Tracing Syntactic Complexity: Exploring the Evolution of Average Dependency Length Across Three Centuries of Scientific English

Marie-Pauline Krielke
Saarland University, Germany
mariepauline.krielke@uni-saarland.de

Diego Alves
Saarland University, Germany
diego.alves@uni-saarland.de

Luigi Talamo
Saarland University, Germany
luigi.talamo@uni-saarland.de

Abstract

We present a diachronic analysis of syntactic change in a corpus of 300+ years (1665 - 1996) of scientific English annotated with Universal Dependencies (UD) and Dependency Length (DL). We trace the development of average Dependency Length (aDL) as a measure of syntactic complexity in scientific English between 1665 and 1996 and describe the corpus construction, and report on the UD annotation evaluation. We find that aDL first decreases toward the 19th c., but then increases significantly in the 20th c. We show that the highly aggregate measure of aDL masks the underlying mechanisms driving shifts in syntactic complexity. A fine-grained analysis of the dependency relations involved in change shows that the increasing use of (multi-word) compounds is a dominant source of long leftward expanded noun phrases, leading to the expansion of syntactic dependencies within and beyond the noun phrase. The results open a new perspective on syntactic complexity, shifting from the sentence to the phrasal level.

1 Introduction

Syntactic change in English in the past 300 years was largely of statistical nature, reflecting shifts in usage rather than structural innovation. These shifts were motivated in response to increasingly changing communicative demands from evolving contexts of usage, such as the advent of new genres like scientific English, beginning with the Scientific Revolution and continuing into the present day.

Syntactic change in the evolution of scientific English has been examined from both qualitative and quantitative perspectives. Ranging from descriptive approaches (e.g. Halliday, 1988; Halliday and Martin, 1993) to early quantitative analyses (Biber and Clark, 2002; Biber and Gray, 2010, 2016; Hundt et al., 2012) report a general shift from subordinate constructions to nominal style. More recent approaches, grounded in cognitive and information-theoretic frameworks, have examined scientific English using measures such as Relative Entropy and Surprisal (e.g. Degaetano-Ortlieb et al., 2016; Bizzoni et al., 2020; Teich et al., 2021), as well as memory-based metrics like Dependency Length (DL, the distance between a syntactic head and its dependent(s)) (Juzek et al., 2020; Krielke, 2024). These studies collectively indicate a long-term trend toward syntactic simplification at the sentence level, characterized by fewer clausal embeddings, increasingly informationally dense noun phrases (NPs), and overall decreasing DL, independent of sentence length.

While these trends are well-documented up to the 19th c., little research has focused on syntactic developments in scientific English throughout the 20th c. Recent findings suggest a marked tendency toward extremely long nominal phrases in this later period, mainly through increased use of premodification strategies such as compounding (Degaetano-Ortlieb, 2021). This phenomenon leads to a pronounced leftward expansion of the NP. Given that aDL is most optimal when syntactic branching is balanced on both sides of the syntactic head (Temperley and Gildea, 2018), we hypothesize that the increasing reliance on leftward expansion in the 20th c. may result in a reversal of the earlier trend, leading to an increase in aDL. This study investigates whether this shift is empirically supported, contributing to a more comprehensive understanding of changes in syntactic complexity in scientific English.

We start by discussing the existing literature on diachronic syntactic change in scientific English and the use of DL as a measure of syntactic complexity (Section 2).

We present our data and methods, i.e. an updated and extended version (+100 years) of the Royal Society Corpus (RSC, Fischer et al., 2020) spanning the years between 1665 and 1996 annotated with Universal Dependencies (UD, de Marneffe et al.,

Proceedings of the Eighth International Conference on Dependency Linguistics (Depling, SyntaxFest 2025), pages 13–23
August 27-28, 2025 ©2025 Association for Computational Linguistics

2021) and DL (Section 3).

Next, we evaluate the dependency annotations using Stanza compared to previous evaluation scores obtained by (Krielke et al., 2022) using UDpipe (Straka and Straková, 2017). Apart from general improvements in the parsing quality by using a more state-of-the-art parser, we test whether parsing accuracy has additionally improved in the last 100 years due to, among other reasons, a notable decrease in average sentence length (SL) over time (i.e. the shorter a sentence, the better the parsing) considering SL impact on the encountered evaluation results (Section 4).

We then analyze the evolution of DL across three centuries. We measure average aDL per SL per 50-year period to gain a first overview of the development of aDL over time. Against common intuition that DL would further decrease over time (i.e. after 1900), we find an increase in aDL normalized by SL in the 20th c. To identify the driving forces behind this upward trend, we analyze the most influential syntactic relations on changes in aDL with SL held stable. The analysis reveals a pivot role of the nominal compounds (Section 5). Section 6 presents a general discussion of the results. We close with conclusions (Section 7) and limitations of this work (Section 8).

2 Related Work

Scientific English has been described to have undergone two notable developments: (i) the creation of specialized terminology (Halliday and Martin, 1993; Wang et al., 2023) and (ii) a shift from clausal to phrasal complexity (Biber and Gray, 2016; Alves et al., 2024), i.e. sentences continuously consist of fewer subclauses and rather long and complex noun phrases (e.g. Halliday, 1988; Halliday and Martin, 1993; Biber and Clark, 2002; Biber and Gray, 2011, 2016; Hundt et al., 2012). This process of syntactic reorganization has been attributed to the need for efficiency achieved by linguistic condensation on the one hand and a reaction to increasing shared expert knowledge that makes it possible to use grammatically implicit encodings (e.g. compounds) instead of explicit ones (e.g. relative clauses) (Biber and Gray, 2010).

