Diachronic Trends in Word Order Freedom and Dependency Length in Dependency-Annotated Corpora of Latin and Ancient Greek

Kristina Gulordava
University of Geneva
kristina.gulordava@unige.ch

Paola Merlo
University of Geneva
paola.merlo@unige.ch

Abstract

One easily observable aspect of language variation is the order of words. In human and machine natural language processing, it is often claimed that parsing free-order languages is more difficult than parsing fixed-order languages. In this study on Latin and Ancient Greek, two well-known and well-documented free-order languages, we propose syntactic correlates of word order freedom. We apply our indicators to a collection of dependency-annotated texts of different time periods. On the one hand, we confirm a trend towards more fixed-order patterns in time. On the other hand, we show that a dependency-based measure of the flexibility of word order is correlated with the parsing performance on these languages.

1 Introduction

Languages vary in myriad ways. One easily observable aspect of variation is the order of words. Not only do languages vary in the linear order of their phrases, they also vary in how fixed and uniform the orders are. We speak of fixed-order languages and free word order languages.

Free word order has been associated in the linguistic literature with other properties, such as richness of morphology, for example. In natural language processing, it is often claimed that parsing freer word order languages is more difficult, for instance, than parsing English, whose word order is quite fixed.

Quantitative measures of word order freedom and investigations of it on a sufficiently large scale to draw firm conclusions, however, are not common (Liu, 2010; Futrell et al., 2015b). To be able to study word order flexibility quantitatively and computationally, we need a syntactic representation that is appropriate for both fixed and flexible word order; we need languages that exhibit genuine optionality of word order, and for which large amounts of text have been carefully annotated in the chosen representation.

In the current choice of hand-annotated treebanks, these requirements are fullfilled by dependency-annotated corpora of Latin and Ancient Greek. These two languages are extensively documented, they are dead languages and are therefore studied in a tradition where careful text editing and curation is a necessity, and have the added advantage that their genealogical children, Romance languages and Modern Greek, are also grammatically well studied, so that we can add a diachronic dimension to our observations.

Both Latin and Ancient Greek allow a lot of freedom in the linearisation of sentence elements. In these languages, this also concerns the noun-phrase domain, which is otherwise typically more constrained than the verbal domain in modern European languages[1]. In this study, we propose syntactic correlates of word order freedom both in the noun phrase and at the sentence level: variability in the directionality of the head-modifier relation, adjacency of the head-modifier relation (also called non-projectivity), and degree of minimisation of dependency length.

First, we look at head directionality, that is, post-nominal versus prenominal placement, of adjectives and numerals. While the variation in adjective placement is a wide-spread and well-studied phenomenon in modern languages, such as Romance languages, for example, the variation in numeral placement is a rarer phenomenon and is particularly interesting to investigate.

Then, we analyse the discontinuity of noun-

[1]Regarding the diachronic change in word order freedom, Tily (2010) found that in the change from Old to Middle and Modern English, the verb-headed clause changed considerably in word order and dependency length, from verb-final to verb initial, while the domain of the noun phrase did not.

Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015), pages 121–130,
Uppsala, Sweden, August 24–26 2015.

Language	Text	Period	#Sentences	#Words
Latin	*Caesar*, Commentarii belli Gallici	58-49 BC	1154	22408
	Cicero, Epistulae ad Atticum & De officii	68–43 BC	3830	44370
	Aetheriae, *Peregrinatio*	4th century AD	921	17554
	Jerome's *Vulgate*	4th century AD	8903	79389
Ancient Greek	*Herodotus*, Histories,	450-420 BC	5098	75032
	New Testament	4th century AD	10627	119371

Table 1: Summary of properties of the treebanks of Latin and Ancient Greek languages, including the historical period and size of each text.

phrases. Specifically, we extract the modifiers that are separated from the noun by some elements of a sentence that are not themselves noun dependents. Example (1) illustrates a non-adjacent dependency between the noun *maribus* and the adjective *reliquis*, separated by the verb *utimur*.

(1) (Caes. Gal. 5.1.2)

... quam quibus in reliquis$_a$ utimur$_v$ maribus$_n$
... than those in other we-use seas
'... than those (that) we use in (the) other seas'

We apply our two indicators to a collection of dependency-annotated texts of different time periods and show a pattern of diachronic change, demonstrating a trend towards more fixed-order patterns in time.

The different word order properties that we detect at different points in time for the same language allow us to set up a controlled experiment to ask whether greater word-order freedom causes greater parsing difficulty. We show that the dependency formalism provides us with a sentence-level measure of the flexibility of word order which we define as the distance between the actual dependency length of a sentence and its optimal dependency length (Gildea and Temperley, 2010). We demonstrate that this robust measure of the word order freedom of the languages reflects their parsing complexity.

