Computational quantitative syntax

The case of Universal 18

Kristina Gulordava & Paola Merlo

Accounting for the constraints on the possible word orders of a sentence in a language and across the world languages is a core challenge for syntactic theory. In the spirit of computational quantitative syntax, in this paper we present quantitative evidence about Universal 18. We show that corpus data confirms a dispreference for the word order combination where adjectives precede but numerals follow the nouns (Adj-N and N-Num). We then investigate if this dispreference is better explained as a constraint expressed at the level of the dominant orders or at the level of individual structures. Corpus counts support the latter interpretation. Finally, we propose a formal model of how this bias against Adj-N-Num orders can be integrated in the grammar.

Keywords: Universal 18, quantitative syntax, corpus counts, adjective-noun order, noun-numeral order, treebanks, Latin, Ancient Greek, modelling

1. Introduction

Languages arrange constituents in a sentence in very different ways. Consider the same sentence in English, Irish and Persian, illustrated in (1)–(3).

(1) James saw her on top of the hill. English

(2) Chonaic Séamus ar bharr an chnoic í. Irish
saw Séamus on top of hill her
'Séamus saw her on top of the hill'

(3) Ali dokhtari ra dar balaye tapeh did. Persian (Farsi)
Ali girl *acc* on top hill saw
'Ali saw a girl on top of the hill'

Looking across all documented languages of the world, it is apparent that not all possible word order arrangements are attested. Moreover, it has long been known that there are correlations between the orders of certain grammatical categories. For example, if, in a language, the verb precedes its complement, as in 'saw a girl',

https://doi.org/10.1075/rllt.16.08gul

then very frequently the adposition will also precede the noun, as in 'on the hill' (Dryer 1992). Accounting for the constraints on the possible word orders of a sentence is a core challenge for syntactic theory.

One of the goals of computational quantitative syntax is to determine the status that frequency plays in this context. The current work investigates the connection between the quantitative aspects of word order variation across languages and the quantitative aspects of word order variation within a language. We take as our starting point the empirical generalisations on word order typology known as word order universals (Greenberg 1963). We study the corresponding patterns in language-internal variation by looking at syntactically-annotated corpora of various languages. Specifically, we investigate in what way Universal 18 – which restricts the combination of word orders between adjective and noun on the one hand, and numeral and noun on the other – is a constraint pertinent to the language system.

Universal 18 states a constraint on cross-linguistic distributions of the combination of two NP-internal orders: the order between adjective and noun and the order between numeral and noun.

> When the descriptive adjective precedes the noun, the demonstrative and the numeral, with overwhelmingly more than chance frequency, do likewise.
> (Greenberg 1963, p. 67–68)

As can be seen, this is already a statistical universal, although its quantification is not precisely numerical. We study the quantitative aspects of this statement, in the framework of current theoretical proposals. Specifically, we analyse two types of quantified observations about word order: *token-level* and *type-level* observations. Token-level observations correspond to the individual instances of word order realisation in a language. All instances of word order alternation at the token level constitute the language-internal variation. Type-level observations correspond to the word order properties of a language as observed in the grammar or in the typological descriptions. For example, the notion of "most frequent word order" (see below, footnote 4) is a type-level property of a language. Typological variation describes the distribution of languages by types, according to such properties. The formulation of Universal 18 above is ambiguous between these two interpretations. The statement reported above could mean that the order of adjective and noun and the order of numeral and noun are very frequently harmonious across the grammars of the world languages or it could mean that if we look at the actual sentences and phrases in a language the cooccurrence of prenominal adjective and postnominal numeral in the same phrase is dispreferred.

We exploit two different types of linguistic resources – syntactically annotated corpus data and typological data collected from large databases – to collect the quantitative distributions of the word order variation expressed by Universal 18, within a language and across languages. In the next section, we discuss new typological data collected from large-scale typological resources. The rest of the paper will concentrate on corpus counts and language-internal variation.

2. Typological data on Universal 18

The World Atlas of Language Structures (WALS) (Dryer & Haspelmath 2011)[1] and the Syntactic Structures of the World's Languages (SSWL) databases[2] are large-scale typological resources that allow us to revisit Universal 18. Assuming that a language has categories such as numerals and adjectives, there are four possible combinations of word orders between adjective and noun and numeral and noun in a noun phrase: L1 – Adjective-Noun & Numeral-Noun; L2 – Adjective-Noun & Noun-Numeral; L3 – Noun-Adjective & Numeral-Noun; L4 – Noun-Adjective & Noun-Numeral.

According to Greenberg's generalisation, the languages of type L2 which exhibit both word orders Adj≺N & N≺Num are very infrequent. This tendency has been shown to be very robust and has been attested in a large sample of languages of the world (Hawkins 1983; Rijkhoff 1998; Hurford 2003). New supporting evidence can be found by consulting the databases. Table 1 presents the number of languages of each word order type as found in the World Atlas of Languages (WALS).[3] Out of four possible language types, the most infrequent one (4%) is Adj≺N & N≺Num type which confirms the generalisation of Universal 18.

Table 1. Counts of languages according to their adjective-noun and numeral-noun word order types, as presented in WALS

	Num≺N	**N≺Num**	**No dominant order**	**Total**
Adj≺N	251	37	12	300
N≺Adj	168	510	35	713
No dominant order	35	22	13	70
Total	**454**	**569**	**60**	**1083**

1. http://wals.info/

2. http://sswl.railsplayground.net/

3. These are counts of all languages in the database, without sampling by genera. See, for example, Hurford (2003) for the counts in an accurately constructed representative sample of languages of the world.

Generative syntacticians have typically paid more attention to the different possible orders within a language, rather than to the dominant order. Using the Syntactic Structures of the World's Languages (SSWL), we can show that a constraint against Adj≺N & N≺Num – Universal 18 – is also observed for combinations of word order variation types, that is, the different grammatically possible word orders that a given language allows. The SSWL database lists grammatical and ungrammatical morpho-syntactic constructions in a language as binary properties. The Adj≺N and N≺Adj, as well as the Num≺N and N≺Num orders, are encoded as properties that tell whether a language allows a certain word order. Languages are therefore divided into three types with respect to adjective-noun word order: Adj≺N, N≺Adj, or both word orders are available, such as Romance languages (Table 2). Analogously, there are three types of languages with respect to numeral-noun word order. In total, there are nine possible types of different grammatical word order combinations. Table 2 presents the counts of these language types in the SSWL sample. A new observation is that the counts for rigid word order languages (in light gray) are distributed very similarly to the corresponding counts of languages extracted from WALS and thus re-state Universal 18. The least frequent type is the Adj≺N & N≺Num (one language, in bold), as expected. Furthermore, the counts for flexible word order languages (at least one word order alternates) also indirectly confirm Universal 18. Languages that combine Adj≺N & N≺Num orders (even if it does not dominate) are less frequent than other types. For example, only six languages allow Adj≺N & N≺Adj and N≺Num and only eight allow Adj≺N and Num≺N & N≺Num (in dark grey) while there are 29 and 14 alternating languages that do not allow Adj≺N & N≺Num combination (Adj≺N & N≺Adj with Num≺N and Num≺N & N≺Num with N≺Adj).