The preference for dense nominal structures over intricate subordinate structures is associated with a general trend towards lower grammatical complexity, as it is connected to the hypothesis that aDL minimizes diachronically (cf. Juzek et al., 2020; Krielke, 2024). This assumption is grounded in the Dependency Length Minimization (DLM) Hypothesis, according to which human languages tend to reduce the distance between syntactically related words due to limited working memory capacity and the principle of least effort (Zipf, 1949) and assuming that shorter DL is easier to produce and comprehend (Hawkins, 1994, 2004; Gibson, 2000; Demberg and Keller, 2008).

DLM is widely recognized as a universal property and has been observed across languages (Gildea and Temperley, 2010; Liu et al., 2017; Futrell et al., 2015), across genres (Wang and Liu, 2017), and diachronically (Tily, 2010; Lei and Wen, 2020; Liu et al., 2022). In particular, for scientific English between 1650 and 1900, Juzek et al. (2020) and Krielke (2024) observe a steady reduction in aDL; both papers attribute this trend to a persistent increase in nominal structures, which create rather short dependency relations, and a strong decrease of clausal subordination, which represent rather long dependency relations.

Diachronic research on other genres (e.g., political speeches, Lei and Wen, 2020; Liu et al., 2022) found a similar trend but did not strictly control for SL. Comparing the aDL in scientific and general English, Krielke (2024) in fact showed that aDL is strongly correlated with SL and, when holding SL stable, only scientific texts showed a significant aDL reduction over time. Liu et al. (2022) find general downward trends in their data but looking at specific dependency relations (deprels), they find that nominal relations (attributive adjectives, possessive modifiers, compounds, determiners, etc.) become longer over time while clausal relations become shorter.

Using information theoretic measures, Degaetano-Ortlieb (2021) showed for the 20thc. that especially composite terminology undergoes expansion in comparison to other nominal pre- and postmodification patterns; this, on the one hand, seems to point to structural compression but, on the other hand, might indicate a further expansion of the core nominal group leading to increasing distances between dependency relations on average.

3 Data and Methods

Our corpus contains texts extracted from the *Philosophical Transactions* and *Proceedings* of the Royal Society of London. It represents an exten-

sion (+76 years) of the open version *The Royal Society Corpus 6.0* (Fischer et al., 2020), covering texts until 1920.

On top of sentence splitting and tokenization using TreeTagger (Schmid, 1994) and spelling normalization, using a trained model of VARD (Baron and Rayson, 2008), we enriched the corpus with two types of token-level annotation layers: (i) UD annotation and (ii) dependency length. Layer (i) was annotated using Stanford Stanza v.1.5 with the English combined model pre-trained on five English treebanks (EWT, GUM, GUMReddit, PUD, and Pronouns) from UD v.2.12[1], performing the following NLP tasks: lemmatization, parts-of-speech, morphological features tagging, dependency parsing, and named entity recognition. We chose Stanford Stanza because it natively supports UD, can be implemented as a Python library, and performs slightly better on English data with respect to similar tools, i.e., UDPipe and spaCy (for a comparison, see Qi et al., 2020). Layer (ii) consists of four different values calculated at the sentence level and excluding punctuation: dependency length (DL), sentence length (SL), sum of all (absolute) dependency lengths (sumDL), and average dependency length per sentence (aDL). For example, in the sentence displayed in Figure 1 the head of the nominal compound *brachiopod shell* is *shell* (token id = 14) and since *shell* depends on *calcite* (token id = 10), has a DL of -4 i.e., $10 - 14$; its only modifier, *brachiopod* (token id = 13), has a DL of 1 i.e., $14 - 13$; both compound constituents have an SL of 14 - the sentence contains only a punctuation symbol - a sumDL of 24 i.e., $1 + 1 + 0 + 1 + 2 + 1 + 2 + 2 + 1 + 3 + 3 + 2 + 1 + 4$ and an aDL of 1.84 i.e., $\frac{1+1+0+1+2+1+2+2+1+3+3+2+1+4}{13}$. For our analyses, we bin periods of 50 years, e.g. 1700 = 1701 – 1750, 1750 = 1751 – 1800.

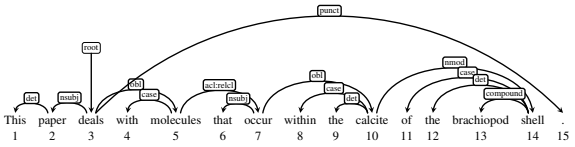


Figure 1: Dependency tree of a sentence with dependency labels and token ids.

[1] https://stanfordnlp.github.io/stanza/combined_models.html

Subcorpus	Sentences	Tokens
1665-1700	42,398	2,195,166
1701-1750	57,915	2,860,083
1751-1800	109,771	5,037,846
1801-1850	169,224	7,004,769
1851-1900	352,259	12,935,866
1901-1950	663,561	21,031,436
1951-1996	2,587,326	78,165,925
All	3,982,454	129,231,091

Table 1: Total number of tokens and sentences per 50-year period.

Subcorpus	Sentences	Tokens	Deprel
1665-1700	20	1093	40
1701-1750	20	1024	40
1751-1800	20	1038	40
1801-1850	20	967	40
1851-1900	20	982	38
1901-1950	20	1011	34
1951-1996	20	789	38
All	140	6904	49

Table 2: Evaluation set description in terms of number of sentences, tokens, and dependency relation tags.

