Samuel Andersson, QP1 (Yale University)   2019-03-26

The Evolutionary Syntax of Universal 20[1]

1 Introduction

This paper discusses a typological framework that I refer to as Evolutionary Syntax, and how it applies to Universal 20 word order patterns: the relative order of demonstrative (Dem), numeral (Num), adjective (A), and noun (N). In recent years, especially after the publication of Cinque (2005), the typology of Universal 20 has become a popular research topic. However, word order patterns in the noun phrase have been studied from a typological perspective ever since Greenberg (1966). This paper builds on the typological work that has already been done, and shows that the word order patterns can be explained diachronically: common word orders are those that many independently-motivated pathways of change converge on, while rare word orders require rare historical changes to arise.

Many different explanations for Universal 20 word order patterns have been proposed, all based on slightly different typological samples, and using different typological frameworks. This paper explores Universal 20 from the perspective of diachronic explanation, an increasingly popular approach to typology in phonology (Blevins 2004), morphology (Anderson 2016), syntax, and semantics (Cristofaro 2010). Following Blevins' (2004) Evolutionary Phonology, and Andersson's (2018a) Evolutionary Morphology, I call the approach taken in this paper Evolutionary Syntax. It is assumed that typological distributions are the consequence of pathways of change, without any influence from goal-oriented pressures (as in functional-typological approaches) or constraints imposed by synchronic linguistic representations (as in most generative work on typology). This paper, therefore, is not primarily concerned with how to represent speakers' synchronic competence, nor with functional pressures towards efficient communication and the like. Our focus will be on diachronic pathways which produce certain types of word orders more often than others, and how these pathways can be grounded in factors such as acquisition, processing, and usage frequencies.

It is argued here that the typological distribution of word orders in the noun phrase follows from changes based in processing. I use the word order principles from Dryer (2018) as a starting point, since these have already been shown to account for the relative frequencies of noun phrase word orders. However, I suggest that these principles, rather than simply being stipulated, are the natural outcome of language change. By making reference to findings from the literature on working memory, syntactic processing, and syntactic priming, I show that language change will tend to produce languages which obey Dryer's principles more often than they produce languages which disobey them. In summary, then, the typology of Universal 20 is constrained by diachrony, and diachrony is constrained by processing factors. This view allows for a purely diachronic explanation of Universal 20, which hopefully serves as a useful case study for future work on Evolutionary Syntax.

The rest of this paper is structured as follows. Section 2 introduces some different approaches to typological patterns, and covers synchronic representational, functional-typological, and diachronic (or evolutionary) theories. Section 3 focuses on previous work on Universal 20, divided according to the typological approaches above. Section 4 discusses Dryer's (2018) approach to Universal 20, and attempts to identify diachronic explanations for his word order principles. This includes new typological results on the connection between word length and word order, as well as discussion of how processing factors constrain word order change. Section 5 concludes.

2 Typological frameworks

This section discusses some of the different ways that linguists have analyzed typological asymmetries, such as one word order being more common than another. As these frameworks are in no way tied to particular subfields of linguistics, the examples and illustrations in this section will not be limited to syntax. We begin by surveying what is arguably the most common way of analyzing typological data within generative theories of linguistics. This will be followed by a discussion of the traditional functional-typological approach. Finally, we will examine the framework of diachronic explanation used in this paper.

2.1 Synchronic representational approaches

The traditional goal of generative linguistics is to provide accounts of the synchronic grammatical competence of speaker-listeners, or I-languages (Chomsky 1965: 3-5, 1986: 21ff). The pursuit of this goal has led to many generative theories of what these representations are, including syntactic approaches such as Phrase Structure Rules, X'-theory, Government and Binding, and proposals within the Minimalist Program. Although these all attempt to explain properties of I-languages, generative linguistics has long attempted to use these same representations to account for typological data, one example being the Principles and Parameters framework (see Chomsky and Lasnik 1993 among others). Below I give what I see as the implicit hypothesis underlying such work:

> The synchronic representational hypothesis: Typological data can
> be explained in terms of the same representations used to provide
> synchronic accounts of I-languages

Subscribing to this hypothesis does not in any way entail that *all* typological data must or can be explained in this way. However, in practice, individual generative analyses typically assume that the synchronic representational hypothesis holds for the data being examined. Examples of influential work of this type include Cinque's (1999) work on adjective ordering, and Caha's (2009) research on case hierarchies.

This hypothesis about the relation between typology and synchronic representation is in no way limited to morphosyntax. The clearest phonological example is Optimality Theory (McCarthy and Prince 1993; Prince and Smolensky 1993). OT is a theory of the synchronic phonological representation of I-languages, but is explicitly designed to account for typological data. Postulating a set of violable and freely-rerankable universal constraints, OT predicts that cross-linguistic variation results from constraint reranking. This means that the set of possible languages in a given domain can be relatively easily identified by seeing what type of language is generated under each possible ranking of the relevant constraints. This property of OT is known as factorial typology, and is often held to be an advantage of the theory (Kager 1999).

3

It is commonly assumed in the synchronic representational research tradition that the set of attested languages in a typological sample is the same as the set of possible languages. This position is clearly stated by Smith et al. (2018) in the context of case and number suppletion, but the view can be found in most generative typological work: "[T]he unattested patterns do not arise as they cannot be generated in a manner consistent with Universal Grammar" (Smith et al. 2018: 3). Here a theory is pursued where all patterns which are unattested in a typological sample are assumed to be impossible in natural language. Cinque (2005: 316, fn. 4) says that while further word orders in the noun phrase may be discovered in the future, his approach assumes that what is currently unattested is underivable. This approach to attested vs possible languages is questioned by the approaches discussed in 2.3 below.

2.2 Functional-typological approaches

In traditional work on functional-typological linguistics, typological data are explained without reference to formal representations of speaker competence. Various factors which fall under performance in generative approaches are key, and the fact that language is often used for communication between human beings with communicative goals is central. This focus on certain patterns being favored over others makes these approaches teleological (also 'goal-oriented' or 'functionalist'). The goals driving typological patterns may be ease of communication, articulation, perception, acquisition, processing, and/or many others. Important work in this tradition has been done by Greenberg (1966), Comrie (1981), Givón (1984), Hawkins (1988), and Croft (1990), to name a few examples.

Work in the functional-typological tradition sometimes agrees with generative analyses about the basic principles and generalizations. For example, when discussing the typology of vowel inventories, both functional-typological and generative treatments agree on the generalization that vowels are preferentially dispersed in acoustic space (see references in Vaux and Samuels 2015). But while generative analyses in Optimality Theory attribute this to synchronic phonological constraints (as in Flemming 1995), the functional-typological approach would be that dispersed systems strike the best balance between articulatory ease and lack of

perceptual confusion (Lindblom 1986). Lindblom's work is explicitly functional, arguing that "language form is forged by the sociobiological conditions of its use" (Lindblom 1986: 13), which in this case amounts to constraints on articulation and perception.

Functional-typological analyses are often stated in terms of a set of principles, or generalizations, which explain the data observed. One example is 'Bartsch's principle of natural serialization' (Vennemann 1975: 80), stating that languages tend to have a consistent order between heads and their modifiers. Vennemann interprets this as a tendency for languages to move towards more consistent head-modifier orders, while Jacques (2013) has interpreted it as an inviolable constraint on change, and argues against it on that basis. A third possibility is to treat principles as synchronic tendencies, as with many of Greenberg's (1966) universals. On this third interpretation, the data are not fully explained, but only summarized in the form of significant generalizations. These can then serve as the basis for more grounded explanations. This view is adopted by Dryer, who calls some of his own word order principles "explanatory only in a superficial sense" (Dryer 2018: 801). We will return to deeper explanations for these principles, which are related to Universal 20, in section 4.

2.3 Diachronic approaches

Diachronic (or evolutionary) approaches to typology are characterized by the assumption that typological patterns are consequences of the way languages change. The explanations proposed are thus typically independent of what one's theory of synchronic analysis is: a diachronic explanation of a syntactic pattern does not depend on whether Minimalism or Construction Grammar is chosen as a theory of synchronic syntax. Patterns of change are themselves explained by factors such as acquisition, usage frequencies, perceptual biases, and so on, depending on the pattern under study. Another noteworthy feature, which distinguishes this framework from the ones discusses earlier in this section, is the lack of goal-oriented reasoning. Final devoicing is not common because voiceless obstruents are preferred, but simply because there are many independently-motivated phonetically-grounded sound changes which cause languages to develop this sound pattern (Blevins 2004: ch. 4). The prevalence of one and the

same output (final devoicing) is a consequence of the individual changes involved: devoicing for aerodynamic reasons (Ohala 1997, Passy 1890), domain-final lengthening which itself leads to voicelessness (Blevins 2004: 104-105), and so on. Convergence on a particular output, then, does not need to explained by incorporating any goal-oriented principles into our typological framework.[2]

As an example of a diachronic explanation, palatalization before high front vowels is common because there are articulatory and perceptual reasons which give rise to palatalizing sound changes in this environment (see Grammont 1933, Anttila 1989, Hock 1991, Ohala 1992, Guion 1998). Thus, over time, we expect many languages without palatalization to develop it through phonetically-motivated sound change. By contrast, there are few or no changes which lead to depalatalization, so that if a language undergoes palatalization, it is not likely to lose it again. This in turn leads to the observed synchronic distribution, where palatalization is a common phonological pattern. The phonological example above reflects the fact that diachronic explanations are particularly well-studied in this subfield. Work which assumes this perspective includes Evolutionary Phonology (Blevins 2004, 2006 et seqq.), and Substance-Free Phonology (Hale and Reiss 2008, Bale and Reiss 2018, among others).

However, diachronic explanation is far from limited to phonology. Anderson (2016) discusses a few case studies from morphology and syntax, as well as the general theory. The paper discusses the relationship between aspect and case systems in languages which have both nominative/accusative and ergative/absolutive alignment. Various paths of grammaticalization converge on the result "in which the perfect is associated with ergative marking and the imperfect with accusative marking" (Anderson 2016: 23; see Anderson 2004 for a detailed discussion, and Cristofaro 2012 for more discussion on the diachrony of alignment systems). Andersson (2017) discusses the typology of gender syncretism, showing that diachronic factors based in usage frequencies can explain the distribution of masculine-feminine syncretism in the languages which have it. Also in the domain of morphology, Andersson (2018a) argues that an apparent *ABA restriction (see Bobaljik 2012) in the domain of Germanic verbs should be

---

2   For simulations showing how typological asymmetries can arise from random phonetically-motivated sound change, see Sayeed and Ceolin (2018).

explained diachronically, with reference to Proto-Germanic verb classes. In semantics, important work on so-called semantic maps (encoding restrictions on which meanings can be expressed with the same word cross-linguistically) has been done by Cristofaro (2010). Rather than treating the patterns identified in the literature as universal constraints on colexification, Cristofaro (2010) shows that many of the patterns can be understood through a diachronic process of metonymization. It is clear, then, that especially in recent years, there is much research on diachronic explanation in a wide variety of linguistic areas.