Table 2. Number of languages in SSWL database having one of the combinations of word order properties

	Num N, N Num	**Num N, –**	**–, N Num**	**no info**	**total**
Adj N, N Adj	16	29	6	4	55
Adj N, –	8	40	1	2	51
–, N Adj	14	16	48	3	81
no info	3	4	4	14	25
Total	**41**	**89**	**59**	**23**	**212**

In summary, counts collected from typological databases that describe the properties of the word orders of whole languages support the quantitative generalisation expressed in Universal 18, both according to the notion of dominant word order expressed in WALS, and also based on data that take optionality of word orders into account, such as those in SSWL.

Table 3. Percentage of prenominal adjectives and number of alternating adjectives in five Romance languages

Language	nb. instance	% prenominal	nb. alternating adjs
Spanish	20809	27	445
Catalan	17367	22	251
Italian	2868	34	147
Portuguese	7565	30	175
French	15806	27	256

It is important to note, however, that Greenberg's Universal 18 and the WALS counts refer to the dominant word orders of languages.[4] The dominant word order values in the WALS database are given according to the two criteria of unmarkedness and frequency and, in our work, we assume that dominant word order is equivalent to the most frequent word order.

This frequency-based notion hides the fact that word orders can vary. Adjectives in Romance languages are a well-known example. While in these languages adjectives occur predominantly after the noun, there is also considerable variation. Some adjective positioning changes meaning (a 'pauvre femme' is pitiable, but a 'femme pauvre' is a pauper) or restricts the number of meanings available ('une ancienne eglise' can be either a former church or an old church, but a 'eglise ancienne' is only old.). Other adjectives can be only postnominal or only prenominal, but many can alternate based on many other factors. (See Gulordava and Merlo (2015); Gulordava et al. (2015) for corpus-based discussion of the factors of these alternations.) Quantitatively, Table 3 shows the frequencies of five Romance languages, in a sample of 64'000 instances of noun-adjective word orders.[5] The table shows clearly that while the dominant position for adjectives is post-nominal

4. The notion of dominant or basic word order has been largely discussed and challenged in the typological literature (Dryer 1989, 1995; Mithun 1992). Commonly, by basic word order, one understands the unmarked word order, specifically, pragmatically unmarked word order. An exception to this view are the languages that were identified as having discourse-determined word order, where all of possible orders are grammatical and there is no clear neutral word order (Mithun 1992). Yet, in his extensive work, Dryer argues that basic word order can be defined in terms of frequency: the basic word order in a language is that which is used most frequently. See also Haspelmath (2006) who argues that all notions of markedness could be replaced with the notion of frequency.

5. The dependency annotated corpora of five Romance languages are used to collet these counts: Catalan, Spanish, Italian (Hajič et al. 2009), French (McDonald et al., 2013), and Portuguese (Buchholz & Marsi 2006). For more details on the data and data collection, please refer to Gulordava and Merlo (2015).

in all the languages, a sizable number of adjectives occur also prenominally for a total of up to 30% of the token occurrences for some languages. The variation in the positioning of adjectives, then, is not a marginal phenomenon. An even finer-grained investigation, then can be based on syntactically-annotated corpora, studying variation at the sentence level in individual languages, as we show in the next sections.

3. Accounts of language universals and Universal 18

Approaches that try to explain the asymmetries of typological distributions documented in the previous section roughly fall in two categories: the approaches that explain language universals as the result of constraints on language or, more broadly, the cognitive system of native speakers, and the explanations that argue exclusively in favour of a source external to cognition. The latter explanations focus on historical development, spread and contact of languages and speaker communities. For example, a much discussed piece of work by Dunn et al. (2011) models computationally the genetic evolution of four language families and claims to show that word order traits in the evolutionary process were developing independently of each other, whereas a linguistic explanation predicts that they must be correlated. Based on the observation that some traits were dependent during the genetic evolution in one language family but not in all language families, Dunn et al. (2011) conclude that word order universals are culturally dependent and not cognitively universal.

We adopt the competing view which argues that linguistic or cognitive factors play at least some role in shaping typological universals. We present in more detail two families of explanations, one corresponding to a token-level interpretation of word order universal 18 (Cinque 2005; Biberauer et al. 2014) and one corresponding to a type-level interpretation (Culbertson et al., 2012). We will then compare their predictions based on corpus data, to select the model that best accounts for the data.

3.1 Structure-level accounts

Typological word order universals are commonly divided in two types: harmonic universals and implicational universals. Harmonic universals constrain the same types of constructions to appear in the same order, for instance, for a dependent to be always on the left of its head.[6] A much cited example is a correlation between

6. We do not adhere to an explicit definition of the terms 'head' and 'dependent'. Most linguistic theories, both formal and functional, define these terms in some specific way, but for

the order of object and verb and the order of noun and adposition. In generative grammar, these harmonic tendencies are captured by the head-directionality parameter (Baker 2001).

In addition to harmonic universals, we can recognise a second type of universals, referred to as *implicational*, such as Universal 18. Implicational universals are generalisations of robust asymmetries between two disharmonic word order options. Consider the four possible combinations of two orders between H_1 and Dep_1 and H_2 and Dep_2 ($\prec$ means 'precedes'):

O1 – $H_1 \prec Dep_1$ & $H_2 \prec Dep_2$; O_2 – $Dep_1 \prec H_1$ & $Dep_2 \prec$ H2;
O3 – $H_1 \prec Dep_1$ & $Dep_2 \prec H_2$; O_4 – $Dep_1 \prec H_1$ & $H_2 \prec Dep_2$.

Orders O1 and O2 are harmonic, while O3 and O4 are disharmonic combinations. An implicational universal is stated when one disharmonic combination (e.g. O3) is observed much more frequently than another disharmonic combination (O4) across languages. These kinds of universals are not predictable under the head-directionality parameter. A recent proposal is to appeal to the Final-over-Final condition (FOFC) (Biberauer et al. 2014), introduced by Holmberg (2000). The FOFC states that syntactic structures where a head-final phrase ($Dep_1 \prec H_1$) takes as its complement a head-initial phrase ($Dep_1 = H_2$, $H_2 \prec Dep_2$) are prohibited in syntax. Consider Figure 1, which illustrates four possible structures between the phrase *α*P with the head α and the phrase *β*P with the head *β* which takes *α*P as its complement. These structures correspond to the four possible word order combinations O1–O4. FOFC bans the disharmonic structure (1d) creating therefore an asymmetric account of the two disharmonic structures (1c) and (1d).

The generalisation discussed in Biberauer et al. (2014) can in principle apply to the asymmetry captured by Universal 18.[7] This explanation depends on the specific structure that is assumed for the noun phrase. It presupposes that the structure of the noun phrase has an additional functional head X^0, associated with a particular adjectival position, and that this position takes the noun as its complement to form an XP phrase. The numeral, in turn, takes XP as its complement to create a NumP. If this is the case, then the Adj≺N≺Num order must

most of the discussion we will assume a simple interpretation of the head-dependent (as well as the head-modifier) relation: we take noun to be the head of noun phrase and adjectives and numerals to be modifiers of the noun; verb to be the head of verb phrase and object to be its complement; adposition to be the head and take a noun as its complement.

