Exploring diachronic syntactic shifts with dependency length: the case of scientific English

Tom S Juzek
Saarland University
Saarbrücken, Germany
tom.juzek@uni-saarland.de

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

Elke Teich
Saarland University
Saarbrücken, Germany
e.teich@mx.uni-saarland.de

Abstract

We report on an application of universal dependencies for the study of diachronic shifts in syntactic usage patterns. Our focus is on the evolution of Scientific English in the Late Modern English period (ca. 1700-1900). Our data set is the Royal Society Corpus (RSC), comprising the full set of publications of the Royal Society of London between 1665 and 1996. Our starting assumption is that over time, Scientific English develops specific syntactic choice preferences that increase efficiency in (expert-to-expert) communication. The specific hypothesis we pursue in this paper is that changing syntactic choice preferences lead to greater dependency locality/dependency length minimization, which is associated with positive effects for the efficiency of human as well as computational linguistic processing. As a basis for our measurements, we parsed the RSC using Stanford CoreNLP. Overall, we observe a decrease in dependency length, with long dependency structures becoming less frequent and short dependency structures becoming more frequent over time, thus marking an overall push towards greater communicative efficiency.

1 Introduction

It is obvious that language use changes dynamically due to external pressures, e.g. new vocabulary emerging continuously, but what is less reflected upon is how despite these pressures language remains communicatively intact. There is accumulating evidence of the regulatory function of grammar in the process of changing language use, helping retain efficiency in communication. Specifically, grammatical variation in word order, taxis and syntactic embedding is an important means to control linguistic complexity and is thus instrumental in levelling out processing effort (see e.g. Hawkins (1994); Hawkins (2004); Hawkins (2014)).

One approach to capture syntactic complexity with a view to cognitive processing is dependency locality theory (Gibson, 2000). This theory says that languages try to minimize syntactic dependencies in order to manage processing effort by reducing the time in which linguistic elements are held in working memory, e.g. by modulating word order (cf. Hahn et al. (2020)). For instance, Temperley (2007) shows for English that syntactic choices exhibit a preference for structures with shorter dependencies on the basis of the WSJ and Brown corpora and Futrell et al. (2015) provide cross-linguistic evidence of dependency length minimization. Recently, there is an increasing interest in dependency length minimization as a diachronic optimization process affecting the language system in general (e.g. Tily (2010); Gulordava and Merlo (2015)) or specific sublanguages/registers in particular (e.g. Lei and Wen (2020)).

In our research, we are taking a diachronic perspective looking for traces of optimization of syntactic complexity by dependency length minimization in the development of Scientific English from the mid 17th century onward. Our overarching hypothesis is that the changes in grammatical choice preferences can, at least partially, be explained by dependency length minimization. Such changes include for instance the shift from paratactic clause complexes to single clauses with complex NPs; see examples (1) and (2) below, both taken from the Royal Society Corpus (Fischer et al., 2020).

(1) *...I found not only one hollowness, but as often as I cut the Nerve asunder, the hollowness still continued therein, and I found not only one cavity...* (1674)

Proceedings of the Fourth Workshop on Universal Dependencies (UDW 2020), pages 109–119
Barcelona, Spain (Online), December 13, 2020

(2) *In contrast to the complete and temporary visual motion blindness which occurs during stimulation of V5, a less-prominent interference with the perception of visual motion occurs at 70–80 ms after the onset of the visual stimulus when TMS is applied to V1.* (1992)

The two examples illustrate two kinds of syntactic organization associated with orality and informal prose (example 1) and writtenness and informativeness (example 2). Example (1) is a historical example illustrating a paratactic clause complex, marked by the coordinating conjunctions *but* and *and*; example (2) is a contemporary example that is syntactically characterized by a simple structure at clause level (*X occurs (at Y)*). Its complexity is in the nominal phrases, which exhibit complex premodification with multiple adjectives as well as compounds (*temporary visual motion blindness*), postmodification with relative clauses (... *motion blindness* WHICH *occurs during stimulation of V5*) and prepositional phrases (*less-prominent interference* WITH *the perception* OF *visual motion*).

In this light, assuming that shorter dependencies result in more efficient code, our overarching research question is whether this is borne out diachronically for Scientific English: If the structures that become typical (more frequent) of Scientific English over time are associated with greater dependency locality, then we may conclude that the sublanguage of Scientific English develops a more efficient code. Specifically, we expect dependency length minimization to arise through a shift from cross-clausal, tactic relations to a preference of single-clause structures with (complex) nominal phrases.

The remainder of the paper is organized as follows. Section 2 discusses more related work. Section 3 provides details on the data set (Royal Society Corpus), Universal Dependency parsing and the measure of dependency length. In Section 4 we present our results, including a discussion of the kinds of syntactic structures involved in dependency length minimization. Section 5 concludes the paper with a summary and sketch of follow-up studies.