It is worth commenting on the division of labor between synchrony and diachrony assumed in this work. As noted above, it is possible to provide diachronic explanations without ever discussing synchronic analyses. However, in many cases a diachronic account accounts for the same data as synchronic representational or functional-typological explanations. In such cases, an argument from Occam's Razor can often be made in favor of a purely diachronic story. Blevins (2004), for example, argues that Evolutionary Phonology captures the same set of phenomena usually attributed to phonological markedness in synchronic representational theories of typology. This means that we can explain markedness phenomena without assuming rich synchronic representations of markedness constraints, hierarchies between these constraints, etc. (see de Lacy 2006 for one such proposal). By Occam's Razor, we should not make these assumptions unless we absolutely have to, and a diachronic explanation gives us the same patterns "for free" (Bybee 2010: 111). For stronger arguments against synchronic representational analyses of typological patterns, see Andersson (2018b) on voicing in obstruents, and Sayeed (2018) on compensatory lengthening.

If we take the approach above, this also entails a different view of attested and possible languages. The set of possible languages is a strict superset of the set of attested languages, and there may be hundreds of millions of possible human languages which simply happen not to be attested in the small sample studied by linguists. The fact that a particular language type is not attested in a typological sample, then, does not tell us much about whether or not it is possible. The strongest diachronic position to take would be that *all* unattested properties in human languages can be explained solely through diachrony, although in reality, few probably have such

a strict interpretation.

3 Previous work on Universal 20

This section is a non-exhaustive survey of some of the work that has been done on Universal 20. We will begin with the classic formulation by Greenberg (1966), who gave this area of research its name. Sections 3.2 and 3.3 describe the subsequent decades of research, with 3.2 covering functional-typological approaches, and 3.3 synchronic representational ones. For sociological reasons, this division yields an almost chronological order.

3.1 Greenberg

Greenberg's work on typology has been hugely influential both empirically and methodologically. One of the most famous articles is "Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements," originally published in 1963, and then again in Greenberg (1966). A sample of 30 languages was used for large parts of this work. Based on the properties of this sample, a set of 45 proposed universals were formulated. Although later typological work has sometimes shown that some of Greenberg's universals were too restrictive, others are still widely accepted today, and continue to spark new research. For example, Universal 34 below is generally accepted, and is incorporated into work on number from many perspectives (see Harley and Ritter 2002 for a feature geometry of pronouns, and Ackerman, Malouf, and Blevins 2016 for an acquisition-based explanation of the universal):

> Universal 34: No language has a trial number unless it has a dual.
>
> No language has a dual unless it has a plural. (Greenberg 1966: 94)

Greenberg's classification and numbering of universals are often retained, even when our understanding of the typological patterns has changed. A good example of this is the topic of this paper, Universal 20, where Greenberg's original formulation was as follows:

Universal 20: When any or all of the items (demonstrative, numeral, and descriptive adjective) precede the noun, they are always found in that order. If they follow, the order is either the same or its exact opposite (Greenberg 1966: 87)

As noted by Hawkins (1983: 118), the wording of this universal is ambiguous, but on the intended interpretation, it permits the following orders:[3]

(1) Greenberg's orders

No post-nominal modifiers (three pre-nominal modifiers)

Dem-Num-A-N

One post-nominal modifier (two pre-nominal modifiers)

Dem-Num-N-A

Dem-A-N-Num

Num-A-N-Dem

Two post-nominal modifiers (one pre-nominal modifiers)

Dem-N-Num-A

Dem-N-A-Num

Num-N-Dem-A

Num-N-A-Dem

A-N-Dem-Num

A-N-Num-Dem

Three post-nominal modifiers (no pre-nominal modifiers)

N-Dem-Num-A

N-A-Num-Dem

---

3   The ambiguity is in the reference of the pronoun 'they.' Specifically, the first 'they' refers only to the modifiers that do precede the noun, and the second 'they' only to the modifiers that follow (Hawkins 1983: 118)

Greenberg's (1966) thirty-language sample is given below. I have added Kikuyu, which he also mentions specifically in relation to Universal 20.

(2) Greenberg's sample

Basque	Serbian	Welsh	Maya (Yucatec?)	Greek
Italian	Finnish	Yoruba	Nubian (unspecified)	Swahili
Fula	Maasai	Kannada	Berber (unspecified)	Turkish
Hebrew	Burushaski	Hindi	Songhai (unspecified)	Japanese
Thai	Burmese	Māori	Zapotec (unspecified)	Luritja
Norwegian	Chibcha	Guaraní	Quechua (unspecified)	Malay
Kikuyu				

It was discovered relatively soon that Greenberg's typology incorrectly ruled out attested word orders (see sections 3.2 and 3.3). However, Greenberg's work still stands out as the first solid attempt at describing restrictions on the order of these elements in the noun phrase. The subsections below describe how others have elaborated on this work by expanding the sample, and by attempting to explain why some orders are unattested.

3.2 The functional-typological framework

Much work has been done on word order typology in the noun phrase from a functional-typological perspective. Here we will survey some of the more important works, beginning with Hawkins (1983). This book covers word order universals, and builds on Greenberg's work in both methodology and data. He uses an expanded sample of 336 languages, more than ten times the sample size used by Greenberg (1966). However, data on all properties studied is not available for all languages. Below we will focus on what Hawkins has to say about Universal 20, and about how noun phrase universals may be explained. He agrees with Greenberg's (1966) general tendency for demonstratives and numerals to be outermost, followed by adjectives closer to the noun. The orders are reported to be the same for prepositional and postpositional languages (Hawkins 1983: 75, 86). The following principle, Universal 20', is proposed, with a

parenthetical remark clarifying an ambiguity in Greenberg's (1966) formulation:

> When any or all of the modifiers (demonstrative, numeral, and descriptive adjective) precede the noun, they (i.e., those that do precede) are always found in that order. For those that follow, no predictions are made, though the most frequent order is the mirror-image of the order for preceding modifiers. In no case does the adjective precede the head when the demonstrative or numeral follow. (Hawkins 1983: 119-120)

Hawkins proposes that the patterns are ultimately grounded in processing. Light (i.e. short) elements such as demonstratives and numerals are preferentially placed before the noun, while the heavier (i.e. longer) adjectives typically follow the noun (the Heaviness Serialization Principle, HSP; Hawkins 1983: 89ff). This is compared to Universal 25: "if the [light - SA] pronominal object follows the verb, so does the [heavy - SA] nominal object." The HSP is argued to underlie English alternations such as *the yellow book* as opposed to *the book yellow with age* (Hawkins 1983: 91). In processing, Hawkins argues that we have to make constant decisions with every incoming word as to its head or non-head status. If long elements stack up pre-nominally, all of these modifiers must be retained in working memory, and we must dismiss them as non-heads in order to correctly parse the sentence. For this reason, we prefer getting to the head noun as quickly as possible, and when elements do precede the head, they should preferentially be short (Hawkins 1983: 98ff.).

Another functional-typological view of the noun phrase is given in Rijkhoff's (2002) Functional Grammar (Dik 1997) analysis. This book discusses the noun phrase from the perspective of Functional Grammar (Dik 1997). Rijkhoff assumes that modifiers come in the following order when moving away from the head noun: Quality (A), Quantity (Num), Location (Dem). In other words, qualifying words such as adjectives appear closest to the nouns, while modifiers specifying a location appear further away, with quantifying modifiers intervening between the two. Note that Rijkhoff formulates his proposal using semantic notions (quality,

11

quantity, location) rather than syntactic categories, something which is discussed further for Dryer's work later in this section. Rijkhoff attempts to ground his theory of basic Location-Quantity-Quality order in scopal relations (the Principle of Scope):

> In an NP like 'those two black dogs on the carpet' it is only the dogs that are black, not the quantity or the location. And the quantifying modifier *two* specifies the number of black dog entities, not the number of locations; and both *those* and *(on) that old blanket* [sic – SA] specify the location of dog entities with all their qualitative and quantitative properties as given in the NP. (Rijkhoff 2002: 223)

Rijkhoff explores Universal 20 patterns using a sample of 26 languages (Rijkhoff 2002: 328). The full sample is given below, with the five bolded languages indicating overlap with Greenberg's (1966) sample.

(3) Rijkhoff's sample

Alamblak	Bambara	**Basque**	Berbice Dutch Creole	Bukiyip
Burushaski	Dutch	Georgian	**Guaraní**	Hittite
Hmong	Hungarian	Arhuaco	Kayardild	Ket
Kisi	Khoekhoe	Naasioi	Ngalakan	Ngiti
Oromo	Náwat Pipil	**Imbabura Quechua**	Tamil	**Turkish**
Wambon				

Several principles are invoked to explain the patterns found. One of these is the Principle of Scope, cited immediately above. Also relevant is the Principle of Head Proximity, which favors configurations where adjectives and nouns are adjacent.[4] Below we see the orders predicted to occur by interactions of Rijkhoff's principles, along with statements of whether or not they are

---

4   Note that Rijkhoff's (2002) conception of head differs from that in Cinque (2005), where all modifiers of the noun have a dedicated syntactic projection, including a head.

attested in Rijkhoff (2002) and/or Hawkins (1983):

(4) New word orders (data from Rijkhoff 2002: 331)

Order	Attested in Rijkhoff	Attested in Hawkins
Dem-Num-A-N	✓	✓
Dem-Num-N-A	✓	✓
Dem-A-N-Num	-	-
Dem-N-A-Num	✓?	✓
Num-A-N-Dem	✓	-
Num-N-A-Dem	✓	✓
A-N-Num-Dem	-	-
N-A-Num-Dem	✓	✓

Rijkhoff's approach is very similar to that of Hawkins (1983), although the theoretical frameworks assumed are different. A set of largely semantic principles are proposed to account for the ordering restrictions, and these are further motivated by functional considerations such as processing.