7. Culbertson et al. (2012), which we discuss later, mention FOFC as a possible explanation for the suppression of Adj≺N & N≺Num order with respect to another disharmonic order Num≺N & N≺Adj.

be the linearisation of a structure (with or without movement) of type [[Adj N] Num], directly violating FOFC.[8]

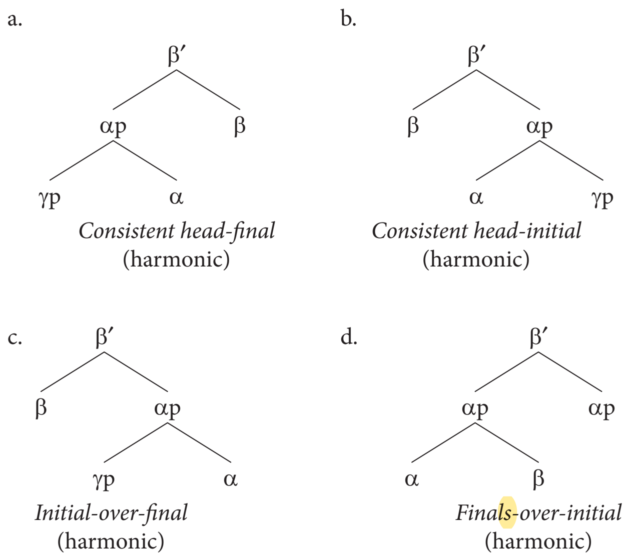

Figure 1. Four possible complementation combinations of two phrases *α*P and *β*P with respect to their head-directionality replicated from Biberauer et al. (2014, p. 117)

FOFC is put forward as a categorical constraint to capture a large set of syntactic phenomena cross-linguistically. The strong categorical assumption of FOFC runs

8. The assumption of the hierarchical structure [[Adj N] Num] for the noun phrase is controversial since, in many syntactic approaches, the adjective phrase is treated as an adjunct to the noun (there could be several adjective phrases modifying the same noun), and not a higher projection which takes a noun as its complement. Interestingly, the alternative of treating adjectives as lower dependents of N was employed to explain a different set of word order universals. The Head-final Filter of Williams (1982) was proposed as a constraint against [[*α γ*P] *β*] structures arising in noun phrases such as **[proud [of his son]] man*. Sheehan et al. (2017) discusses the close relation between FOFC and the Head-final Filter. The main motivation for the Num > Adj > N hierarchical structure comes from the notion of semantic scope (Rijkhoff, 1998). Adjectives typically modify the internal properties of the noun, while numerals do not. In English, while both Adj≺Num≺N ("heavy three boxes") and Num≺Adj≺N ("three heavy boxes") are possible, the first option has a special interpretation with the adjective having its scope over the numeral: the three boxes are heavy together (Ouwayda 2014; Schwarzschild 2009). The iconicity principle suggests that the semantic hierarchy Num > Adj > N should be mapped onto the syntactic structure (Rijkhoff 1998; Culbertson & Adger 2014).

into empirical problems when confronted with the gradual nature of constraints observed in typological distributions and language variation. Hawkins (1994), for example, suggested a processing constraint targeting [[α γP] β], involving memory or computational constraints, while Abels (2013) proposes that this structure is disfavored because of its misalignment with prosodic boundaries. For our purposes, the important common part of these accounts is that they all postulate a (hard or soft) constraint against a particular structural realisation of the complete NP phrase.

3.2 Grammar-level accounts

Experimental work by Culbertson et al. (2012) gives a type-level account of Greenberg's Universal 18. It takes as a starting point the original formulation by Greenberg involving the dominant adjective-noun and numeral-noun word order properties of a language. The results of an artificial learning experiment where participants learn a language of one of four possible word order types L1–L4 show, they argue, that the markedness of Adj≺N & N≺Num type is a result of a cognitive learning bias.

The experiment by Culbertson et al. (2012) consists in learning a small artificial language with a grammar defined by two probabilistic phrasal rules, which specify a dominant combination of adjective-noun and numeral-noun word orders. Specifically, participants receive as learning input instances of noun phrases of two types: adjective plus noun phrases, i.e. the artificial language equivalent of "green apple", or numeral plus noun phrase, i.e. the equivalent of "three apples". The lexicon is taught to participants using a visual interface. Four groups of participants are taught four different types of languages: each language has a grammar corresponding to one of four possible dominant word order combinations L1–L4. These grammars G_1–G_4 are presented in Table 4.

Table 4. Four types of grammars taught in the artificial language learning experiment of Culbertson et al. (2012)

	Adj N	**N Adj**
Num N	$G_1 = G_{0.7,0.7}$	$G_2 = G_{0.3,0.7}$
N Num	$G_3 = G_{0.7,0.3}$	$G_4 = G_{0.3,0.3}$

Each grammar G_i is probabilistic, with parameters α and β corresponding to the percentages of instances of dominant adjective-noun and numeral-noun orders which participants receive as an input, as in the following:

$$G_{\alpha,\beta} = \left\{ \begin{matrix} S_{adj} \xrightarrow{\alpha} \text{Adj N} & S_{num} \xrightarrow{\beta} \text{Num N} \\ S_{adj} \xrightarrow{1-\alpha} \text{N Adj} & S_{num} \xrightarrow{1-\beta} \text{N Num} \end{matrix} \right\}. \quad (1)$$

For example, a participant that learns G_1 will see the dominant Adj≺N word order 70% of the time and the alternative N≺Adj order 30% of the time, and dominant Num≺N order 70% of the time together with N≺Num 30% of the time. After the learning phase, a participant is asked to produce adjective phrases and numeral phrases describing the pictures of objects.

By introducing learners to input with variation, this mixture-shift paradigm allows the experimenter to observe the extent to which the learners alter the language grammar to bring it in line with hypothesised learning biases. Culbertson and colleagues hypothesise several biases which predict different learning outputs. First of all, following many psycholinguistic studies in language learning, they hypothesise a regularisation bias – a general learning bias which constrains the acquired grammar to minimise its variation. This bias predicts that learners will alter a grammar with variation rate (α,β) to a grammar where word order between a modifier and a noun is more fixed, a grammar $G_{\alpha',\beta'}$ where $|\alpha' - 0.5| > |\alpha - 0.5|$ and $|\beta' - 0.5| > |\beta - 0.5|$. For example, if the input grammar has $\alpha = 0.7$, in production, the dominant Adj≺N word order will be used more than 70% of the time $(\alpha' > 0.7)$ and if $\alpha = 0.3$, the dominant N≺Adj order will be used more than 70% of time $(\alpha' < 0.3)$.

Another type of bias considered in the study is a substantive bias – a bias that drives learners away from some particular structure. Culbertson and colleagues's central hypothesis is that the typological Universal 18 is indeed a reflex of a substantive bias against the Adj≺N & N≺Num word order combination. If this is the case, then the learners' output will reflect this dispreference by showing a different behaviour from standard regularisation behaviour, predicted by the general regularisation bias. The findings of the study confirm these predictions. While participants that were learning languages with grammars G_1, G_3, G_4 showed regularizing behaviour (average use of majority order was significantly greater than 70%), the group that was learning the language of the "dispreferred" type with grammar G_2 did not (its average use of the majority order was below 70%). Both a mixed-effect analysis and Bayesian modeling confirm statistically that these experimental data are best explained by postulating a learning bias against the Adj≺N & N≺Num word order combinations. The conclusion of the authors is that a substantive learning bias against a particular combination of word orders does exist and plays a role in shaping typological distribution of word orders. Moreover, they argue that the fact that it was observed during the short-time of a learning experiment rules out accounts of word order universals that place the origins of typological asymmetries in factors external to cognition.

