Parataxis and hypotaxis in historical corpora

George Walkden
University of Konstanz

December 29, 2025

Parataxis and hypotaxis in historical corpora

This is a pre-copyedited, author-produced postprint version of an article published in *Language* (*Research Reports*) following peer review. The version of record is available online at `https://doi.org/10.1353/lan.2025.a978279`; if you cite this paper, please use the pagination in that version.

ABSTRACT

A recurring narrative in studies of syntactic change is that hypotaxis gains ground at the expense of parataxis. This report shows that this claim – if construed as a quantitative claim about the frequency of different types of clause-combining – finds little support in parsed diachronic corpora of seven languages. Genre appears to be a factor that substantially influences the proportion of hypotaxis found in a text in consistent ways, but time does not.[*]

Keywords: hypotaxis, parataxis, parsed corpora, syntactic change

[*]Versions of this paper were presented at CamCoS 7 (Cambridge, 2018), DiGS 20 (York, 2018), ConSOLE XXVII (HU Berlin, 2019), and in talks at the universities of Aarhus (2018), Tübingen (2018), Oslo (2019), Konstanz (2019), FU Berlin (2019), Stony Brook (2021), and Zurich (2021). I am grateful to audiences there, and in particular to Dora Alexopoulou, Theresa Biberauer, Bettina Braun, Anne Breitbarth, Massimiliano Canzi, Gerhard Jäger, Henri Kauhanen, Frans Plank and Richard Zimmermann, as well as to the anonymous reviewers for *Language*, to associate editor Graeme Trousdale, and to editor John Beavers, for their comments and suggestions. The usual disclaimers apply.

1. INTRODUCTION. The view that parataxis precedes hypotaxis has a long pedigree. As early as 1883, Gildersleeve felt justified in writing that '*Nihil est in hypotaxi quod non prius fuerit in parataxi* [there is no hypotaxis that was not formerly parataxis] is the motto of recent investigators' (1883: 419). Yet the notions of parataxis and hypotaxis involved often remain vague and disputed (Harris & Campbell 1995: 283), as do the details of the idea that there is a directional progression in language history from the former to the latter.

In this report I first make the case that there are multiple different versions of the parataxis-precedes-hypotaxis claim, and that these need to be kept distinct. I then revisit one version of the claim in detail: the idea that, quantitatively, we see an increase in hypotactically embedded clauses – as a proportion of all clauses – over time in the history of languages. Advances in the development of parsed diachronic corpora over the past quarter of a century have made it possible to conduct a broad-based quantitative evaluation of this claim. Using diachronic corpora of Chinese, English, French, Icelandic, Irish, Low German and Portuguese, I show that the claim receives little support from the histories of these languages. The evidence I present does not bear on other versions of the claim.

The structure of this report is as follows. Section 2 discusses the definition of the terms parataxis and hypotaxis and theories about the diachronic relations between them. In section 3 the data sources are presented, along with the methods used to retrieve quantitative information about parataxis and hypotaxis. Section 4 presents and discusses the findings, and section 5 concludes.

2. UNDERSTANDINGS OF PARATAXIS AND HYPOTAXIS.

2.1. DEFINING PARATAXIS AND HYPOTAXIS. Hypotaxis is defined in the OED (3rd edn, s.v. hypotaxis) as a synonym for subordination, while parataxis is defined as 'The placing of propositions or clauses one after another, without indicating by connecting words the relation (of coordination or subordination) between them' (OED, 3rd edn, s.v. parataxis). These definitions will not do for our purposes, as they allow for hypotaxis to be parataxis (in the case of subordination without a 'connecting word'), and the two terms are usually construed in opposition to one another – even if linguists rarely attempt to define them explicitly. Here I define them (in 1) in a way that relies on a notion of constituency and on the idea that some constituents are clausal.

(1) **Hypotaxis**
Two clausal constituents α and β stand in a relation of hypotaxis iff either α dominates β or β dominates α.

By 'clausal constituent' or, more simply, 'clause' is meant a single phrase-structural node that dominates everything else within the clause; this would be, for instance, S in traditional phrase

structure grammar, or CP in modern generative syntax. Whenever a clausal node dominates another clausal node, the two stand in a relation of hypotaxis. This includes the relation between verbal complement clauses and the clause in which the embedding verb is immediately contained, for instance, but also the relation between e.g. adverbial clauses and the clause they modify, and the relation between relative clauses and the clause containing the nominal they modify (if there is one), up to an arbitrary level of embedding. In clausal adjunction, too, the two clauses stand in a relation of hypotaxis by this definition.

Parataxis is then defined as in 2.

(2) **Parataxis**
Two clausal constituents α and β stand in a relation of parataxis iff they are linearly adjacent and do not stand in a relation of hypotaxis.

This negative definition reflects the fact that parataxis is typically taken to encompass not just non-subordinative relations (e.g. coordination) but also instances in which there is no syntactically definable relation between adjacent clauses at all, e.g. fully independent clauses (of the type *I came, I saw, I conquered*).

The definitions in 1 and 2 will not capture everyone's intuitions about what these terms mean. Hopper & Traugott (2003: 177–178), for instance, define hypotaxis as an intermediate (–embedded, +dependent) stage on a cline between parataxis (–embedded, –dependent) and subordination (+embedded, +dependent). Thus, for them, subordinate clauses are actually not instances of hypotaxis. On the other hand, Donoghue & Mitchell (1992), in critiquing the general confusion in use of the two terms, argue that parataxis and hypotaxis are not grammatical terms, but rather belong to the domain of literary stylistics: for them, 'although subordination is essential for hypotaxis, the two terms are not synonymous' (1992: 169). The advantage of the definitions in 1 and 2 is that they are relatively straightforward to operationalize in corpus research (see section 3).

Further issues with these definitions concern the status of non-finite clauses and of coordination. Non-finite clauses, insofar as they are clausal constituents, will typically be instances of hypotaxis (or, more precisely, a non-finite clause will stand in a relation of hypotaxis with a finite clause that dominates it) – at least in languages like English in which they rarely stand alone. In the literature this does not go without saying: Harris & Campbell (1995: 283) state that 'the term hypotaxis seems usually to include only finite dependent clauses, not reduced (non-finite) clauses'; they themselves include both types, but distinguish between them, and this is what I also do in this report.

As for coordination, under the definition in 2, coordinated clauses and independent clauses both instantiate parataxis. As with non-finite clauses, to get a clearer picture of the empirical

facts it will be useful to make a distinction here: this time between overtly coordinated clauses (‘syndetic parataxis’; Donoghue & Mitchell 1992) and all other clauses not introduced by an overt coordinator (‘asyndetic parataxis’).

2.2. WHAT DOES IT MEAN FOR PARATAXIS TO PRECEDE HYPOTAXIS?. It is not only the definition of hypotaxis and parataxis that has given rise to multiple interpretations: the diachronic understandings of parataxis preceding hypotaxis have also been varied. At least four interpretations have emerged in the literature.

One of these is with respect to evolutionary time. Progovac (2014: ch. 4) argues that, in the biological evolution of the language faculty, a paratactic operation Conjoin, which does not create hierarchical structure or headedness, precedes the hierarchy-creating operation Merge (see also Karlsson 2009: 195). Since my concern in this report is with the historical record, which begins very many millennia after the emergence of cognitively modern humans, I will set aside this biological interpretation of parataxis-to-hypotaxis.[1] As Newmeyer (2002: 369, fn. 13) puts it, with reference to all attested languages, ‘It is important to stress that nobody – at least one would hope nobody – has claimed that there exists a language for which subordination is literally impossible ... All languages seem to allow the possibility.’

The other three understandings of parataxis preceding hypotaxis both relate more directly to the historical record. One is that hypotactic structures originate in paratactic structures, e.g. via reanalysis. The classic example adduced to support this claim is the view that the English complementizer *that* (and its Germanic cognates) comes into existence through direct reanalysis of two independent clauses, one with a cataphoric demonstrative pronoun: *I saw that. He is content.* becomes *I saw that he is content*. This scenario is proposed in Paul 1920: §211 and survives essentially without modification into Hopper & Traugott 2003: 192. In recent years, however, Axel-Tober (2012, 2017) has shown that this diachronic scenario is extremely unlikely to be correct: instead, complement clauses developed from relative clauses (see also Axel 2009, Weiß 2020 and Walkden 2024 for discussion). Harris & Campbell (1995: ch. 10) discuss this understanding of parataxis preceding hypotaxis – which they call the Parataxis Hypothesis – in great cross-linguistic detail, coming to a sceptical conclusion. This understanding is also not the subject of the present report, and is set aside here.

The third understanding of parataxis preceding hypotaxis is the idea that the maximal possible level of hypotaxis increases over time. This understanding has been particularly widespread in the literature on the history of German since the work of Wladımir Admoni (e.g. 1980; see Ágel 2000: 1879–1888 for discussion and further references). Admoni (1980) shows that, in the data he investigates, clauses with the most levels of embedding are typically found circa 1700, in chancery prose texts. Given that there is, in principle, no upper limit on the

number of levels of grammatical embedding at any stage of the history of German (or any other language), and given the highly genre-specific distribution of this phenomenon, it seems very unlikely that we are dealing with a true change in the distribution of mental grammars among the population of German-speakers of the time; a genre-specific development in stylistic norms is a more likely candidate for explanans, and indeed Admoni frames his findings as pertaining to the *Literatursprache* 'literary language' of the period. Of note is that this understanding of parataxis preceding hypotaxis does not involve arguing for a monotonic increase in levels of embedding, but rather for a peak in the Early Modern period. Relatedly, Karlsson (2009: 202) claims that 'It is a well-known fact that ... German and English were syntactically most complex in the 17th century and Swedish in the 19th century'; by syntactic complexity Karlsson is referring here to clauses with high levels of embedding. This third understanding will also be put aside for the rest of this report.

The fourth understanding of parataxis preceding hypotaxis, and the one that this report focuses on, is that the proportion of hypotaxis (as opposed to parataxis) increases across the historical record. This too is found in many sources, including at least the following from the last forty years: 'as societies 'modernize', the use of subordination becomes more frequent' (Newmeyer 2002: 369, fn. 13); 'the proportion of hypotaxis versus parataxis tends to increase in the course of time' (Jucker 1991: 203); 'The earliest written texts in a language are usually highly paratactic ... while later texts typically show more use of subordination. The historical increase in the frequency of subordination is gradual' (Dąbrowska 2015: 230). For individual languages this is often assumed without argument: Bloem et al. (2015: 24), for instance, state that 'subordinate clauses have become more prevalent' in the history of Dutch, and build an agent-based simulation that incorporates this assumption.

These are quantitative claims, and it should be straightforward in the corpus era to provide empirical evidence that bears on them. Yet none of the references cited in the previous paragraph as regards this fourth understanding actually does so. The aim of this short report is to remedy this situation.

Before proceeding, a contextual note is in order. Historically the idea that parataxis precedes hypotaxis has sometimes gone hand in hand with ideas about 'primitive' culture and cognition that are no longer accepted today, and with good reason. Small (1924: 125), for instance, writes 'It may be laid down as a general principle that in the progress of language parataxis precedes hypotaxis. The former is associated with the uncultivated mind; the latter, with the cultivated mind of civilized peoples.' Similarly, Andrew (1940: 87) describes early Old English as characterized by 'simply a lack of grammatical subordination such as we find in the language of children and some primitive people'. The fact that ideas of parataxis preceding hypotaxis have historically been associated with racist and other chauvinist positions does not, of course, entail

that specific claims about parataxis preceding hypotaxis are false. But it is important to bear in mind that claims (in science as elsewhere) may persist because of ideology rather than merit.

3. DATA AND METHODS.

3.1. PARSED DIACHRONIC CORPORA. This study investigates the claim that the proportion of hypotaxis increases at the expense of parataxis in the historical record in seven languages: Chinese, English, French, Icelandic, Irish, Low German and Portuguese. The choice of these languages was determined by the fact that for each of them there exists a constituency-parsed corpus of prose texts spanning at least four hundred years and following the general annotation principles of the Penn Parsed Corpora of Historical English, enabling comparable searches to be carried out. Details of the corpora used are provided in Table 1.

< INSERT TABLE 1 ABOUT HERE >

Of the seven languages for which the above criteria hold, all but Chinese belong to the Indo-European family; three are Germanic and two are Romance. These corpora also vary in size and in temporal resolution. It would therefore be a mistake to generalize the results of this report to the history of the world's languages in general; in particular, more parsed corpora of historically attested non-Indo-European languages are needed.