2 Materials

Before discussing our measures in detail, we take a look at the resources that are available and that are used in our study.

2.1 Dependency-annotated corpora

The dependency treebanks of Latin and Ancient Greek used in our study come from the PROIEL project (Haug and Jøhndal, 2008). Compared to other treebanks, such as the Perseus treebanks (Bamman and Crane, 2011), previously used in the parsing literature, the PROIEL corpus contains exclusively prose and is therefore more appropriate for a word order variation study than other treebanks, which also contain poetry. Moreover, the PROIEL corpus allows us to analyze different texts and authors independently of each other. This, as we will see, provides us with interesting diachronic data. Table 1 presents the texts included in the corpus with their time periods and the size in sentences and number of words.

The texts in Latin range from the Classical Latin period (Caesar and Cicero) to the Late Latin of 4th century (Vulgate and Peregrinatio). Jerome's Vulgate is a translation from the Greek New Testament. The two Greek texts are Herodotus (4th century BC) and New Testament (4th century AD). The sizes of the texts are uneven, but include at least 17000 words or 900 sentences.

2.2 Modifier-noun dependencies in the corpus

We use the dependency and part-of-speech annotations of the PROIEL corpus to extract adjective-noun and numeral-noun dependencies and their properties.

Both Latin and Ancient Greek are annotated using the same guidelines and tagsets. We identify adjectives by their unique (fine and coarse) PoS tag "A-". The PoS annotation of the PROIEL corpora distinguishes between cardinal and ordinal numerals ("Ma" and "Mo" fine tags correspondingly). Cardinal numerals differ in their structural and functional properties from ordinal numerals; current analysis includes only cardinals to ensure the homogeneity of this class of modifiers.

For our analysis, we consider only adjectives and numerals which directly modify a noun, that is, their dependency head must be tagged as a noun ("Nb" and "Ne" fine tags). Such dependencies must also have an “atr” dependency label, for attribute.

The overall number of extracted adjective dependencies ranges from 600 (Peregrinatio) to 1700 (Herodotus and NewTestament), with an average of 1000 dependencies per text. The overall number of extracted numeral dependencies ranges from 83 (Peregrinatio) to 400 (New Testament and Vulgate), with average of 220 dependencies per text.

2.3 Measures

Our indicators of word order freedom are based on the relationship between the head and the dependent.

Head-Dependent Directionality Word order is a relative positional notion. The simplest indicator of word order is therefore the relative order of head and dependent. We say then that a language has free(r) word order if the position of the dependents relative to the head, before or after, is less uniform than for a fixed order language. In traditional linguistic studies, this is the notion that is most often used. However, it is a measure that is often too coarse to exhibit any clear patterns.

Head-Dependent Adjacency A more sensitive measure of freedom of word order will take into account adjacency to the head. Dependents can be adjacent to the head or not. Dependents that are not adjacent to the head can be separated by elements that belong to the same subtree or not. If dependents are not adjacent and are separated by a different subtree, we talk of non-projectivity.

The notion of non-projectivity encodes therefore both a notion of linear order and a notion of structural relation. It is this last notion that we consider relevant as a correlate of free word order.

The non-projectivity measure can be encoded in two ways: either as a simple indicator, a binary variable that tells us if a dependency is projective or not, or a distance measure that counts the distance of non-adjacent elements, as long as they are crossed by a non-projective dependency.

In this paper, we present an adjacency analysis for the noun phrase. More precisely, we identify modifiers which are separated from their head noun by at least one word which does not belong to the subtree headed by the noun. For instance, as can be seen from the dependency tree in Figure 1, the adjective *reliquis* is separated from its head *maribus* by the verb *utimur*, which does not belong to the subtree of *maribus* (which comprises only *reliquis* and *maribus*, in this example). We calculate the proportion of such non-projective adjectives over all adjectives whose head is a noun. In addition, we report the average distance of non-projective adjectives from their head. The same values are also computed and reported for numerals.

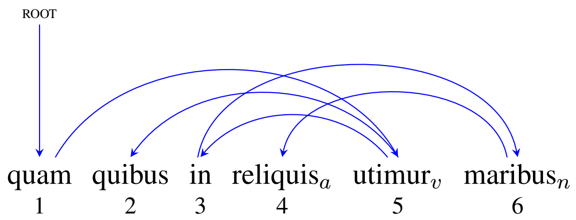

Figure 1: The dependency tree of the sentence from Example (1), extracted from the original PROIEL treebank.

3 NP-internal word order variation

We begin our investigation of word order variation by looking at word order in the noun phrase, a controlled setting potentially influenced by fewer factors than sentential word order.