Very importantly, a Bayesian model is used to provide an explicit account of Universal 18 (Culbertson & Smolensky 2012). This account assumes, following the experimental design, that the learning bias affects two word order properties of the grammar under study, represented by parameters α and β. These parameters are independent in the grammar, and their joint distribution is only affected by the substantial bias against Adj≺N & N≺Num combination. Specifically, they model the process of acquisition as learning the values of the two parameters α and β. In this model, the bias is represented as a prior probability on the α and β values. Universal 18 is therefore accounted for in terms of the probabilistic constraint on two type-level properties.

4. Our approach to Universal 18

The studies presented in the previous section argue for the existence either of a structural linguistic constraint or of a cognitive language learning bias to explain the cross-linguistic dispreference for the Adj≺N & N≺Num word order combination. We provide further evidence in favour of this view by studying quantitative patterns of NP-internal word order variation across languages in corpora.

We follow here the line of work which connects *quantitative* patterns of syntactic variation to underlying grammatical properties (see for example Merlo and Stevenson (2001); Bresnan et al. (2001); Merlo and Stevenson (2004); Bresnan et al. (2007); Hawkins (2009); Samardžić and Merlo (2012, 2018)). Frequency is known to be the product of many factors external to the grammar, such as the topic of discourse, the style of a speech or a text, and social conventions. However, these pieces of work convincingly show that at least in some cases frequency does reflect grammatical properties. Our investigation of the accounts of Universal 18 applies this same reasoning. We group the accounts of Universal 18 reviewed in the previous section into two types, according to their basic underlying assumptions: structural, token-level accounts, and grammar-level, type-level accounts.

The first group includes, e.g., Biberauer et al. (2014), Svenonius (2007) and Cinque (2005). All these accounts assume that the constraint blocks out some particular structure of the noun phrase, for example the [[Adj N] Num] structure. It is the presence of all of the elements – adjective, numeral, noun – and their arrangement in this particular structure which triggers the constraint. A noun phrase without a numeral will not meet this structural condition and will not be affected by a structural constraint. We can say, therefore, that the constraint operates at the token level as it restricts only certain realisations of a noun phrase.

The second type of accounts include Culbertson et al. (2012), and also Cysouw (2010), and corresponds to the grammar-level interpretation of the Universal 18. When discussing this work, we underlined that their probabilistic Bayesian account assumes a bias against a combination of two grammar-level surface properties and not against a particular noun phrase structure. The two word order properties were defined by parameters $\alpha = p(\text{Adj} \prec \text{N})$ and $\beta = p(\text{Num} \prec \text{N})$. A grammar is assumed to have one of four possible combinations of these properties and a learning bias is assumed to restrict grammars that have $\alpha > 0.5$ and $\beta < 0.5$. Therefore, we call this account a type-level account, as the bias operates on aggregated word order properties of a language and does not affect every instance of a noun-phrase directly.

These two types of accounts differ substantially in their treatment and interpretation of Universal 18, and provide different predictions for the word order variation patterns within a language. We formalise the difference between the structural and the grammar-level accounts in terms of probability distributions of word orders. Two basic assumptions underlie our account: (i) the dominant word order corresponds to the most probable word order; (ii) the effect of a constraint or bias is reflected in the probability distribution of word order alternatives. The first assumption is largely motivated by the typological work on basic word order discussed above. The second assumption reflects our view that grammatical properties induce frequency effects in corpora of language use.[9]

We start by defining two contexts, C_1 and C_2, that distinguish between the two types of accounts.

1. C_1 when adjectives and numerals occur independently in different noun phrases: C_1 = {noun phrases: adjective + noun = "old + book", or noun phrases: numeral + noun = "three + books"}
2. C_2 when adjectives and numerals co-occur in a noun phrase: C_2 = {noun phrases: adjective + numeral + noun = "three + old + books" }

9. This view is also articulated in Bresnan et al. (2001), where it is argued that grammatical constraints that apply categorically in certain languages, also show up as soft constraints in other languages. Universal 18 appears as a categorical restriction in certain languages, yielding the typological distribution discussed in Section 2, and it shows up as a more gradual constraint on frequencies in other languages, such as Latin and Ancient Greek. We will express this notion of gradation as probability distributions below, although this is not the only way of expressing gradation.

Following the notation from the previous sections, we define the parameters of word order distributions to be $\alpha = pc_1(\text{Adj} \prec \text{N})$ and $\beta = pc_1(\text{Num} \prec \text{N})$. These are word order probabilities for the context C_1. We define the four-way distribution of possible word orders in context C_2 as $\boldsymbol{\gamma}$. If the two contexts C_1 and C_2 are equivalent, then γ should be equal to a joint distribution of the two word order distributions $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ defined by two scalar parameters α and β:

$$\boldsymbol{\gamma}(x, y) = \boldsymbol{\alpha}(x) \cdot \boldsymbol{\beta}(y) \quad (2)$$

(where x is the word order between adjective and noun and y is the word order between numeral and noun); and in particular:

$$\gamma(\text{Adj} \prec \text{N}, \text{Num} \prec \text{N}) = \alpha \cdot \beta \quad (3)$$

Then $\boldsymbol{\alpha}(x) \cdot \boldsymbol{\beta}(y)$ is the probability (and the expected frequency) of word orders in C_2 under the hypothesis that the two word orders are independent and the two contexts are equivalent.

Structural accounts predict that the constraint underlying Universal 18 applies in context C_2, but not in context C_1. We should therefore observe an effect of this constraint as a categorical or statistical difference between word order distributions of context C_1 and context C_2. Specifically, in case of statistical bias, structural accounts predict that the observed frequency of the Adj≺N≺Num word order in the context C_2 must be smaller than the expected frequency (under the assumption that Adj≺N and N≺Num phrases could freely combine), that is:

$$\boldsymbol{\gamma}(\text{Adj} \prec \text{N} \prec \text{Num}) < \boldsymbol{\alpha}(\text{Adj} \prec \text{N}) \cdot \boldsymbol{\beta}(\text{N} \prec \text{Num}) \quad (4)$$

Type-level accounts offer a different prediction. Specifically, they assume that the bias affects the combination of the two parameters α and β in the language, but not word order combinations in each sentence of a language. These types of accounts do not distinguish therefore between the two contexts C_1 and C_2. Consequently, the two distributions $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha} \cdot \boldsymbol{\beta}$ should be equal.[10]

Given these two different predictions, we can use corpus data to test which theory complies better with it. To observe a possible effect of the structural constraint, we need to look at languages where both Adj≺N and N≺Num are possible grammatical orders (as dominant or alternative orders). While Romance languages show substantial variation in the prenominal or postnominal positioning of adjectives, as discussed above, the numeral almost exclusively precedes the noun. We must look into the predecessor of Romance languages, Latin, to find

10. In principle, C_1 and C_2 could be different and give rise to different word order distributions, however, it is important for us here that this difference is not due to the bias against Adj≺N≺Num.

substantial variation of both these orders. The languages in our sample that show the required variations are Latin and Ancient Greek. We quantitatively analyse the corpora of these languages to confirm the predictions on observed data. Specifically, we collect counts corresponding to $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$,$\boldsymbol{\gamma}$. Using these counts, we test the null hypothesis that observed counts $\boldsymbol{\gamma}$ correspond to expected counts sampled from the distribution $\boldsymbol{\alpha}\cdot\boldsymbol{\beta}$. Accepting the null hypothesis will give evidence in favour of grammar-level accounts; rejecting the null hypothesis will provide evidence in favour of structural accounts. In the next sections, we describe our data, the procedure to collect the counts and present our results.