3.2. IDENTIFYING PARATAXIS AND HYPOTAXIS. In the annotation scheme of the Penn Parsed Corpora of Historical English and related corpora, clause-sized categories are consistently annotated as IP, with an extended label depending on what type of clause they are. Unembedded clauses are IP-MAT ('matrix'), embedded clauses are IP-SUB ('subordinate'), and non-finite clauses of all types are IP-INF ('infinitive'); these annotations are manually checked, and represent clausal constituency in the judgment of the annotator. Using these phrase-level labels in combination with part-of-speech tags, it is possible to define and identify four types of clause that are relevant to the research question:

1. Conjoined unembedded clauses: instances of IP-MAT where the first element is a coordinating conjunction[2]
2. Independent unembedded clauses: all other instances of IP-MAT, plus direct questions and exclamatives[3]
3. Finite embedded clauses: all instances of IP-SUB except direct questions and exclamatives[4]
4. Non-finite embedded clauses: all instances of IP-INF[5]

As a reviewer observes, there are many edge cases for a classification like this one, including instances in which the correct theoretical treatment is unclear. For instance, if the theory of coordination in Johannessen (1998) is on the right track, then all coordination arguably reduces to hypotaxis by the definition adopted here. I will instead assume that conjoined clauses should be kept separate, since this is what the literature on parataxis and hypotaxis diachronically tends to assume. Similarly, there is no consensus on whether speech reports are hypotactic or paratactic – either universally, or in individual languages (see e.g. Sauerland et al. 2020 on reported speech in Teiwa). Since the scope of a study like this one, across seven languages, does not allow for making detailed analytic decisions about every construction in every language, I am forced to rely on the annotators' decisions on whether a clause or clause-type should be annotated as IP-MAT or IP-SUB. This is problematic insofar as annotation decisions are not theoretical claims, and different classifications could lead to different conclusions; a reader who wishes to take issue with the classification as operationalized here is welcome to look at the supplementary materials and reclassify where they deem it appropriate.

We define the HYPOTAXIS LEVEL of a text as simply the proportion of all clauses that are embedded according to those definitions. The hypotaxis level is calculated by counting all instances of 3. and 4. and dividing the sum by the total number of clauses, i.e.

$$\frac{embedded}{embedded + unembedded} \tag{3}$$

Thus, in a text in which two of every five clauses are embedded, the hypotaxis level is 0.4, or 40%.

The clause types in 1.-4. were extracted using search and coding queries in CorpusSearch 2 (Randall et al. 2005–2018), a program designed for searching Penn-style parsed corpora; see the supplementary materials at `https://github.com/gwalkden/hypotaxis` for the technical details.

Each text was also annotated for genre and date. Both were based on information in the corpus manual wherever possible. In the case of texts that are only dated approximately, e.g. to the range 1150–1250 or to the middle of the fourteenth century, the midpoint of the date range was used, e.g. 1200 and 1350 respectively for the examples previously given.

4. RESULTS. The full datasets along with quantitative results are provided in the supplementary materials. Section 4.1 visualizes the findings, and section 4.2 presents inferential statistics.

4.1. DESCRIPTIVE FINDINGS. The scatterplots in this section present hypotaxis level on the y axis and time on the x axis. Points are individual texts in the corpus, scaled by the total number of clauses in the text; text genre is represented by the colour of the point. A LOESS local

regression line is also provided, with 95% confidence intervals in grey, as implemented by the GEOM_SMOOTH function in the package GGPLOT2 (Wickham 2016).

For each language, a stacked area chart is also provided, breaking embedded and unembedded clauses down into the four categories conjoined, independent, finite embedded and non-finite embedded. These charts bin texts into time periods; bin size was determined manually. R code to reproduce all graphs from the datasets is available in the supplementary materials.

CHINESE. < INSERT FIGURE 1 ABOUT HERE >

The ChiPaHC (Li 2017) is small, but covers a very long time span. Figure 1 shows that there has been no discernible change in the distribution of clause types over the history of the language, at least as far as this dataset is concerned. There is also no clear effect of genre, though the sparsity of texts makes it difficult to know for sure.

Note that there is a debate as to whether finiteness is a relevant grammatical category for Chinese (see e.g. Hu et al. 2001) In this report, following Li 2017, I assume that it is; see also Sybesma 2017, Grano 2017, and (with respect to historical stages of Chinese) Li 2018 for evidence and arguments supporting this conclusion.

ENGLISH. < INSERT FIGURE 2 ABOUT HERE >

For English there is substantially more data. However, it is difficult to identify an overall tendency in Figure 2. This is in part due to the corpus composition: the PPCEME and PPCMBE2 contain substantially more, and more diverse, texts than the earlier periods. Glancing at the smooth line in Figure 2a, it appears that there is a peak in hypotaxis between 1500 and 1650. But the genre composition of the corpora appears to be a factor with a clearer and more consistent role to play than date. The texts which show the highest levels of hypotaxis – above 80% – are statutes and (to a lesser extent) letters, two text types which are barely represented before 1500. Conversely, other text types, such as biblical texts, drama, and diaries, generally display low levels of hypotaxis regardless of period. The only temporal effect of which one can be reasonably confident is the decline in hypotaxis level after 1600, and even this is mitigated by the fact that legal texts are not well represented in the corpus post-1750.

Overall there is nothing here to support the idea that hypotaxis becomes more frequent over time. Instead we see intricate patterns of genre effects that should be subject to further research in future (see also Walkden 2024 for more detailed discussion of the English data). On the other hand, the peak between 1500 and 1650, if it is not an artefact of corpus composition, is roughly compatible with Karlsson's (2009: 202) claim that English was most hypotactic in

the seventeenth century (though the sixteenth century has just as much claim to this title). In Figure 2b we also see a gentle rise in the proportion of non-finite clauses, especially between the late Old and Early Modern English periods, which is consistent with Los's 2005 argument that *to*-infinitives in the history of English take over some of the functional load of finite *that*-clauses over time.

FRENCH. < INSERT FIGURE 3 ABOUT HERE >

In French, too, we see no evidence for a consistent diachronic trend. The apparent dip between the earliest texts and 1200 is attributable to the fact that this corpus – alone among the Penn-style parsed historical corpora used in this study – contains several verse texts, and that these texts are clustered in this period. Due to the relatively strict metrical requirements of Old French verse, it cannot be ruled out that authors' syntactic planning reflected the needs of the metre rather than autochthonous natural syntax: see e.g. Rainsford & Scrivner 2014 on syntax and metre in treebanks of historical French. On the other side of the coin, legal texts show consistently high levels of hypotaxis, as they do in English: from the Strasbourg Oaths of the ninth century onwards, legal texts and letters show hypotaxis levels of circa 75%, to which no other text comes close. The apparent dip in Figure 3 is entirely driven by these differences, since the earliest text – the Strasbourg Oaths – is a very hypotactic legal text, but the other pre-1100 texts are almost exclusively very paratactic verse texts; this part of the smooth line should thus not be viewed as reflecting a population-level change. After 1200, when the corpus coverage is prose-based and more balanced, no change is visible at all.

ICELANDIC. < INSERT FIGURE 4 ABOUT HERE >

What we see in the history of Icelandic resembles what we see in the history of French: apparent fluctuations, but no overall trend. An early dip around 1300 can be attributed to a cluster of saga texts, which typically show a low level of hypotaxis; this has long been observed to be a stylistic feature of saga syntax (cf. e.g. the quantitative study in Bouman 1958). Saga texts from between 1600 and 1900 are still relatively paratactic, but constitute a lower proportion of all texts during this period. Overall, the hypotaxis level of Icelandic texts hovers around 50% for the entire duration of the period covered by the corpus, over nine hundred years. Figure 4b shows a gentle increase in the relative frequency of non-finite clauses, just as for English.

IRISH. < INSERT FIGURE 5 ABOUT HERE >

Like the Chinese corpus, the Irish corpus is small. Little can be concluded from Figure 5, but there is no evidence for a change in the relative frequencies of hypotaxis and parataxis, or of clause types in general.

LOW GERMAN. < INSERT FIGURE 6 ABOUT HERE >

In the context of the results for the previous five languages investigated, the results for Low German should not come as a surprise: we see fluctuations, but no consistent diachronic trend. The CHLG (Booth et al. 2020) comprises a small number of fairly extensive texts. Charters and other legal texts have a comparatively high level of hypotaxis, in line with what we have seen for the other languages, and these are all there is in the corpus for the period before 1400. The apparent rise in hypotaxis post-1550 is driven by the fact that a single, small text is all we have for this period. As for Icelandic, the hypotaxis level of Low German hovers around 50% for the entire period covered by the corpus, with apparent deviations driven by genre and sampling effects.

PORTUGUESE. < INSERT FIGURE 7 ABOUT HERE >

The final language investigated, Portuguese, shows the results that are perhaps the most difficult to interpret. The four WOChWEL texts, pre-1400, show hypotaxis levels of just over 50%. The Tycho Brahe texts, on the other hand, start circa 1500 with a high hypotaxis level of around 80%, then show a steady decrease over the period covered by the corpus. Recall that an increase is what is predicted by the quantitative parataxis-to-hypotaxis hypothesis. This decrease is not obviously explicable in terms of genre, or indeed in terms of anything else. Since non-finite clauses increase gently in relative frequency over the history of Portuguese, the decrease mainly involves finite subordinate clauses becoming much less frequent and independent clauses becoming more frequent (Figure 7b).

4.2. MODELS. If the quantitative parataxis-to-hypotaxis claim is on the right track, we should expect to see a positive linear effect of time emerging as significant. This is not to say that a simple linear model will provide the best fit for the data – merely that such a linear effect ought to be detectable. Indeed, eyeballing the graphs in section 4.1 suggests that the effects of time might well be decidedly nonlinear. Thus, in 4.2 I fit a linear model to the data, and in 4.2 I explore possible nonlinear effects using generalized additive mixed models (GAMMs).

LINEAR MODELS. The package LME4 (Bates et al. 2015) was used to fit a mixed-effects regression model with date as a fixed effect and genre as a random intercept. The choice to treat

genre as a random effect was motivated by the facts that a) this study does not formulate explicit hypotheses about the putative effect of genre and (more importantly) b) the range of genres included for a given language does not exhaust the variability found in the population (Johnson 2009: 365), i.e. these are not exhaustive lists of possible text types but rather what happens to be found in the corpus.

Table 2 shows a summary of the effect sizes for date, as well as *p*-values for those effects calculated using Satterthwaite's degrees of freedom method as provided in the package LMERTEST (Kuznetsova et al. 2017), and R^2 values calculated using the Nakagawa & Schielzeth 2013 and Nakagawa et al. 2017 method as implemented in the package MUMIN (Bartoń 2023). Full model outputs can be found in the supplementary materials. The effect of date is plotted for each language in Figure 8.[6]

< INSERT TABLE 2 ABOUT HERE >

< INSERT FIGURE 8 ABOUT HERE >

The *p*-values only indicate significance for the effect of date in the English and French datasets, with a relatively small effect size. The R^2 values are more interesting. R^2 can be interpreted as a measure of the proportion of variance explained by the model, ranging from 0 to 1 (see Nakagawa & Schielzeth 2013); marginal R^2 is the proportion of variance explained when only fixed effects are taken into account, and conditional R^2 is the proportion of variance explained when random effects are also considered. For our purposes, then, marginal R^2 is the proportion of variance explained by date as a predictor, while conditional R^2 is the proportion of variance explained by both date and genre taken together. What is striking is that the marginal R^2 values are extremely low across the board, ranging from nearly 10% for Irish to less than 0.001% for Icelandic. These values suggest that date is a very poor predictor of hypotaxis level. Genre, on the other hand, has the potential to explain much more, though this is dependent on the language. For Chinese, the conditional R^2 is no better than the marginal R^2, suggesting that genre explains nothing in this dataset. At the other end of the spectrum, in French only 7% of the data can be explained by date, but over 80% of the data can be explained by date and genre together. Finally, consider English, which has by far the largest number of data points in the corpus. Here, although the effect of date is significant according to the *p*-value, it turns out that it explains less than 1% of the variance in the data. A model incorporating genre, on the other hand, can explain 45% of the variance.

Another way of making the same point is to compare the linear models represented in Figure 8 with models that are identical in all respects except that the fixed effect of time is removed, i.e.

models with no fixed effects specified at all. The Akaike Information Criterion (AIC) then allows us to compare, for instance, a model for English that contains a linear effect of time with a model that does not: AIC permits model comparison that rewards goodness of fit while penalizing models that are overly complex. Table 3 provides the AIC values for all seven languages to two decimal points, both with time and without time as a predictor. In all seven cases, the AIC for the simpler model without time as a predictor is lower than the AIC for the more complex model with time as a predictor, suggesting that the simpler model without time as a predictor should be preferred in all cases.