2 Related work

There is a good deal of research on Scientific English from a synchronic as well as diachronic perspective. From a diachronic perspective, important descriptive work is found in Halliday (1988) and Halliday and Martin (1993) and corpus-based analyses are pursued e.g. by Biber and Gray (2011); Biber and Gray (2016). Existing works converge on the observation of a basic linguistic shift to heavy noun phrases and simpler clause structure resulting in higher lexical density and simpler sentence structures, such as single relational or passive clauses (cf. also Atkinson (1999), Banks (2008)). More recently, Degaetano-Ortlieb and Teich (2019) forward the hypothesis that Scientific English, as it evolves, develops towards an optimal code for communication. Using information-theoretic measures (relative entropy, average surprisal), they find that over time Scientific English becomes more distinct from "general language", due to a drift towards nominal style and distinctive syntactic usage at clause level. These insights from previous work build our starting point.

The operationalization of syntactic complexity we draw upon is dependency length, assuming along with others that shorter dependency length makes grammars more efficient, both computationally as well as in human language processing (Hahn et al., 2020). Dependency locality theory (Gibson, 1998; Gibson, 2000) claims that dependencies between words should be as short as possible rendering sentence processing more efficient through a lower cost in working memory in incremental comprehension and production (Gibson et al., 2019). Vice versa, experimental studies have shown that long dependencies lead to comprehension difficulty (Grodner and Gibson, 2005; Bartek et al., 2011). Large-scale corpus studies show this to be true across languages (Liu et al., 2017; Futrell et al., 2015) with only small cross-linguistic differences (e.g. Gildea and Temperley (2010) comparing English and German). Several other corpus studies find dependency locality driving clause level word order (Liu, 2008; Gildea and Temperley, 2010; Wasow, 2002; Rijkhoff, 1990). Gulordava et al. (2015) show that dependency locality is involved in noun phrase structuring and Wasow (2002) illustrates its relevance for word order in the postverbal region. For a recent, comprehensive review of research on dependency length see also Temperley and Gildea (2018). Particularly relevant in our context are studies of English that inspect a range of syntactic phenomena, such as Rajkumar et al. (2016), Temperley (2007) or Levshina (2019). Also, studies that highlight dependency length minimization from a diachronic perspective, such as Gulordava et al. (2015) or Lei and Wen (2020) are directly relevant. Here, we pursue similar goals as Levshina (2019), taking note of the specific syntactic phenomena marked by relatively short dependency lengths that occur with high frequencies and thus contribute a lot to average dependency length minimization and overall efficiency as well as those with relatively large dependency lengths that are highly frequent and therefore have a negative impact on efficiency. Correspondingly, low frequency phenomena with short/large dependency lengths will not have a big (positive or negative) impact overall.

3 Data and Methods

3.1 Corpora

Royal Society Corpus. The Royal Society Corpus (RSC) is a diachronic corpus of Scientific English covering the period from 1665 until 1996 drawn from the *Philosophical Transactions* and *Proceedings* of the Royal Society of London (Fischer et al., 2020). In total, the corpus contains 295,895,749 tokens in 47,837 texts. The corpus comes tokenized, lemmatized and PoS-tagged. Here, we use RSC version 4.0, which covers the first 200 years of publications of the Royal Society (1665-1869; ca. 32 million tokens), roughly corresponding to the late Modern period.[1]

Penn Parsed Corpus of Modern British English. For comparative purposes, we employ the Penn Parsed Corpus of Modern British English (PPCMBE). PPCMBE consists of prose texts from 1700 to 1914 and contains about 1 million tokens (Kroch et al., 2010) covering a (unbalanced) mix of genres and registers[2] While much smaller than the RSC, it roughly spans the same time period. We use PPCMBE to be able to check whether the observed tendencies in dependency length development are specific to scientific language or also hold for language use more broadly in the given time period.

We start from the assumption that in the course of becoming a language for experts, scientific writing is under more pressure to develop an efficient code for communication than many other domains or genres, e.g. literary texts or travel reports. On the other hand, in the given time period, literacy becomes more wide spread and general writing standards develop that will reflect an overall communicative optimization for written language (McIntosh, 1998).

3.2 Universal Dependencies parsing

For the analysis of dependency lengths, we employ the framework of Universal Dependencies (UD), which expresses syntactic relations through dependencies: each element depends on another element, its head (Nivre et al., 2019). In UD, in contrast to most other dependency frameworks, the head is taken to be the semantically salient element and the dependent modifies the head. The top-level head is the root of a sequence, which is typically the main verb of the matrix clause. For instance, in the sentence *The great sailor is waiving*, *the* and *great* modify and depend on *sailor*, while *sailor* and *is* modify and depend on *waiving*, which is the root of the sentence. Figure 1, created with spaCy's visualizer[3], illustrates a dependency analysis of this example. The UD framework aims to be universal, i.e. suitable for all of the world's languages, and there are a great number of resources and tools available.[4] Importantly, UD parsing labels nodes with syntactic functions such as *nominal subject, adverbial modifier*, etc. This is crucial for exploring the functions that are associated with dependency length minimization over time (see Section 4).