4.1 Cross-linguistic corpus data

Corpus data constitute important evidence for studying syntactic variation. To induce reliable generalisations from corpora, one needs to ensure that the sample of extracted constructions is both accurate and representative of the language's grammar and variation. We use gold syntactically annotated corpora, since we analyse certain structural constructions in a language and not just collocations of words. A drawback of using gold-annotated corpora is their limited size since the manual syntactic annotation is expensive to produce. Most of the recent corpora come in dependency format. The features necessary for our task include part-of-speech (PoS) tags of the words, the head of the current token (the dependent of the head), and the type of dependency relation between the token and the head. These basic features are illustrated in Figure 2, for two typologically different languages. For our investigation, we extract all noun phrases that contain either adjective or numeral modifiers (or both) from our corpora.

Corpus data for Latin and Ancient Greek

The corpora for Latin and Ancient Greek come from the Perseus project (Bamman & Crane 2011) and from the PROIEL project (Haug & Jøhndal 2008). The Latin Perseus corpus is small, and in our study of the co-occurrences of adjectives and numerals in a noun phrase, we observed only 16 such cases. The Ancient Greek Perseus corpus consists for the most part of poetic works (including Homer's Iliad and Odyssey, Aeschylus' Agamemnon and others). Rhythmic constraints in poetic work significantly affect the choice of word order and it could therefore influence our results. The PROIEL Ancient Greek corpus contains exclusively prose and is therefore more appropriate for a word order variation study. We therefore base our analyses and conclusions primary on the PROIEL sources. The Perseus treebanks remain nevertheless important additional data which corroborate our results.

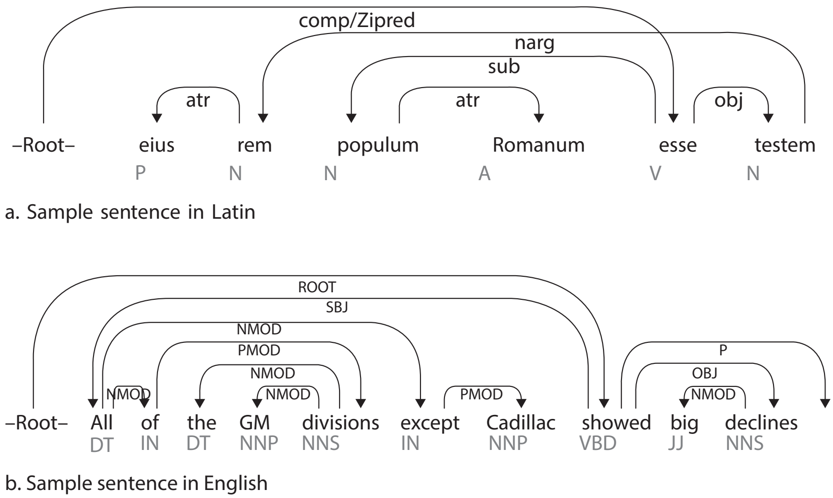

Figure 2. Illustration of dependency annotation, visualised as dependency arcs and their functional labels

Preprocessing and collection of counts

We preprocess the PROIEL corpora to obtain the universal PoS tags (Petrov et al. 2012) and dependencies as our basic annotation information. The original annotation of the PROIEL corpora distinguishes between cardinal and ordinal numerals. The case of Universal 18 concerns only cardinal numerals, which differ in their structural and functional properties from ordinal numerals. The preprocessing ensures that only numerals tagged as cardinals in the original corpora are included in the analysis.

We collect two types of counts for our analysis. First of all, to estimate the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ distributions, we extract the frequencies of pre- and post-nominal adjective-noun and numeral-noun word orders. In addition, to estimate the $\boldsymbol{\gamma}$ distribution, we extract the same pair-wise word order frequencies in the restricted context of noun phrases that contain both an adjective and a numeral. The total counts for word order combinations are then computed from the output.

Collected counts

Table 5 displays raw counts and relative frequencies – α and β parameters – for adjective-noun and numeral-noun word order pairs collected from the Ancient Greek and Latin PROIEL corpora. We present frequencies for each literary work separately.

Table 5. The percentages of pre-nominal and post-nominal placement of numerals and adjectives in the Ancient Greek and Latin PROIEL corpus. Parameter α is estimated as the percentage of Adj≺N and β is estimated as the percentage of Num≺N. Abbreviations: Chr. = Chroniche; NT = New Testament; Her. = Herodotus; Per. = Peregrinatio; Vul. = Vulgate

	Chr.	**NT**	**Her.**	**Total**	**Caesar**	**Cicero**	**Per.**	**Vul.**	**Total**
Adj≺N	285	648	768	1701	614	629	320	412	1975
N≺Adj	111	1042	978	2131	238	462	272	794	1766
α	**72.0**	**38.3**	**44.0**	**44.4**	**72.1**	**57.7**	**54.1**	**34.2**	**52.8**
Num≺N	26	287	215	528	88	71	65	257	481
N≺Num	39	146	101	286	58	26	18	118	220
β	**40.0**	**66.3**	**78.0**	**64.9**	**60.3**	**73.2**	**78.3**	**68.5**	**68.6**

Table 6 displays observed co-occurrence counts for adjective-noun and numeral-noun word orders – the counts used to estimate the γ distribution in context C_2. The tables also include the respective *expected* co-occurrence counts: the counts that would be produced under the joint distribution $\alpha \cdot \beta$ (α and β are estimated from Table 5) given that the context C_2 of a noun phrase containing both an adjective and a numeral was not different from the general context C_1. Each expected count for word orders x and y is calculated proportionally to $\alpha(x) \cdot \beta(y)$ and the observed total number of noun phrases where two modifiers co-occur. We compare the observed and the expected counts to draw the generalisations about constraints on adjective and numeral syntactic distributions.

Table 6. Co-occurrence counts in the Ancient Greek and Latin PROIEL corpus, divided by source text. Abbreviations: Chr. = Chroniche; NT = New Testament; Her. = Herodotus; Cae. = Caesar; Cic = Ciceroo; Per. = Peregrinatio; Vul.=Vulgate.