< INSERT TABLE 3 ABOUT HERE >

It is worth emphasizing again that the Irish and Chinese datasets are an order of magnitude smaller than the others. As is well known, statistical power is closely related to sample size, and low power is associated with Type II errors, i.e. failing to reject the null hypothesis when in fact it should be rejected. Thus – although it is by no means impossible to detect effects of time in diachronic datasets of this size (see e.g. the case studies in Kroch 1989) – the absence of demonstrable effects of time in these two languages in particular may be an artefact of small sample size, and cannot be taken as conclusive evidence that there is no such effect. To assess this possibility in more detail, I conducted a power analysis for Irish and Chinese (as the 'smallest' languages in the dataset) and English (as the 'largest') using the package SIMR (Green & MacLeod 2016). This process allows one to estimate the power of a study (i.e. the probability that the test will correctly reject the null hypothesis if it is false) based on properties of the sample and the model, by simulating values for the dependent variable, refitting the model, and testing the simulated fit. Keeping the model structure and random effects constant, I tested this for a range of effect sizes, starting from 0.000005 per year (equivalent to an increment of 1% in hypotaxis level every 2,000 years, i.e. a very weak effect) and rising to 0.001 per year (equivalent to an increment of 1% every ten years, i.e. a moderate effect). For each effect size, 100 simulations were conducted. The conventional threshold for a well-powered test is a power of 0.8 (Green & MacLeod 2016: 494), meaning that there is an 80% chance of correctly rejecting the false null hypothesis. For English, this power threshold is reached with an effect size of 0.000055 (1% every 182 years), and for Chinese, with an effect size of 0.000075 (1% every 133 years); for Irish, probably due to the uneven distribution of texts over time, the power threshold is only reliably reached with an effect size of around 0.000665 (1% every 15 years). For English and Chinese, therefore, we can be relatively confident that any non-trivial linear effect of time would be detected; for Irish, it is possible that a weak effect might fly under the radar. For all languages, carrying out comparable studies on larger corpora (insofar as this is possible at all) is of course a desideratum for future research.[7]

In summary, the linear models presented in this subsection provide a poor fit to the data, and time as a predictor explains very little of the variance in hypotaxis level encountered. Thus, these linear models provide little support for the quantitative parataxis-to-hypotaxis claim.

GA(M)Ms. Generalized additive mixed models are similar to linear mixed models but without the linearity. While a linear model predicts the value of a dependent variable y in terms of a linear function of a predictor variable x, i.e. $y = \alpha + \beta x$, generalized additive models allow any function of x without the requirement that it be linear, i.e. $y = f(x)$. In practice, GA(M)Ms are constructed by stitching together a number of basis functions (e.g. linear or polynomial functions) and smoothing the line that results. GAMMS differ from GAMS in allowing for random as well as fixed effects. See Sóskuthy (2017) for an introduction in the context of linguistics.

Since there exists (to my knowledge) no hypothesis that predicts hypotaxis level as a specific function of time, the use of GA(M)Ms here must be viewed as exploratory rather than inferential. GAMs were fitted using the BAM function from the MGCV package in R (Wood 2011); the plots in Figure 9 were generated using the PREDICT_GAM function from the TIDYGAM package (Coretta 2023). Full model details can be found in the supplementary materials.

< INSERT FIGURE 9 ABOUT HERE >

For Chinese and Irish, the languages with the least data, the GAM reduces to a linear model, and hence what we see in Figures 8 and 9 is basically the same. These languages will be left out of consideration in the remainder of this section.

The other five languages permit more complex model fits. For these, in addition to the GAMs, a GAMM with a random intercept for genre was also fitted. The key properties of the models are summarized in Table 4. For Chinese and Irish, the *p*-values for the predictor Date in this table are taken from the GAM; for the other five languages, the *p*-values from the GAMM are used.

< INSERT TABLE 4 ABOUT HERE >

By virtue of their greater flexibility, the GAM(M)s permit a closer fit to changing diachronic trends, but the findings in Figure 4 do surprisingly little to alter the overall picture obtained from the linear models in section 4.2. As in Table 2, we see that the effect of date is not significant for Chinese, Icelandic, Irish or German, while it is significant for English and French. The only change is with regard to Portuguese, where we see a significant effect of date in the GAMM but not in the linear model. The random effect of genre, on the other hand, was significant in every language in which it was tested, except Portuguese. Turning to the variance explained, Portuguese is also the only language for which genre explains less of the variance than date does.

As in section 4.2, it is possible to conduct AIC comparisons for all of the languages except Chinese and Irish. The AICs of the GAMs and GAMMs presented in Table 4 are given in Table 5, as well as the AIC of a GAMM with genre as a random effect but no fixed effect of time, again rounded to two decimal points. Here the picture is more mixed than in section 4.2. For Icelandic and Low German, the best model according to AIC is the one with only genre as a random effect, and no effect of time; the worst model is the one with only time as a predictor. For English, French and Portuguese, the best model is the most complex model, with both time and genre as a predictor; for English and French, the worst model is the model with only time as a predictor, and for Portuguese the worst model is the one with only genre as a random effect.

< INSERT TABLE 5 ABOUT HERE >

I interpret these findings as follows. The quantitative parataxis-to-hypotaxis claim is that hypotaxis levels increase over time, which I operationalized as meaning that a positive effect of time should be detectable in a linear model. As seen in section 4.2, of the seven languages investigated, this was the case only for English and French, and the explanatory power of the effect was very weak in both instances. The results in this section go beyond this and suggest that time is IN GENERAL a mediocre to poor predictor of hypotaxis levels, even when the precise function linking time to hypotaxis level is allowed to vary dramatically. For Icelandic and Low German, models with time as a predictor were strictly worse than models with only genre as a random effect. For English, French and Portuguese, a GAMM could be fit with time as a predictor which outperformed a GAMM with only genre as a random effect, but the peaks and troughs of these models are varied, with no obvious generalization to be found across the three languages as to when hypotaxis is at its most prevalent.

5. DISCUSSION AND CONCLUSION. The aim of this report was to evaluate the claim that, in language history, the proportion of hypotaxis will increase at the expense of parataxis. Based on data from parsed corpora of Chinese, English, French, Icelandic, Irish, Low German and Portuguese, I found little support for this claim. A statistically significant increase was only found for English and French, and in these two languages the effect accounts for very little of the variance in the data. In addition, results from generalized additive mixed models (GAMMs) suggest that modelling hypotaxis level as a function of time is quite generally not likely to be successful: the proportion of variance accounted for is still low, and there is a risk of fitting models which have no sensible diachronic linguistic interpretation, as well as a risk of overfitting to the corpora of historical texts that we happen to have annotated or attested. An important caveat, of course, is that the results in this paper only pertain to the periods actually covered by the corpora: if, for instance, parataxis lost ground in the history of Low German before 1200 or

after 1650, this would not be captured in the results reported here, and the same holds for the remaining attested histories of Irish and Portuguese. This limitation means that further research on different chronological stages is warranted, at least where texts exist.

Genre and text type, on the other hand, seems to exert a substantial influence on hypotaxis levels. This study was not designed to test hypotheses about the effects of different genres, but the fact that genre differences exist should come as no surprise in view of the literature: the papers in Whitt 2018 make the case for including genre as a predictor in diachronic corpus linguistics,[8] and differences in the distribution of syntactic properties across genres in present-day English have been demonstrated by Biber and colleagues (e.g. Biber 1995, Biber & Gray 2016, Biber & Conrad 2019), among others. The Appendix to this paper provides graphs breaking down the diachronic developments by genre in the better-attested languages. This report, then, contributes to a literature suggesting that what at first blush appears to be diachronic change may in fact reflect the shifting availabilities of different genres, registers and text types in the corpora available to us (see e.g. Warner 2005).

Some generalizations about genre effects can be made, albeit tentatively. Legal texts, in the languages for which they are robustly attested in the corpora (English, French, Low German), consistently exhibit more hypotaxis than other genres. By contrast, text types such as diaries, dramas, and verse generally seem to exhibit low levels of hypotaxis. It is tempting to arrange the attested genres on a scale from least to most prototypically oral: Chafe (1982: 44) presents data showing that complement clauses and relative clauses in English are about twice as common in formal written production than in informal spoken production, attributing this to the fact that speaking is faster than writing and that there is hence more pressure against structures (like at least some hypotactic relations) that are harder to process. It is important to avoid circularity in this sort of argument, though: one should not simultaneously assume that a genre is more oral simply because it has more parataxis. Ideally, orality itself would be inferred on the basis of a set of properties that can be independently measured in corpora, as is done for historical Spanish by Rosemeyer (2019) and McCarley (to appear). Future work should explore this possibility, though we should probably not expect genre effects to be TOO universal: genres/text types themselves can vary and change over time (see e.g. the papers in Taavitsainen & Palander-Collin 2011 on medical and scientific writing in the history of English), and genres that exist for one language or culture – e.g. Icelandic sagas – may simply fail to have a counterpart in others.

All a study like this can do is fail to reject the null hypothesis; the parataxis-to-hypotaxis claim finds little support here, but neither has it been falsified. Moreover, it could be that this report is looking in the wrong place. One possibility is that these specific seven languages happen to be exceptions to a more general trend; in this connection it is worth emphasizing again that six of the seven are Indo-European, and that the Chinese corpus – the only non-Indo-European

language represented – is small enough that the absence of an effect could be due to lack of power. Another possibility is that the corpora do not contain the relevant evidence: for instance, perhaps the texts included in the corpora are not representative of broader language use. A third possibility is that the annotation is systematically misleading: perhaps the methods used to parse the Penn-style corpora drawn on here do not succeed in capturing the distinction between parataxis and hypotaxis, or are systematically biased towards one or another. Still another possibility is that the claim pertains to an earlier stage of linguistic history: only the Chinese corpus contains data from before 700 CE, leaving room for the possibility that a quantitative shift from parataxis to hypotaxis took place during classical antiquity or in prehistory. If so, the question arises of whether the general claim is testable at all, given how few languages are robustly attested from the first millennium CE or earlier, and how sparse and fragmentary the surviving texts from these languages are. Still, none of these four possibilities can be ruled out.

Moreover, as discussed in section 2, there exist other understandings of the parataxis-to-hypotaxis claim that this type of study does not directly bear on at all. These include the idea that clausal complementation structures are reanalysed from juxtaposition of unembedded clauses (see the discussion in Weiß 2020 and Walkden 2024: 12–16), the idea that the maximum possible level of embedding increases over time, and even the idea that parataxis precedes hypotaxis in the evolution of the human language faculty (Progovac 2014). These ideas stand and fall irrespective of the evidence presented in this paper.

Still, the paucity of the evidence FOR the parataxis-to-hypotaxis claim stands in stark contrast to the frequency and confidence with which it has been asserted in the literature. Hopefully the findings presented in this report serve to cast doubt on a venerable piece of received wisdom, or at least contribute to shifting the burden of proof.

A. APPENDIX: VISUALIZATION OF DIACHRONIC DEVELOPMENTS BY GENRE. Chinese and Irish are not included here due to data sparsity.

< INSERT FIGURE 10 ABOUT HERE >

< INSERT FIGURE 11 ABOUT HERE >

< INSERT FIGURE 12 ABOUT HERE >

< INSERT FIGURE 13 ABOUT HERE >

< INSERT FIGURE 14 ABOUT HERE >

REFERENCES

ADMONI, WLADIMIR G. 1980. *Zur Ausbildung der Norm der deutschen Literatursprache im Bereich des neuhochdeutschen Satzgefüges (1470–1730). Ein Beitrag zur Geschichte des Gestaltungssystems der deutschen Sprache*. Berlin: De Gruyter.

ANDREW, S. O. 1940. *Syntax and style in Old English*. Cambridge: Cambridge University Press.

AXEL, KATRIN. 2009. Die Entstehung des *dass*-Satzes: ein neues Szenario. *Koordination und Subordination im Deutschen*, ed. by Veronika Ehrich, Christian Fortmann, Ingo Reich, and Marga Reis, 21–41. Hamburg: Buske.

AXEL-TOBER, KATRIN. 2012. *(Nicht-)kanonische Nebensätze im Deutschen: Synchrone und diachrone Aspekte*. Berlin: Mouton de Gruyter.

AXEL-TOBER, KATRIN. 2017. The development of the declarative complementizer in German. *Language* 93.e29–e65.

BARTOŃ, KAMIL. 2023. Package 'MuMIn'. Version 1.47.5.

BATES, DOUGLAS; MARTIN MÄCHLER; BEN BOLKER; and STEVE WALKER. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67.1–48.

BIBER, DOUGLAS. 1995. *Dimensions of register variation: a cross-linguistic comparison*. Cambridge: Cambridge University Press.

BIBER, DOUGLAS, and SUSAN CONRAD. 2019. *Register, genre, and style*. 2 edn. Cambridge: Cambridge University Press.

