

Hess CW, Sefton KM, Landry RG: **Sample Size and Type-Token Ratios for Oral Language of Preschool Children.** *J. Speech Lang. Hear. Res.* 1986; **29**(1): 129–134.
Publisher Full Text

Jarvis S: **Capturing the Diversity in Lexical Diversity.** *Lang. Learn.* 2013; **63**(s1): 87–106.
Publisher Full Text

Johnson W: **Studies in language behavior: A program of research.** *Psychol. Monogr.* 1944; **56**(2): 1–15.
Publisher Full Text

Karsdorp FB, Manjavacas E, Fonteyn L: **Introducing functional diversity: A novel approach to lexical diversity in (historical) corpora.** *CHR 2022: Computational Humanities Research Conference, Antwerp, Belgium: CEUR Workshop Proceedings (CEUR-WS.org.* 2022; pp. 114–126).
Reference Source

Kelly HA: **Inquisition and the prosecution of heresy: Misconceptions and abuses.** *Church History.* 1989; **58**(4): 439–451.
Publisher Full Text

Kleiber C, Kotz S: *Statistical Size Distributions in Economics and Actuarial Sciences.* John Wiley & Sons; 2003.
Publisher Full Text

Koplenig A: **Using the parameters of the Zipf–Mandelbrot law to measure diachronic lexical, syntactical and stylistic changes – a large-scale corpus analysis.** *Corpus Linguist. Linguist. Theory.* 2018; **14**(1): 1–34.
Publisher Full Text

Korkiakangas T: **A linguist's viewpoint: formulaic language as a challenge for historical linguistics.** *Formulaic Language in Historical Research and Data Extraction.* Amsterdam: REPUBLIC; 2024.
Publisher Full Text

Kotzé G, *et al.*: Data supplement for: Did the language of medieval inquisition records become richer over time? A diachronic lexical richness perspective on heresy trial proceedings from Languedoc and Italy, 1231–1400. [Dataset]. *Zenodo.* 2026.
Publisher Full Text

Kras P: *The system of the inquisition in medieval Europe.* Berlin: Peter Lang (Geschichte - Erinnerung - Politik; 2020; 36).
Publisher Full Text

Laufer B, Nation P: **Vocabulary Size and Use: Lexical Richness in L2 Written Production.** *Appl. Linguis.* 1995; **16**(3): 307–322.
Publisher Full Text

Le Roy Ladurie E: *Montaillou, village occitan de 1294 à 1324.* Paris: Gallimard; 1975.

Lerner RE: *The heresy of the Free spirit in the later Middle Ages.* Berkeley: University of California Press; 1972.
Publisher Full Text

Linnarud M: *Lexis in Composition. A Performance Analysis of Swedish Learners' Written English.* Malmö: Liber Förlag/CWK Gleerup (Lund Studies in English, 74); 1986.

Lissón P, Ballier N: **Investigating Lexical Progression through Lexical Diversity Metrics in a Corpus of French L3.** *Discours. Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics.* 2018; **23**. [Preprint].
Publisher Full Text

Long JS, Ervin LH: **Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model.** *Am. Stat.* 2000; **54**(3): 217–224.
Publisher Full Text

Maas H-D: **Über den Zusammenhang zwischen Wortschatzumfang und Länge eines Textes - ProQuest.** *Zeitschrift für Literaturwissenschaft und Linguistik.* 1972; **8**: 73–79.

MacKinnon JG, White H: **Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties.** *J. Econ.* 1985; **29**(3): 305–325.
Publisher Full Text

MacWhinney B: *The CHILDES project: Tools for analyzing talk I: Transcription format and programs.* Psychology Press; 2014.

McCarthy PM, Jarvis S: **MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment.** *Behav. Res. Methods.* 2010; **42**(2): 381–392.
PubMed Abstract | **Publisher Full Text**

McKee G, Malvern D, Richards B: **Measuring vocabulary diversity using dedicated software.** *Literary and Linguistic Computing.* 2000; **15**(3): 323–338.
Publisher Full Text

Merlo GG: *Eretici e inquisitori nella società piemontese del Trecento.* Torino: Claudiana; 1977.

Montgomery DC, Peck EA, Vining GG: *Introduction to linear regression analysis.* Hoboken, New Jersey: Wiley; 2021.

Müngen AA: **Exploring lexical richness in English-language theses across disciplines: A comparative analysis.** *J. Sci. Res.* 2023; **12**(2): 459–468.
Publisher Full Text

Palès-Gobilliard A, (ed.). *L'inquisiteur Geoffroy d'Ablis et les cathares du comté de Foix (1308–1309).* Paris: Centre National de la Recherche Scientifique; 1984.