4 Dependency Parsing Evaluation

To evaluate Stanza's parsing performance on RSC texts, we carried out a quality check over 140 sentences randomly selected from the parsed corpus (i.e., 20 from each 50-year period[2]). These sentences were manually corrected by a linguist and a student with previous experience with Universal Dependencies; the cases of disagreement were discussed until an agreement was reached. Then, the sentences parsed with Stanza were compared to the manually corrected ones using the DependAble tool (Choi et al., 2015).

Table 2 presents the overall characteristics of the evaluation set manually corrected in terms of the number of tokens and dependency relation labels.

The model used to parse the RSC corpus is the combined English one from the Stanza repository. When combined, these training sets have 53 deprels. Thus, our evaluation set covers the vast majority of the dependency parsing labels excluding `csubj:outer`, `reparandum`, `list`, and `dislocated`.

[2] The parsed sentences along with the manual corrections can be accessed at https://tinyurl.com/4j5a99dx.

Subcorpus	UAS	LAS
1665-1700	88.29	85.09
1701-1750	90.33	87.21
1751-1800	91.52	89.02
1801-1850	90.49	87.49
1851-1900	88.80	84.93
1901-1950	94.86	92.88
1951-1996	93.92	90.87
All	91.06	88.11

Table 3: UAS and LAS results of the evaluation set.

4.1 Evaluation Results

Table 3 presents the results concerning unlabeled and labeled attachment scores (UAS and LAS respectively).

We notice a tendency of better parsing scores for more recent texts (i.e., from the 20th c.), as one would expect regarding the texts composing the training corpora of the Stanza model. Compared to the similar analysis conducted by (Krielke et al., 2022) where RSC texts from 1665 to 1899 were parsed with UDPipe 1.0 (Straka and Straková, 2017), Stanza seems to provide better parsing scores, an improvement of around 7 points regarding the overall LAS. However, it is important to mention that the number of analyzed sentences concerning the evaluation set is quite small, so further analyses are necessary for a more complete and statistically valid evaluation of the parsed corpus.

As was the case in a previous parsing analysis of the RSC (cf. Krielke et al., 2022), during the manual correction of the new version of the corpus we found that many parsing errors are due to OCR errors and tokenization issues (e.g., "simul taneous" instead of "simultaneous" and "bem" instead of "been"). Furthermore, when equations were part of sentences, they were usually tokenized in random ways and labeled with erroneous deprels.

Besides overall UAS and LAS, the DependAble tool also provides detailed information regarding scores in relation to sentence length (Table 4) and distance between heads and dependents (Tables 5 and 6). These analyses were conducted considering all 140 sentences of the evaluation set.

Sent. Length	≤ 20	≤ 30	≤ 40	≤ 50	≥ 50
No. Sentences	5	16	12	13	21
LAS	91.13	90.49	90.04	87.67	87.26

Table 4: LAS in relation to sentence length.

In terms of SL, we can observe in Table 4 that LAS results tend to deteriorate for longer sentences. Figure 2 shows that SL tends to decrease over time, thus, this may influence the better LAS scores observed in more recent texts.

Dist.	< -5	-5	-4	-3	-2	-1
LAS	81.16	84.85	86.49	93.92	95.52	95.63

Table 5: LAS by distance when the head is after the dependent.

Dist.	1	2	3	4	5	>5
LAS	77.94	83.09	88.04	80.31	70.34	69.99

Table 6: LAS by distance when the head precedes the dependent.

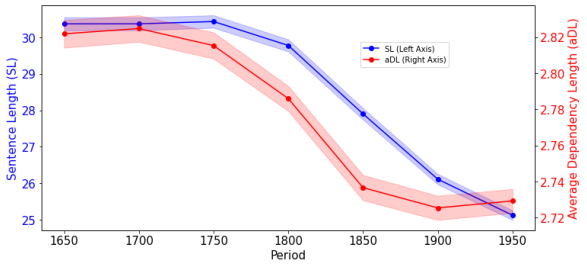


Figure 2: Development of mean SL and mean aDL with 95% CI.

Table 5 shows that the closest the dependent is to the head, the better are LAS results in the cases where the dependent precedes the head in the sentence. However, as shown in Table 6, it is not the case when the dependent follows the head. These counter-intuitive results concerning dependents with a distance of 1 or 2 from their heads can be explained by recurrent OCR and tokenization errors, leading to several occurrences of `goeswith` in the manually annotated evaluation corpus. For example, in one sentence the token *itself* is split into two tokens, *it* and *self*; accordingly, the second token is labeled `goeswith` in the corrected file. However, from distance 3 to >5, the results are as expected, the further the dependent, the lower the LAS.

A qualitative analysis of the dependency parsing evaluation shows that, of the deprels occurring more than 20 times in the evaluation set, the following ones present LAS below 80: `parataxis`, `acl:relcl`, `goeswith`, `acl`, `advcl`, and `conj`. Moreover, we find that `obl`, `cc`, and `conj` tend to have better LAS values in more recent texts.