The final functional-typological treatment of Universal 20 that we consider here is Dryer's work, which has been presented many times in slightly different forms during the last decade. We will focus here on the recent *Language* article, Dryer (2018), as opposed to earlier conference presentations (Dryer 2019, 2011, 2013). The typological sample comprises 576 languages, the largest sample used for investigating Universal 20 in the published literature. Although the sample has good areal and family coverage, this is a convenience sample, meaning that it is not fully balanced. The dataset is available in the supplementary materials for Dryer (2018). The goal of this work is ambitious: to propose a set of principles which account not only for the attested vs unattested orders, but also for their relative frequency. Dryer (2018: 799) reports the following frequency data:

(5) Dryer's frequency counts

Order	Number of languages	Number of genera
N-A-Num-Dem	182	85
Dem-Num-A-N	113	57
Num-N-A-Dem	67	40
Dem-N-A-Num	53	32
Dem-Num-N-A	40	19
N-A-Dem-Num	36	27
N-Dem-A-Num	13	10
Dem-A-N-Num	12	11
Dem-N-Num-A	12	9
N-Num-A-Dem	11	6
Num-A-N-Dem	8	7
N-Dem-Num-A	8	5
A-N-Num-Dem	5	3
A-N-Dem-Num	5	3
Num-N-Dem-A	5	3
Dem-A-Num-N	3	2
Num-Dem-A-N	2	2
N-Num-Dem-A	1	1
A-Num-Dem-N	0	0
Num-A-Dem-N	0	0
A-Dem-Num-N	0	0
Num-Dem-N-A	0	0
A-Dem-N-Num	0	0
A-Num-N-Dem	0	0

The following five principles are proposed to explain the data (Dryer 2018: 801):

1. Iconicity principle 1: The demonstrative tends to occur outside the adjective and numeral (i.e. further from the noun) when the demonstrative and the adjective or numeral (or both) occur on the same side of the noun.

2. Iconicity principle 2: The adjective tends to occur inside the numeral (i.e. closer to the noun) when they occur on the same side of the noun.

3. Asymmetry principle: The iconicity principles apply more strongly to pre-nominal modifiers than they do to post-nominal modifiers; exceptions to the iconicity principles will occur only with post-nominal modifiers.

4. Post-nominal adjective preference: Noun-adjective order is preferred over adjective-noun order

5. Intracategorial harmony: The demonstrative, numeral, and adjective tend to all occur on the same side of the noun.

Dryer (2018: 801) notes that these principles are themselves not fully explanatory, since no reason is given for why they would hold (see section 2.2). An advantage of this is that it is easy to quantify how often each principle is violated given a sample of languages. In other words, the principles are easily falsifiable (see discussion of Futrell, Levy, and Dryer 2017 in section 4.1).

Similarly to Rijkhoff (2002), Dryer stresses that he uses the words 'demonstrative,' 'numeral,' and 'adjective' as semantic notions rather than syntactic categories (Dryer 2018: 798). This makes the prediction that, for example, words which function semantically as adjectives will show the same word order patterns independently of their syntactic realization. This prediction is confirmed in a subsample of 267 languages. When semantic adjectives are syntactically verbs, they are closer to the noun than numerals 90% of the time (Iconicity

principle 2). When semantic adjectives are non-verbal syntactically, the same percentage is 93% (Dryer 2018: 812). In other words, Iconicity principle 2 can apply to syntactic verbs, if they function semantically as adjectives. Dryer cites similar data to show that the other four principles also apply independently of syntactic category. This is a strong empirical argument that the explanation for Universal 20 word order patterns cannot be phrased in terms of syntactic categories. This argument was first made by Dryer (2009), but has not been responded to in any subsequent literature that I am aware of.

3.3 The synchronic representational framework

This section will discuss representational approaches to Universal 20, all of which have been syntactic in the literature up to this point. We begin by covering the most famous analysis, that of Cinque (2005), which sparked a large interest in these typological patterns among generative syntacticians. Some of the work which has been done with Cinque (2005) as their starting point are covered in 3.3.2. However, because Cinque's treatment and description of Universal 20 are so detailed, we will focus on his analysis here.

3.3.1 Cinque (2005)

It is situated firmly within the synchronic representational approach to typology, as can be seen from the abstract: "it is proposed that /.../ none of the unattested [orders - SA] are derivable" (Cinque 2005: 315). The syntactic representations which drive Cinque's analysis are those of Antisymmetry (Kayne 1994), although, as we will see, additional restrictions also have to be imposed. Cinque (2005) does not provide a single sample of languages, but has instead performed the very useful task of compiling the attested word orders from various typological sources. Beyond this, additional languages that Cinque is aware of are also incorporated (see especially the footnotes in Cinque 2005: 319-320).

In the languages and sources consulted by Cinque, 14 out of the 24 logically possible orders of demonstrative, numeral, adjective, and noun are attested. These are a strict subset of the 18 orders found by Dryer (2018). Like Dryer (2018), Cinque sets himself the ambitious goal of

deriving the relative frequencies of the attested orders, in addition to ruling out the unattested ones. A four-way division into "very many languages," "many languages," "few languages," and "very few langauges" is used for these purposes. The table below shows a comparison between Dryer's (2018) genera frequency counts and the classification of Cinque (2005).

(6) Comparing Dryer and Cinque

Order	Dryer (2018), number of genera	Cinque (2005), frequency
N-A-Num-Dem	85	Very many
Dem-Num-A-N	57	Very many
Num-N-A-Dem	40	Few
Dem-N-A-Num	32	Many
Dem-Num-N-A	19	Many
N-A-Dem-Num	27	Few
N-Dem-A-Num	10	Very few
Dem-A-N-Num	11	Very few
Dem-N-Num-A	9	Very few
N-Num-A-Dem	6	Few
Num-A-N-Dem	7	Very few
N-Dem-Num-A	5	Few
A-N-Num-Dem	3	Very few
A-N-Dem-Num	3	Very few
Num-N-Dem-A	3	Unattested
Dem-A-Num-N	2	Unattested
Num-Dem-A-N	2	Unattested
N-Num-Dem-A	1	Unattested
A-Num-Dem-N	0	Unattested
Num-A-Dem-N	0	Unattested
A-Dem-Num-N	0	Unattested
Num-Dem-N-A	0	Unattested
A-Dem-N-Num	0	Unattested
A-Num-N-Dem	0	Unattested

Frequencies are assumed to be inversely correlated with the number of marked syntactic movements needed to derive a particular word order. The syntactic assumptions made by Cinque are discussed immediately below.

Cinque works in the framework of Antisymmetry, where differences in word order across languages are explained by applying different syntactic movement operations to a universal Merge order. Moreover, languages cannot differ in which linear order they assign to a particular syntactic structure (Kayne 1994). Cinque (2005: 321) proposes that this universal base order is [Dem [Num [A [N]]]], which agrees with the proposed scope order in semantic treatments of Universal 20 (Lu 1998, Rijkhoff 2002). The following movement constraints are then imposed (Cinque 2005: 321):

(7) Movement constraints

- There is no head movement (nor remnant movement; Cinque 2005: 326)

- All relevant movements must move the NP[5]

- Not moving anything is unmarked

- Movement of [NP [XP]] (*whose picture* pied-piping) is unmarked

- Raising the NP across all modifiers is unmarked, while raising the NP across only some modifiers is slightly marked

- Movement of NP without pied-piping is marked

- Movement of [XP [NP]] (*picture of who* pied-piping) is highly marked

Some of these assumptions appear motivated only by the fact that they derive the correct patterns and relative frequencies, and there is "no clear independent reason" for some of the markedness assumptions made (Cinque 2005: 325).[6] The fifth assumption in the list above is perhaps worth highlighting in this regard: it seems strange that moving NP across AP is marked, while moving

---

5   For Cinque, the NP is the constituent containing the noun, but excluding all of the modifiers.
6   This is somewhat puzzling given the abstract of the paper, which claims that the patterns are derivable "from independent conditions on phrasal movement" (Cinque 2005: 315).

NP across AP, NumP, and DemP is unmarked. Cinque (2005) does not speculate on why this might be the case. Wherever these markedness patterns come from, Cinque (2005: 321-324) goes through each of the 24 orders, and discusses the (im)possibility and markedness status of the derivations needed. This broadly agrees with the frequencies of the word orders, and all unattested word orders in the sample are underivable.

The final parts of the article discuss extending the analysis to other parts of speech, including the rest of the DP (quantifiers, relative clauses, etc.), and possibly even word order variation in other parts of the clause.

3.3.2 Other work

After the publication of Cinque (2005), a number of other synchronic representational accounts of Cinque's data have been proposed. These assume that the crosslinguistic patterns reported in Cinque (2005) are valid, i.e. that only 14 out of 24 possible orders are attested. Unlike Cinque (2005), however, the approaches we will look at here typically do not attempt to derive the relative frequencies of those orders (but see Abels and Neeleman 2012: 29, fn. 4, and Medeiros 2012). It would take too long, and take us too far afield, to give a detailed presentation of all representational approaches to Universal 20. A brief summary is presented here, in order to give an understanding of the breadth of research on this universal.

Steddy and Samek-Lodovici (2011) adopt Cinque's universal Merge order, and assume that deviations from the base order are derived by movement. Their proposal is couched within Optimality Theoretic syntax, meaning that the proposal involves ranked, violable constraints on syntactic structures. All of their constraints are in the ALIGN family, specifying that the edge of one constituent must be aligned with the edge of another one. Specifically, they assume four constraints, one for each of DemP, NumP, AP and NP, which require the left edge of these constituents to be aligned with the left edge of the highest DP projection. These constraints are written here in simplified form as LEFT(DemP), LEFT(NumP), LEFT(AP), and LEFT(NP). The effect of this constraint set is that every constituent prefers to be as far left as possible. For example, Cinque's (2005) base order satisfies only the DemP constraint: the DemP is the

structurally highest constituent, and so its left edge is aligned with the left edge of the highest constituent. The NumP constraint is violated once for this order, since one phrase (DemP) intervenes between it and the highest constituent. The AP constraint is violated twice, because both DemP and NumP intervene. And the NP constraint is violated three times. In structures with movements, it is crucial for their proposal that traces also count as interveners for the purposes of assigning violations.

In Optimality Theory, languages differ in their rankings of constraints. Steddy and Samek-Lodovici (2011) show that the structures required by Cinque's (2005) ten unattested orders are harmonically bounded by structures which lead to different orders. Harmonic bounding refers to a scenario where a structure A has a strict superset of the violations of another structure B. This means that A is never better than B on any constraint ranking, since for any constraint, the number of violations for A will always be greater than or equal to the number of violations for B. In other words, the ten unattested orders cannot be derived, since no ranking of the alignment constraints will make them better than other structures/orders. Steddy and Samek-Lodovici (2011) assume different constraints on movement than Cinque (2005). They do not require that a constituent containing NP is moved, allowing also cases of remnant movement where a constituent containing an NP trace is moved. Most of the time, remnant movement does not improve the violation profile of a structure, but in some specific circumstances, it can create better structures and thus becomes important to their derivations (see discussion of N-Dem-A-Num in Steddy and Samek-Lodovici 2011: 452-455). This proposal, then, relies heavily on assumptions from Cinque (2005), but sometimes contains deviations, and is implemented in an entirely different framework than that assumed by Cinque.

Abels and Neeelman (2012) argue that a correct account of Universal 20 crucially relies on languages differing in how they linearize a certain structure, making their theory fundamentally incompatible with Antisymmetry. Despite this difference, their proposal is otherwise similar to Cinque's, assuming the validity of several of his movement constraints, as well as the base order Dem-Num-A-N.[7] Different ways of linearizing this structure immediately

---

7   Dryer (2018: 824, fn. 28) claims that Cinque's base order is never assumed. However, Abels and Neeleman (2012: 30) explicitly give the "underlying hierarchy in the extended projection of the noun" in Cinque's order.

gives eight of the 14 attested orders (one binary linearization choice for each of Dem, Num, and A, and $2^3 = 8$). The remaining six are derived straightforwardly by leftward movement of a constituent containing the noun, while the 10 unattested orders are underivable.