3.1 Head-Dependent Directionality

For each of the texts in our corpus, we computed the percentage of prenominal versus post-nominal placement for two modifiers — adjectives and numerals. To avoid interference with size effects, these counts include only simple one-word modifiers.

If languages are sensitive to complexity, and tend to reduce it, our expectation for the diachronic trend is straight-forward. We expect the amount of prenominal-postnominal variation to be reduced. Also, we expect it to take the Latin grammar in the direction of the Romance-like grammar and Ancient Greek grammar in the direction of the Modern Greek grammar. Specifically, we expect adjective order to be more post-nominal in Latin in the course of time and more prenominal in Ancient Greek (Modern Greek has rigid prenominal adjective placement). For numerals, both Latin and Ancient Greek are expected to show more prenominal orders in the more recent texts (no post-nominal numerals are possible at all either in Romance languages or Modern Greek).

Table 2, left panel, shows the results. For adjectives in Latin, the observed percentages of prenominal adjectives exhibit the expected diachronic trend, moving from 73% to 36% of

		Head-Directionality				Adjacency			
		Adjective		Numeral		Adjective		Numeral	
Language	Text	#	%	#	%	%	Dist	%	Dist
Latin	Caesar	784	73	110	68	17	1.21	15	1.17
	Cicero	1064	60	104	80	11	1.14	12	1.31
	Peregrinatio	533	58	69	78	5	1.10	6	1.06
	Vulgate	1088	36	352	72	4	1.05	3	1.03
Ancient	Herodotus	1409	49	282	69	27	1.38	16	1.20
Greek	NewTestament	1257	49	400	70	9	1.10	4	1.04

Table 2: Quantitative summary of the variation in placement of two noun modifiers — adjectives and numerals in the Latin and Ancient Greek treebanks. The number of modifier-noun pairs and the percentage of prenominal order is given on the left; the percentage of non-adjacent modifiers (out of the total number) and the average distance from the noun head is given on the right.

prenominal adjectives. In terms of magnitude of the head-directionality measure, the shift from head-initial to head-final in Latin is of roughly the same size around the mean, which does not yet support strong regularisation. We know however, from statistics on modern Romance languages that this trend has converged to post-nominal patterns that range around 70% (Spanish 73%; Catalan 79%; Italian 67%; Portuguese 71%; French 74%)[2]. Adjective placement in Ancient Greek does not show any regularisation. For numerals, we do not observe a strong regularisation pattern for either language.

Since our expectations about trends of head-dependent directionality are only confirmed by adjectives in Latin, we conclude that this measure is weak and might not be sensitive to small changes in word order freedom.

3.2 Head-dependent adjacency

A more interesting diachronic observation comes from the number of non-adjacent versus adjacent modifiers (Table 2, right panel). Similar to the head-directionality patterns, our expectation is that the number of non-adjacent modifiers will decrease over time to eventually converge to the modern language situation, where such dependencies practically do not exist. The observed pattern is very sharp. This change is clear from the decline in percentage: from 17% to 4% for adjectives in Latin and 27% to 9% for adjectives in Ancient Greek. For numerals, the non-projectivity decreases from 15% to 3% in Latin and from 16% to 4% in Ancient Greek. It is important to notice that this decline can be made apparent only through a quantitative study, as it requires a full-fledged syntactic analysis of the sentence covering the non-projective dependencies. This phenomenon is relatively infrequent and the difference in percentages might not be perceived in traditional descriptive work.

Our results on head-directionality and adjacency for noun modifiers, summarised in Table 2, show that the two measures of word order freedom which we proposed do not pattern alike. While head-directionality does not show much change (with the exception of adjectives in Latin), the results on adjacency measure confirm our expectation that both languages converged with time towards a more fixed word order.

The tendency for non-projectivity and for preferences of head-adjacency of one-word modifiers are often explained as a tendency to minimise dependency-length, tendency that languages use to facilitate processing and production (Hawkins, 2004). In the next two sections, we study this more general principle of dependency length minimisation. We extend our investigation from the limited, controlled domain of the noun phrase to the more extended context of sentences. We investigate whether the dependency length measure at the sentence level correlates with our findings so far, and whether it is a good predictor of parsing complexity. We expect to see that, as languages have more and more fixed word order patterns, they become easier to parse.

4 Minimising Dependency Length

Very general, intuitive claims, both in human sentence processing and natural language processing,

[2]These counts are based on the dependency treebanks of these languages, available from Zeman et al. (2012).

state that free word order and long dependencies give rise to greater processing complexity. As such, languages should show patterns of regularisation, diachronic and synchronic, towards shorter dependencies and more homogeneous word orders. Notice, however, that these two pressures are in contradiction, as a reduction in dependency length can be obtained by placing modifiers at the two sides of the head, increasing variation in head directionality. How exactly languages develop, then, is worthy of investigation.