BIBER, DOUGLAS, and BETHANY GRAY. 2016. *Grammatical complexity in academic English: linguistic change in writing*. Cambridge: Cambridge University Press.

BLOEM, JELKE; ARJEN VERSLOOT; and FRED WEERMAN. 2015. An agent-based model of a historical word order change. *Proceedings of the sixth workshop on cognitive aspects of computational language learning*, 22–27. Stanford, CA: Association for Computational Linguistics.

BOOTH, HANNAH; ANNE BREITBARTH; AARON ECAY; and MELISSA FARASYN. 2020. A Penn-style treebank of Middle Low German. *Proceedings of the 12th conference on Language Resources and Evaluation (LREC 2020)*, 766–775.

BOUMAN, ARI C. 1958. An aspect of style in Icelandic sagas. *Neophilologus* 42.50–67.

CHAFE, WALLACE. 1982. Integration and involvement in speaking, writing, and oral literature. *Spoken and written language: Exploring orality and literacy*, ed. by Deborah Tannen, 35–53. New Jersey, NJ: Ablex.

CORETTA, STEFANO. 2023. *tidygam: Tidy prediction and plotting of generalised additive models*. https://github.com/stefanocoretta/tidygam, https://stefanocoretta.github.io/tidygam/.

DONOGHUE, DANIEL, and BRUCE MITCHELL. 1992. Parataxis and hypotaxis: A review of some terms used for Old English syntax. *Neuphilologische Mitteilungen* 93.163–183.

DĄBROWSKA, EWA. 2015. Language in the mind and in the community. *Change of paradigms – new paradoxes: recontextualizing language and linguistics*, ed. by Jocelyne Daems, Eline Zenner, Kris Heylen, Dirk Speelman, and Hubert Cuyckens, 221–236. Berlin: de Gruyter.

FARASYN, MELISSA; GEORGE WALKDEN; SHEILA WATTS; and ANNE BREITBARTH. 2018. The interplay between genre variation and syntax in a historical Low German corpus. *Diachronic corpora, genre, and language change*, ed. by Richard Jason Whitt, 181–200. Amsterdam: John Benjamins.

FOX, JOHN, and SANFORD WEISBERG. 2019. *An R companion to applied regression*. 3rd edn. Thousand Oaks CA: Sage. Online: `https://socialsciences.mcmaster.ca/jfox/Books/Companion/index.html`.

GALVES, CHARLOTTE; AROLDO LEAL DE ANDRADE; and PABLO FARIA. 2017. Tycho Brahe parsed corpus of historical Portuguese. Unicamp. Online: `https://www.tycho.iel.unicamp.br/corpus/en/index.html`.

GILDERSLEEVE, BASIL L. 1883. On the final sentence in Greek. *The American Journal of Philology* 4.416–444.

GRANO, THOMAS. 2017. Finiteness contrasts without Tense? A view from Mandarin Chinese. *Journal of East Asian Linguistics* 26.259–299.

GREEN, PETER, and CATRIONA J MACLEOD. 2016. SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution* 7.493–498.

GÄRTNER, HANS-MARTIN. 2014. On covert modality in German root infinitives. *Proceedings of the 31st West Coast Conference on Formal Linguistics*, 199–206.

HARRIS, ALICE, and LYLE CAMPBELL. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.

HOPPER, PAUL, and ELIZABETH C. TRAUGOTT. 2003. *Grammaticalization*. 2 edn. Cambridge: Cambridge University Press.

HU, JIANHUA; HAIHUA PAN; and LIEJIONG XU. 2001. Is there a finite vs. nonfinite distinction in Chinese? *Linguistics* 39.1117–1148.

JOHANNESSEN, JANNE B. 1998. *Coordination*. Oxford: Oxford University Press.

JOHNSON, DANIEL EZRA. 2009. Getting off the GoldVarb standard: Introducing Rbrul for mixed effects variable rule analysis. *Language and Linguistics Compass* 3.359–383.

JUCKER, ANDREAS H. 1991. Between hypotaxis and parataxis: clauses of reason in Ancrene Wisse. *Historical English syntax*, ed. by Dieter Kastovsky, 203–220. Berlin: de Gruyter.

KARLSSON, FRED. 2009. Origin and maintenance of clausal embedding complexity. *Language complexity as an evolving variable*, ed. by Geoffrey Sampson, David Gil, and Peter Trudgill, 192–202. Oxford: Oxford University Press.

KROCH, ANTHONY. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1.199–244.

KROCH, ANTHONY, and BEATRICE SANTORINI. 2021. Penn-BFM parsed corpus of Historical French. Version 1.0. Online: `https://github.com/beatrice57/mcvf-plus-ppchf`.

KROCH, ANTHONY; BEATRICE SANTORINI; and LAUREN DELFS. 2004. Penn-Helsinki parsed corpus of early Modern English. University of Pennsylvania. Online: `https://www.ling.upenn.edu/ppche/ppche-release-2016/PPCEME-RELEASE-3`.

KROCH, ANTHONY; BEATRICE SANTORINI; and C. E. A. DIERTANI. 2016. Penn-Helsinki Parsed Corpus of Modern British English. 2nd edition. Online: `https://www.ling.upenn.edu/ppche/ppche-release-2016/PPCMBE2-RELEASE-1`.

KROCH, ANTHONY, and ANN TAYLOR. 2000. Penn-Helsinki Parsed Corpus of Middle English Prose. 2nd edition. Online: `https://www.ling.upenn.edu/ppche/ppche-release-2016/PPCME2-RELEASE-4`.

KUZNETSOVA, ALEXANDRA; PER B. BROCKHOFF; and RUNE H. B. CHRISTENSEN. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82.1–26.

LASH, ELLIOTT. 2014. The parsed Old and Middle Irish corpus (POMIC). Version 0.1. Online: https://www.dias.ie/celt/celt-publications-2/celt-the-parsed-old-and-middle-irish-corpus-pomic/.

LI, MAN. 2017. Chinese Parsed Historical Corpus (ChiPaHC). Newcastle University.

LI, YIN. 2018. The diachronic development of the Chinese passive: From the *wei ... suo* passive to the long passive. *Language (Historical Syntax)* 94.e74–e98.

LOS, BETTELOU. 2005. *The rise of the* to-*infinitive*. Oxford: Oxford University Press.

MARTINEAU, FRANCE; PAUL HIRSCHBÜHLER; ANTHONY KROCH; and YVES CHARLES MORIN. 2021. MCVF Corpus. Version 2.0. Online: https://github.com/beatrice57/mcvf-plus-ppchf.

MARTINS, A. M.; SANDRA PEREIRA; and ADRIANA CARDOSO. 2015a. Parsed Demanda do Santo Graal. WOChWEL by Centro de Linguística da Universidade de Lisboa. Online: http://alfclul.clul.ul.pt/wochwel/oldtexts.html.

MARTINS, A. M.; SANDRA PEREIRA; and ADRIANA CARDOSO. 2015b. Parsed José de Arimateia. WOChWEL by Centro de Linguística da Universidade de Lisboa. Online: http://alfclul.clul.ul.pt/wochwel/oldtexts.html.

MARTINS, A. M.; SANDRA PEREIRA; and ADRIANA CARDOSO. 2015c. Parsed legal documents. WOChWEL by Centro de Linguística da Universidade de Lisboa. Online: http://alfclul.clul.ul.pt/wochwel/oldtexts.html.

MCCARLEY, GEMMA. 2025. Orality and overtness: effects on Spanish subject use. *Journal of Historical Sociolinguistics*, Online ahead of print: doi = 10.1515/jhsl-2024-0017.

NAKAGAWA, SHINICHI; PAUL C. D. JOHNSON; and HOLGER SCHIELZETH. 2017. The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface* 14.20170213.

NAKAGAWA, SHINICHI, and HOLGER SCHIELZETH. 2013. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in ecology and evolution* 4.133–142.

NEWMEYER, FREDERICK J. 2002. Uniformitarian assumptions and language evolution research. *The transition to language*, ed. by Alison Wray, 359–375. Oxford: Oxford University Press.

PAUL, HERMANN. 1920. *Prinzipien der Sprachgeschichte*. 5 edn. Halle: Max Niemeyer.

PEREIRA, SANDRA. 2017. Parsed Crónica Geral de Espanha. WOChWEL by Centro de Linguística da Universidade de Lisboa. Online: `http://alfclul.clul.ul.pt/wochwel/oldtexts.html`.

PROGOVAC, LJILJANA. 2014. *Evolutionary syntax*. Oxford: Oxford University Press.

RAINSFORD, THOMAS, and OLGA SCRIVNER. 2014. Metrical annotation for a verse treebank. *The 13th international workshop on treebanks and linguistic theories (TLT13)*, 149–159.

RANDALL, BETH; ANTHONY KROCH; and ANN TAYLOR. 2005–2018. CorpusSearch 2. Online, University of Pennsylvania. Online: `http://corpussearch.sourceforge.net/`.

ROSEMEYER, MALTE. 2019. Actual and apparent change in Brazilian Portuguese *wh*-interrogatives. *Language Variation and Change* 31.165–191.

SAUERLAND, ULI; BART HOLLEBRANDSE; and FRANTIŠEK KRATOCHVÍL. 2020. When hypotaxis looks like parataxis: embedding and complementizer agreement in Teiwa. *Glossa* 5.89.

SMALL, GEORGE W. 1924. *The comparison of inequality*. Baltimore: University Press.

SÓSKUTHY, MÁRTON. 2017. Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction. arXiv preprint arXiv:1703.05339.

SYBESMA, RINT. 2017. Finiteness. *Encyclopedia of Chinese language and linguistics*, ed. by Rint Sybesma. Leiden: Brill.

TAAVITSAINEN, IRMA, and MINNA PALANDER-COLLIN (eds.) 2011. *Medical writing in Early Modern English*. Cambridge: Cambridge University Press.

TAYLOR, ANN; ANTHONY WARNER; SUSAN PINTZUK; and FRANK BETHS. 2003. York-Toronto-Helsinki Parsed Corpus of Old English Prose. 1st edition. Online: `http://www-users.york.ac.uk/~lang22/YCOE/YcoeHome.htm`.

TRUSWELL, ROBERT. 2016. Book review: *Evolutionary Syntax*, by Ljiljana Progovac. *Journal of Language Evolution* 1.168–170.

WALKDEN, GEORGE. 2024. Parataxis and hypotaxis in the history of English. *Keys to the history of English: diachronic linguistic change, morpho-syntax and lexicography. Selected papers from the 21st ICEHL*, ed. by Thijs Porck, Moragh S. Gordon, and Luisella Caon, 10–33. Amsterdam: John Benjamins.

WALLENBERG, JOEL C.; ANTON KARL INGASON; EINAR FREYR SIGURÐSSON; and EIRÍKUR RÖGNVALDSSON. 2011. Icelandic Parsed Historical Corpus (IcePaHC), version 0.9. http://linguist.is/icelandic_treebank.

WARNER, ANTHONY. 2005. Why DO dove: evidence for register variation in Early Modern English negatives. *Language Variation and Change* 17.257–280.

WEISS, HELMUT. 2020. Where do complementizers come from and how did they come about? A re-evaluation of the parataxis-to-hypotaxis hypothesis. *Evolutionary Linguistic Theory* 2.30–55.

WHITT, RICHARD JASON (ed.) 2018. *Diachronic corpora, genre, and language change*. Amsterdam: John Benjamins.

WICKHAM, HADLEY. 2016. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Online: `https://ggplot2.tidyverse.org`.

WOOD, S. N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)* 73.3–36.

ÁGEL, VILMOS. 2000. Syntax des Neuhochdeutschen bis zur Mitte des 20. Jahrhunderts. *Sprachgeschichte: Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*, ed. by Werner Besch, Anne Betten, Oskar Reichmann, and Stefan Sonderegger, vol. 2, 2 edn., 1855–1903. Berlin: Walter de Gruyter.

NOTES

[1]See Truswell 2016 for critical discussion of Progovac's thesis, observing that the proposal potentially conflates biological with cultural evolution.

[2]This conjunction is tagged KON in CHLG, CONJ0 in the French corpora, and CONJ in all the other corpora. Asyndetic coordination – i.e. coordination without an overt conjunction – is another possibility, but the Penn annotation scheme does not represent it, nor is it likely that there exists any consistent method for distinguishing asyndetic coordination from truly independent clauses, at least in the seven languages in question. Thus all potential instances of asyndetic coordination are treated as independent unembedded clauses for the purposes of this study.

[3]Direct questions and exclamatives are annotated as IP-SUB immediately dominated by a CP-QUE or CP-EXL. The latter is always unembedded, while the former is annotated as an unembedded question if it does NOT contain anything tagged C (for subordinating conjunction).