Palès-Gobilliard A, (ed.). *Le livre des sentences de l'inquisiteur Bernard Gui, 1308–1323.* Paris: Centre National de la Recherche Scientifique (Sources d'histoire médiévale, 30); 2002.

Paolini L, Orioli R, (eds.). *Acta S. Officii Bononie ab anno 1291 usque ad annum 1310.* Roma: Istituto storico italiano per il Medio Evo (Fonti per la storia d'Italia, 106); 1982.

Parmeggiani R: **Un secolo di manualistica inquisitoriale (1230-1330): intertestualità e circolazione del diritto.** *Rivista internazionale di diritto comune.* 2002; **13**: 229–270.

Patschovsky A: **Gli eretici davanti al tribunale: A proposito dei processi-verbali inquisitoriali in Germania e in Boemia nel XIV secolo.** *La parola all'accusato.* 1991; 242–267.

Paul J: **Jacques Fournier inquisiteur.** Guillemain B, (ed.). *La papauté d'Avignon et le Languedoc (1316–1342).* Toulouse: Privat (Cahiers de Fanjeaux, 26); 1991; pp. 39–67.

Pazé P: **Gli interrogatori di Antonio Galosna.** Pazé P, Patria L, (eds.). *Valdo e Francesco: inizi e sviluppi di due movimenti religiosi: dai conflitti alla convivenza.* Perosa Argentina: LAReditore (Convegni del Laux, 10); 2016a; pp. 131–251.

Pazé P: **Gli interrogatori di Giacomo Bech.** Pazé P, Patria L, (eds.). *Valdo e Francesco: inizi e sviluppi di due movimenti religiosi: dai conflitti alla convivenza.* Perosa Argentina: LAReditore (Convegni del Laux, 10); 2016b; pp. 252–286.

Pegg MG: *The corruption of angels: the great inquisition of 1245–1246.* Princeton; Oxford: Princeton University Press; 2001.
Publisher Full Text

Pegg MG: **Questions about questions: Toulouse 609 and the great inquisition of 1245-6.** Bruschi C, Biller P, (eds.). *Texts and the repression of medieval heresy.* Woodbridge; Rochester: York Medieval Press (York studies in medieval theology, 4); 2003; pp. 111–125.

Pegg MG: **Historians and inquisitors: testimonies from the early inquisitions into heretical depravity.** Rosenthal JT, (ed.). *Understanding medieval primary sources: using historical sources to discover medieval Europe.* London; New York: Routledge (Routledge guides to using historical sources); 2012; pp. 98–113.

Pihko S: **The construction of information in medieval inquisition records: a methodological reconsideration.** *I quaderni del m.æ.s. /Journal of Mediæ Ætatis Sodalicium.* 2024; **22**(1): 165–189.
Publisher Full Text

Popescu II, Altmann G: **Some aspects of word frequencies.** *Glottometrics.* 2006; **13**: 23–46.

Read J: *Assessing Vocabulary.* Cambridge: Cambridge University Press (Cambridge Language Assessment); 2000.
Publisher Full Text

Rehr J-P, (ed.). *The registry of the great inquisition at Toulouse, 1245–46: edition and translation of Bibliothèque municipale de Toulouse MS 609.* 2019. (Accessed: May 20, 2019).
Reference Source

Rey SJ: **pysal/inequality: Measures of spatial (and non-spatial) inequality.** 2018. (Accessed: April 15, 2026).
Reference Source

Sackville LJ: **The Ordo Processus Narbonensis: The earliest inquisitor's handbook, lost and refound.** *Aevum.* 2019; **93**(2): 363–396.
Publisher Full Text

Shannon CE: **A mathematical theory of communication.** *Bell Syst. Tech. J.* 1948; **27**(3): 379–423.
Publisher Full Text

Shaw RLJ, Hampejs T, Zbíral D: **Narratives of religious dissidence in medieval inquisition records: computing the representation and sequencing of crimes in Peter Seila's register of sentences (1241–2).** *Hist. Res.* 2026; **99**(284): 235–258.
Publisher Full Text

Shen L: **LexicalRichness: A small module to compute textual lexical richness.** 2022.
Reference Source

Shi Y, Lei L: **Lexical Richness and Text Length: An Entropy-based Perspective.** *Journal of Quantitative Linguistics.* 2022; **29**(1): 62–79.
Publisher Full Text

Simpson EH: **Measurement of diversity.** *Nature.* 1949; **163**(4148): 688–688.
Publisher Full Text

Somers HH: **Statistical methods in literary analysis.** Leed J, (ed.). *The computer and literary style.* Kent: Kent State University Press; 1966; pp. 128–140.