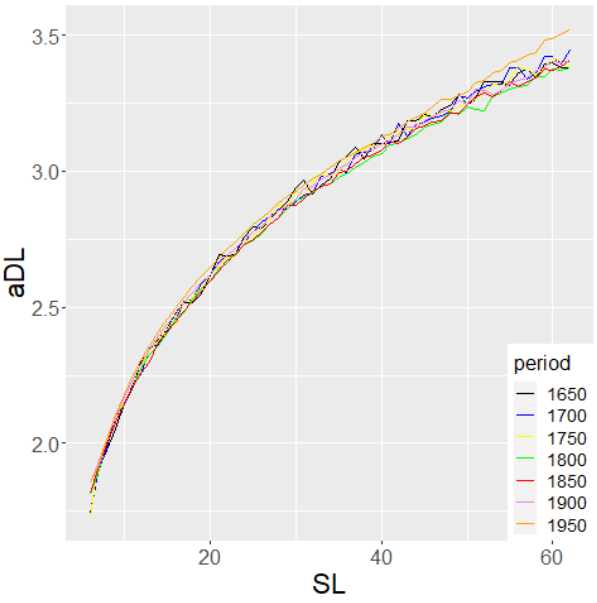


Figure 3: aDL per SL per period.

5 Analysis

We conduct a quantitative analysis concentrating on the development of aDL as a measure of syntactic complexity in scientific English texts over the past 300 years. First, our goal is to verify results from earlier studies, especially whether the use of a different parser on the same data influences the outcome of DL calculations. For this comparison, we refer to the results presented by (Juzek et al., 2020) obtained from Stanford Parser (Klein and Manning, 2003). Second, we are interested in the aDL development in the 20th c., since earlier studies only cover the years until 1900. Regarding changes in the 20th c., we expect to find an increase in aDL due to leftward NP expansion (compare example in Figure 7).

5.1 Dependency Length across periods

aDL overall decreases steadily until 1900 together with SL and increases slightly after 1950 (Figure 2). The trends of both measures are moderately correlated (Pearson Correlation Coefficient: 0.5560, p-value < 0.01). Hence, aDL values without controlling for periodically dominant SLs paint a skewed picture. We calculate mean aDL per SL per 50-year period (Figure 4, (a)) and only display SLs occurring ≤ 250 times (i.e., SL ≥ 6 and SL ≥ 62). 1950 (orange) consistently shows the highest aDL per SL among all periods. Only 1650 (black) sporadically shows peaks surpassing the aDL of the latest period. aDL per SL is lowest in the periods 1800 (green) and 1850 (red).

Since the lines in Figure 3 overlap and are therefore hard to interpret, we conduct independent Two-Sample t-tests to compare the means of aDL per SL between two adjacent periods for each SL and to identify significant differences between our results of aDL per SL per period. Figure 4 shows that aDL in the period 1900 vs. 1950 significantly differs for nearly all SLs. Also, the aDL in the period 1850 vs. 1900 shows significant differences for most (especially the shorter) SLs. Among the earlier periods, only 1750 vs. 1800 shows significant differences for nearly half of all SLs. For the other comparisons, we find only very few significant differences; aDL only seems to differ in these periods sporadically at specific SLs.

The results indicate an SL-independent (for most SLs non-significant) decrease of aDL between the 18th c. and the 19th c., followed by a highly significant increase throughout the 20th c. The development until 1900 is in line with that found by Juzek et al. (2020), who showed that aDL went gradually down between 1650 and 1889. The discovery of an increase in aDL in the 20th c. points to a possible impact of increasingly leftward expansion of nominal phrases.

However, this comparison represents a highly aggregated view of the situation. To observe changes in the full range of aDL values over time and still be able to control for SL, we choose a SL that is strongly represented in all periods. For this, we take the arithmetic mean between the observed SLs (6 - 62): $(6 + 62)/2 = 68/2 = 34$. We thus use SL 34 for further inspection.

5.2 Development of aDL at SL 34

aDL at SL 34 first shows a slight increase between 1650 and 1700, and a downward trend towards the 19th c. In the 20th c., aDL increases steeply and reaches its highest mean value in the period 1950. Independent Two-Sample t-tests with *period* as the predictor and *aDL* as the response variable show a significant decrease between 1750 and 1800 as well a highly significant increase in aDL in the last three periods, i.e. 1850 vs. 1900 and 1900 vs. 1950 (Table 7). To verify whether the restructuring of the NP (decreasing subordination between the 18th and 19thc. and increasing premodification in the 20thc.) is the driving force behind the observed aDL changes, we further analyze the development of individual deprels.

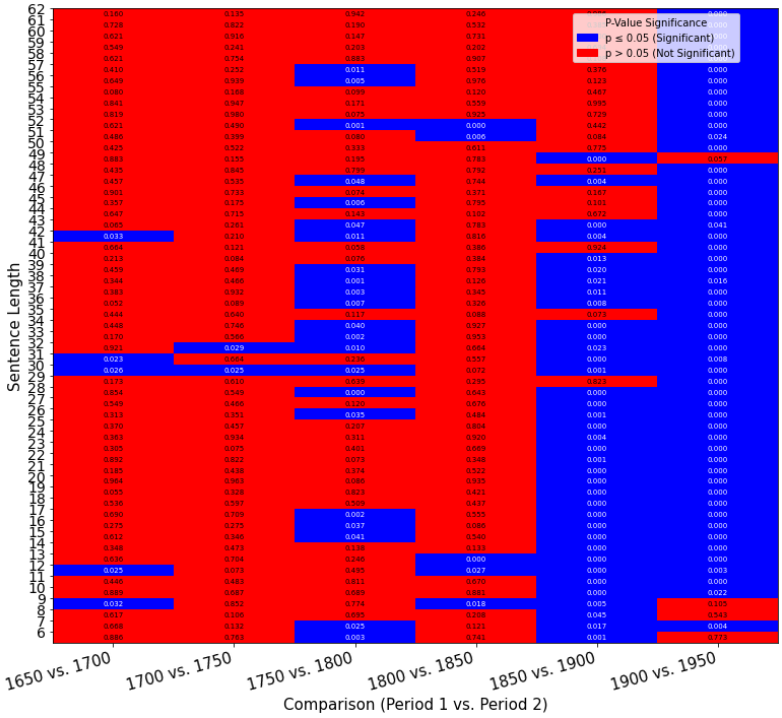


Figure 4: P-values for two-sided t-tests comparing the aDLs per sentence of each SL (6–62) between all adjacent periods.