[4]The IP-SUB category is not restricted to clauses introduced by an overt subordinator. As such, this category includes asyndetic embedded clauses such as verb-first conditional (*Should you have problems, ...*) and argument clauses without an overt subordinator (*He said he was sick*).

[5]As alluded to in section 2.1, there exist rare exceptions like *Oh, to be in Paris in the springtime!*; see e.g. Gärtner 2014 and references cited there on German root infinitives. These typically have an elliptical flavour, and pilot searches conducted on the languages in question confirmed that they are quantitatively marginal. As such, they are not considered further in this report. Also not included are other, 'smaller' types of non-finite clause such as participial absolutes (IP-ABS) and small clauses (IP-SMC). These types are a) not hugely common, b) not consistently annotated across all of the Penn-style parsed corpora, and c) are usually not in competition with finite clauses in the sense of the Labovian envelope of variation.

[6]Plots of the linear models are generated using the EFFECTS function from the EFFECTS package in R (Fox & Weisberg 2019), along with GGPLOT2.

[7]Information on how to replicate this power analysis, as well as plots of power against effect size for the Chinese, English and Irish datasets, can be found in the supplementary materials at `https://github.com/gwalkden/hypotaxis`.

[8]For instance, Farasyn et al. (2018) show that genre is a significant factor in the use of referential null subjects and different types of relativization strategies in historical Low German.

B. LIST OF TABLES.

C. LIST OF FIGURES.

Language	Corpora	Words	Dates covered
Chinese	ChiPaHC (Li 2017)	50k	C3rd BCE – C17th
English	YCOE (Taylor et al. 2003), PPCME2 (Kroch & Taylor 2000), PPCEME (Kroch et al. 2004), PPCMBE2 (Kroch et al. 2016)	7.2 mil	c. 800 – 1913
French	MCVF2 (Martineau et al. 2021), PPCHF (Kroch & Santorini 2021)	1.6 mil	C9th – C18th
Icelandic	IcePaHC (Wallenberg et al. 2011)	1 mil	c. 1150 – 2008
Irish	POMIC (Lash 2014)	30k	c. 700 – c. 1100
Low German	CHLG (Booth et al. 2020)	1.45 mil	1200 – 1650
Portuguese	WOChWEL (Martins et al. 2015a,b,c, Pereira 2017), Tycho Brahe (Galves et al. 2017)	1.8 mil	C13th – C19th

TABLE 1. Corpora used in the study

Language	Effect size (date)	p-value (date)	Marginal R^2	Conditional R^2
Chinese	1.651e-05	0.454	0.0522226	0.0522226
English	4.524e-05	0.0122*	0.005637543	0.4518298
French	2.830e-04	0.000958**	0.06928952	0.8028335
Icelandic	6.907e-07	0.991	2.403185e-06	0.405462
Irish	-0.0002921	0.1961	0.09869311	0.4916362
Low German	1.039e-04	0.686	0.00284414	0.779462
Portuguese	-0.0001309	0.206	0.05181601	0.2153938

TABLE 2. Effect sizes, p-values for date, and R^2 values. * = significant at the $p < 0.05$ level; ** = significant at the $p < 0.01$ level.

Language	Model AIC with time	Model AIC without time
Chinese	-9.56	**-30.65**
English	-863.39	**-879.13**
French	-110.94	**-118.79**
Icelandic	-84.22	**-103.85**
Irish	-0.58	**-16.17**
Low German	-6.00	**-22.59**
Portuguese	-34.06	**-51.01**

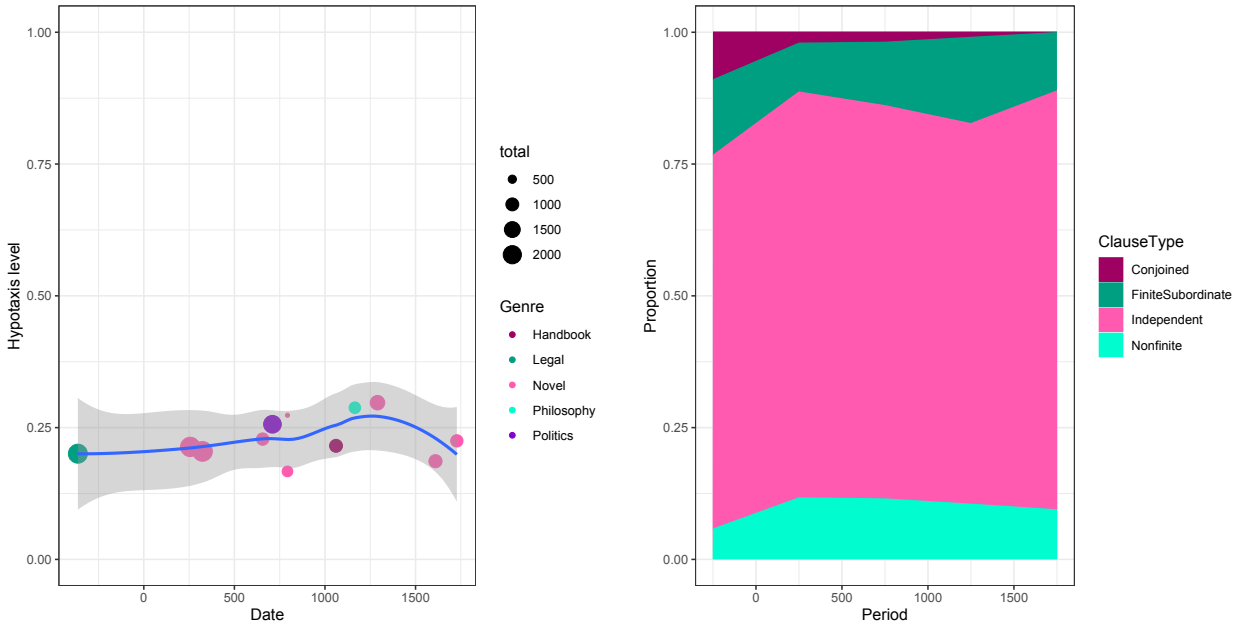
TABLE 3. AIC values for models with and without time as a predictor of hypotaxis level. The best AIC is bolded.

Language	p (date)	p (genre)	Variance explained (GAM)	Variance explained (GAMM)
Chinese	0.454	–	5.71%	–
English	<2e-16 **	<2e-16 **	17.7%	37.4%
French	7.49e-06 **	<2e-16 **	32.7%	83.2%
Icelandic	0.854	8.7e-06 **	20.9%	42.7%
Irish	0.324	–	8.82%	–
Low German	0.645789	0.000655 **	37.7%	85.3%
Portuguese	0.00384 **	0.07657	42.5%	59.3%

TABLE 4. GAM(M)s: p-values and variance explained. ** = significant at the $p < 0.01$ level.

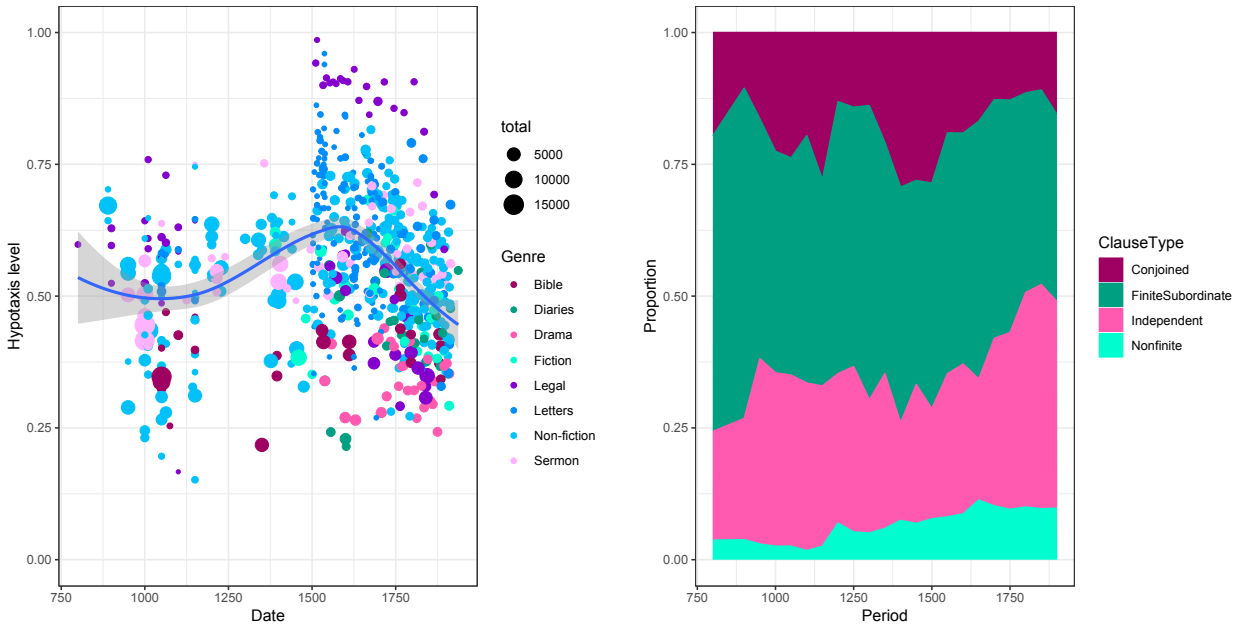
Language	GAM AIC with time	GAMM AIC with time and genre	GAMM AIC without time
English	-825.93	**-996.03**	-909.15
French	-102.82	**-175.73**	-143.04
Icelandic	-100.65	-114.47	**-118.14**
Low German	-19.74	-37.96	**-40.00**
Portuguese	-69.16	**-70.92**	-58.91

TABLE 5. AIC values for models with and without time as a predictor of hypotaxis level. The best AIC is bolded.



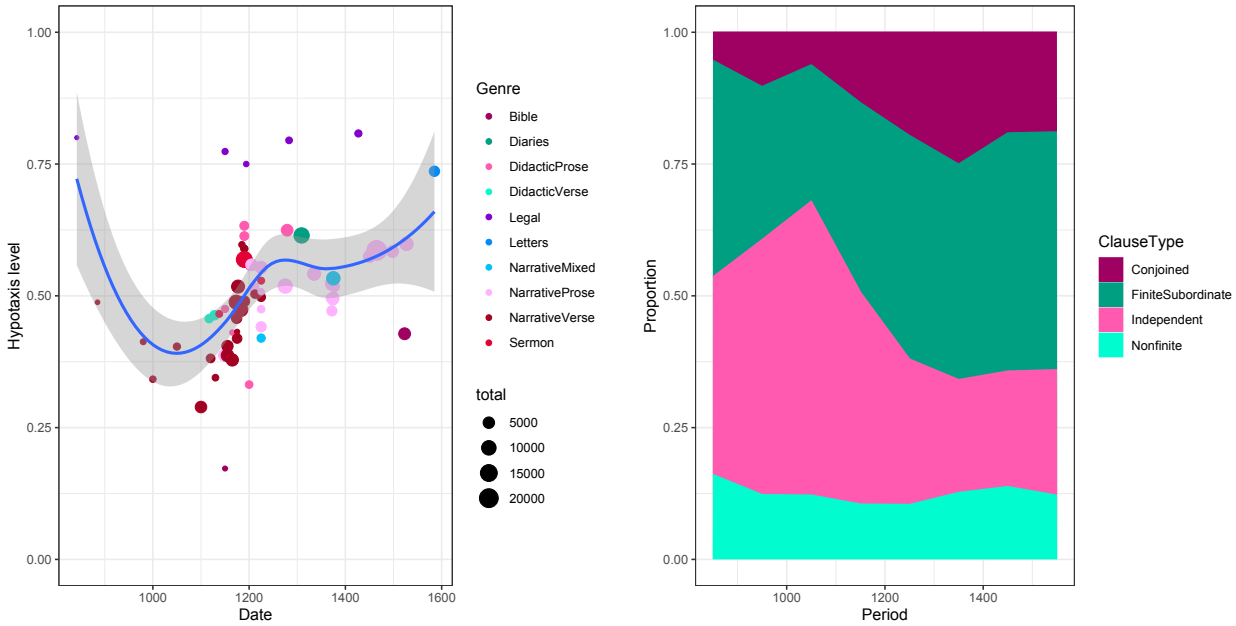
(A) Scatterplot by text (B) Clause types (bin width 500 years)