Experimental and theoretical language research has yielded a large and diverse body of evidence for dependency length minimisation (DLM). Gibson (1998, 2000) argues that structures with longer dependencies are more difficult to process, and shows that this principle predicts a number of phenomena in comprehension. One example is the finding that subject-extracted relative clauses are easier to process than object-extracted relative clauses.

Dependency length minimisation also concerns phenomena of syntactic choice. Hawkins (1994, 2004) shows, through a series of corpus analyses, that syntactic choices generally respect the preference for placing short elements closer to the head than long elements. This choice minimises overall dependency length in the tree. For example, in cases where a verb has two prepositional-phrase dependents, the shorter one tends to be placed closer to the verb. This preference is found both in head-first languages such as English, where PPs follow verbs and the shorter of two PPs tends to be placed first, and in head-last languages such as Japanese. Hawkins (1994, 2004) also shows that, in languages in which adjectives and relative clauses are on the same side of the head noun, the adjective, which is presumably generally shorter than the relative clause, is usually required to be closer to the noun. Temperley (2007) finds evidence for DLM in a variety of syntactic choice phenomena in written English. For example, subject NPs tend to be shorter than object NPs: as the head of an NP tends to be near its left end, a long subject NP creates a long dependency between the head of the NP and the verb, while a long object NP generally does not.

Recently, global measures of dependency length on a larger scale have been proposed, and cross-linguistic work has used these measures. Gildea and Temperley (2010) look at the overall dependency length of a sentence given its unordered structure to study whether languages tend to minimize dependency length. In particular, they observe that German tends to have longer dependencies compared to English, which they attribute to greater freedom of word order in German.

Their study, however, suffers from the shortcoming that they are comparing different annotations and different languages. From a methodological point of view, our experimental set up is more controlled because we compare several texts of the same language (Latin or Ancient Greek) and these texts belong to the same corpus and are annotated using the same annotation scheme. This means that the annotation scheme assumes the same underlying head-dependent relations in all texts for a given pair of parts-of-speech. From the linguistic point of view, the comparison of different amounts of word order freedom comes not from comparing different languages — a comparison where many other factors could come into play — but from comparing the same language over time as its word order properties were changing. The possible differences in DLM in these texts can be therefore directly attributed to the flexibility of their orders with respect to each other, since neither language nor annotation changes.

We test, then, whether a coarse dependency length measure (Gildea and Temperley, 2010) can capture the rate of the flexibility of word order in our controlled setting.

The dependency length of a sentence is simply defined as the sum of the lengths of all of its dependencies. The length of a dependency is taken to be the difference between position indices of the head and the dependent. To illustrate, for the subtree in Figure 1, the overall dependency length is equal to 14 for five dependencies. This is a particularly high value because there are two non-projective dependencies in the sentence. Dependency length is therefore conditioned both on the unordered tree structure of the sentence and the particular linearisation of this unordered graph, the order of words.

Following Gildea and Temperley (2010) and Futrell et al. (2015a) we also compute the optimal and random dependency length of a sentence, based on its unordered dependency tree available from the gold annotation. More precisely, to compute the random dependency length, we permutate the positions of the words in the sentence and cal-

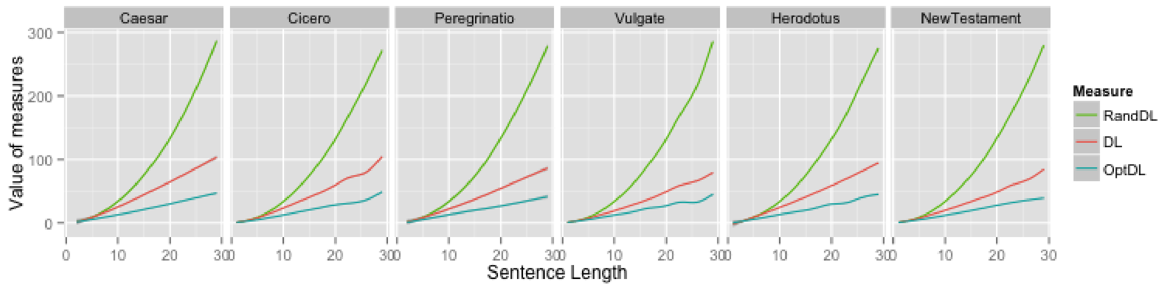

Figure 3: Average random, average optimal and actual dependency lengths of sentences by sentence length for each text.