We use the Stanford Parser (Klein and Manning, 2003) with out-of-the-box settings to parse our corpora. To reduce end-of-sentence errors, we collected and escaped additional abbreviations in the RSC. An evaluation of 1500 head tags taken from randomly sampled sentences gives an accuracy rate of 91.1% for data from 1665 to 1699 and 91.5% for 1850 to 1869. We also analysed parsing accuracy by sentence length. An analysis of 1500 tokens shows that accuracy for sentences with 13 tokens comes out at 91.4%, at 91.6% for sentences with 26 tokens (the median sentence length in the RSC), and at 92.6% for sentences with 52 tokens. Sentences with less than three tokens were excluded, as they typically lack a verb

[1] `https://fedora.clarin-d.uni-saarland.de/rsc_v4/URLRSC4.0`

[2] `https://www.ling.upenn.edu/ppche/ppche-release-2016/PPCMBE2-RELEASE-1/`

[3] `https://spacy.io/`

[4] See `https://universaldependencies.org/tools.html` for details.

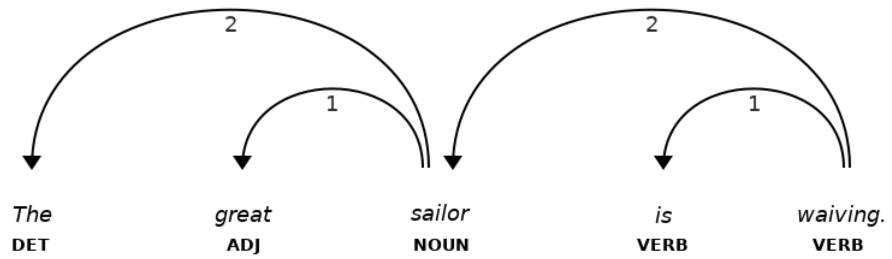

Figure 1: Visualising a simple sequence in the Universal Dependencies framework. The edges represent a dependency relation pointing from head to dependent; the numbers denote dependency distances between tokens/nodes. The distance is measured in number of tokens, starting at 1.

and thus a root. Punctuation was included for parsing but excluded for the analyses below, by excluding tokens with the dependency tag *punct*. We put both the RSC and PPCMBE through the same pipeline, to ensure comparability.

3.3 Dependency measure

We measure the dependency length of sentences summing all the words' distances to their heads averaged by sentence length. The measure is taken from Futrell et al. (2015) and Gibson et al. (2019) and builds on work by Liu (2008). The sum of distances is measured in word/token distance between a head and its dependent, with a minimal distance of 1. The formula is as follows.

$$\sum d_{s(n)} = |t_1.id - t_2.hd| + |t_2.id - t_2.hd| + ... + |t_n.id - t_n.hd| \quad (3)$$

For any sentence *s* of length *n*, the distances for all tokens t_1 to t_n are summed up, where a distance is calculated between a token's position, *t.id*, and its head *t.hd*. For illustration, consider the example in Figure 1. The dependency distances between two tokens/nodes are marked by the numerals below a dependency edge. For instance, the distance from *the* to its head *sailor* is 2; the sum of distances for the sentence is 6 (2+1+2+1). We average the summed distances per sentence length, which gives us the *averaged summed distances* per sentence length.

The value of the measure comes from the fact that it can be interpreted as a proxy of (cumulative) processing difficulty/complexity of a sequence, where long distance structures are regarded as more complex than structures with shorter distances. The analysis of particle verbs illustrate this point: In *Mary picked the guy that we met the other day up*, the particle *up* has a long distance to its head *picked* and the sequence is relatively hard to process. This contrasts to *Mary picked up the guy that we met the other day*, where the dependency between the particle and its head is short – which reduces cognitive load.

4 Analyses and results

We here present the results of the overall diachronic tendency regarding dependency lengths (Section 4.1), followed by a micro-analysis of the types of syntactic structure that contribute to dependency length minimization (Section 4.2).

4.1 Results for dependency length

Figure 2 illustrates the results for the averaged sums of dependency length by sentence length and 50-year periods. For ease of interpretation, the measure is normalised by sentence length. For the RSC, we observe a tendency towards lower dependency length over time, pointing to greater efficiency overall. Comparing the first larger period (1665-1699) with the latest (1850-1869), the difference is highly significant with $p \leq 0.01$. The trend for PPCMBE is not significant ($p > 0.05$). Recall that PPCMBE, while containing scientific texts as well, is a mixed-register corpus. On average, dependency lengths are slightly higher in the RSC, Figure 2 (left), in comparison to PPCMBE, Figure 2 (right), the difference being significant at $p \leq 0.01$.