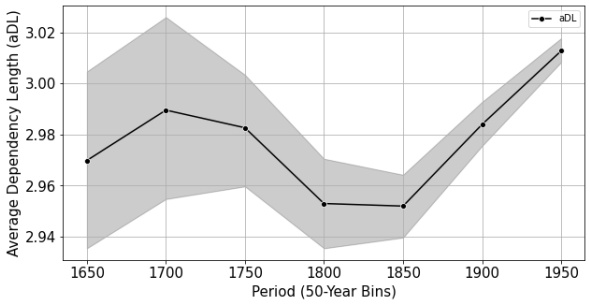


Figure 5: aDL at SL 34 per 50-year period.

Comparison	P-Value	Significant
1650 vs. 1700	0.45	False
1700 vs. 1750	0.74	False
1750 vs. 1800	0.04	True
1800 vs. 1850	0.93	False
1850 vs. 1900	0.00	True
1900 vs. 1950	0.00	True

Table 7: T-Test results for AVD changes in sentence length 34 across time periods

5.3 Individual deprels at SL 34

To detect influential deprels for the decrease in aDL between 1750 and 1800, and the increase in the 20th c., we take a closer look at two factors influencing the overall aDL (holding SL stable at 34) in a period: changes in aDL per deprel per period (x-axis), and changes in fpm (y-axis) per deprel per period (see Figure 6), formalized as in Equation 1 for Δfpm and Equation 2 for ΔaDL.

$$\Delta fpm(\text{deprel}, t_i, t_j) = fpm_{t_i}(\text{deprel}) - fpm_{t_j}(\text{deprel}) \quad (1)$$

$$\Delta aDL(\text{deprel}, t_i, t_j) = aDL_{t_i}(\text{deprel}) - aDL_{t_j}(\text{deprel}) \quad (2)$$

Figures 8a and 8b show deprels occurring $> 10,000$ times per million tokens and display the frequency and aDL differences of each deprel between two periods; the colors indicate the average aDL of each deprel ranging from short-distance (< 3) over mid-distance (> 3) to long-distance functions (> 6).

5.3.1 Comparison 1750 vs. 1800

Between 1800 and 1750 (Figure 8a), we find a notable increase in two noun phrase (NP) premodifying elements: `amod` (attributive adjective) and `det` (determiner), and in two postmodifying elements: `case` (adposition) and `nmod` (nominal modifier). This increase in both NP pre- and postmodifiers indicates that during this period, NPs become increasingly loaded with additional information, while showing an equal distribution of pre- and postmodification on both sides of the nominal head, possibly leading to the observed decrease in aDL.

The most substantial frequency decrease is found for `nsubj` (nominal subject), `cc` (coordinating con-

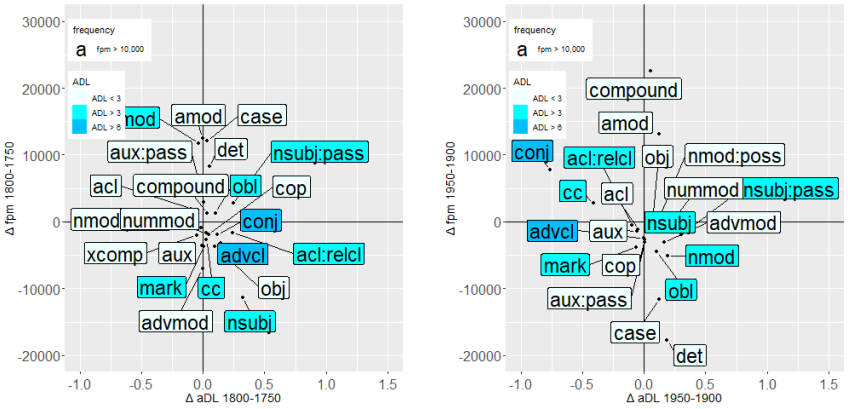


Figure 6: Relative frequency and aDL development of high-frequency deprels between 1750 and 1800 (a) and between 1900 and 1950 (b).

junction), `conj` (conjunct), indicating a decrease in clausal subordination being an additional factor of Dependency Length Minimization.

In terms of aDL, we can observe that almost all deprels become slightly longer or do not change at all over time. The significant decrease in aDL between 1750 and 1800 could be owed to an equal frequency increase in typically short deprels functioning as NP pre- and postmodifying constituents and creating an equal distribution of information on either side of the nominal head. Additionally, longer deprels typically indicating the presence of a subclause, e.g. `mark` (word marking a clause as subordinate to another clause), `advcl` (adverbial clause modifier) or `acl:relcl` (relative clause) drop in frequency, further contributing to the observed aDL reduction.