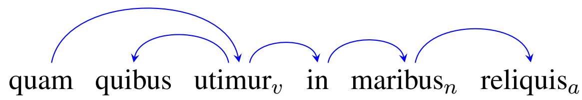

Figure 2: A word ordering of the sentence from Example (1) which yields minimal dependency length.

culate the new random dependency length preserving the original unordered tree structure.[3]

The optimal dependency length is calculated using the algorithm proposed by Gildea and Temperley (2007). Given an unordered dependency tree spanning over a sentence, the algorithm outputs the ordering of words which gives the minimal overall dependency length. Roughly, the algorithm implements the DLM tendencies widely observed in natural languages: if a head has several children, these are placed on both sides of the head; shorter children are closer to the head than longer ones; the order of the output is fully projective. Gildea and Temperley (2007) prove the optimality of the algorithm. For instance, the optimal ordering of the tree in Figure 1 would yield the dependency length of 6, as can be seen from the Figure 2.

Note that two sentences with the same unordered tree structure will have the same optimal dependency lengths.[4] If such sentences have different actual dependency lengths, this must then be directly attributed to the differences in their word order. We can generalise this observation to the structural descriptions of languages that are known to have similar grammatical structures. This similarity will be necessarily reflected by similar average values of the optimal dependency lengths in the treebanks. For such languages, systematic differences in actual dependency lengths observed across many sentences can be consequently attributed to their different word order patterns.

Our Latin and Ancient Greek texts show exactly this type of difference in their dependency lengths. Figure 3 illustrates the random, optimal and actual dependency lengths averaged for sentences of the same length.[5] First of all, we can observe that languages do optimise dependency length to some extent as their dependency lengths (indicated as DL) are lower than random. However, they are also not too close to the optimal values (indicated as $OptDL$). As can be also seen from Figure 3, the optimal dependency lengths across the texts are very similar. Their actual dependency lengths, on the contrary, are more variable. If we define the DLM score as the difference between the optimal and the actual dependency length, $DL - OptDL$, we observe a diachronic pattern aligned with the non-projectivity trends from the previous section. The patterns are shown in Figures 4 and 5, where for the sake of readability, we have plotted $DL - OptDL$ against the sentence length in log-log space.

For each language, we tested whether the pairwise differences between $DL - OptDL$ trends are significant by fitting the linear regressions $\log(DL - OptDL + 1) \sim \log(Sent)$ for two texts

[3] We do not impose any constraints on the random permutation of words. See Park and Levy (2009) for an empirical study of different randomisation strategies for the estimation of minimal dependency length with projectivity constraints.

[4] Also, two sentences with the same number of words will have the same random dependency lengths (on average).

[5] Since the optimal and random dependency length values depend (non-linearly) on the sentence length n, it is customary to analyse them as functions $DL(n)$ (and $E[DL(n)]$) and not as global averages over all sentences in a treebank (Ferrer-i-Cancho and Liu, 2014).

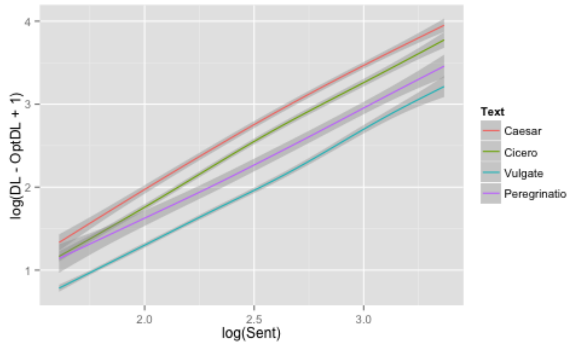

Figure 4: Rate of DLM for Latin texts, measured as $DL - OptDL$ and mapped to sentence length (in log-log space).

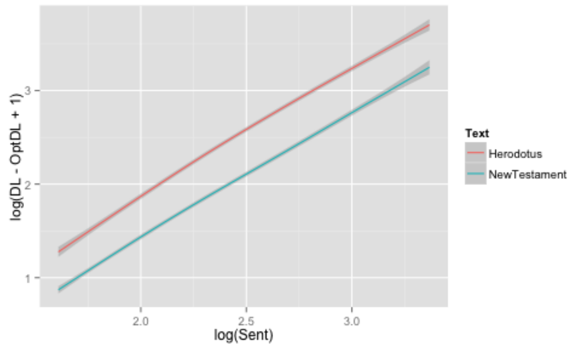

Figure 5: Rate of DLM for Greek texts, measured as $DL - OptDL$ and mapped to sentence length (in log-log space).

and comparing their intercepts[6]. These were significant at the $p < 0.001$ level for all pairs of texts.

So we can conclude that for Latin, older manuscripts of Caesar and Cicero show less minimisation of dependency length than later Latin texts of Vulgate and Peregrinatio. For Ancient Greek, Herodotus, which is the oldest test in the collection, has the smallest minimisation of dependency length. Since modern Romance languages and modern Greek have dependency lengths very close to optimal (Futrell et al., 2015a), we expect that Latin and Ancient Greek minimise the dependency length over time. Our data confirm this expectation.