FIGURE 1. Parataxis and hypotaxis in the history of Chinese



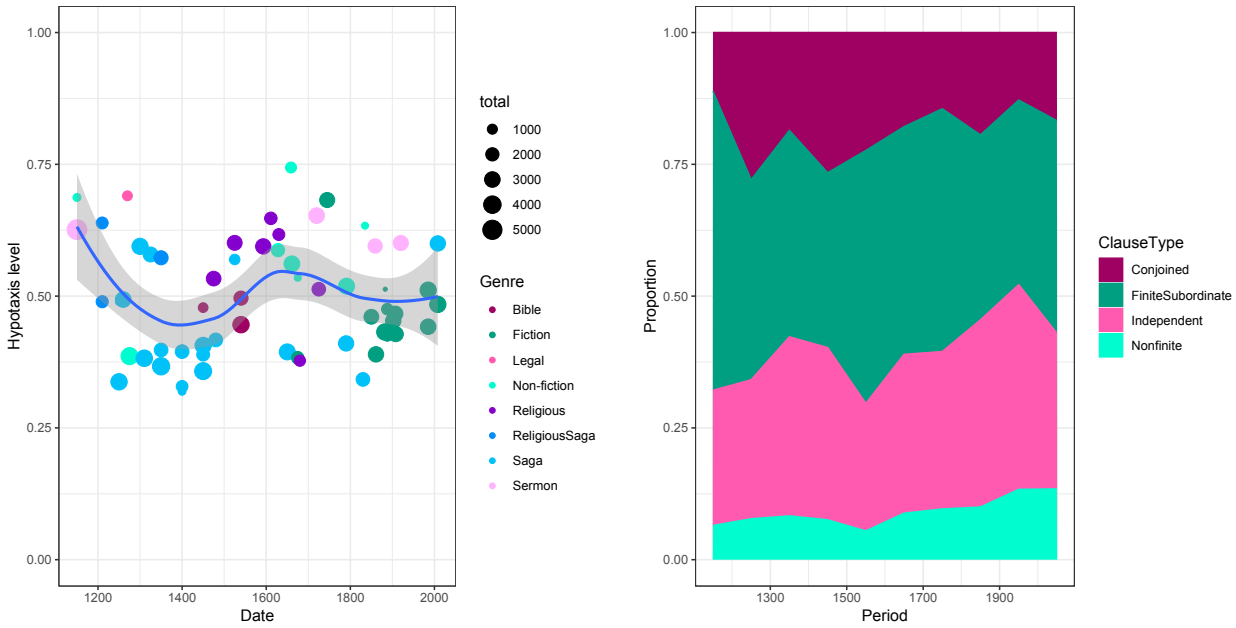
(A) Scatterplot by text (B) Clause types (bin width 50 years)

FIGURE 2. Parataxis and hypotaxis in the history of English



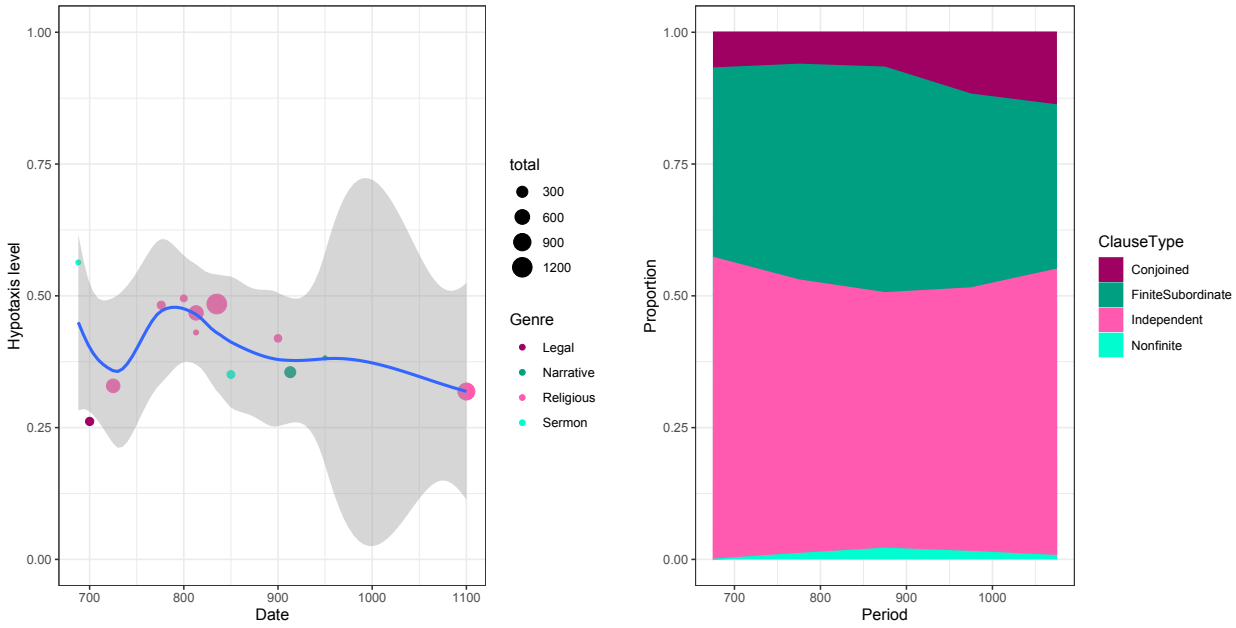
(A) Scatterplot by text (B) Clause types (bin width 100 years)

FIGURE 3. Parataxis and hypotaxis in the history of French



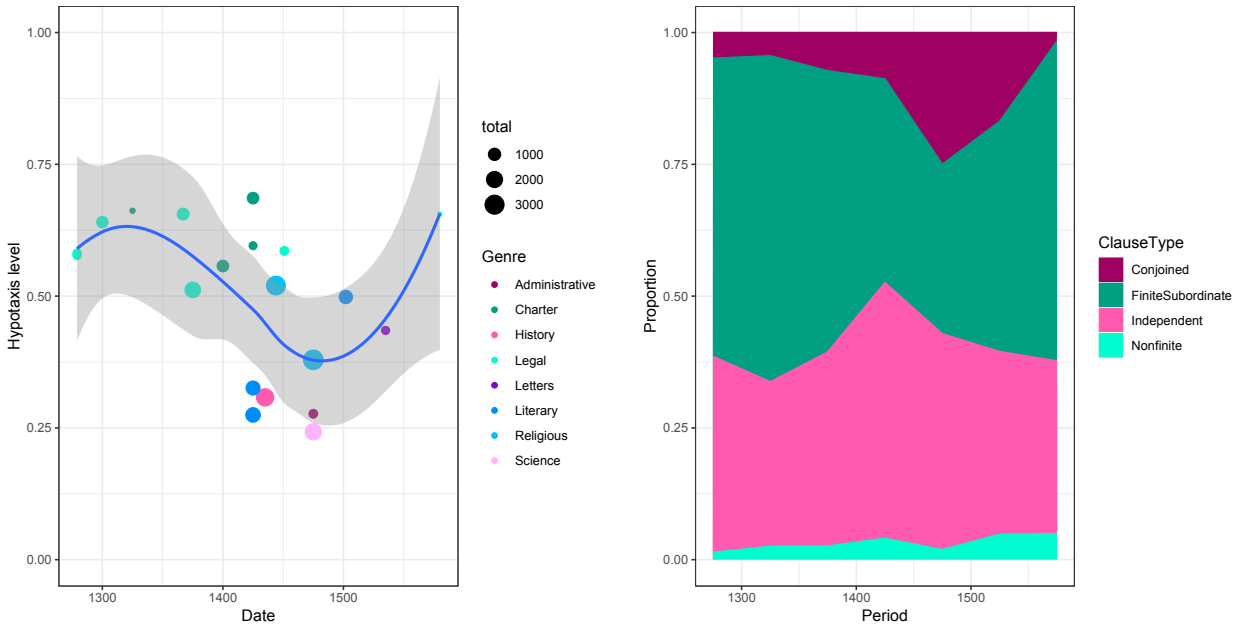
(A) Scatterplot by text (B) Clause types (bin width 50 years)

FIGURE 4. Parataxis and hypotaxis in the history of Icelandic



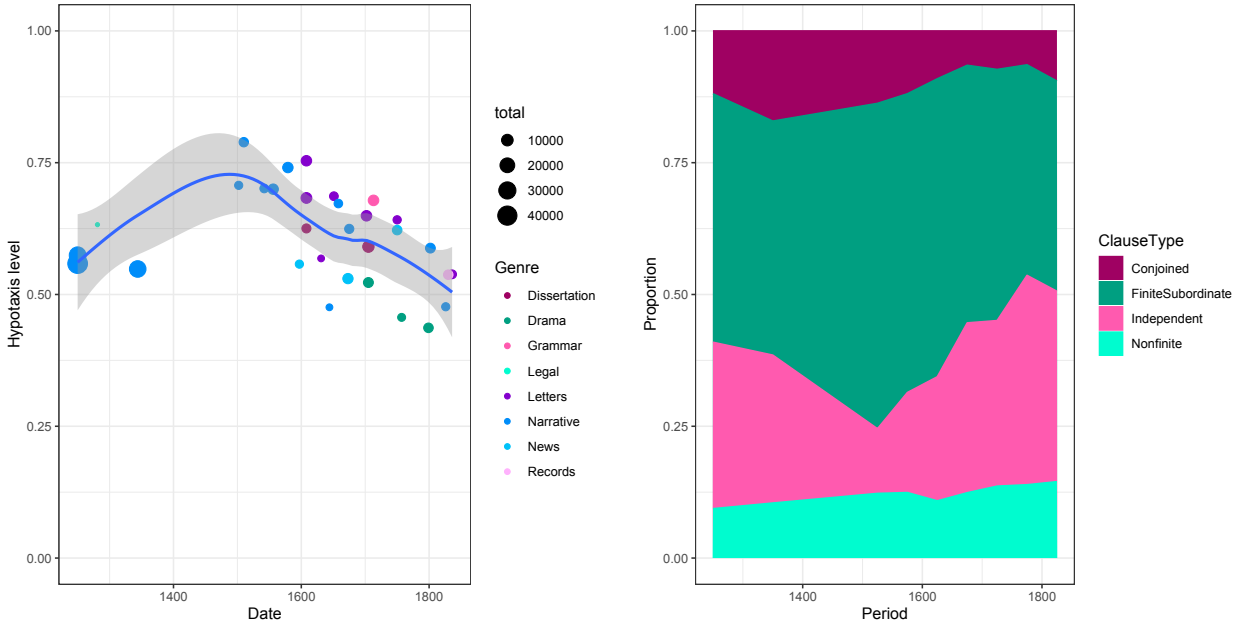
(A) Scatterplot by text (B) Clause types (bin width 100 years)

FIGURE 5. Parataxis and hypotaxis in the history of Irish



(A) Scatterplot by text (B) Clause types (bin width 50 years)

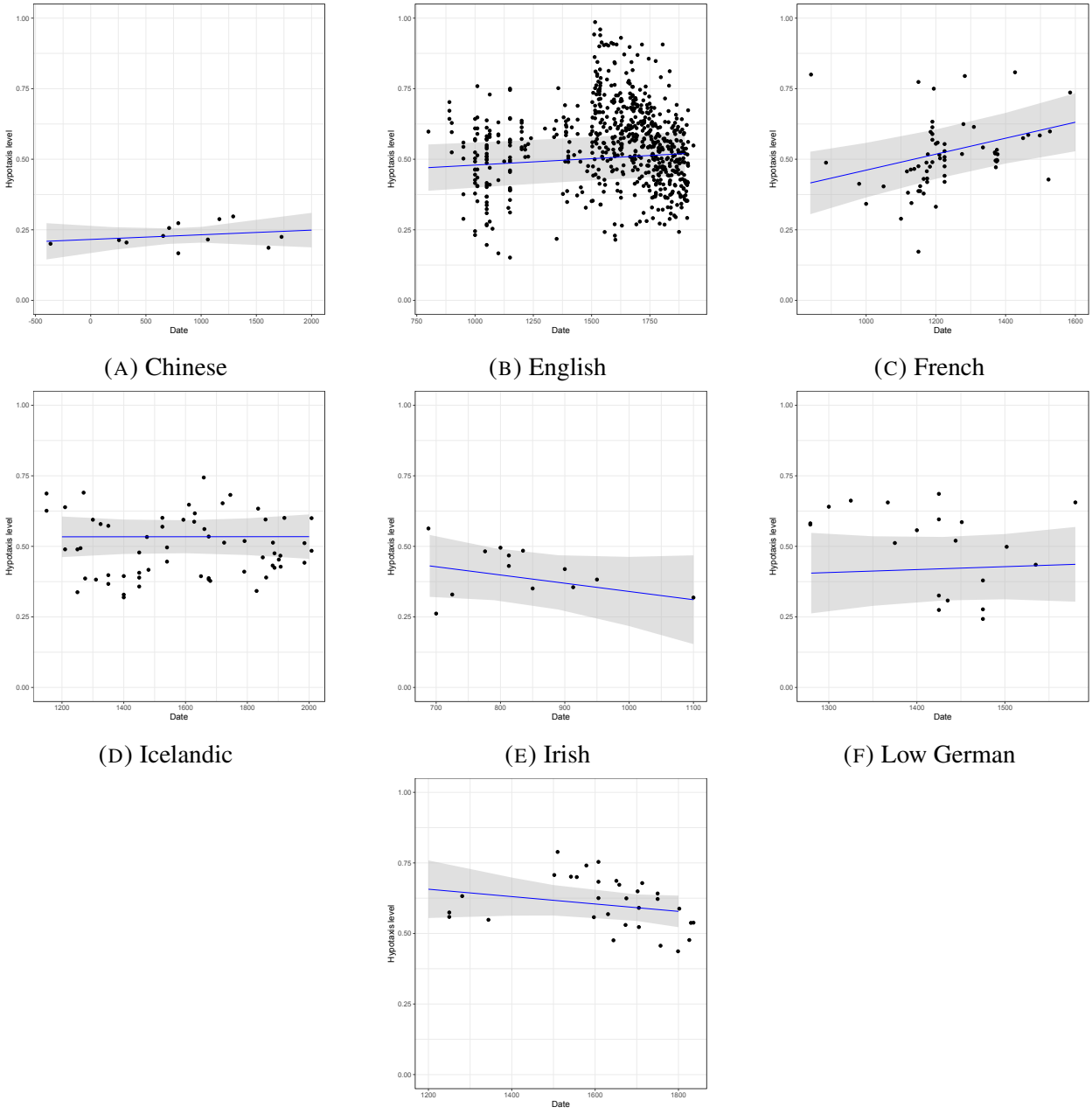
FIGURE 6. Parataxis and hypotaxis in the history of Low German