5.3.2 Comparison 1900 vs. 1950

Comparing 1900 and 1950 (Figure 8b), the most notable difference to 8a is the increase of `compounds`, followed by a rise in `amod`, both indicating leftward expansion of the NP. The strongest decrease in frequency can be observed for determiners (`det`), followed by case markers (`case`) and nominal modifiers (`nmod`), typically indicating an NP postmodifying element (prepositional phrases). The frequency development points to a shift of NP modification towards the left side of the nominal head, confirming our hypothesis that leftward expansion is associated with the aDL increase observed in the 20th c. We observe an increase in aDL in most of the deprels categorized as belonging to *nominals* according to the UD-framework [3]. More specifically, the deprels belong to the group of nominal premodifiers. The aDL increase of these deprels can be explained by the increase in compounds since their presence extends the distance between premodifiers (e.g., determiners, numeric modifiers, adjectives) and the nominal head (as in e.g, *the supersonic convective Mach number shear layer*).

aDL decreases for a few clausal deprels, such as relative clauses (`acl:relcl`) and deprels indicating coordination. Interestingly, conjuncts (`conj`) increase in frequency but become shorter on average. In UD, conjuncts can refer to both nominal coordination (*Mary and John*) and clausal coordination (*John slept and Paul left*). The aDL decrease of conjuncts suggests a shift towards nominal coordination and away from clausal coordination, which is in line with the literature (e.g. Halliday, 1988).

5.3.3 The influence of compounds on aDL

To conclusively verify the influence of compounds on the overall extension of aDL in the 20th c., we recalculate aDL per SL for the last 96 years (1900 - 1996), using the code[4] provided by Dyer (2023), treating compounds as single words vs. compounds treated as multi-word expressions (MWEs, cf. Figure 7).

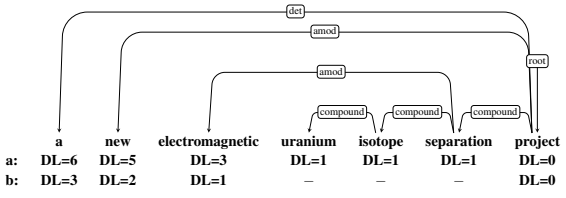


Figure 7: Calculation of DL for compounds as MWEs (a) vs. as single words (b).

[3] https://universaldependencies.org/u/dep/all.html

[4] https://github.com/andidyer/DependencyLengthSurvey

Although SL changes when compounds are counted as single words, the visualization of aDL against SL (Figure 8a) shows that, aDL with compounds as single words drops visibly below the curve of compounds treated as MWEs. This shows that the strong presence of compounds in the 20th c. leads to a substantial increase in aDL.

Compared to earlier periods, this effect is also visible when comparing aDL calculated with compounds as single words: aDL in the 20th c. stays below the aDL in other periods (Figure 8b).

6 Discussion

DL is a well-established measure of syntactic complexity that impacts memory-based processing costs. Our results may suggest that an increase in aDL in the 20th c. makes scientific English increasingly difficult to process. However, this interpretation should be approached with caution.

First, it is important to recognize that sentence length (SL) is the most influential factor in aDL. The overall decreasing trend in both mean SL and mean aDL (without controlling for their interaction) suggests that sentence complexity has primarily decreased due to shorter sentence structures. Shorter sentences are generally easier to process than longer ones.

Second, averaging DL per sentence conceals the distinct factors contributing to this aggregate measure. In our case, subordination and premodification both increase aDL on average per sentence. This reflects a broader principle: Any construction that creates an uneven distribution of syntactic dependents relative to the syntactic head tends to increase DL and thus aDL. However, from a processing standpoint, it remains unclear which of these syntactic configurations is inherently more difficult to process.

From a qualitative perspective, we argue that subordination typically conveys more explicit syntactic relations between constituents (e.g., *a new project that focuses on the separation of electromagnetic uranium isotopes*), which may make such structures more accessible to readers. In contrast, highly premodified noun phrases (e.g., *a new electromagnetic uranium isotope separation project*) may be more compact but cognitively demanding due to covert relations that need to be inferred relying on expert knowledge. While we can assume that experts and non-experts differ widely in processing these constructions, ultimately, behavioral research is required to determine the relative processing difficulty of these configurations.

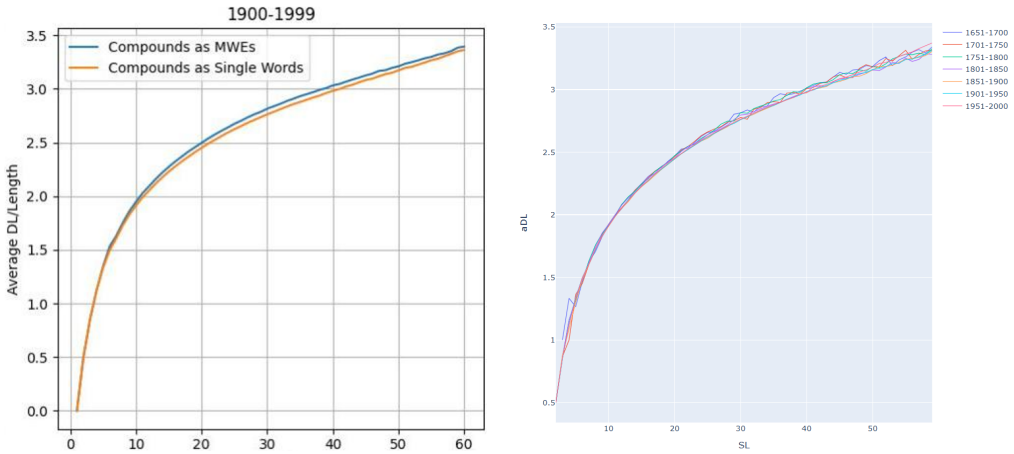
7 Conclusion

We presented a corpus spanning 300+ years of scientific English annotated with UD and DL. We discussed Stanza's accuracy on our historical data, finding increasing accuracy for more recent data, as well as more accurate parses for shorter sentences.