We have also observed that the smaller percentage of non-projective arcs aligns with the higher rate of DLM across texts. This result confirms empirically a theoretical observation of Ferrer-i-Cancho (2006).

5 Word order flexibility and parsing performance

The previous section confirms through a globally optimised measure, what is already visible in the diachronic evolution of the adjacency measure in Table 2: older Latin and Ancient Greek texts exhibit longer dependencies and freer word order than later texts.

It is often claimed that parsing freer-order languages is harder. Specifically, parsers learn locally contained structures better and have more problems recovering long distance dependencies (Nivre et al., 2010). Handling non-projective dependencies is another long-standing problem (McDonald and Satta, 2007). We investigate the source of these difficulties, by correlating parsing performance on our texts from different time periods to our free word order measures. It is straight-forward to hypothesise that a tree with a small overall dependency length will be easier to parse than a tree with a large overall dependency length, and that a projective tree will be easier than a non-projective tree. Given our corpus, which is annotated with the same annotation scheme for all texts, we have an opportunity to test this hypothesis on texts that constitute truly controlled minimal pairs for such analysis.

The parsing results we report here are obtained using the Mate parser (Bohnet, 2010). Graph-based parsers like Mate do not have architectural constraints on handling non-projective trees and have been shown to be robust at parsing long dependencies (McDonald and Nivre, 2011). Given the high percentage of non-projective arcs and the number of long dependencies in the Latin and Ancient Greek corpora, we expect a graph-based parser to perform better than other types of dependency parsers. On a random training-testing split for all our, Mate parser shows the best performance among several of the dependency parsers we tested, including the transition-based Malt parser (Nivre et al., 2006).

We test several training and testing configurations. Since it is not clear how to evaluate a parser to compare texts with different rates of word order freedom, we used two different set-ups: training and testing within the same text and across different texts.

For the "within-text" evaluation, we apply a

[6] More precisely, we fitted a linear regression $\log(DL - OptDL + 1) = \beta \cdot Text + \log(Sent)$, where $Text$ is a binary indicator variable, on the combined data for two texts. We compare this model to the null model with $\beta = 0$ by means of an ANOVA to test whether two texts are best described by linear regressions with different or equal intercepts.

Lang	Configuration	Train. Size	UAS
Latin	Caesar	18k	66.46
	Cicero	18k	63.11
	Peregr.	18k	74.35
	Vulgate	18k	83.92
	all texts	155k	78.30
Greek	Herodotus	75k	69.76
	NewTest	75k	88.01
	all texts	195k	79.94

Table 3: Parsing accuracy for random-split training (90%) and test (10%) configurations for each language and for each text independently.

Lang	Training	Test	Train. Size	UAS
Latin	BC	AD	67k	67.27
	AD	BC	106k	57.72
Greek	Herodotus	NewTest	75k	76.05
	NewTest	Herodotus	120k	61.27

Table 4: Parsing accuracy for period-based training and test configurations for Latin and Ancient Greek.

standard random split, 90% of the corpus assigned to training and 10% assigned to testing, for each text separately. We eliminated potentially confounding effects due to different training sizes by including only around 18'000 words for each text in Latin (the size of the Peregrinatio corpus), and around 100'000 in Ancient Greek. We also report a strong baseline for each language, calculated by training and testing on all texts combined and split randomly with 90%/10% proportion. We evaluate the parsing performance using Unlabelled Accuracy Scores (UAS). The use of the unlabelled, rather than labelled, accuracy scores is the appropriate choice in our case because we seek to correlate the dependency length minimisation measure, a structural measure based on unlabelled dependency trees, to the parsing performance. The results for these experiments are reported in Table 3. First, the cumulative parsing accuracy on both Latin and Ancient Greek is relatively high as seen from the 'all texts' random split configuration[7]. Importantly, we can also observe that the older varieties of both Latin and Ancient Greek have lower UAS scores than their more recent counterparts.

We also evaluate parsing performance across time periods. Our intuition is that it is harder to generalise from a more fixed-order language to a freer-order language than vice versa. In addition, this setup allows us to use larger training sets for a more robust parsing evaluation. For this experiment, for Latin, we divide the four texts into two diachronic groups, where they naturally belong, BC for Caesar and Cicero and AD for Vulgate and Peregrinatio. We then train the parser on texts from one group and test on texts from the other. For Greek, as we do not have several texts from the same period, we test a similar configuration by training on one text and testing on the other. The results of these configuration are presented in Table 4. These results confirm our hypothesis and suggest that it is better to train the parser on a freer word order language. Despite the fact that it is harder to parse freer word order languages, as shown in Table 3, they provide better generalisation ability.