		Chr.	**NT**	**Her.**	**Tot**	**Cae.**	**Cic.**	**Per.**	**Vul.**	**Tot**
					Observed counts, γ					
Adj≺N	Num≺N	1	10	3	14	5	3	2	2	12
Adj≺N	**N≺Num**	**0**	**2**	**2**	**4**	**1**	**1**	**0**	**2**	**4**
N≺Adj	N≺Num	1	3	13	17	7	4	3	3	17
N≺Adj	Num≺N	1	9	15	25	1	5	12	15	33
	Total	3	24	33	60	14	13	17	22	66
					Expected counts, $\alpha \cdot \beta$					
Adj≺N	Num≺N	0.9	6.1	9.9	17.3	6.1	5.5	7.2	5.2	23.9
Adj≺N	**N≺Num**	**1.3**	**3.1**	**4.6**	**9.4**	**4.0**	**2.0**	**2.0**	**2.4**	**10.9**
N≺Adj	N≺Num	0.5	5.0	5.9	11.7	1.6	1.5	1.7	4.6	9.8
N≺Adj	Num≺N	0.3	9.8	12.6	21.6	2.4	4.0	6.1	9.9	21.4

4.2 Results and discussion

Overall there were only 60 cases for Ancient Greek and 66 for Latin of noun phrases with both an adjective and a cardinal numeral modifier. If we take each work separately, the counts are insufficient to allow statistically significant tests. However, combining observations for multiple texts, and especially for the two languages together, we can perform a relevant cumulative statistical analysis. An appropriate analysis would include a multilevel statistical model, such as a mixed-effect model, with word order distribution parameters varying for each text and a bias parameter equal for all texts and languages. Unfortunately, this approach is not applicable given the small number of observations for individual texts. We assume that the cumulative χ^2 statistical tests we present here provide a sufficient approximation to the actual levels of statistical significance.

The main observation arising from the data in Table 6 is that the word order combination Adj≺N≺Num is always underobserved with respect to the expected counts. It is true for both languages and for all texts, despite the few counts that are available for analysis. Cumulatively, while the observed four-way distribution for Ancient Greek is not statistically different from the expected distribution (χ^2 = 6.6, p = 0.085), the difference between counts is significant for Latin (χ^2 = 22, p < 0.001). Most importantly, we evaluate a possible bias against the Adj≺N≺Num word order by comparing statistically expected and observed two-way distributions: Adj≺N≺Num word order versus all other word order combinations. For Ancient Greek we therefore compute the χ^2 test on pairs of counts: observed counts (4; 56) versus expected counts (9.4; 50.6), which is marginally significant (χ^2 = 3.6, p = 0.056). For Latin, the difference between observed counts and expected counts, (4; 62) v (10.9; 61.1), is also significant (χ^2 = 5.3, p = 0.02). Furthermore, combining counts for the two languages and testing the two-way distribution yields a significance level of p = 0.003. Despite the limited number of co-occurrences of adjectives and numerals, our data gives a very interesting and at least marginally significant observation: the Adj≺N≺Num word order, corresponding to the dispreferred order in Universal 18, is under-observed in all cases.

Note that counts for other word order options do not show a similarly constant pattern. In Latin, Adj≺N & Num≺N order is also significantly under-observed, but this does not hold for Ancient Greek. We also controlled for the size of numeral and adjective phrases as a possible explanatory factor in word order variation. The distribution of simple and complex (longer than one word) adjective phrases and numeral phrases were the same for both samples: the sample of pair-wise word orders and the sample of word orders in co-occurrence. This factor therefore could not be the reason for the persistent differences in two distributions of word orders.

On the basis of the statistical tests and observations above, we therefore reject the null hypothesis corresponding to the prediction of the grammar-level accounts. The distributions for modifiers on their own and together differ in a way which confirms that they are not independent and, moreover, that there is a bias against one particular word order combination: we have observed that when adjectives and numerals co-occur there is a statistical tendency to skew the distributions away from the typologically dispreferred Adj≺N≺Num word order.

Our final conclusion is two-fold: first, the word order variation data support the accounts of Universal 18 that argue for its language-internal origin; secondly, our data speaks in favour of the token-level accounts and against the type-level accounts. Specifically, the data points to a linguistic bias which constrains token-level distributions and not type-level properties. Recall that, although Universal 18 is stated typologically for dominant word orders, we have observed, based on the data in SSWL, that its structural interpretation is also confirmed for a number of languages. Therefore, there seems to be converging evidence in favour of structural accounts.

Nevertheless, this view faces problems when attempting to explain both the experimental results of Culbertson et al. (2012) and our corpus data under a single theory. Culbertson et al. (2012) mention FOFC as a potential explanation for Universal 18. However, some additional considerations are necessary to unify type-level and token-level accounts. In the learning experiment, participants received input adjective phrases and numeral phrases independently. They never observed the realisations of the dispreferred Adj≺N≺Num order. The learners, however, still showed a bias against the language type L2 (Adj≺N & N≺Num). To claim that this is expected under FOFC or any other structural account, one needs to specify how the learners induced the underlying grammar and particularly, how they induced the structure of the complete noun phrase from the input data they were given. One possibility is to say that the complete NP structure is entertained even if only a partial overt realisation (e.g., adjective plus noun) is observed. Such an assumption is in line with frequent adherence to empty projections in generative syntax and an extended NP phrase structure (with different adjective functional projections). If this is the case, then this explains the results of Culbertson et al. (2012), but we still need to answer why in our corpus data there is an additional bias observed in noun phrases with both an adjective and a numeral compared to distributions for partial NPs. We could speculate that a structural constraint is triggered when all the projections are overt.

5. Towards a model explaining Universal 18

In this section, we sketch the reasoning needed to develop an explanatory model of the observed corpus distributions of nominal modifiers for Latin and Ancient Greek. We would like to statistically compare a model which incorporates a bias predicted by the Universal 18 and a model without such bias. Following the discussion in the previous sections, crucially, our problem includes two types of data – adjective-noun and numeral-noun occurrences (call these data D_1) and co-occurrences of adjectives and numerals inside a single noun phrase (call these data D_2). Given these data of two distinct but related types, we define the grammatical phenomenon (or process) we want to model as what is expressed in D_2 and Universal 18. In particular, our abstract models must specify what their parameters are, how the parameters are estimated from the data, and which data are used to test the competing models.

The third question receives a natural answer right away, as it is D_2 which shows the biased distribution of modifiers and, therefore, is the natural choice for testing the significance of the bias effect. In the following, we will try to narrow down the possible answers for the first two questions.

The first two intuitive parameters that we assume at the moment are α and β, which are the probabilities for the Adj<N and Num<N orders. These are also the parameters that Culbertson and colleagues use in their experiment and its Bayesian modelling. A word order model that has only these two parameters (we can call it $G(\alpha,\beta)$) is a minimal model. Figure 3(a) (Model 1) depicts this minimal model with two parameters α and β. Importantly, this model assumes that D_1 and D_2 are generated by the same grammar. This also means that we can choose to estimate the model parameters from D_1, D_2 or both.