Importantly, Abels and Neeleman (2012) point out a generalization which is captured by their theory but which could not be captured in Cinque's (2005) system. They divide up the 24 possible orders into two classes: the 'asymmetric' class covers orders where any two post-nominal modifiers appear in the *same* order as in the universal base Dem-Num-A-N, while the 'symmetric' class covers all other orders (Abels and Neeleman 2012: 32):

(8) Abels and Neeleman's typology

Symmetric orders		Asymmetric orders	
Dem-Num-A-N	N-A-Num-Dem	Dem-N-Num-A	*A-Num-N-Dem
Dem-Num-N-A	A-N-Num-Dem	N-Dem-Num-A	*A-Num-Dem-N
Dem-A-N-Num	Num-N-A-Dem	A-N-Dem-Num	*Num-Dem-N-A
Dem-N-A-Num	Num-A-N-Dem	N-Num-A-Dem	*Dem-A-Num-N
*A-Dem-Num-N	*N-Num-Dem-A	N-Dem-A-Num	*Num-A-Dem-N
*A-Dem-N-Num	*Num-N-Dem-A	N-A-Dem-Num	*Num-Dem-A-N

The attested symmetric word orders share the property that the mirror image of these orders (i.e. reading them right-to-left instead of left-to-right) are also attested, while the unattested word orders never have an attested mirror image. This is shown in the two columns under the symmetric orders, which show the mirror image of each order. The attested word orders of the asymmetric class are the opposite: the mirror image is never attested. Again this is shown in the two columns under the asymmetric orders. Moreover, it just so happens that the attested symmetric orders are all and only the orders derived by different linearizations, while the attested asymmetric orders are all and only the ones derived by movement. Generalizations of this type cannot be captured in any theory which does not distinguish between movement-derived and linearization-derived orders, such as that of Cinque (2005).

Most of the analyses above have been based on relatively familiar theoretical machinery.

Medeiros (2012) represents the first analysis which proposes new theoretical devices in order to capture Universal 20 patterns. The following new proposal is made: syntactic movement aims to decrease the number of c-command relations in the tree. This is the only constraint Medeiros (2012) uses to derive the Universal 20 patterns. In order to illustrate what this means intuitively, Medeiros discusses 'spiny' trees and 'bushy' trees, where the latter have fewer c-command relations between leaves of the tree. We can think of bushy trees as ones where each left branch contains approximately the same number of leaves as the right branch. Movement is thus an operation which serves to balance tree branches.

The only movements allowed are those which move from spine-like shapes toward bush-like shapes. Assuming a basic Dem-Num-A-N, all and only the 14 attested orders can be derived. Moreover, the number of distinct ways in which each order can be derived seems to be positively correlated with the crosslinguistic frequency of the order. Medeiros (2012) thus presents a very different way of thinking about Universal 20. But the proposal appears to be empirically sound, and requires only one constraint on movement. It can also account for the frequency patterns of the different orders without stipulating, as Cinque (2005) did, that certain movements are more marked than others.

Five years after Medeiros (2012), the same author presented a new proposal for Universal 20 patterns. The new proposal has, as far as I can tell, only one thing in common with the old one: Dem-Num-A-N is still assumed to be a universal base order. Medeiros (2017, 2018) uses an algorithm for sorting elements known as Knuth's stack-sorting algorithm (described fully in the papers, and in Knuth 1968). The algorithm uses stack memory (last-in, first-out), and moves through each item in {Dem, Num, A, N} one at a time. The algorithm determines its decisions to push to, or pop from, the stack by whether or not the current element considered c-commands the item on top of the stack in the universal base hierarchy. It is applied to Universal 20 in the following way: give DemP, NumP, AP, and NP numbers corresponding to their depth of embedding: DemP = 1 (least embedded), NumP = 2, AP = 3, and NP = 4 (most embedded). There are 24 possible ways of ordering these numbers. Consider now a hearer, exposed to a particular order (such as 1 4 2 3). They want to figure out the syntactic structure underlying the linear order

of elements that they have just heard. So they run the order through the algorithm, and find, for 1 4 2 3, that the algorithm outputs 4 3 2 1. This corresponds to the universal base order (from most to least embedded). You would find the same output for all 14 attested word orders: they all spit back a base order consistent with the one given by UG. By contrast, all 10 unattested word orders produce a different result when run through the algorithm, one not consistent with the UG-given base order: "[b]y hypothesis, this is why such orders are unavailable: they are mapped to gibberish, an uninterpretable order of composition not matching the universal fseq [functional sequence - SA]" (Medeiros 2017: 7).

Medeiros (2017, 2018) shows that by inserting brackets at specific points in the derivation, we can create a unique bracketing for each order, corresponding to the constituency structure. We can store these bracketings, and use them in production to produce specific word orders. In other words, the theory is not just a theory of parsing. Since there is no crosslinguistic variation in the steps a learner has to take when following this algorithm, the proposal can be straightforwardly applied to languages with multiple word orders: all one needs to do is run the algorithm, and store its output for production purposes. This can be done with one or multiple word orders without altering any mechanisms or features. Moreover, Medeiros' (2017, 2018) theory assigns a unique structure to each word order, whereas earlier approaches had strings which were structurally ambiguous. Since such strings never appear to be *interpretationally* ambiguous, "[w]ith a grammatical device built around stack-sorting, there is no ambiguity of structure without ambiguity of meaning" (Medeiros 2017: 16). As a theory of Universal 20, this approach appears to be empirically equivalent to others. But for a general theory of syntax, it is not yet clear what role sorting algorithms may have to play.

3.4 Conclusions

We have now seen several approaches to the typology of noun phrase word order. Some are functional-typological, while others follow the synchronic representational framework. We have also surveyed some of the entirely typological literature on Universal 20, including Greenberg's original work on universals. Because so much work has been done, the approaches that have

been used are incredibly diverse. However, we have not seen any diachronic approaches to Universal 20, and this reflects a gap in the current literature. In the next section, this will be addressed, as we attempt to formulate an Evolutionary Syntax account for Universal 20.

4 An Evolutionary Syntax account of Universal 20

In this section we will see a possible diachronic explanation for Universal 20 word order patterns. At the core of this explanation will be Dryer's (2018) principles, since these are known to account for the crosslinguistic patterns of word orders and their frequencies (Futrell et al. 2017). If each of Dryer's principles can be explained diachronically, this also provides a diachronic account of the word order patterns themselves. We will begin by considering what principles Dryer proposes, in section 4.1, followed by diachronic explanations of each of them in sections 4.2-4.4. These explanations will incorporate insights from earlier work on Universal 20, as well as new typological results about interactions between word length and part of speech.

4.1 Dryer's principles

Below I repeat the word order principes from Dryer (2018: 801):

> 1. Iconicity principle 1: The demonstrative tends to occur outside the adjective and numeral (i.e. further from the noun) when the demonstrative and the adjective or numeral (or both) occur on the same side of the noun.

> 2. Iconicity principle 2: The adjective tends to occur inside the numeral (i.e. closer to the noun) when they occur on the same side of the noun.

3. Asymmetry principle: The iconicity principles apply more strongly to pre-nominal modifiers than they do to post-nominal modifiers; exceptions to the iconicity principles will occur only with post-nominal modifiers.

4. Post-nominal adjective preference: Noun-adjective order is preferred over adjective-noun order

5. Intracategorial harmony: The demonstrative, numeral, and adjective tend to all occur on the same side of the noun.

Dryer (2018) proposes that the number of principles violated by a given language is inversely proportional to its frequency in typological samples (cf. Cinque 2005). It has been shown by Futrell et al. (2017) that these principles provide a good fit to the typological data. They conducted a statistical analysis using Poisson regression, which showed that the number of principles violated by a particular word order is indeed inversely proportional to the number of languages with that word order. The analyses by Dryer and Cinque performed equally well in this test, so we know that both Dryer's (2018) principles and Cinque's (2005) system are equally good at capturing the typological patterns of Universal 20. Dryer rightly emphasizes that this is a significant result: it is well worth our time as linguists to find ways of generalizing over large sets of data, especially if this results in a small set of principles which can easily be verified or falsified (Dryer 2018: 801).

Although descriptive principles such as the ones above represent progress, I agree with Dryer that explanations of such principles should also be found. What reasons could there be that languages show noun-adjective order more often than adjective-noun? In his section 5, Dryer (2018) considers some possible ways of motivating his proposal on independent grounds. However, some patterns are unexplained, such as the asymmetry principle, and others only raise more questions: if the iconicity principles are due to the relationship between proximity to the noun and 'inherentness,' why is it the case that modifiers occurring closer to the noun tend to

denote more inherent properties? Questions such as these do not receive clear answers in Dryer (2018). However, for many of the principles, promising directions for future work are laid out, and the section functions as an overview of which explanations have already been proposed in the literature. Thus, the sections below often build directly on insights mentioned by Dryer (2018).

4.2 Some additional typology

In the sections that follow, we will often make use of a connection between word order and word length, a connection which I argue to be mediated by working memory constraints on processing (see section 4.3). The word order facts are relatively well-known, but there is, as far as I can tell, no typological work comparing the word lengths of demonstratives, numerals, and adjectives. For this reason, I have conducted two typological surveys to investigate the typology of word length as it relates to Universal 20. The main finding that I report is that demonstratives tend to be shorter than numerals, which in turn tend to be shorter than adjectives.