Our analyses show a generally expected decrease in aDL in scientific English until 1900, while scientific texts from the 20th c. display a significant increase in aDL. The initial aDL decrease can be attributed to a decline in subordinate clauses and coordination, as shown by decreasing usage of long (mostly clausal) relations such as relative and adverbial clauses (including their core arguments) as well as coordinate structures becoming much less frequent, paired with an even distribution of information left and right of the nominal head. These results are largely in line with prior observations based on different parsers by Juzek et al. (2020) and Krielke (2024).

The significant increase in aDL in the 20th c. seems to be attributed to a rising usage of compounds as well as attributive adjectives expanding the entire length of the nominal group. It is precisely the leftward expansion with premodifying elements increasing in frequency and length that drive the overall aDL increase observed, which was confirmed by comparing aDL calculated with compounds as MWEs and as single words revealing a substantial increase for compounds as MWEs even when controlling for SL.

Our results reflect a shift of complexity from clausal into phrasal structures. Being perfectly in line with previous work on the development of scientific English (e.g. Biber and Gray, 2011, 2016), the shift towards phrasal complexity is also associated with optimization through densification and Dependency Length Minimization, while clausal complexity is rather associated with longer DL. However, our study has shown that, in fact, both clausal and phrasal complexity ultimately have similar effects on aDL. The expansion effect of extremely premodified noun phrases due to excessive compound usage on the overall aDL represents a valuable insight into the development of scientific English syntax and its implications for aDL. We have thus shown that syntactic complexity can be triggered by different syntactic renderings, i.e. on



(a) (b)

Figure 8: a) aDL per SL in the 20th c. with compounds as MWEs vs. single words. b) aDL per 50-year period calculated with compounds as single words (cf. Figure 7).

the phrasal and the causal level Biber and Gray (2016, p.62), and both affect aDL. Hence, phrasal compression through premodification of the nominal head does not necessarily minimize aDL but can rather lead to aDL expansion if used excessively.

As future work, we intend to go beyond DL to analyze trade-offs in syntactic complexity over time (e.g., tree depth, intervener complexity), and to incorporate measures from constituency-parsed corpora (e.g., average branching factor).

8 Limitations

This study includes only time as a predictor, despite considerable variation in text type, author, and topic. Future work should control for these factors to assess the robustness of temporal effects. Additionally, the analysis is limited to scientific texts; comparing multiple genres would clarify whether the observed trend is genre-specific or indicative of a broader pattern in English.

Acknowledgements

This research is funded by *Deutsche Forschungsgemeinschaft* (DFG, German Research Foundation) – Project-ID 232722074 – SFB 1102.

References

Diego Alves, Stefan Fischer, Stefania Degaetano-Ortlieb, and Elke Teich. 2024. Multi-word expressions in English scientific writing. In *Proceedings of the 8th joint SIGHUM workshop on computational linguistics for cultural heritage, social sciences, humanities and literature (LaTeCH-CLfL 2024)*, pages 67–76.

David Banks. 2008. *The development of scientific writing: Linguistic features and historical context*. Equinox.

Alistair Baron and Paul Rayson. 2008. VARD 2: A tool for dealing with spelling variation in historical corpora. In *Proceedings of the Postgraduate Conference in Corpus Linguistics*.

Douglas Biber and Victoria Clark. 2002. Historical shifts in modification patterns with complex noun phrase structures. *Teresa Fanego, Maria López—Couso and Javier Perez—Guerra (eds.). English Historical Morphology. Selected Papers from*, 11:43–66.

Douglas Biber and Bethany Gray. 2010. Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9(1):2–20.

Douglas Biber and Bethany Gray. 2011. Grammatical change in the noun phrase: the influence of written language use. *English Language and Linguistics*, 15(2):223–250.

Douglas Biber and Bethany Gray. 2016. *Grammatical Complexity in Academic English: Linguistic Change in Writing*. Studies in English Language. Cambridge University Press.

Yuri Bizzoni, Stefania Degaetano-Ortlieb, Peter Fankhauser, and Elke Teich. 2020. Linguistic Variation and Change in 250 years of English Scientific

Writing: A Data-driven Approach. *Frontiers in Artificial Intelligence, section Language and Computation*.

Jinho D Choi, Joel Tetreault, and Amanda Stent. 2015. It depends: Dependency parser comparison using a web-based evaluation tool. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 387–396.

Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, 47(2):255–308.

Stefania Degaetano-Ortlieb. 2021. Measuring informativity: The rise of compounds as informationally dense structures in 20th-century scientific English. In *Corpus-based Approaches to Register Variation*, pages 291–312. John Benjamins Publishing Company.

Stefania Degaetano-Ortlieb, Hannah Kermes, Ashraf Khamis, and Elke Teich. 2016. An information-theoretic approach to modeling diachronic change in scientific English. In Carla Suhr, Terttu Nevalainen, and Irma Taavitsainen, editors, *Selected papers from VARIENG – from data to evidence (d2e)*, Language and computers. Brill.