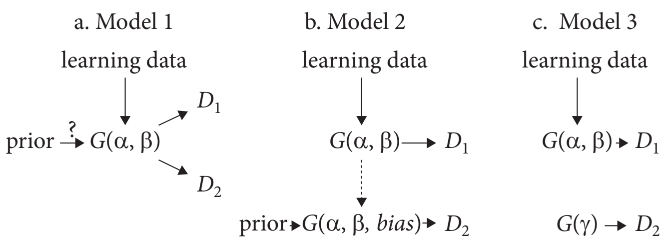

Figure 3. Three models applicable to the data D1 and D2

On the opposite side of the modelling spectrum, we can devise a maximal model (Figure 3(c), Model 3). This model assumes that the two types of data are generated by two different grammars. While D_1 is generated by $G(\alpha,\beta)$ (which is a maximal model for D_1 in the sense that it uses two parameters to explain two independent binary distributions), D_2 is generated by a four-way distribution γ. The parameters are straight-forwardly estimated from D_1 and D_2 correspondingly. Model 3 has a higher number of free parameters than Model 1, and therefore can fit the data better. For our data set the resulting log-likelihood of Model 3 was equal to −24.13. However, this model does not reflect our intuition that $G(\alpha,\beta)$ and $G(\gamma)$ are related and does not really correspond to any existing theory. Even usage-based theories would predict, we think, a link between α and β and γ, possibly with a data dependency from γ to α and β. For this reason we do not further discuss Model 3.

From this point of view, the sketch of Model 2 (Figure 3 (b)) presents a more satisfactory model. The property of such model would be the coupling of the parameters of $G(\alpha,\beta)$ and $G(\gamma)$. In particular, we can assume that $G(\gamma)$ is a combination of $G(\alpha,\beta)$ and a structural bias which could be modelled with as few as one parameter. While the α and β parameters can be estimated from D_1, the bias parameter(s) must be necessarily estimated from D_2.

Formalized in this way, the models are easily compared to each other. Comparison of Model 1 and Model 2 can answer the question whether there is a significant bias effect in the co-occurrence of modifiers.

5.1 Comparison of models 1 and 2

We assumed that both models are probability distributions (corresponding to probabilities of four possible word order combinations: $p1 + p2 + p3 + p4 = 1$). We computed the likelihoods of the two models on D_2 data given their parameters estimated from D_1 data for Model 1 and both D_1 and D_2 data for Model 2.

Specifically, in both cases we calculate the likelihoood as a multinomial dependent on the four parameters, ($C(\cdot)$ indicates frequency counts):

$$p(D_2 \mid \text{Model 1}) = Mult\left(C(D_2) \mid p_1, p_2, p_3, p_4\right) \tag{5}$$

For Model 1, the probabilities of the distribution come from the $\alpha\cdot\beta$ joint probabilities:

$$p_1 = \alpha\cdot\beta;\ p_2 = (1-\alpha)\cdot\beta;\ p_3 = \alpha\cdot(1-\beta);\ p_4 = (1-\alpha)\cdot(1-\beta). \tag{6}$$

For Model 2, we introduce one more parameter for the bias as follows, where b is a normalisation parameter, so that $p1 + p2 + p3 + p4 = 1$:

$$\begin{aligned} p_3 &= p(\text{Adj} \prec \text{N \& N} \prec \text{Num}) = \alpha\cdot(1-\beta)\cdot bias \\ p_1 &= \alpha\cdot\beta\cdot b \\ p_2 &= (1-\alpha)\cdot\beta\cdot b \\ p_4 &= (1-\alpha)\cdot(1-\beta)\cdot b \end{aligned} \tag{7}$$

The bias parameter was estimated on D_2 data using MLE estimation. In other words, the likelihood of Model 2, $p(D_2|Model2)$, is a function of the parameter *bias* and we choose the value of this parameter which maximises the likelihood:

$$bias^* = arg \max_{bias\in[0,1]} Mult_{bias}(D_2 \mid p_1, p_2, p_3, p_4). \tag{8}$$

Note that *bias* is defined to be less than 1, which corresponds to a decreased probability of the dispreferred word order Adj≺N & N≺Num with respect to Model 1.

Model 2 is therefore an extension of Model 1 with one added parameter. The resulting log-likelihoods on D_2 for the two models were log-Likelihood(Model 1) = −55.31 and log-Likelihood(Model 2) = −50.37. We use these log-likelihood values to test the fit of the models. Note that Model 2 has a higher number of free parameters than Model 1, and therefore can fit the data better. Model 2 uses two free parameters per text, overall fourteen parameters, plus one global bias parameter, resulting in fifteen parameters, while Model 1 has the same number of free parameters but no bias, that is fourteen for our set of seven texts. According to a χ^2-squared test, a test that takes into account the degrees of freedom, the difference between the fit of Model 1 and the better fit of Model 2 is statistically significant at $p < 0.01$.

6. Conclusions

Our results, combined with previous theoretical proposals and experimental findings, can provide new evidence for the debate about language universals, in particualr for the competing explanations provided by generative syntax and usage-based theories of language. The experimental results can be seen as evidence that learners acquire complete noun-phrase structure (including the hierarchy of relevant adjectival and numeral heads) without seeing actual realisations of them. This conclusion is not surprising under current assumptions in generative syntax, which postulate the innate knowledge of basic grammatical structure. Our corpus counts show a clear numerical under-representation corresponding to Universal 18, thereby corroborating this hypothesis. Preliminary definition and testing of explanatory models show that a model where the grammar-level parameters are modulated by a structure-specific bias is more strongly supported by the evidence than a model without bias.

The study of Universal 18 demonstrates that quantitative corpus-based approaches to syntactic variation can inform formal syntactic approaches and in particular can provide more precise specification of the models, thereby providing the means for precise models selection. We showed that by studying quantitative patterns of NP-internal word order variation in Latin and Ancient Greek we can provide new evidence that discriminates between two types of theoretical accounts, thereby teasing apart fine-grained difference in proposals about a typological universal by looking at language-internal word order variation.

References

Abels, Klaus. 2013. “On 2-3-1.” *UCL Working Papers in Linguistics* 25.

Baker, Mark. 2001. *The Atoms Of Language: The Mind’s Hidden Rules of Grammar*. New York, NY: Basic Books.

Bamman, David, and Gregory R. Crane. 2011. “The Ancient Greek and Latin Dependency Treebanks.” In *Language Technology for Cultural Heritage* (Theory and Applications of Natural Language Processing), ed. by Caroline Sporleder, Antal Bosch, and Kalliopi Zervanou, 79–98. Berlin Heidelberg: Springer. https://doi.org/10.1007/978-3-642-20227-8_5

Biberauer, Theresa Anders Holmberg, and Ian Roberts. 2014. “A Syntactic Universal and its Consequences”. *Linguistic Inquiry* 45 (2): 169–225. https://doi.org/10.1162/LING_a_00153

Bresnan, Joan, Anna Cueni, Tatiana Nikitina, and Harald Baayen. 2007. “Predicting the Dative Alternation.” In *Cognitive Foundations of Interpretation*, ed. by G. Boume, I. Kraemer, and J. Zwarts, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Bresnan, Joan, Shipra Dingare, and Christopher D. Manning. 2001. “Soft Constraints Mirror Hard Constraints: Voice and Person in English and Lummi.” In *Proceedings of the LFG 01 Conference*, 13–32. CSLI.