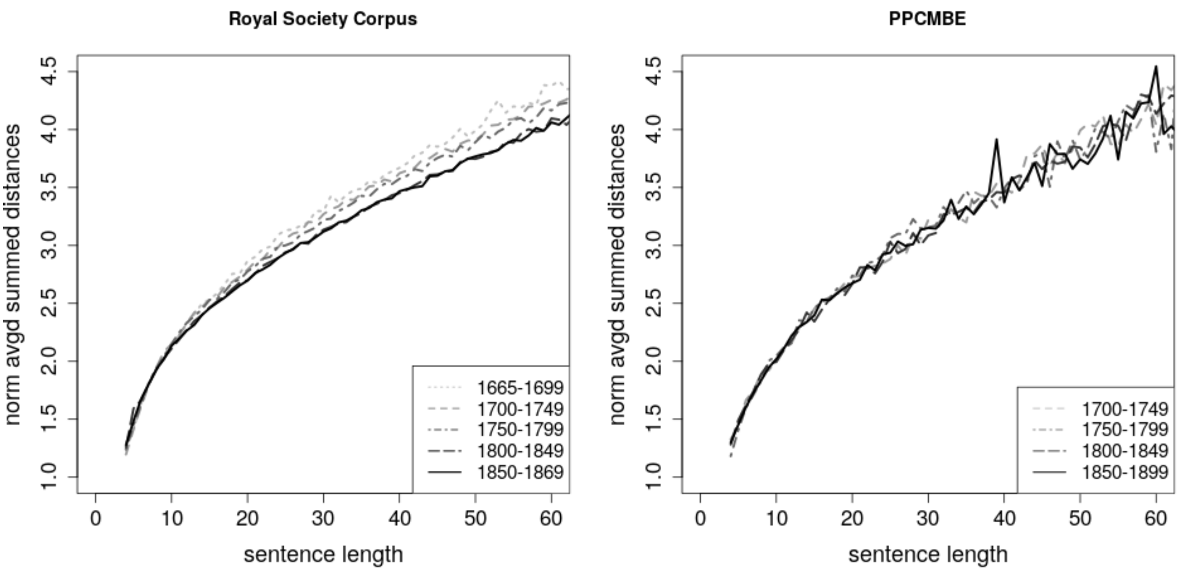

Figure 2: Averaged summed dependency lengths (y-axis) per sentence length (x-axis) over time for the RSC (left) and for the PPCMBE (right). The lengths are normalised by sentence length.

4.2 Types of syntactic functions marked by dependency length

In addition to the general trends, we are interested in the syntactic phenomena involved in dependency length minimization in scientific language. For this, we focus on sentences of 26 tokens in length (the median sentence length in the RSC) assuming that this length is representative of the majority of the data contained in our corpus. For comparison we replicate our analyses on parses of another, randomly chosen and non-extreme sentence length of 52 tokens. Considering the average sum of dependency lengths per 50 years (see Figure 3), we see a decreasing general trend for both sentence lengths. Sentences with 26 tokens decrease in their sum of dependencies from 81.6 in the first period (1650-1700) to 77.1 in the period between 1800-1850, then slightly rising again to 77.2. Sentences with 52 tokens fall from 249 to 239.

To find out what this dependency length reduction is owed to, we look at average dependency length of each syntactic function defined by the UD annotation guidelines[5] with a frequency of at least 1000 per million tokens, altogether 31 different functions. Consider Figure 4 showing all functions[6] and their average dependency length, ranging between 1 (for multi word expressions and phrasal verb particles) and 10.8 (for parataxis). We analyse the relative frequency of each function in 1850 compared to 1650 in relation to their average dependency length, finding that functions with dependency length > 5 decrease in frequency (see Figure 4: dark bars show decreasing frequency, light grey bars increasing frequency). Functions responsible for long dependencies are conjuncts (Example 4), adverbial clauses (Example 5), clausal complements (Example 6) relative clauses (Example 7) and parataxis (Example 8), which is plausible, since they involve cross-clausal dependencies (a dependency between the sentence root and the main verb of the subordinate clause). Conjuncts, however are a special case, since they can represent coordinate structures between lexical, phrasal or clausal categories. An equivalent development is shown for sentence length 52 with appositions (as one more nominal function) becoming more frequent over time (see Figure 7 in the appendix).

[5] `https://universaldependencies.org/guidelines.html`

[6] We exclude the functions punct and ROOT.