Vera Demberg and Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2):193–210.

Andrew Thomas Dyer. 2023. Revisiting dependency length and intervener complexity minimisation on a parallel corpus in 35 languages. In *Proceedings of the 5th Workshop on Research in Computational Linguistic Typology and Multilingual NLP*, pages 110–119.

Stefan Fischer, Katrin Menzel, Jörg Knappen, and Elke Teich. 2020. The Royal Society Corpus 6.0 providing 300+ years of scientific writing for humanistic study. In *Proceedings of the conference on language resources and evaluation (LREC)*.

Richard Futrell. 2019. Information-theoretic locality properties of natural language. In *Proceedings of the First Workshop on Quantitative Syntax (Quasy, SyntaxFest 2019)*, pages 2–15, Paris, France. Association for Computational Linguistics.

Richard Futrell, Kyle Mahowald, and Edward Gibson. 2015. Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, 112(33):10336–10341.

Edward Gibson. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1):1–76.

Edward Gibson. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain*, pages 95–126.

Edward Gibson, Richard Futrell, Steven P. Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen, and Roger Levy. 2019. How Efficiency Shapes Human Language. *Trends in Cognitive Sciences*, 23(5):389–407.

Daniel Gildea and David Temperley. 2010. Do Grammars Minimize Dependency Length? *Cognitive Science*, 34(2):286–310.

Kristina Gulordava, Paola Merlo, and Benoit Crabbé. 2015. Dependency length minimisation effects in short spans: a large-scale analysis of adjective placement in complex noun phrases. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 477–482. Association for Computational Linguistics.

Michael A. K. Halliday. 1988. On the language of physical science. In Mohsen Ghadessy, editor, *Registers of written English: Situational factors and linguistic features*, pages 162–177. Pinter.

Michael A.K. Halliday and James R. Martin. 1993. *Writing science: Literacy and discursive power*. Falmer Press.

John A. Hawkins. 1994. *A performance theory of order and constituency*. Cambridge University Press.

John A. Hawkins. 2004. *Efficiency and Complexity in Grammars*. Oxford University Press.

Marianne Hundt, David Denison, and Gerold Schneider. 2012. Relative complexity in scientific discourse. *English Language and Linguistics*, 16(2):209–240.

Tom S Juzek, Marie-Pauline Krielke, and Elke Teich. 2020. Exploring diachronic syntactic shifts with dependency length: the case of scientific English. In *Proceedings of the fourth workshop on universal dependencies (UDW 2020)*, pages 109–119.

Dan Klein and Christopher D Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st annual meeting of the association for computational linguistics*, pages 423–430.

Marie-Pauline Krielke. 2024. Cross-linguistic Dependency Length Minimization in scientific language: Syntactic complexity reduction in English and German in the Late Modern period. *Languages in Contrast*, 24(1):133–163.

Marie-Pauline Krielke, Luigi Talamo, Mahmoud Fawzi, and Jörg Knappen. 2022. Tracing Syntactic Change in the Scientific Genre: Two Universal Dependency-parsed Diachronic Corpora of Scientific English and German. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 4808–4816. European Language Resources Association.

Lei Lei and Ju Wen. 2020. Is dependency distance experiencing a process of minimization? A diachronic study based on the State of the Union addresses. *Lingua*, 239:102762.

Haitao Liu, Chunshan Xu, and Junying Liang. 2017. Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews*, 21:171–193.

Xueying Liu, Haoran Zhu, and Lei Lei. 2022. Dependency distance minimization: a diachronic exploration of the effects of sentence length and dependency types. *Humanities and Social Sciences Communications*, 9(1):1–9.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A Python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*.

Jesús Romero-Barranco. 2020. Linguistic Complexity across Two Early Modern English Scientific Text Types. *Atlantis*, 42(2):50–71.

Helmut Schmid. 1994. Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of the International Conference on New Methods in Language Processing*, pages 44–49.

Milan Straka and Jana Straková. 2017. Tokenizing, pos tagging, lemmatizing and parsing ud 2.0 with UDpipe. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 88–99.

Elke Teich, Peter Fankhauser, Stefania Degaetano-Ortlieb, and Yuri Bizzoni. 2021. Less is More/More Diverse: On The Communicative Utility of Linguistic Conventionalization. *Frontiers in Communication*, 5:142.

David Temperley. 2008. Dependency-length minimization in natural and artificial languages. *Journal of Quantitative Linguistics*, 15(3):256–282.

David Temperley and Daniel Gildea. 2018. Minimizing syntactic dependency lengths: Typological/cognitive universal? 4(1):67–80.

Harry Tily. 2010. The role of processing complexity in word order variation and change.

Gui Wang, Hui Wang, Xinyi Sun, Nan Wang, and Li Wang. 2023. Linguistic complexity in scientific writing: A large-scale diachronic study from 1821 to 1920. *Scientometrics*, 128(1):441–460.

Yaqin Wang and Haitao Liu. 2017. The effects of genre on dependency distance and dependency direction. *Language Sciences*, 59:135–147.

Himanshu Yadav, Ashwini Vaidya, and Samar Husain. 2017. Understanding constraints on non-projectivity using novel measures. In *Proceedings of the fourth international conference on dependency linguistics (depling 2017)*, pages 276–286.

George Kingsley Zipf. 1949. *Human behavior and the principle of least effort: An introduction to human eoclogy*. Addison-Wesley Press.