Buchholz, Sabine, and Erwin Marsi. 2006. “CoNLL-X Shared Task on Multilingual Dependency Parsing.” In *Proceedings of the Tenth Conference on Computational Natural Language Learning, CoNLL-X ‘06*, 149–164. Stroudsburg, PA, USA: Association for Computational Linguistics. http://dl.acm.org/citation.cfm?id=1596276.1596305.
https://doi.org/10.3115/1596276.1596305

Cinque, Guglielmo. 2005. “Deriving Greenberg’s Universal 20 and its Exceptions”. *Linguistic Inquiry* 36 (3): 315–332. https://doi.org/10.1162/0024389054396917

Culbertson, Jennifer, and David Adger. 2014. “Language Learners Privilege Structured Meaning Over Surface Frequency”. *Proceedings of the National Academy of Sciences of the United States of America* 111 (16): 5842–5847. https://doi.org/10.1073/pnas.1320525111

Culbertson, Jennifer, and Paul Smolensky. 2012. “A Bayesian Model of Biases in Artificial Language Learning: The Case of a Word-Order Universal”. *Cognitive Science* 36 (8): 1468–1498.
https://doi.org/10.1111/j.1551-6709.2012.01264.x

Culbertson, Jennifer, Paul Smolensky, and Géraldine Legendre. 2012. “Learning Biases Predict a Word Order Universal”. *Cognition* 122 (3): 306–329.
https://doi.org/10.1016/j.cognition.2011.10.017

Cysouw, Michael. 2010. “Dealing with Diversity: Towards an Explanation of NP Word Order Frequencies”. *Linguistic Typology* 14 (2): 253–287. https://doi.org/10.1515/lity.2010.010

Dryer, Matthew S. 1989. “Discourse-Governed Word Order and Word Order Typology”. *Belgian Journal of Linguistics* 4: 69–90. https://doi.org/10.1075/bjl.4.05dry

Dryer, Matthew S. 1992. “The Greenbergian Word Order Correlations”. *Language* 68: 81–138. https://doi.org/10.1353/lan.1992.0028

Dryer, Matthew S. 1995. “Frequency and Pragmatically Unmarked Word Order.” In *Word Order in Discourse*, ed. by Mickey Noonan, and Pamela Downing, 105–135. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.30.06dry

Dryer, Matthew S., and Martin Haspelmath (eds.). 2011. *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. http://wals.info/

Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson, and Russell D. Gray. 2011. “Evolved Structure of Language Shows Lineage-Specific Trends in Word-Order Universals”. *Nature* 473 (7345): 79–82. https://doi.org/10.1038/nature09923

Greenberg, Joseph H. 1963. “Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements.” In *Universals of Language*, ed. by Joseph H. Greenberg, 73–113. MIT Press.

Gulordava, Kristina, and Paola Merlo. 2015. "Structural and Lexical Factors in Adjective Placement in Complex Noun Phrases across Romance Languages". In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*, 247–257. Beijing, China.

Gulordava, Kristina, 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), 477–482. Beijing, China.

Hajič, Jan, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, Maria Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, Pavel Straňák, Mihai Surdeanu, Nianwen Xue, and Yi Zhang. 2009. "The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages." In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task, CoNLL '09*, 1–18. Stroudsburg, PA, USA: Association for Computational Linguistics. http://dl.acm.org/citation.cfm?id=1596409.1596411

Haspelmath, Martin. 2006. "Against Markedness (and What to Replace it With)". *Journal of Linguistics* 42 (01): 25–70. https://doi.org/10.1017/S0022226705003683

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

Hawkins, John A. 1983. *Word Order Universals*. New York: Academic Press.

Hawkins, John A. 1994. *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.

Hawkins, John A. 2009. "Language Universals and the Performance-Grammar Correspondence Hypothesis." In *Language Universals*, Chapter 4, ed. by M. H. Christiansen, C. Collins, and S. Edelman, 54–79. Oxford University Press.
https://doi.org/10.1093/acprof:oso/9780195305432.003.0004

Holmberg, Anders. 2000. "Deriving OV Order in Finnish." In *The Derivation of VO and OV*, 123–152. Amsterdam: John Benjamins. https://doi.org/10.1075/la.31.06hol

Hurford, James R. 2003. "The Interaction between Numerals and Nouns." In *Noun Phrase Structure in the Languages of Europe*Volume 20–7 of Empirical Approaches to Language Typology, ed. by F. Plank. Mouton de Gruyter.

McDonald, Ryan, Joakim Nivre, Yvonne Quirmbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, Claudia Bedini, Núria Bertomeu Castelló, and Jungmee Lee. 2013. "Universal dependency annotation for multilingual parsing." In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (Volume 2: Short Papers), 92–97. Association for Computational Linguistics. http://aclweb.org/anthology/P13-2017

Merlo, Paola, and Suzanne Stevenson. 2001. "Automatic Verb Classification Based on Statistical Distributions of Argument Structure". *Computational Linguistics* 27 (3): 373–408.
https://doi.org/10.1162/089120101317066122

Merlo, Paola, and Suzanne Stevenson. 2004. "Structure and frequency in verb classification." In *Proceedings of Incontro di Grammatica Generativa XXX*, ed. by Laura Brugé, Giuliana Giusti, Nicola Munaro, Walter Schweikert, and Giuseppina Turano, Venice, Italy.

Mithun, Marianne. 1992. "Is Basic Word Order Universal? In *Pragmatics of Word Order Flexibility*, ed. by Doris L. Payne, 16–61. Amsterdam: John Benjamins.

Ouwayda, Sarah. 2014. “Where Number Lies: Plural Marking, Numerals, and the Collective-Distributive Distinction.” Ph.D. thesis, University of Southern California.

Petrov, Slav, Dipanjan Das, and Ryan McDonald. 2012. “A Universal Part-of-Speech Tagset.” In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, ed. by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, 2089–2096. Istanbul, Turkey, European Language Resources Association (ELRA).

Rijkhoff, Jan. 1998. “Order in the Noun Phrase of the Languages of Europe.” In *Constituent Order in the Languages of Europe*, Volume 20–1 of Empirical Approaches to Language Typology, ed. by Anna Siewierska, 321–382. Mouton de Gruyter.

Samardžić, Tanja, and Paola Merlo. 2018. “Probability of External Causation: An Empirical Account of Cross-Linguistic Variation in Lexical Causatives”. *Linguistics* 56 (5): 895–938. https://doi.org/10.1515/ling-2018-0001

Samardžić, Tanja, and Paola Merlo. 2012. “The Meaning of Lexical Causatives in Cross-Linguistic Variation”. *Linguistic Issues in Language Technology* 7 (1).

Schwarzschild, Roger. 2009. “Stubborn Distributivity, Multiparticipant Nouns and the Count/Mass Distinction.” In *Proceedings of the 40th Annual Meeting of the North East Linguistic Society*, Volume 39, 1–18. Mass.: MIT.

Sheehan, Michelle, Theresa Biberauer, Ian Roberts, and Anders Holmberg. 2017. *The Final-Over-Final Condition: A Syntactic Universal*, chapter The Final-Over-Final Condition and the Head-Final Filter. MIT Press. https://doi.org/10.7551/mitpress/8687.001.0001

Svenonius, Peter. 2007. “1…3-2.” In *Oxford Handbook of Linguistic Interfaces*, ed. by Gillian Ramchand, and Charles Reiss, 239–288. Oxford University Press.

Williams, Edwin. 1982. “Another Argument that Passive is Transformational”. *Linguistic Inquiry* 13: 160–163.

Author Query
Please provide an affiliation (if present) for the authors of this chapter.