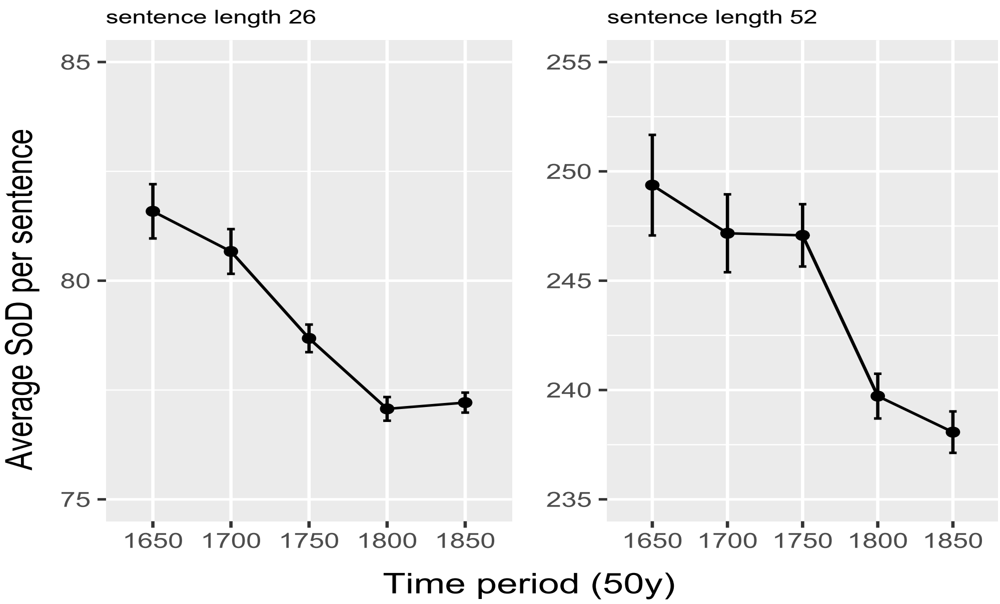

Figure 3: Average sum of distances per sentence of 26 (left) and 52 (right) tokens length by 50 years.

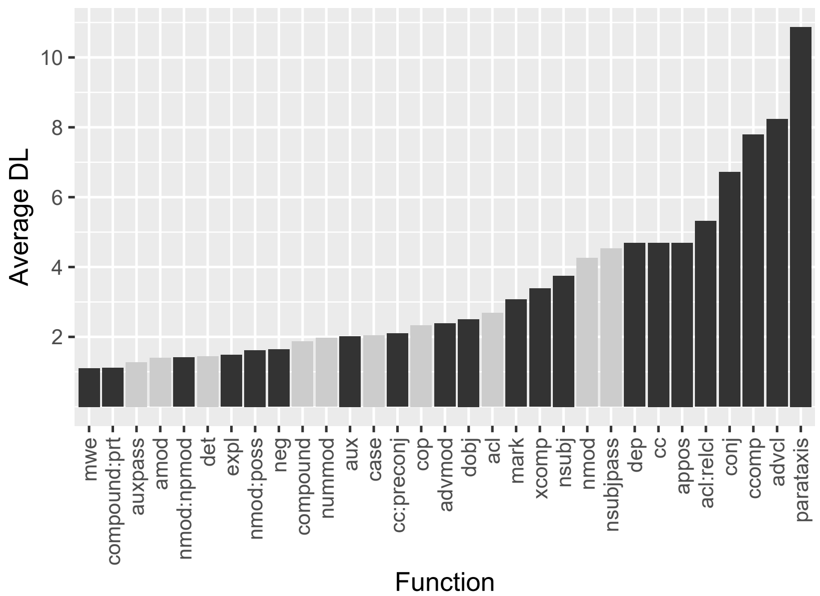

Figure 4: Syntactic functions and their average dependency length. Black denotes decreasing frequency of a function over time, light grey marks increasing frequency.

(4) *When they have done, they **lick*** (ROOT) *their fingers, and as often as they have a hot dish, they **wash*** (conj) *their hands afresh.* (1699)

(5) *I shall **give*** (ROOT) *you an Abstract of the Accounts I received in Answer to these Letters, after I have **described*** (advcl) *our Observations at Edinburgh.* (1737)

(6) ***I think*** (ROOT) *it may be inferred, therefore, that when other vegetable and animal substances are similarly treated they will also **yield*** (ccomp) *analogous results.* (1850)

(7) *Indeed, we find many **Plants*** (head) *mentioned by the same Author, which either are not **known*** (acl:relcl) *to us at this present, or neglected.* (1671)

(8) *Thus this poor creature* ***lived*** (ROOT) *without any other considerable complaint above thirty years, the most remarkable circumstance, I* ***think*** (parataxis)*, in her case.* (1714)

Looking at the diachronic development of these long distance functions (Figure 5 left), the longest dependencies are formed by parataxis. Overall, dependency length does not vary significantly over time ($p > 0.9$). The relative frequencies (see Figure 5 right), however, show that all long functions decrease significantly ($p < 0.05$). The same is observed for the longer sentences (see Figure 8 left in the appendix). Over time, the highly frequent conjuncts decrease remarkably. The less frequent functions (adverbial and relative clauses, parataxis and clausal complements) instead, decrease to a smaller extent making conjuncts the principal suspects to contribute to a shorter dependency length on average in the later time periods.