Two typological samples are used in this section. The first, smaller sample is derived from corpora available through SketchEngine (Kilgarriff et al. 2014). An attempt was made to create a geographically and genealogically balanced sample. However, corpora for understudied languages are often not tagged for part of speech, and could therefore not be used for this sample. The resulting sample, which is diverse though not fully balanced, is given below (classification data closely based on that reported by Dryer 2018). The ten most frequent numerals and adjectives were extracted from the corpora, keeping only cardinal numbers and the positive forms of adjectives wherever possible (cf. Cinque's 2005: 328 suggestion that cardinals and ordinals obey different ordering restrictions). The corpora typically give the lemma form of each word. In almost all cases, demonstratives were not tagged separately from other pronouns. These were therefore extracted from grammars of each language instead, references to which are also given in the table below. All words were transcribed so as to have one symbol corresponding to one phonological segment.[8]

---

8   I am grateful to Chris Geissler for help with the phonological transcription of Tibetan.

(9) The sample

Language	Genus	Family	SketchEngine reference	Grammar reference
Amharic	Semitic	Afro-Asiatic	Rychlý and Suchomel (2016), Kilgarriff et al. (2010)	Leslau (1995)
Mandarin	Chinese	Sino-Tibetan	Jakubíček et al. (2013), Suchomel and Pomikálek (2012)	Ross and Sheng Ma (2006)
Estonian	Finnic	Uralic	Jakubíček et al. (2013), Suchomel and Pomikálek (2012)	Harms (1962)
Indonesian	Malayo-Sumbawan	Austronesian	Kilgarriff et al. (2010)	Sneddon et al. (2010)
Japanese	Japanese	Japonic	Srdanović et al. (2013), Jakubíček et al. (2013)	McGovern (1942)
Korean	Korean	Korean	Jakubíček et al. (2013), Suchomel and Pomikálek (2012)	Chang (1996)

Russian	Slavic	Indo-European	Jakubíček et al. (2013), Suchomel and Pomikálek (2012)	Wade (2011)
Somali	Lowland East Cushitic	Afro-Asiatic	Kilgarriff et al. (2010)	Saeed (1999)
Spanish	Romance	Indo-European	Kilgarriff and Renau (2013), Jakubíček et al. (2013)	Bradley and MacKenzie (2004)
Swahili	Bantoid	Niger-Congo	No references given	Thompson and Schleicher (2001)
Tibetan	Tibetic	Sino-Tibetan	Meelen et al. (2017)	Hannah (1912)
Turkish	Oghuz	Turkic	Baisa and Suchomel (2012), Jakubíček et al. (2013), Suchomel and Pomikálek (2012)	Göksel and Kerslake (2005)
Vietnamese	Vietic	Austro-asiatic	Baroni et al. (2009), Kilgarriff et al. (2010)	Thompson (1965)

For each part of speech (demonstrative, numeral, and adjective), the average length was calculated. Three predictions were then tested for each language:

> Are demonstratives shorter than numerals (Dem < Num)?
> Are numerals shorter than adjectives (Num < A)?
> Are demonstratives shorter than adjectives (Dem < A)?[9]

The average word lengths, in number of segments, and the evaluated predictions, are as follows:

(10) Word length data and predictions

Language	Dem	Num	A	Dem < Num	Num < A	Dem < A
Amharic	4.5	6	5.7	Yes	No	Yes
Mandarin	2.5	2.8	4.1	Yes	Yes	Yes
Estonian	3	4.3	5.2	Yes	Yes	Yes
Indonesian	4.5	4.3	5.2	No	Yes	Yes
Japanese	3.67	3.1	5.2	No	Yes	Yes
Korean	1.67	2.6	3.5	Yes	Yes	Yes
Russian	3.5	4.4	7.8	Yes	Yes	Yes
Somali	3.33	4.2	5.2	Yes	Yes	Yes
Spanish	4	4.4	6	Yes	Yes	Yes
Swahili	4	4.8	5.4	Yes	Yes	Yes
Tibetan	2	2.4	4.7	Yes	Yes	Yes
Turkish	1.67	3.3	4.3	Yes	Yes	Yes
Vietnamese	3	3.8	3.7	Yes	No	Yes

The first two predictions are true in 11/13 languages, while the third is exceptionless in this small sample. If numbers for the word lengths were generated randomly, each prediction would be true 50% of the time. In a two-tailed one-sample t-test, the three observed ratios of Yes:No in the table above were all significantly greater than 0.5 (p < 0.01 in all cases). All three predictions are

---

9   The answer to this question is necessarily yes if the answer to the previous two questions was yes.

thus confirmed.

Because of the limited sample used above, I also evaluated the second prediction (Are numerals shorter than adjectives?) using the ASJP Database (Wichmann, Holman, and Brown 2018). This database contains partial Swadesh lists for thousands of languages, including entries for 'one' and 'two,' as well as semantic adjectives 'full' and 'new.' The data are transcribed in the system described by Brown et al. (2008). The four entries above ('one,' 'two,' 'full,' 'new') were extracted from the 7,608 doculects present in version 18 of the database. The following predictions were evaluated:

'full' is longer than 'one'

'full' is longer than 'two'

'new' is longer than 'one'

'new' is longer than 'two'

The average length of 'full' and 'new' is greater than the average

length of 'one' and 'two'[10]

'longer' was defined as 'strictly longer' rather than 'longer than or equal in length.' Cases where two items were equal in length were not counted, as these neither confirm nor falsify any predictions. For many of the doculects under consideration, one or more of the relevant entries are missing. This means that some of the predictions cannot be tested against all doculects. Below I report the results of the five predictions:

---

10 Or, equivalently: The combined length of 'full' and 'new' is greater than the combined length of 'one' and 'two.'

(11) ASJP results

Prediction	N	Mean proportion of confirmations (μ)	μ > 0.5? (two-tailed one-sample t-test)
'full' > 'one'	5,705	0.5227	Yes, p = 0.001***
'full' > 'two'	5,462	0.5447	Yes, p = 0.000***
'new' > 'one'	5,562	0.5380	Yes, p = 0.000***
'new' > 'two'	5,385	0.5044	Yes, p = 0.522
average(A) > average(Num)	3,059	0.5691	Yes, p = 0.000***

Four of the five predictions were highly significant (p < 0.001). Only one prediction was disconfirmed: 'new' was not significantly longer than 'two' in the languages in this sample. This is line with Hawkins' (1983: 90) conjecture that adjectives tend to be the longest modifiers. Of course, these data should be taken with a large grain of salt. The citation forms of numerals and semantic adjectives may differ from the form(s) they take when modifying nouns, and there are no data on the frequencies of the four words considered. Moreover, the sample is not balanced, but will instead show biases towards well-documented languages, which often come from large language families in Eurasia and Africa. Despite all of these problems, the data support the conclusions from the smaller but more responsible 13-language sample above. It does seem to be the case that demonstratives are shorter than both numerals and adjectives, and that numerals are shorter than adjectives.

4.3 The post-nominal adjective preference

Dryer's (2018) fourth principle states that noun-adjective order is more frequent than adjective-noun order. In his sample, orders where N precedes A are found in 428 out of 576 languages investigated, or just over 74% of the time. Here I follow Hawkins (1983), who has provided a functional-typological motivation for this asymmetry. I first decsribe Hawkins' (1983) work, and then restate it in diachronic terms, instead of treating it as a synchronic functional principle.

Hawkins (1983) appeals to what he calls the Heaviness Serialization Principle (HSP),

which states that prosodically heavy (i.e. long) constituents preferentially occur after the head that they modify, more so than prosodically light (i.e. short) constituents (Hawkins 1983: 89ff). This is meant to be grounded in processing factors: as speech comes in, we have to make decisions on where the head of the current phrase is. If long elements keep stacking up pre-nominally, we have to keep all of those in working memory (see section 4.4 for discussion and references), and we have to keep making headedness decisions, until we find the noun and can determine the dependencies. For this reason, we prefer anything coming before the noun to be as short as possible, leaving longer modifiers in the post-nominal position.

Hawkins (1983) claims that adjectives tend to be longer than numerals, and that this explains why we have a post-nominal adjective preference rather than, say, a post-nominal numeral or demonstrative preference. Dryer (2018: 827) is skeptical of this typological generalization about word length, but as we have seen above, it appears to hold a significant proportion of the time.

We will now consider how these differences in word length give rise to differences in word order diachronically, mediated by the processing factors pointed out by Hawkins (1983). Consider first that many languages permit variation in whether adjectives are pre-nominal or post-nominal (see Cinque 2010 for detailed discussion). According to Hawkins' (1983) processing explanation, N-A orders should be favored since they impose less of a processing load. If N-A is easier to process, this should lead to speakers failing to parse N-A orders less often than A-N orders. A learner exposed to 50% N-A and 50% A-N should therefore show a slight bias towards N-A, since some of the A-N orders will not be parsed correctly due to their processing load. Since speakers tend to overgeneralize more frequent patterns at the expense of less frequent variants (see Culbertson and Adger 2014 for experimental evidence in the context of noun phrase word order), this should create a feedback loop, where speakers gradually increase the proportion of N-A orders in their language. This could lead to a dominant or even obligatory N-A order within a language over time.

4.4 Iconicity and asymmetry

In this section I will give a possible explanation for the iconicity principles (1-2) and the asymmetry principle (3). It is argued that there is a connection between word length and word order, such that shorter modifiers tend to precede longer modifiers. Based on the typological data on word length in section 4.2, this can account for the iconicity principles. This tendency is grounded in constraints on working memory, which have been experimentally tested for spoken and signed languages, as well as in language acquisition. Modifiers before the head must be kept in working memory until the dependencies with the head are determined. However, once the head is reached, these dependencies can be resolved. For post-nominal modifiers, dependencies can be resolved immediately, meaning that there is no predicted working memory effect on post-nominal modifiers. This will be shown to explain the asymmetry principle.

The order of pre-nominal modifiers is assumed to be Dem-Num-A-N in theories with a basic universal hierarchy (Cinque 2005, Steddy and Samek-Lodovici 2011, Abels and Neeleman 2012, Medeiros 2012, 2017, 2018). Rijkhoff (2002) claims this is the basic scope hierarchy of the modifiers (see section 3.2), and Lu (1998: 32-34) defends the same hierarchy in terms of iconicity. Here I propose a different explanation in terms of word length, recalling from section 4.2 above that demonstratives tend to be shorter than numerals, which tend to be shorter than adjectives: Dem-Num-A.

I follow Hawkins (1983) in assuming that modifiers must be kept in working memory until the head is reached (see also Ford 1983, Hawkins 1990, Gibson 1998, 2000, Gordon et al. 2001, McElree et al. 2003, Hawkins 2004, Grodner and Gibson 2005, Lewis and Vasishth 2005, Lewis et al. 2006, Vasishth and Lewis 2006, and Hawkins 2014). Until this point, the dependence relations between the incoming words are not known to the listener. When the listener processes the head, the relations between the head and its modifiers can be worked out, and the modifiers no longer need to be kept in short-term memory. Word length is known to affect working memory load, and the "[m]emory span for words is inversely related to [the - SA] spoken duration of the words" (Baddeley 1992: 558). In other words, adjectives are more costly for working memory than numerals, which are in turn more costly than demonstratives. This word

length effect has been shown to be present in L1 acquisition (Hulme and Tordoff 1989), as well as in American Sign Language (Wilson and Emmorey 1998).

Consider now the working memory load when processing the common word order Dem-Num-A-N, and the inverse (unattested) word order A-Num-Dem-N. For the sake of simplicity, we will assign Dem a processing load of 1, Num 2, and A 3. These numbers are purely expository, and are not meant to imply that adjectives require three times as much working memory as demonstratives, nor that they are typically three times as long as demonstratives. After the first modifier, Dem-Num-A-N has a memory load of 1, while A-Num-Dem-N is already at 3. Num adds 2 to both numbers, which are now 3 and 5 respectively. After the final modifier, both languages would have a memory load of 6. After each modifier, the load on working memory for A-Num-Dem-N is greater than or equal to that of Dem-Num-A-N.