To summarise, in our experiments we see that the accuracy for older texts written in Latin in the BC period is much lower than the accuracy for late Latin texts written in the AD period. This pattern correlates with the previously observed smaller degree of dependency length minimisation of BC texts compared to AD texts. Similarly, for Greek, Herodotus is much more difficult to parse than the New Testament text, which corresponds to their differences in the rate of DLM as well as the non-projectivity in the noun phrase. The presented results confirm, therefore, the postulated hypothesis that freer order languages are harder to parse. In combination with the results from the previous sections, we can conclude that this difficulty is particularly due to longer dependencies and non-projectivity.

6 Related work

Our work has both similarities and differences with traditional work on Classical languages. Much work on word order variation using traditional, scholarly methods relies on unsystematically chosen text samples. Conclusions are often made about the Latin language in general, based on relatively few examples extracted from as few as one literary work. The analyses and the conclusions could therefore be subject to both well-known kinds of sampling errors: bias error due to a skewed sample and random error due to small

[7] These performance values are especially high compared to the previous results reported on the LDT and AGDT corpora, 61.9% and 70.5% of UAS, respectively (Lee et al., 2011). This increase in accuracy is likely due to the the fact that our texts are prose and not poetry.

sample sizes.

In particular, word order variation is one of the most studied syntactic aspects of Latin. For example, much descriptive evidence is dedicated to show the change from SOV to SVO order. However, starting from the work of Panhuis (1984), the previously assumed OV/VO change has been highly debated. At present, there is no convincing quantitative evidence for the diachronic trend of this pattern of variation in Classical Latin. In general, such coarse word order variation patterns are often bad cues of diachronic change and a more accurate syntactic and pragmatic analysis is required.

Non-projectivity goes under the name of *hyperbaton* in the classical literature. Several pieces of work address this phenomenon. Some of the authors give estimations of the number of discontinuous noun phrases, based on their analysis of particular texts (see Bauer (2009, 288-290), and the references there). These estimations range from 12% to 30% and are admittedly controversial because the counting procedure is not clearly stated (Pinkster, 2005, 250).

We are aware of only very few pieces of work that make use of syntactically-annotated treebanks to study diachronic word order variation. Bamman and Crane (2008) present some statistics on SVO order and on adjective-noun order, extracted from their Perseus treebanks for several subcorpora. Their data shows very different patterns of observed SVO variation across different texts. These patterns change from author to author and are hard to analyse in a systematic way. The work described in Tily (2010) is the closest to ours. The order of Old English is analysed using the same dependency length measure proposed by Gildea and Temperley (2010). On a large sample of texts, it is shown that there is a clear decrease in overall dependency length (averaged across sentences of all lengths in a corpus) from 900 to 1500 AD.

Another very relevant piece of work by Futrell et al. (2015a) also concerns dependency length minimisation. The general results of this study over thirty-four languages is that languages minimise dependency length over a random baseline. In these results, Latin and Ancient Greek are exceptions and do not appear to show greater than random dependency length minimisation. This is in contrast to our results. We conclude that this is an effect of the corpus used in Futrell's study, which contains a lot of poetry, while our texts are prose. Our results show a more coherent picture with their general results.

Finally, in this work, we address word order variation in the noun phrase and the DLM principle applied at the sentence level independently. Gulordava et al. (2015) investigate how these two properties interact and whether DLM modulates the variation in the placement of adjectives.

7 Conclusions

This paper has presented a corpus-based, quantitative investigation of word order freedom in Latin and Ancient Greek, two well-known and well-documented free-order languages. We have proposed two syntactic correlates of word order freedom in the noun phrase: head-directionality and head-dependent adjacency, or non-projectivity. If applied to a collection of dependency-annotated texts of different time periods, the non-projectivity measure confirms an expected trend toward closer adjacency and more fixed-order patterns in time. On the contrary, the head-directionality measure is a weak indicator of the fine-grained changes in freedom of word order. We have then extended the investigation to the sentence level and applied another dependency-based indicator of free word order, the rate of dependency length minimisation. The trend toward more fixed word orders is confirmed by this measure.

Another main result of the paper correlates dependency length minimisation with parsing performances on these languages, thereby confirming the intuitive claim that free-order languages are harder to parse. As a side result, we train parsers for Latin and Ancient Greek with good performance, showing, for future directions, that it will be possible to extend the data for the analysis of these languages by automatically parsing unannotated texts.

Acknowledgements

We gratefully acknowledge the partial funding of this work by the Swiss National Science Foundation, under grant 144362. We thank Lieven Danckaert and Séverine Nasel for pointing relevant Latin and Ancient Greek references to us.