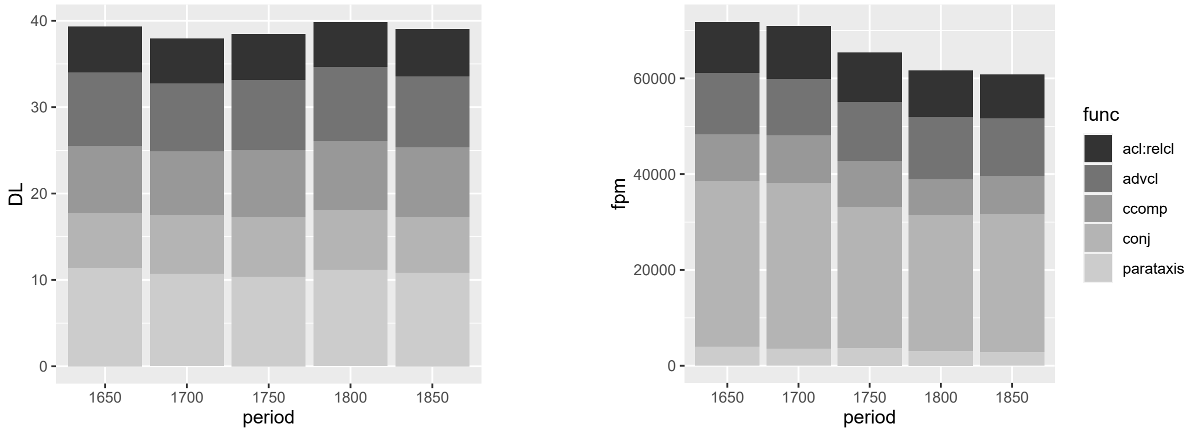

Figure 5: Development of **average dependency lengths** (left) and **fpm** (right) of long distance functions by 50 years.

Looking at the syntactic functions with the shortest dependencies (dependency length between 1 and 2) confirms the intuition that these are relationships with very fixed intra-clausal and intra-phrasal dependency relations (e.g. determiners). They are not flexible in their syntactic placement and can therefore hardly undergo dependency length minimization. We split these functions into nominal (see Figure 6 left) and verbal (see Figure 6 right) modifiers for further inspection, to see whether they indicate the assumed development from an oral/verbal to a written/nominal style, tracking their relative frequencies over time. We exclude multi word expressions since they are known to be tricky relations and not strictly classifiable into either of the groups above. The short nominal modifiers show significant variation over time ($p < 0.05$). Determiners being by far the most frequent function, indeed increase in frequency over time. This corroborates the results of Degaetano-Ortlieb and Teich (2019), ranking the definite determiner as the most distinctive feature over time in the RSC and a clear pointer to nominal style. Adjectival, numeric modifiers and compounds become more frequent over time as well, supporting the expected increasing tendency towards complex premodified noun phrases and again pointing to an expansion of the noun phrase overall. Possessive modifiers as well as the extremely infrequent noun phrases as adverbial modifiers (nmod:npmod, < 2000 per 1 million tokens), however, decrease over time in frequency. In Figure 8 (appendix) we show that the same is true for sentences with 52 tokens with the differences being significant as well. Relative frequency of verbal modifiers decreases in all cases but for passive auxiliaries, indicating the complementary tendency of a dispreference for verbal structures and a preference for a simple clause level structure, as illustrated in Example (2). The comparison between parses of sentences with length 26 and 52 has shown that the observed trends do not only hold for one single sentence length but seem to represent a trend for sentences of relatively common sentence lengths.

5 Summary and conclusions

Using the framework of Universal Dependencies and Dependency Locality Theory, we studied the diachronic development of dependency lengths in a diachronic corpus of Scientific English in the Late

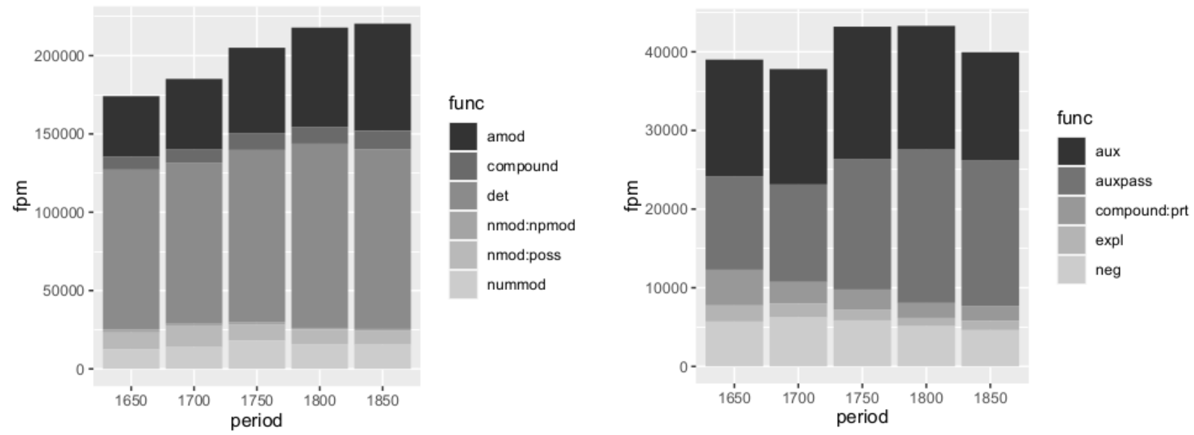

Figure 6: FPM of **nominal** short dependency functions (left) and of **verbal** short dependency functions (right) by 50 years.