In contrast to the above, post-nominal modifiers can have their dependencies resolved instantly. The listener already knows what the head noun is, and so once the entire modifier has been parsed, the speaker can figure out that it is modifying the noun, and then move on. There is no need to wait in order to work out the correct dependencies. In other words, the word order tendencies above, where short modifiers tend to precede long modifiers, only apply to pre-nominal orders. This means that the explanation above captures the iconicity principles (Dem-Num-A is preferred), as well as the context where they apply (only pre-nominally; the asymmetry principle).

The diachronic interpretation of these effects is the same as for the post-nominal adjective preference. No language shows only one noun phrase word order rigidly in all contexts. When multiple word orders are in competition, some will be misparsed more often than others because of the word length effect on working memory. There is now a small bias against the word orders which are more likely to be misparsed, and this bias can spread to subsequent generations. Because of frequency-based overgeneralization (again see Culbertson and Adger 2014), the bias is expected to grow over time, and may eventually lead to establishing a new word order as dominant.

We can see now that the explanation I have proposed is independent of the fact that we

are dealing with noun phrases. It is possible, for example, that the same ordering restrictions hold for the ordering of tense, aspect, and mood markers before the verb (as argued by Cinque 2014). If the word length tendencies align with the ordering preferences, this means we could use this approach to explain these ordering patterns in the verb phrase as well. It is thus only partially true that Dryer's "principles are specific to the nominal phrase and cannot be directly extended to the clausal domain" (Cinque 2017: 1). Finding a deeper explanation for Dryer's principles has allowed us to generalize beyond the nominal phrase in exactly the way Cinque describes.

Another extension of the reasoning in this section might be the well-known difficulties people have with multiple center embedding (for experiments finding a correlation between working memory and ability to parse center-embedded structures, see King and Just 1991). Particularly relevant here is the recent research by Janet Fodor and colleagues (Fodor, Nickels, and Schott 2017), showing that people's ability to process center-embedded sentences depends in part on constituent length. Specifically, multiple center embedding is easier to parse when the middle constituents are short (examples from Fodor et al. 2017; according to Kluender 1998, this effect was first reported by Bever 1970):

(12) Center embedding

> The rusty old ceiling pipes that **the plumber my dad trained fixed** continue to leak occasionally.

> The pipes that **the unlicensed plumber the new janitor reluctantly assisted tried to repair** burst.

Speakers can only resolve the dependencies between the matrix subject and the matrix verb once they see the matrix verb. This means that the constituents between the matrix subject and matrix verb must be kept in working memory while the center-embedded relative clauses are being parsed. In the first sentence above, the distance (in terms of phonological segments, or in terms of time) is relatively short. Listeners have to spend relatively little time keeping track of three subject-verb dependencies. The dependencies between subjects 2 and 3 and verbs 2 and 3 are

resolved quickly, since the verbs appear so soon after their subjects. By contrast, in the second sentence above, a longer time must be spent keeping multiple dependencies in working memory, leading to a parsing failure in most cases. This could be seen as another effect of the working memory constraints that I have used to explain the iconicity principles and the asymmetry principle.

4.5 Intracategorial harmony

In this section we take a closer look at Dryer's fifth principle, intracategorial harmony. Why is it the case that demonstratives, numerals, and adjectives tend to occur on the same side of the noun? I will suggest that this is a consequence of structural priming, a well-documented effect where speakers are primed to use a certain syntactic structure through repeated exposure. If speakers are often exposed to a modifer-noun order, this will bias them towards producing that order more often. This can lead to modifier-noun order becoming dominant in the language, even if at previous stages some modifiers followed the noun. We begin by overviewing some of the literature on structural priming, and continue to show how intracategorial harmony can be derived from this effect.

Speakers' choices for which syntactic structure to use for their utterances can be biased by recent exposure to particular structures (Bock 1986). This effect is known as structural or syntactic priming. A well-studied phenomenon in this area is dative shift in English, which concerns the alternation between *They gave the teachers the apple* (double object construction) and *They gave the apple to the teachers* (prepositional object construction). If presented with a double object construction, speakers are more likely to produce subsequent sentences using this syntactic structure, rather than the prepositional object construction. Gries and Koostra (2017) write that:

> [s]uch priming effects are robust and widespread: they have been
> obtained with both observational and experimental methods, even
> over long distances between the first use of a construction (the
> prime) and the subsequent use (the target), both from production to
> production and from comprehension to production, in various tasks
> (picture description, sentence completion, dialog tasks...), in many
> languages, and, even more interestingly, between languages (see
> e.g., Pickering & Ferreira, 2008, for a review). (Gries and Koostra
> 2017: 235)

Bock (1986) noted that "all three experiments found evidence that the effects of priming were specific to features of sentence form, independent of sentence content" (Bock 1986: 355). In later work, Pickering et al. (2013) have found that priming of the interpretation of ambiguous relative clauses persists even when the verb used in the prime and target are different. Although there appear to be a so-callled lexical boost when words are repeated in prime and target (Coyle and Kaschak 2008), these effects are small (Kaschak and Borreggine 2008). As mentioned above, structural priming can be transferred from perception to production, and across language tasks (Kaschak 2007). The real-world interpretation of this experimental result is that learners can be biased towards a particular construction that they often hear from others.

A number of properties of structural priming are worth pointing out here. It has been shown that there is long-term structural priming, persisting for at least a week after the initial presentation or production of the biasing sentences (Kaschak et al. 2011). These biases, therefore, do not instantly disappear shortly after exposure to the primes, and can affect longer-term language production. Additionally, while priming relies on a similarity between prime and target, structural priming does not require identical syntactic constructions. For example, German and English have dative shift alternations in common, but there are nevertheless morphosyntactic differences across the two languages, so that the syntactic derivation of each construction would be slightly different. Despite this, cross-linguistic structural priming can be observed here, so that primes in English affect targets in German (Loebell and Bock 2003). Structural priming can thus

be transferred to syntactically and semantically similar but non-identical constructions.

It seems that structural priming has all the right properties needed to explain the tendency towards intracategorial harmony. Consider a child exposed to primarily Dem-N and A-N orders, but who usually hears N-Num instead of Num-N. Structural priming creates a bias for modifier-noun orders in such a language, which could affect the numerals as well as solidifying the tendencies for Dem and A to occur pre-nominally. Priming militates against the less common noun-modifier order, which loses ground to the more frequent pattern (for experimental evidence for overgeneralization to more frequent word orders in the noun phrase, see Culbertson and Adger 2014).

Depending on one's syntactic theory, demonstratives, numerals, and adjectives may be different syntactic objects associated with different syntactic positions. However, it is clear that they are all syntactically modifying the noun, and semantically often serve to restrict or qualify the interpretation of that noun. In some languages there are also morphological similarities between some or all of these categories. For example, Swedish demonstratives appear to be articles with the same endings as those found on attributive adjectives: article *den* but demonstrative *denna* 'this (one),' paralleling predicative *fin* but attributive *fina* 'nice, pretty.'[11] In other languages, it appears that numerals are really a subclass of adjectives, as in Swahili where numerals and adjectives take the same noun class markers when modifying nouns (Dryer 2018: 806). Structural priming should therefore be able to operate in the way outlined above across these syntactically, semantically, and sometimes morphologically similar modifers.

4.6 Conclusions on diachronic explanation

Above we have seen possible diachronic explanations for the word order principles proposed by Dryer (2018), which in turn explain the typology of Universal 20 word order patterns. The explanations considered here are based on processing considerations such as constraints on working memory, as well as structural priming. In some cases, notably the post-nominal adjective preference, I have relied on proposals from the existing literature (Hawkins 1983). However, with respect to post-nominal adjectives, and with other word order principles, I have

---

11 The apparent gemination contrast in <denna> vs <den> is purely orthographic: both forms have a geminate [nː].

also needed to rely on new typological investigations of word length.

It would be possible to state all of my explanations as functional-typological principles, which are synchronically active in the minds of individual speakers. However, I hope to have shown that it is not necessary to enrich individual speakers' mental representations in this way. In all of the cases above, we can understand Dryer's word order principles as diachronic outcomes of changes which themselves take place for processing-based reasons. Speakers need only be equipped with a brain and mind which is subject to the processing constraints that I have discussed, in order for the diachronic explanations to go through. There is no need for additional functional principles, nor for any additional syntactic constraints, to explain the typological patterns that we observe.

I will conclude this section with some ideas for how to falsify the explanations that I have provided. I hope that this is useful to researchers working on the same patterns from different theoretical frameworks. First of all, it is possible to question the typological data on word length which underlies many of my explanations. A more balanced typological sample may reveal that my results were due to areal biases, for example. No matter what the results of a larger survey might show, it would be useful for linguists of all stripes to have a better idea of any possible connections between word length and part of speech. Secondly, one could question that the processing pressures I use are real. In practice, this option seems unlikely, since they are supported by much experimental work in psycholinguistics. Nevertheless, the possibility should not be dismissed, and it may be that other explanations are responsible for the effects reported in the experiments I have cited. Finally, one could question the connection between processing factors and language change. Perhaps the processing load imposed by certain word orders is too small to have an appreciable effect on the next generation of speakers. This could fruitfully be explored through artificial grammar learning (Pothos 2007), agent-based simulations (Harrison et al. 2002), iterated learning paradigms (Smith et al. 2003), and other techniques. In any case, I welcome future work on any or all of these issues, which would deepen our understanding of Universal 20 word order patterns.

5 Concluding remarks

This paper has had several goals. First of all, it is a contribution to the diachronic approach to typological asymmetries, where the patterns we see in the world's languages today are argued to be consequences of pathways of historical change. And secondly, I have attempted to give diachronic explanations for Dryer's (2018) word order principles, which in turn explain the observed Universal 20 restrictions. Below I summarize the paper, including what I have said about each of these points.

In section 2 we saw several possible ways of approaching typological questions. One perspective, which I call synchronic representational, assumes that typological patterns are constrained by the same representations that are at the basis of a speaker's synchronic linguistic competence. Another common assumption is that patterns which are unattested are also unattestable, which is to say that they cannot be part of a possible human language. In the domain of Universal 20, this view is most famously found in the work of Cinque (2005). Another popular approach has been functional-typological, stating principles which are effectively descriptive generalizations over the typological data. Functional-typological work is often goal-oriented, and often assumes that languages change in particular directions in order to reach those goals. The positions above can be contrasted with diachronic explanations for typology. This approach is devoid of any functionalist, goal-oriented reasoning, and grounds typological patterns directly in pathways of change, which are themselves explained in a wide variety of ways: processing, acquisition, frequency, and so on. Whichever synchronic representations speakers use is assumed to be irrelevant. However, independently of what synchronic grammars look like, we independently need a theory of diachrony. If it is the case that diachronic pathways can explain the typological patterns we are interested in, then by Occam's Razor, no additional synchronic machinery needs to be assumed. The diachronic approach, therefore, often allows us to have a much simpler view of synchronic representations, sometimes without any restrictions at all on which synchronic patterns constitute possible human languages.