(A) Scatterplot by text

(B) Clause types (bin width 100/50 years)

FIGURE 7. Parataxis and hypotaxis in the history of Portuguese

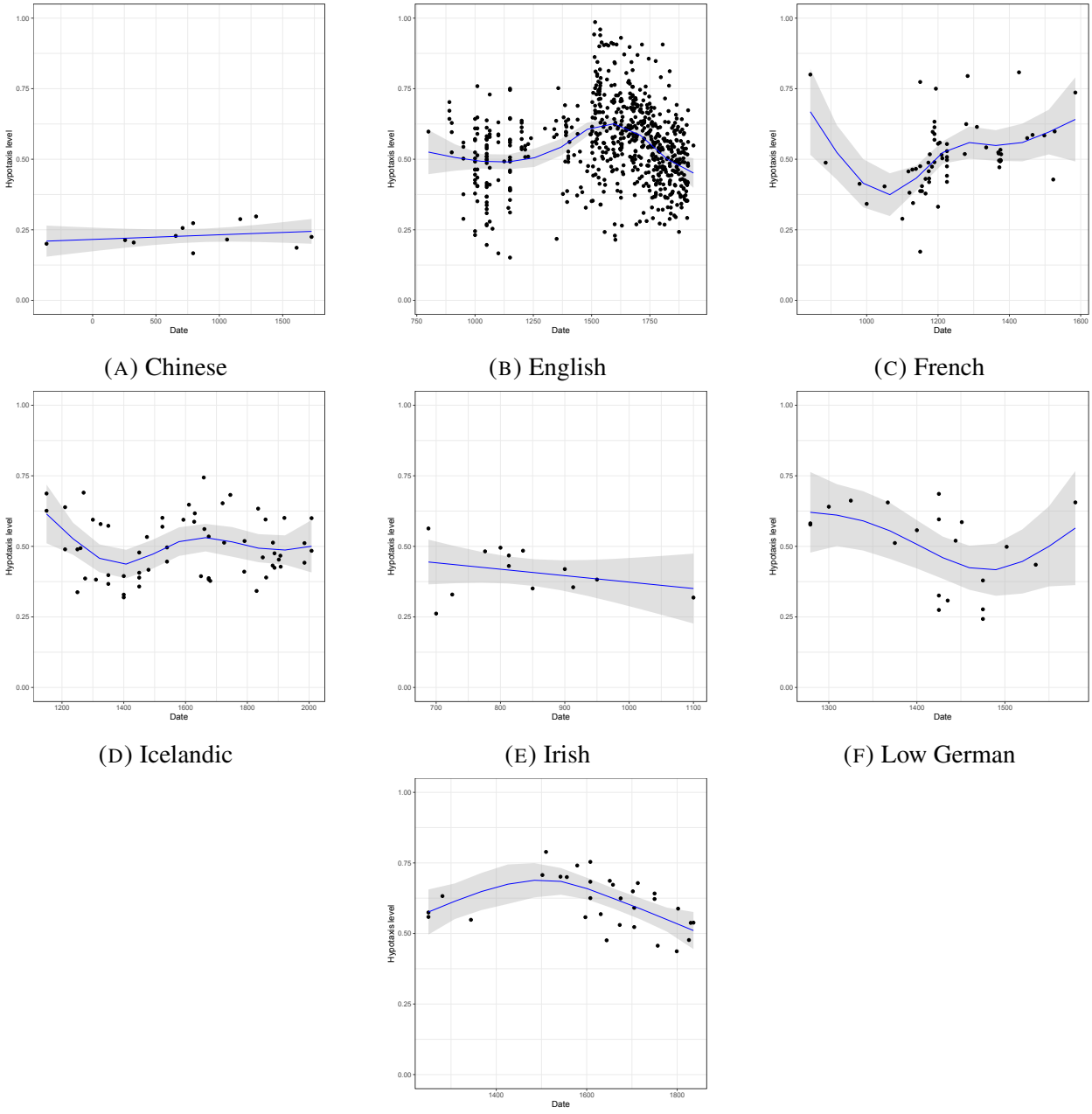


(A) Chinese (B) English (C) French

(D) Icelandic (E) Irish (F) Low German

(G) Portuguese

FIGURE 8. Linear effects of date for all seven languages, with 95% confidence intervals