References

David Bamman and Gregory R. Crane. 2008. Building a dynamic lexicon from a digital library. In *Procs of* *the 8th ACM/IEEE-CS Joint Conference on Digital libraries (JCDL'08)*, 11–20, New York, NY.

David Bamman and Gregory R. Crane. 2011. The Ancient Greek and Latin Dependency Treebanks. In Caroline Sporleder, Antal Bosch, and Kalliopi Zervanou, editors, *Language Technology for Cultural Heritage*, pages 79–98. Springer, Berlin/Heidelberg.

Brigitte L. M. Bauer. 2009. Word order. In Philip Baldi and Pierluigi Cuzzolin, editors, *New Perspectives on Historical Latin Syntax. Vol. 1. Syntax of the Sentence*, 241–316, Berlin. Mouton de Gruyter.

Bernd Bohnet. 2010. Very high accuracy and fast dependency parsing is not a contradiction. In *Procs of the 23rd Int'l Conf. on Computational Linguistics*, COLING '10, 89–97, Stroudsburg, PA.

Ramon Ferrer-i-Cancho and Haitao Liu. 2014. The risks of mixing dependency lengths from sequences of different length. *Glottotheory*, 5(2):143–155.

Ramon Ferrer-i-Cancho. 2006. Why do syntactic links not cross? *EPL (Europhysics Letters)*, 76(6):12–28.

Richard Futrell, Kyle Mahowald, and Edward Gibson. 2015a. Large-Scale Evidence of Dependency Length Minimization in 37 Languages. (Submitted to Proceedings of the National Academy of Sciences of the United States of America).

Richard Futrell, Kyle Mahowald, and Edward Gibson. 2015b. Quantifying Word Order Freedom in Dependency Corpora . In *Proceedings of the Third Int'l Conf. on Dependency Linguistics (DepLing 2015)*, Uppsala, Sweden.

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*, 95–126.

Daniel Gildea and David Temperley. 2007. Optimizing Grammars for Minimum Dependency Length. In *Procs of the Association for Computational Linguistics (ACL'07)*, 184–191, Prague, Czech Republic.

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 *Procs of the Association for Computational Linguistics: Short Papers (ACL'15)*.

Dag T. T. Haug and Marius L. Jøhndal. 2008. Creating a Parallel Treebank of the Old Indo-European Bible Translations. In *Proc of the 2nd Workshop on Language Technology for Cultural Heritage Data*, 27–34, Marrakech, Morocco.

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

John Lee, Jason Naradowsky, and David A. Smith. 2011. A Discriminative Model for Joint Morphological Disambiguation and Dependency Parsing. In *Procs of the Association for Computational Linguistics: Human Language Technologies*, 885–894, Portland, Oregon.

Haitao Liu. 2010. Dependency direction as a means of word-order typology: A method based on dependency treebanks. *Lingua*, 120(6):1567–1578.

Ryan McDonald and Joakim Nivre. 2011. Analyzing and Integrating Dependency Parsers. *Computational Linguistics*, 37(1):197–230.

Ryan McDonald and Giorgio Satta. 2007. On the complexity of non-projective data-driven dependency parsing. In *Procs of the 10th Int'l Conference on Parsing Technologies*, 121–132.

Joakim Nivre, Johan Hall, and Jens Nilsson. 2006. MaltParser: A data-driven parser-generator for dependency parsing. In *Procs of the 5th International Conference on Language Resources and Evaluation (LREC'06)*, 2216–2219.

Joakim Nivre, Laura Rimell, Ryan McDonald, and Carlos Gómez-Rodríguez. 2010. Evaluation of dependency parsers on unbounded dependencies. In *Procs of the Int'l Conference on Computational Linguistics (COLING'10)*, pages 833–841, Stroudsburg, PA.

Dirk Panhuis. 1984. Is Latin an SOV language? A diachronic perspective. *Indogermanische Forschungen*, 89:140–159.

Albert Y. Park and Roger Levy. 2009. Minimal-length linearizations for mildly context-sensitive dependency trees. In *Procs of the North American Chapter of the Association for Computational Linguistics (NAACL'09)*, 335–343.

Harm Pinkster. 2005. The language of Pliny the Elder. In *Proceedings of the British Academy*, volume 129, pages 239–256. OUP.

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

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

Daniel Zeman, David Mareček, Martin Popel, Loganathan Ramasamy, Jan Štěpánek, Zdeněk Žabokrtský, and Jan Hajič. 2012. HamleDT: To Parse or Not to Parse? In *Procs of the Int'l Conference on Language Resources and Evaluation (LREC'12)*, 23–25, Istanbul, Turkey.