Section 3 gave an overview of previous work on Universal 20, i.e. the order of demonstrative, numeral, adjective, and noun. The work by Cinque (2005, 2017) and Dryer

(2009, 2011, 2013, 2018) has been especially important for the purposes of this paper. Cinque's account is purely synchronic representational, and assumes a universal base hierarchy Dem-Num-A-N, along with certain constraints on syntactic movement. By postulating that certain movements are more marked than others, Cinque is able to derive the cross-linguistic frequencies of the different word orders. Dryer (2018) instead gives a functional-typological explanation, with five word order principles. Word orders which violate more of these principles are rarer in the world's languages. He points out that any purely syntactic approach to Universal 20 will undergenerate, since the same word order patterns are observed independently of which syntactic categories are used: words which function semantically like adjectives may be verbs syntactically, but this does not change their behavior with respect to Universal 20.

Finally, in section 4, diachronic explanations for Dryer's word order principles were proposed. The observation that Dem-Num-A order is preferred, and is only preferred pre-nominally, was explained in terms of word length and working memory constraints. Cross-linguistically, demonstratives tend to be shorter than numerals, which tend to be shorter than adjectives. Pre-nominal word orders which have the modifiers in ascending order of length impose less of a processing load on working memory. Listeners are thus more likely to fail to parse orders such as A-Num-Dem-N (unattested) than, say, Dem-Num-A-N (very common). This assumes that word orders which impose more of a processing load are more difficult to process correctly. Over time, this means that learners have fewer examples of (correctly parsed) A-Num-Dem-N orders than Dem-Num-A-N, even if both are equally frequent in the input. Over time, this bias in favor of ordering modifiers by word length should grow stronger, and may establish a new word order pattern as dominant in a language. A very similar explanation was available for the post-nominal adjective preference. Longer modifiers are easier to process in post-nominal position, since the dependencies with the noun can be resolved straight away, without any need to keep a long word or phrase in working memory. Since adjectives tend to be longer than both demonstratives and numerals, this explains why it is adjectives and not demonstratives or numerals that preferentially occur after the noun.

The final principle to be explained was intracategorial harmony: the tendency for

modifiers to occur on the same side of the noun. A plausible explanation for this is structural priming, where speakers are biased towards a particular word order (e.g. modifier-noun) through repeated exposure. If some modifiers in a language are very often pre-nominal, structural priming can bias speakers towards placing other modifiers before the noun as well. This relies on some abstraction away from demonstratives, numerals, and adjectives to a category of modifiers. This type of abstraction seems to be well supported empirically by the psycholinguistic experiments that have been done on structural priming. Not only can we prime particular structures independently of the words used to realize that structure, but we can also see priming effects for similar structures across different languages.

In conclusion it seems as if Dryer's (2018) principles can be relatively straightforwardly explained diachronically. There are processing pressures which will over time cause languages to exhibit certain word order patterns more often than others. These word order patterns are the same as those predicted by Dryer's principles, which we know in turn explain the frequencies of different individual word orders in the world's languages. To the extent that the explanations I have given are successful, there is no need to impose additional constraints on synchronic representations, or additional functional pressures, in order to derive the Universal 20 typology. I hope that this work will be seen as a positive contribution to the study of diachronic explanation in typology, and to Evolutionary Syntax. There are many well-known typological patterns in syntax, few of which have been examined through a diachronic lens. By extending the approach I have taken here, we will likely find that many more syntactic patterns are simply the consequence of how each pattern came to be diachronically.

7 Bibliography

Abels, K., & Neeleman, A. (2012). Linear Asymmetries and the LCA. *Syntax*, *15*(1), 25–74.

Ackerman, F., Malouf, R., & Blevins, J. P. (2016). Patterns and discriminability in language analysis. *Word Structure*, *9*(2), 132–155. https://doi.org/10.3366/word.2016.0091

Anderson, S. R. (2005). Morphological universals and diachrony. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 2004* (pp. 1–17). Dordrecht: Springer Netherlands. https://doi.org/10.1007/1-4020-2900-4_1

Anderson, S. R. (2016). Synchronic Versus Diachronic Explanation and the Nature of the Language Faculty. *Annual Review of Linguistics*, *2*, 11–31. https://doi.org/10.1146/annurev-linguistics-011415-040735

Andersson, S. (2017). *He, he, it: Against universal restrictions on gender syncretism*. Talk presented at the SYNC2017, Yale University.

Andersson, S. (2018a). (*)ABA in Germanic verbs. *Glossa*, *3*(1). https://doi.org/10.5334/gjgl.733

Andersson, S. (2018b). *The Limits of Factorial Typology: For extralinguistic explanation in phonology*. Talk presented at the NAPhC X, Montreal.

Anttila, R. (1989). *Historical and comparative linguistics*. Philadelphia, PA: Benjamins.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Baisa, V., & Suchomel, V. (2012). Large Corpora for Turkic Languages and Unsupervised Morphological Analysis. *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, 28–32.

Bale, A., & Reiss, C. (2018). *Phonology: A Formal Introduction*. Cambridge, MA: MIT Press.

Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, *43*(3), 209–226. https://doi.org/10.1007/s10579-009-9081-4

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and Language Development* (pp. 279–362). New York, NY: John Wiley and Sons.

Blevins, J. (2004). *Evolutionary Phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Blevins, J. (2006). A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics*, *32*(2), 117–166.

Bobaljik, J. D. (2012). *Universals in Comparative Morphology: Suppletion, superlatives, and the structure of words*. Cambridge, MA: MIT Press.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387. https://doi.org/10.1016/0010-0285(86)90004-6

Bradley, P. T., & Mackenzie, I. (2004). *Spanish - An essential Grammar*. London/New York, NY: Routledge.

Brown, C. H., Holman, E. W., Wichmann, S., & Velupillai, V. (2009). Automated classification of the world′s languages: a description of the method and preliminary results. *STUF - Language Typology and Universals Sprachtypologie Und Universalienforschung*, *61*(4), 285–308. https://doi.org/10.1524/stuf.2008.0026

Bybee, J. L. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.

Caha, P. (2009). *The nanosyntax of case*. PhD Dissertation: University of Tromsø.

Chang, S.-J. (1996). *Korean*. Amsterdam/Philadelphia, PA: John Benjamins Publishing Company.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1986). *Knowledge of Language: Its nature, origin, and use*. New York, NY: Praeger.

Chomsky, N., & Lasnik, H. (1993). *Principles and Parameters Theory, in Syntax: An International Handbook of Contemporary Research*. Berlin: de Gruyter.

Cinque, G. (2005). Deriving Greenberg's Universal 20 and Its Exceptions. *Linguistic Inquiry*, *36*(3), 315–332.

Cinque, G. (2010). *The Syntax of Adjectives*. Cambridge, MA: MIT Press.

Cinque, G. (2014). Again on Tense, Aspect, Mood Morpheme Order and the "Mirror Principle." In P. Svenonius (Ed.), *Functional Structure from Top to Toe: The Cartography of Syntactic Structures* (pp. 232–265). Oxford: Oxford University Press.

Cinque, G. (2017). *On the movement account of Greenberg's Universal 20: Refinements and Replies*. Ms.

Comrie, B. (1981). *Language universals and linguistic typology.* Chicago, IL: Chicago University Press.

Coyle, J. M., & Kaschak, M. P. (2008). Patterns of experience with verbs affect long-term cumulative structural priming. *Psychonomic Bulletin & Review*, *15*(5), 967–970. https://doi.org/10.3758/PBR.15.5.967

Croft, W. (1990). *Typology and universals*. Cambridge: Cambridge University Press.

Cristofaro, S. (2010). Semantic Maps and Mental Representation. *Linguistic Discovery*, *8*(1), 35–52. https://doi.org/10.1349/PS1.1537-0852.A.345

Cristofaro, S. (2012). Cognitive explanations, distributional evidence, and diachrony. In N. Gisborne & W. B. Hollmann (Eds.), *Theory and Data in Cognitive Linguistics* (pp. 645–670). Amsterdam/Philadelphia, PA: John Benjamins.

Culbertson, J., & Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *PNAS*, *111*(16), 5842–5847.

de Lacy, P. (2006). *Markedness: Reduction and Preservation in Phonology.* Cambridge: Cambridge University Press.

Dik, S. C. (1997). *The theory of Functional Grammar*. Berlin: Mouton de Gruyter.

Dryer, M. S. (2009). *On the order of demonstrative, numeral, adjective, and noun: An alternative to Cinque*. Presented at the Theoretical approaches to disharmonic word orders, Newcastle University.

Dryer, M. S. (2011). *On the order of demonstrative, numeral, adjective, and noun: An alternative to Cinque*. University of Pavia.

Dryer, M. S. (2013). *On the order of demonstrative, numeral, adjective, and noun: An alternative to Cinque*. University of Toronto.

Dryer, M. S. (2018). On the order of demonstrative, numeral, adjective, and noun. *Language*, *94*(4), 798–833. https://doi.org/10.1353/lan.2018.0054

Flemming, E. (1995). *Auditory representation in phonology* (PhD). University of California, Los Angeles, CA.

Fodor, J. D., Nickels, S., & Schott, E. (2017). Center-Embedded Sentences: What's Pronounceable Is Comprehensible. In R. G. de Almeida & L. R. Gleitman (Eds.), *On Concepts, Modules, and Language: Cognitive Science at Its Core* (pp. 139–168). Oxford: Oxford University Press.

Ford, M. (1983). A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*(2), 203–218. https://doi.org/10.1016/S0022-5371(83)90156-1

Futrell, R., Levy, R., & Dryer, M. S. (2017). *A Statistical Comparison of Some Theories of NP Word Order*. Ms. Retrieved from https://arxiv.org/abs/1709.02783

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, *68*(1), 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1

Gibson, E. (2000). The Dependency Locality Theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, Language, Brain: Papers from the First Mind Articulation Project Symposium* (pp. 95–112). Cambridge, MA / London: MIT Press.

Givón, T. (1984). *Syntax: A functional-typological introduction*. Amsterdam: Benjamins.

Göksel, A., & Kerslake, C. (2005). *Turkish: A comprehensive grammar*. London/New York, NY: Routledge.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-Load Interference in Syntactic Processing. *Psychological Science*, *13*(5), 425–430. https://doi.org/10.1111/1467-9280.00475

Grammont, M. (1933). *Traité de phonétique*. Paris: Delagrave.

Greenberg, J. (1966). *Universals of language*. Cambridge, MA: MIT Press.