(A) Chinese

(B) English

(C) French

(D) Icelandic

(E) Irish

(F) Low German

(G) Portuguese

FIGURE 9. GAMs for all seven languages, with 95% confidence intervals

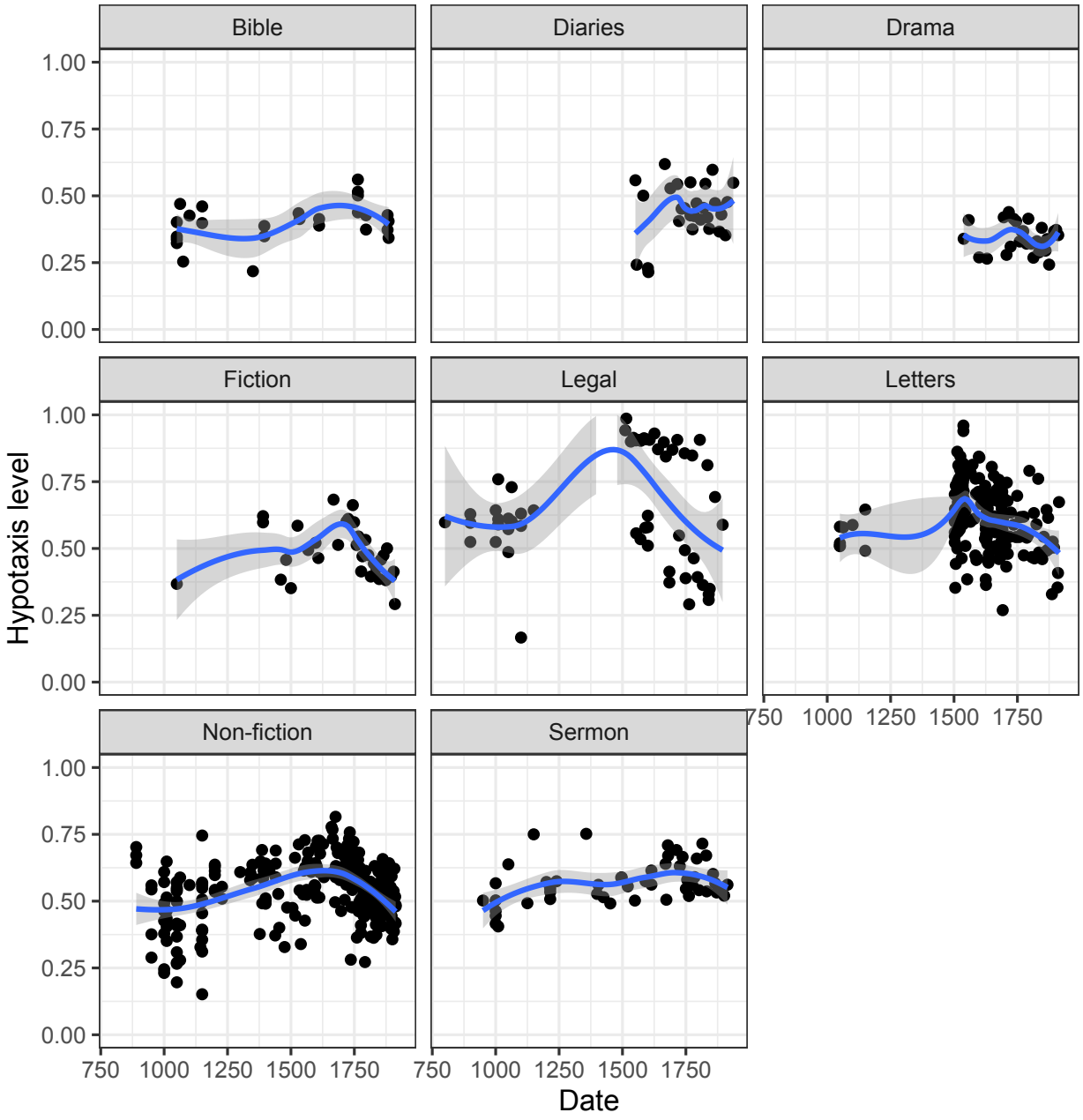


FIGURE 10. Hypotaxis level by genre in English

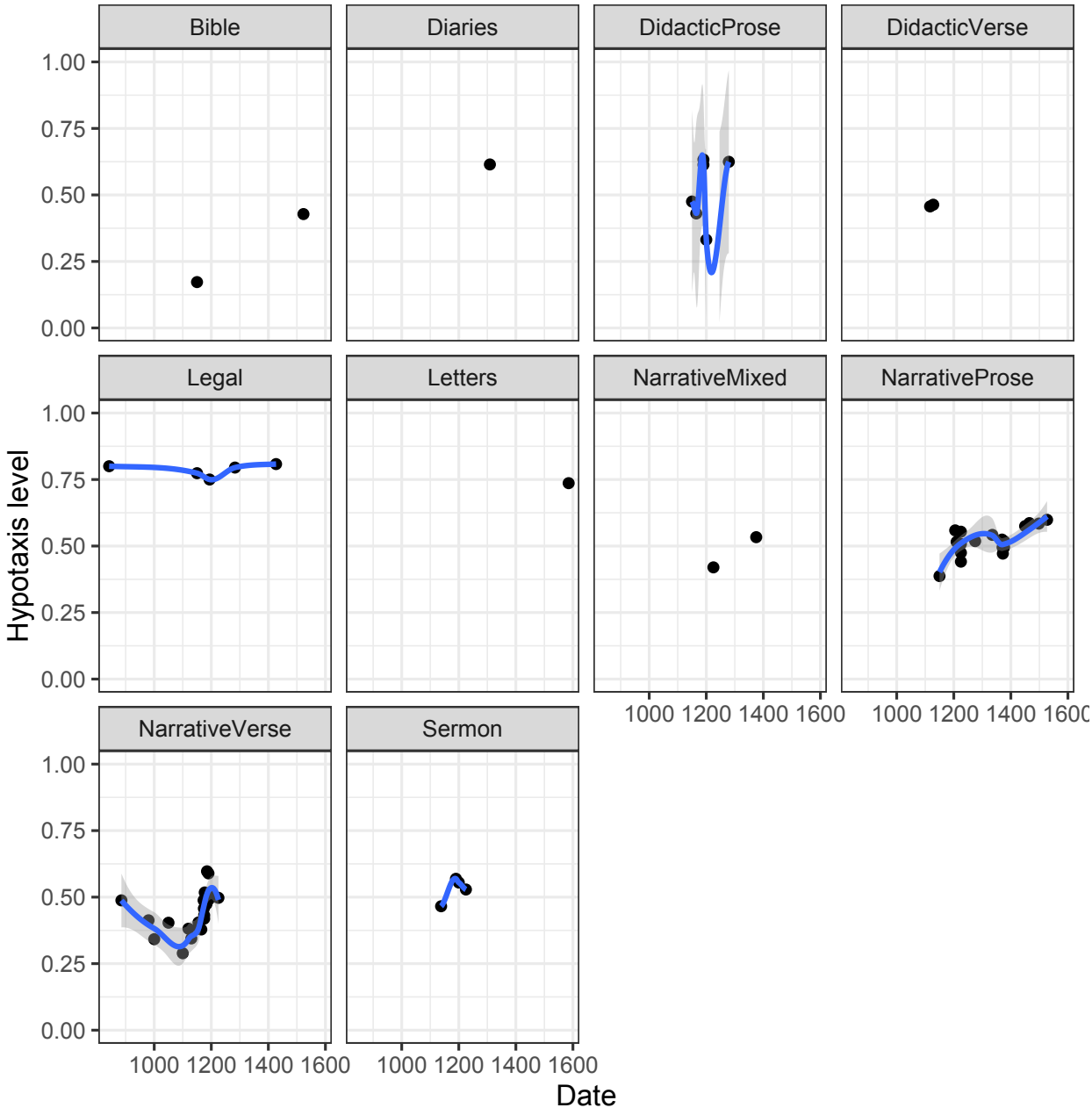


FIGURE 11. Hypotaxis level by genre in French

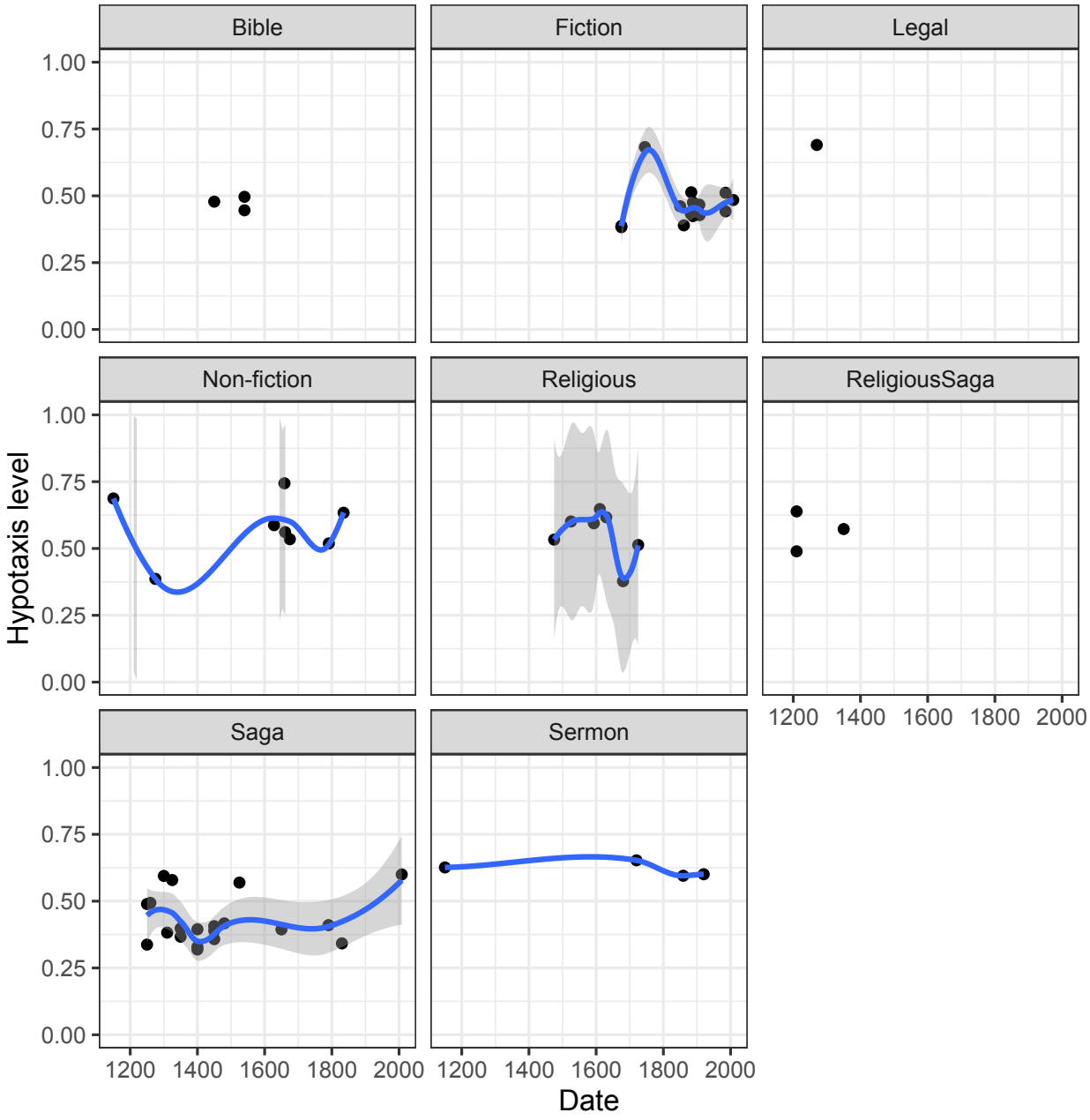


FIGURE 12. Hypotaxis level by genre in Icelandic

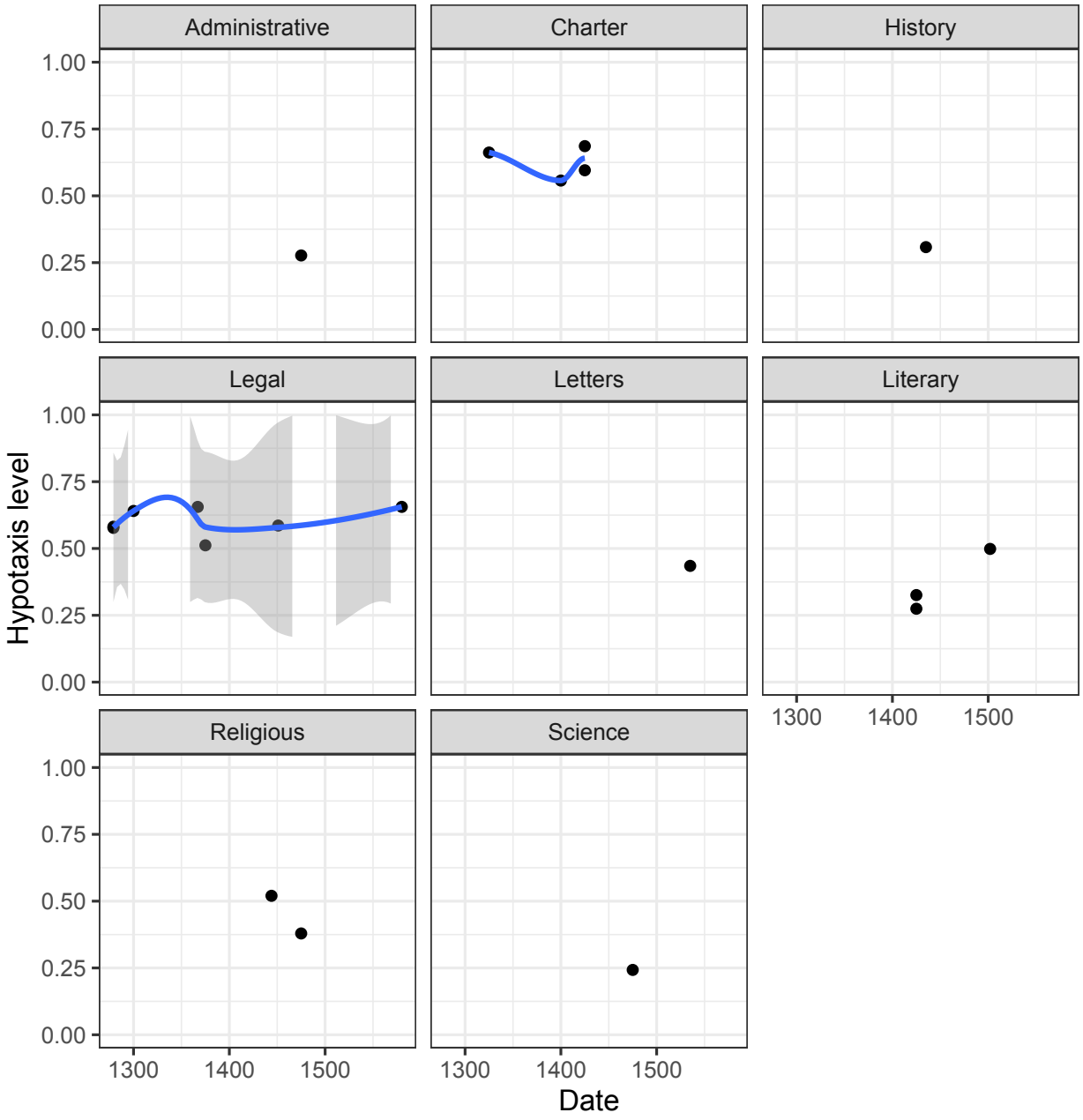


FIGURE 13. Hypotaxis level by genre in Low German

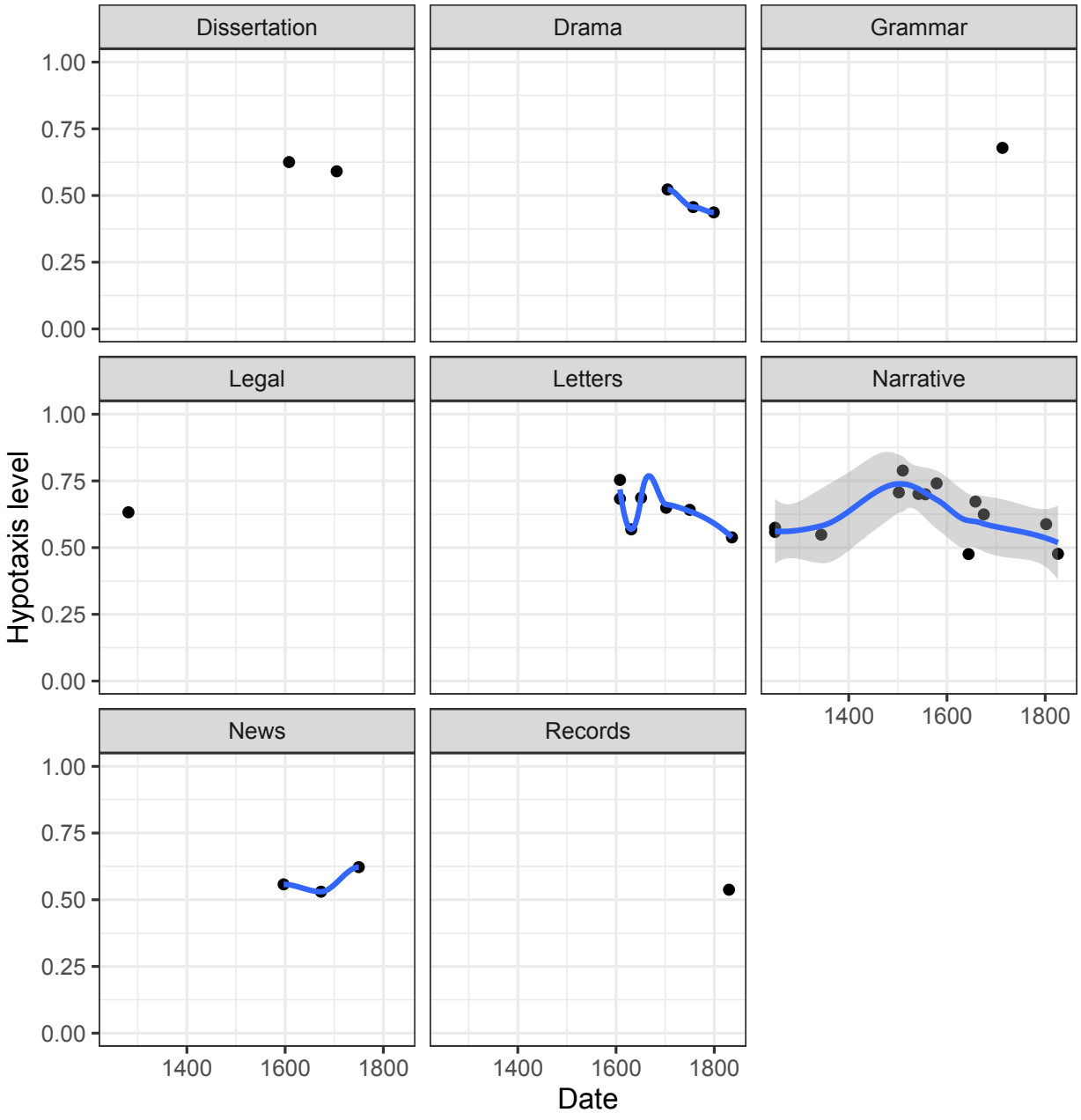


FIGURE 14. Hypotaxis level by genre in Portuguese