Modern period. Our starting assumption was that Scientific English develops a more efficient code over time reflected in a statistical preference of syntactic structures with shorter dependency lengths over time. We found that overall, the average sum of dependencies per sentence goes down over time, if only slightly, and the relative frequency of longer dependencies decreases and particular shorter dependencies become more frequent, especially nominal premodification. This confirms the observed trend towards a nominal, informational style (cf. Section 1). Comparing the trends in sentences of 26 tokens length with sentences of the double length we found that the above mentioned trends seem to generalize across different sentence lengths. In future work we plan to extend this comparison to further (more extreme) sentence lengths to see whether the trends are stable.

Regarding the utility of the UD framework for the present purpose of analysis, the crucial aspect was the labeling with functions, which we found to be descriptively adequate overall with the exception of parataxis and coordination which are notorious problems in dependency parsing (cf. Ahrenberg (2019)). This impacted negatively on our analysis, since we could not interpret the values for the parataxis function.

In our ongoing research, we apply other measures of syntactic complexity, including *dependency depth*, to capture shifts in complexity in terms of syntactic embedding. Recall that we observe nominal phrases becoming more complex over time by multiple pre- and postmodification. This cannot be captured with dependency length alone. Also, we are looking to complement dependency-based analyses of complexity with information-theoretic measures, such as surprisal. For instance, we have estimated the average surprisal of nouns and verbs on a subsample of the RSC and found that while the information content of nouns vs. verbs stays fairly stable over time, nouns have higher surprisal (>9 bits) than verbs (<8 bits) on average. If nouns are generally higher in informativity than other lexical words, then we have additional evidence that "nominal style" can be explained on the grounds of efficient communication.

Acknowledgments

This work is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project-ID 232722074 - SFB 1102. We thank the anonymous reviewers for their detailed comments and suggestions for improving the paper.

References

Lars Ahrenberg. 2019. Towards an adequate account of parataxis in universal dependencies. In *Proceedings of the Universal Dependencies Workshop 2019*, Paris.

Dwight Atkinson. 1999. *Scientific discourse in sociohistorical context: The Philosophical Transactions of the Royal Society of London, 1675-1975*. Erlbaum, New York.

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

Brian Bartek, Richard L. Lewis, Shravan Vasishth, and Mason R. Smith. 2011. In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5):1178.

Douglas Biber and Bethany Gray. 2011. The historical shift of scientific academic prose in English towards less explicit styles of expression. *Researching specialized languages*, 47:11.

Douglas Biber and Bethany Gray. 2016. *Grammatical complexity in academic English: Linguistic change in writing*. Studies in English Language. Cambridge University Press.

Stefania Degaetano-Ortlieb and Elke Teich. 2019. Toward an optimal code for communication: The case of scientific English. *Corpus Linguistics and Linguistic Theory (open access)*, pages 1–33.

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)*. ELRA.

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, Richard Futrell, Steven Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen, and Roger Levy. 2019. How efficiency shapes human language. *Trends in Cognitive Sciences*, 23(5):389–407.

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*, 2000:95–126.

Daniel Gildea and David Temperley. 2010. Do grammars minimize dependency length? *Cognitive Science*, 34(2):286–310.

Daniel Grodner and Edward Gibson. 2005. Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive science*, 29(2):261–290.

Kristina Gulordava and Paola Merlo. 2015. Diachronic trends in word order freedom and dependency length in dependency-annotated corpora of Latin and Ancient Greek. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, pages 121–130, Uppsala, Sweden. Uppsala University.

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.

Michael Hahn, Dan Jurafsky, and Richard Futrell. 2020. Universals of word order reflect optimization of grammars for efficient communication. *Proceedings of the National Academy of Sciences*, 117(5):2347–2353.

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

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, London.

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.

John A. Hawkins. 2014. *Cross-linguistic variation and efficiency*. OUP Oxford.

Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, ACL '03, pages 423–430, Stroudsburg, PA, USA. Association for Computational Linguistics.

Anthony Kroch, Beatrice Santorini, and Ariel Diertani. 2010. Penn Parsed Corpus of Modern British English.

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.

Natalia Levshina. 2019. Universal dependencies in a galaxy far, far away... what makes Yoda's English truly alien. In *Proceedings of the Third Workshop on Universal Dependencies (UDW, SyntaxFest 2019)*, pages 35–45, Paris, France, August. Association for Computational Linguistics.

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.

Haitao Liu. 2008. Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, 9(2):159–191.

Carey McIntosh. 1998. *The evolution of English prose, 1700-1800: style, politeness and print culture*. Cambridge University Press.

Joakim Nivre, Mitchell Abrams, and Željko Agić. 2019. Universal Dependencies 2.4. *LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University*.

Rajakrishnan Rajkumar, Marten Van Schijndel, Michael White, and William Schuler. 2016. Investigating locality effects and surprisal in written English syntactic choice phenomena. *Cognition*, 155:204–232.