Gries, S. T., & Kootstra, G. J. (2017). Structural priming within and across languages: a corpus-based perspective. *Bilingualism: Language and Cognition*, *20*(2), 235–250. https://doi.org/10.1017/S1366728916001085

Grodner, D., & Gibson, E. (2005). Consequences of the Serial Nature of Linguistic Input for Sentenial Complexity. *Cognitive Science*, *29*(2), 261–290. https://doi.org/10.1207/s15516709cog0000_7

Guion, S. G. (1998). The Role of Perception in the Sound Change of Velar Palatalization. *Phonetica*, *55*, 18–52.

Hale, M., & Reiss, C. (2008). *The Phonological Enterprise*. Oxford: Oxford University Press.

Hannah, H. B. (1912). *A Grammar of the Tibetan Language*. Calcutta: Baptist Mission Press.

Harley, H., & Ritter, E. (2002). Person and Number in Pronouns: A Feature-Geometric Analysis. *Language*, *78*(3), 482–526. https://doi.org/10.1353/lan.2002.0158

Harms, R. T. (1962). *Estonian Grammar*. New York, NY: Indiana University, Bloomington.

Harrison, K. D., Dras, M., & Kapicioglu, B. (2002). Agent-Based Modeling of the Evolution of Vowel Harmony. In M. Hirotani (Ed.), *Proceedings of the Northeast Linguistic Society 32* (pp. 217–236). Amherst, MA: GLSA.

Hawkins, J. A. (1983). *Word Order Universals*. New York, NY: Academic Press.

Hawkins, J. A. (Ed.). (1988). *Explaining language universals*. Oxford: Basil Blackwell.

Hawkins, J. A. (1990). A Parsing Theory of Word Order Universals. *Linguistic Inquiry*, *21*(2), 223–261.

Hawkins, J. A. (2004). *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

Hawkins, J. A. (2014). *Cross-Linguistic Variation and Efficiency*. Oxford: Oxford University Press.

Hock, H. H. (1991). *Principles of historical linguistics* (2nd ed.). New York, NY: Mouton de Gruyter.

Hulme, C., & Tordoff, V. (1989). Working memory development: The effects of speech rate, word length, and acoustic similarity on serial recall. *Journal of Experimental Child Psychology*, *47*(1), 72–87. https://doi.org/10.1016/0022-0965(89)90063-5

Jacques, G. (2013). Harmonization and disharmonization of affix ordering and basic word order. *Linguistic Typology, 17*, 187–215. https://doi.org/10.1515/lingty-2013-0009

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The TenTen Corpus Family. *Proceedings of the 7th International Corpus Linguistics Conference CL 2013*, 125–127.

Kager, R. (1999). *Optimality Theory*. Cambridge: Cambridge University Press.

Kaschak, M. P. (2007). Long-term structural priming affects subsequent patterns of language production. *Memory & Cognition*, *35*(5), 925–937. https://doi.org/10.3758/BF03193466

Kaschak, M. P., & Borreggine, K. L. (2008). Is long-term structural priming affected by patterns of experience with individual verbs? *Journal of Memory and Language*, *58*(3), 862–878. https://doi.org/10.1016/j.jml.2006.12.002

Kaschak, M. P., Kutta, T. J., & Schatschneider, C. (2011). Long-term cumulative structural priming persists for (at least) one week. *Memory & Cognition*, *39*(3), 381–388. https://doi.org/10.3758/s13421-010-0042-3

Kayne, R. S. (1994). *The antisymmetry of syntax*. Cambridge, MA: MIT Press.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., … Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, *1*, 7–36.

Kilgarriff, A., & Renau, I. (2013). esTenTen, a Vast Web Corpus of Peninsular and American Spanish. *Procedia - Social and Behavioral Sciences*, *95*, 12–19. https://doi.org/10.1016/j.sbspro.2013.10.617

Kilgarriff, A., Reddy, S., Pomikálek, J., & P.V.S., A. (2010). A Corpus Factory for many languages. *LREC Workshop on Web Services and Processing Pipelines 2010*, 904–910.

King, J., & Just, M. A. (1991). Individual Differences in Syntactic Processing: The Role of Working Memory. *Journal of Memory and Language; New York*, *30*(5), 580–602.

Kluender, R. (1998). On the distinction between strong and weak islands: A processing perspective. *Syntax and Semantics*, *29*, 241–279.

Knuth, D. (1968). *The Art of Computer Programming 3: Sorting and Searching*. Reading, MA: Addison-Wesley.

Leslau, W. (1969). *An Amharic Reference Grammar*. Los Angeles, CA.

Lewis, R. L., & Vasishth, S. (2005). An Activation-Based Model of Sentence Processing as Skilled Memory Retrieval. *Cognitive Science*, *29*(3), 375–419. https://doi.org/10.1207/s15516709cog0000_25

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454. https://doi.org/10.1016/j.tics.2006.08.007

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental Phonology* (pp. 13–44). Orlando, FL: Academic Press.

Loebell, H., & Bock, K. (2003). Structural priming across languages. *Linguistics*, *41*(5), 791–824. https://doi.org/10.1515/ling.2003.026

Lu, B. (1998). *Left-Right Asymmetries of Word Order Variation*. PhD Dissertation: University of Southern Caliifornia (USC).

McCarthy, J. J., & Prince, A. (1993). Prosodic Morphology: Constraint Interaction and Satisfaction. University of Massachusetts, Amherst, and Rutgers University. RuCCS-TR-3.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, *48*(1), 67–91. https://doi.org/10.1016/S0749-596X(02)00515-6

McGovern, W. M. (1942). *Colloquial Japanese*. London: Kegan Paul, Trench, Trubner & Co. Ltd.

Medeiros, D. P. (2012). *Movement as tree-balancing: an account of Greenberg's Universal 20*. Presented at the LSA Annual Meeting, Portland, Oregon.

Medeiros, D. P. (2017). *Stack-sorting: Universal 20 and beyond*.

Medeiros, D. P. (2018). ULTRA: Universal Grammar as a Universal Parser. *Frontiers in Psychology*, *9*(155). https://doi.org/doi:10.3389/fpsyg.2018.00155

Meelen, M., Hill, N., & Handy, C. (2017). *The Annotated Corpus of Classical Tibetan (ACTib), Part II - POS-tagged version, based on the BDRC digitised text collection, tagged with the Memory-Based Tagger from TiMBL* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.822537

Ohala, J. J. (1992). What's cognitive, what's not, in sound change. In M. Kellermann (Ed.), *Diachrony within Synchrony: Language history and cognition* (pp. 309–355). Frankfurt: Lang.

Ohala, J. J. (1997). Aerodynamics of phonology. *Proceedings of the 4th Seoul International Conference on Linguistics*, 92–97.

Passy, P. (1890). *Étude sur les changements phonétiques*. Paris: Firmin-Didot.

Pickering, M. J., McLean, J. F., & Branigan, H. P. (2013). Persistent Structural Priming and Frequency Effects During Comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 890–897. https://doi.org/10.1037/a0029181

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin*, *133*(2), 227–244. https://doi.org/10.1037/0033-2909.133.2.227

Prince, A., & Smolensky, P. (1993). Optimality Theory: Constraint Interaction in Generative Grammar. Center for Cognitive Science, Rutgers University, New Brunswick.

Rijkhoff, J. (2002). *The Noun Phrase*. Oxford: Oxford University Press.

Ross, C., & Sheng Ma, J. (2006). *Modern Mandarin Chinese Grammar: A practical guide*. London/New York, NY: Routledge.

Rychlý, P., & Suchomel, V. (2016). Annotated Amharic Corpora. In P. Sojka, A. Horák, I. Kopeček, & K. Pala (Eds.), *Text, Speech, and Dialogue* (pp. 295–302). Springer International Publishing.

Saeed, J. (1999). *Somali*. Amsterdam/Philadelphia, PA: John Benjamins Publishing Company.

Sayeed, O. (2018). >≠→! Poster presented at the NAPhC X, Montreal.

Sayeed, O., & Ceolin, A. (2018). *"Markedness" is an epiphenomenon of random phonetically grounded sound change*. Poster presented at the 6th Annual Meeting on Phonology (AMP), San Diego, CA.

Smith, K., Kirby, S., & Brighton, H. (2003). Iterated Learning: A Framework for the Emergence of Language. *Artificial Life*, *9*(4), 371–386. https://doi.org/10.1162/106454603322694825

Smith, P. W., Moskal, B., Kang, J., Xu, T., & Bobaljik, J. D. (2015). Pronominal suppletion: Case and number. In T. Bui & D. Özyıldız (Eds.), *Proceedings of the 45th annual meeting of the North East linguistic society 3* (pp. 69–78). Amherst: GLSA, UMass.

Sneddon, J. N., Adelaar, A., Djenar, D. N., & Ewing, M. C. (2010). *Indonesian Reference Grammar* (2nd ed.). Crows Nest, NSW: Allen & Unwin.

Srdanović, I., Suchomel, V., Ogiso, T., & Kilgarriff, A. (2013). Japanese Language Lexical and Grammatical Profiling Using the Web Corpus JpTenTen. *Proceeding of the 3rd Japanese Corpus Linguistics Workshop.*, 229–238.

Steddy, S., & Samek-Lodovici, V. (2011). On the Ungrammaticality of Remnant Movement in the Derivation of Greenberg's Universal 20. *Linguistic Inquiry*, *42*(3), 445–469.

Suchomel, V., & Pomikálek, J. (2012). Efficient Web Crawling for Large Text Corpora. *Proceedings of the Seventh Web-as-Corpus Workshop, Lyon*, 39–43.

Thompson, L. C. (1965). *A Vietnamese Grammar*. Seattle, WA: University of Washington Press.

Thompson, K. D., & Schleicher, A. F. (2001). *Swahili Learners' Reference Grammar.* University of Wisconsin-Madison: NALRC Press.

Vasishth, S., & Lewis, R. L. (2006). Argument-Head Distance and Processing Complexity: Explaining Both Locality and Antilocality Effects. *Language*, *82*(4), 767–794.

Vaux, B., & Samuels, B. (2015). Explaining vowel systems: dispersion theory vs natural selection. *The Linguistic Review*, *32*(3), 573–599. https://doi.org/10.1515/tlr-2014-0028

Vennemann, T. (1975). Analogy in Generative Grammar: The origin of word order. In L. Heilmann (Ed.), *Proceedings of the Eleventh International Congress of Linguists, Bologna-Florence, Aug. 28 - Sept. 2, 1972* (Vol. 2, pp. 79–83). Bologna: Società editrice il Mulino.

Wade, T. (2011). *A Comprehensive Russian Grammar* (3rd ed.). Oxford: Wiley-Blackwell.

Wichmann, S., Holman, E. W., & Brown, C. H. (Eds.). (2018). *The ASJP Database (version 18)*.

Wilson, M., & Emmorey, K. (1998). A "word length effect" for sign language: Further evidence for the role of language in structuring working memory. *Memory & Cognition*, *26*(3), 584–590. https://doi.org/10.3758/BF03201164