Štajner S, Mitkov R: **Using comparable corpora to track diachronic and synchronic changes in lexical density and lexical richness.** *The 5th workshop on building and using comparable corpora.* 2012; pp. 88–97.

Stevens JP: *Applied multivariate statistics for the social sciences.* Routledge; 2012.
Publisher Full Text

Taylor C: **Evidence for dualism in inquisitorial registers of the 1240s: a contribution to a debate.** *History: The Journal of the Historical Association.* 2013; **98**(3): 319–345.
Publisher Full Text

Théry-Astruc J: **Cléricalisme et hérésie des bons hommes: l'exemple d'Albi et de l'Albigeois (1276–1329).** *L'anticléricalisme en France méridionale.* Toulouse: Privat (Cahiers de Fanjeaux; 2003; **38**): pp. 471–508.

Toti D: **François Graverol e un manoscritto perduto dell'inquisizione di Tolosa.** *Riforma e movimenti religiosi.* 2020; **8**: 216–289.

Tweedie FJ, Baayen RH: **How variable may a constant be? Measures of lexical richness in perspective.** *Comput. Hum.* 1998; **32**(5): 323–352.
Publisher Full Text

Van Gijsel S, Speelman D, Geeraerts D: **A variationist, corpus linguistic analysis of lexical richness.** *Proceedings of Corpus Linguistics.* 2005.

Vilandrau C: **Inquisition et 'sociabilité cathare' d'après le registre de l'inquisiteur Geoffroy d'Ablis (1308-1309).** *Heresis.* 2001; **34**: 35–66.
Publisher Full Text

Virtanen P, Gommers R, Oliphant TE, *et al.*: **SciPy 1.0: fundamental algorithms for scientific computing in Python.** *Nat. Methods.* 2020; **17**(3): 261–272.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Wakefield WL: **Some unorthodox popular ideas of the thirteenth century.** *Medievalia et humanistica: Studies in medieval and Renaissance culture.* 1973; **4**: 25–35.

Wallnerová D, Zbíral D, (eds.). *A transcription of four inquisition depositions (1301-1305) on beliefs and the way of life of the Autier group (fonds Doat, vol. 34, fol. 93v-103r).* Zenodo [Preprint]; 2025a.
Publisher Full Text

Wallnerová D, Zbíral D, (eds.). *A transcription of sentences of Ferrer and William Raymond (fonds Doat, vol. 21, fol. 313r-323v).* Zenodo [Preprint]; 2025b.
Publisher Full Text

Wallnerová D, Zbíral D, (eds.). *A transcription of sentences of William Arnold and Stephen of Saint-Thibéry (fonds Doat, vol. 21, fol. 143v-185r).* Zenodo [Preprint]; 2025c.
Publisher Full Text

Wallnerová D, Zbíral D, (eds.). *A transcription of the inquisition register of John Galand and William of Saint-Seine, 1283-1302 (fonds Doat, vol. 26, fol. 79r-316v).* Zenodo [Preprint]; 2025d.
Publisher Full Text

Yule GU: *The statistical study of literary vocabulary.* Cambridge: Cambridge University Press; 1944.

Zanella G, (ed.). **Appendice I.** *Itinerari ereticali: Patari e catari tra Rimini e Verona.* Roma: Istituto storico italiano per il Medio Evo; 1986; pp. 47–102.

Zanella G: **Malessere ereticale in Valle Padana (1260–1308).** Zanella G, (ed.) *Hereticalia: Temi e discussioni.* Spoleto: Centro italiano di studi sull'Alto Medioevo (Collectanea, 7); 1995; pp. 15–66.

Zbíral D, Shaw RLJ: **Hearing voices: reapproaching medieval inquisition records.** *Religions.* 2022; **13**(12): 1175.
Publisher Full Text

Zheng W: **Lexical richness viewed through lexical diversity, density, and sophistication.** *Digital Scholarship in the Humanities.* 2025a; **40**(2): 692–708.
Publisher Full Text

Zheng W: **Text length requirements for stable and genre-sensitive lexical diversity measurement.** *Cogent Arts & Humanities.* 2025b; **12**(1): 2584418.
Publisher Full Text

Zipf GK: *The psycho-biology of language.* Boston: Houghton Mifflin Co; 1935.