Jan Rijkhoff. 1990. Explaining word order in the noun phrase. *Linguistics*, 28(1):5–42.

David Temperley and Daniel Gildea. 2018. Minimizing syntactic dependency lengths: typological/cognitive universal? *Annual Review of Linguistics*, 4:67–80.

David Temperley. 2007. Minimization of dependency length in written English. *Cognition*, 105(2):300–333.

Harry Tily. 2010. *The role of processing complexity in word order variation and change*. Ph.D. thesis, Stanford University.

Thomas Wasow. 2002. *Postverbal behavior*. CSLI Stanford, CA.

A Appendix

A.1 Comparison with Sentence Length 52

In the following we replicate the figures provided in Section 4.2 for parses of sentences with a 52 tokens length.

A.1.1 Average DL and frequency trend

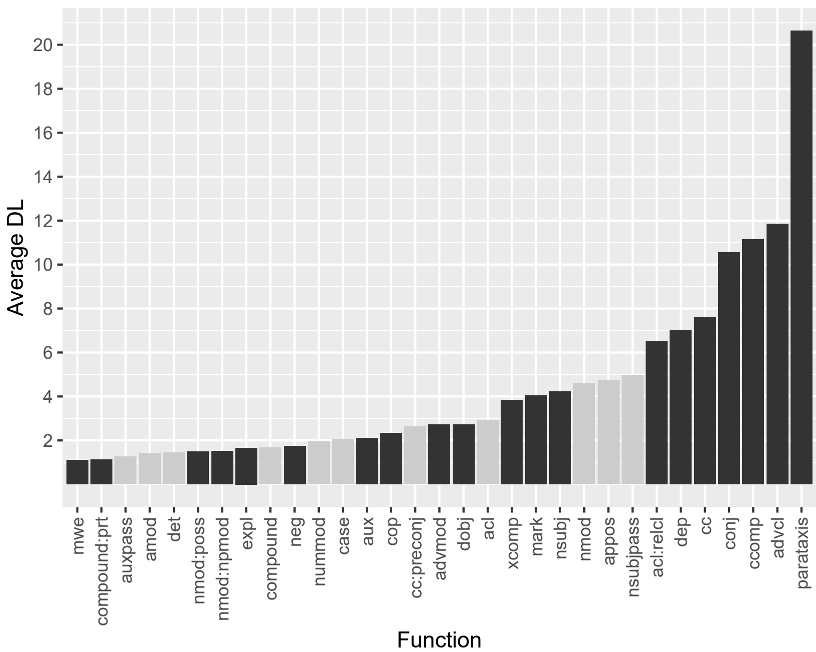

Figure 7: Syntactic functions and their average dependency length (SL 52). Black denotes decreasing frequency of a function over time, light grey marks increasing frequency.

A.1.2 Frequency of Long Distance Functions v. Short Distance Nominal Functions

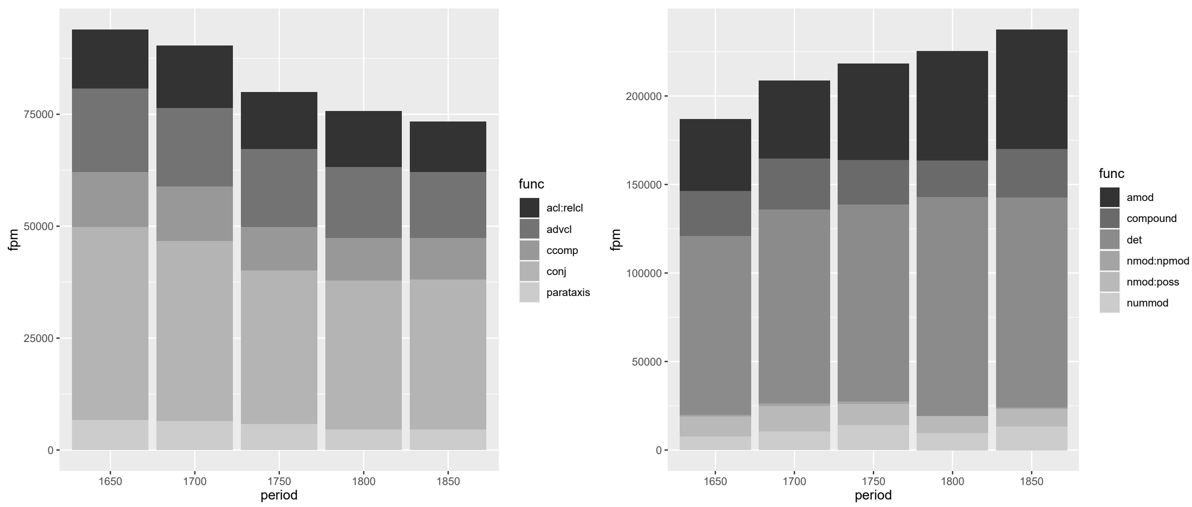

Figure 8: FPM of **long** distance functions (left) and **short nominal** functions (right) (SL 52) by 50 years.