

Review

Natalia Levshina*, Savithry Namboodiripad*,
Marc Allasonnière-Tang, Mathew Kramer, Luigi Talamo,
Annemarie Verkerk, Sasha Wilmoth, Gabriela Garrido Rodriguez,
Timothy Michael Gupton, Evan Kidd, Zoey Liu, Chiara Naccarato,
Rachel Nordlinger, Anastasia Panova and Natalia Stoyanova

Why we need a gradient approach to word order

<https://doi.org/10.1515/ling-2021-0098>

Received May 13, 2021; accepted April 9, 2022; published online April 25, 2023

***Corresponding authors: Natalia Levshina**, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands, E-mail: natalevs@gmail.com; and **Savithry Namboodiripad**, Department of Linguistics, University of Michigan, 409 Lorch Hall, 611 Tappan Street, Ann Arbor, MI 48109-1220, USA, E-mail: savithry@umich.edu. <https://orcid.org/0000-0002-7685-5895>

Marc Allasonnière-Tang, MNHN/CNRS/Université Paris Cité, Paris, France, E-mail: marc.allasonniere-tang@mnhn.fr. <https://orcid.org/0000-0002-9057-642X>

Mathew Kramer, University of Michigan, Ann Arbor, MI, USA, E-mail: arkram@umich.edu. <https://orcid.org/0000-0002-0509-1453>

Luigi Talamo and Annemarie Verkerk, Saarland University, Saarbrücken, Germany, E-mail: luigi.talamo@uni-saarland.de (L. Talamo), annemarie.verkerk@uni-saarland.de (A. Verkerk). <https://orcid.org/0009-0009-4640-3052> (L. Talamo). <https://orcid.org/0000-0002-3351-8362> (A. Verkerk)

Sasha Wilmoth and Rachel Nordlinger, University of Melbourne, Melbourne, Australia, E-mail: sasha.wilmoth@unimelb.edu.au (S. Wilmoth), racheln@unimelb.edu.au (R. Nordlinger). <https://orcid.org/0000-0002-6626-9104> (S. Wilmoth). <https://orcid.org/0000-0003-4126-8022> (R. Nordlinger)

Gabriela Garrido Rodriguez, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, E-mail: Gabriela.Garrido@mpi.nl

Timothy Michael Gupton, University of Georgia, Athens, GA, USA, E-mail: gupton1@uga.edu. <https://orcid.org/0000-0003-4970-823X>

Evan Kidd, MPI for Psycholinguistics, Nijmegen, The Netherlands; The Australian National University, Canberra, Australia; and ARC Centre of Excellence for the Dynamics of Language, Canberra, Australia, E-mail: Evan.Kidd@mpi.nl. <https://orcid.org/0000-0003-4301-2290>

Zoey Liu, University of Florida, Gainesville, FL, USA, E-mail: liu.ying@ufl.edu

Chiara Naccarato, Anastasia Panova and Natalia Stoyanova, National Research University Higher School of Economics, Moscow, Russia, E-mail: ch1naccarato@gmail.com (C. Naccarato), anastasia.b.panova@gmail.com (A. Panova), stoyanova@yandex.ru (N. Stoyanova). <https://orcid.org/0000-0003-0017-6316> (C. Naccarato). <https://orcid.org/0000-0003-0793-671X> (A. Panova). <https://orcid.org/0000-0001-8979-3788> (N. Stoyanova)

Abstract: This article argues for a gradient approach to word order, which treats word order preferences, both within and across languages, as a continuous variable. Word order variability should be regarded as a basic assumption, rather than as something exceptional. Although this approach follows naturally from the emergentist usage-based view of language, we argue that it can be beneficial for all frameworks and linguistic domains, including language acquisition, processing, typology, language contact, language evolution and change, and formal approaches. Gradient approaches have been very fruitful in some domains, such as language processing, but their potential is not fully realized yet. This may be due to practical reasons. We discuss the most pressing methodological challenges in corpus-based and experimental research of word order and propose some practical solutions.

Keywords: continuous variables; entropy; gradience; typology; variability; word order

1 Aims of this article

1.1 What do we mean by a gradient approach?

In this article we argue for a gradient approach to word order. By advocating for a gradient approach, we put forth two main theoretical stances. First, we argue for the presumption of variability in word order research, or for treating variability as the null hypothesis. Second, from a crosslinguistic perspective, we argue for a presumption of gradience; by default, we expect that languages should vary in degree but not kind when it comes to word order variability. From the perspective of description, a gradient approach means that word order patterns should be treated as a continuous variable. For example, in addition to – or instead of – labeling a language as SO (with the dominant Subject-Object order) or OS (with the dominant Object-Subject order), we can compute and report the proportion of SO and OS based on behavioral data (from sources such as corpora or experiments). Similarly, instead of or in addition to labeling languages as having fixed, flexible, or free word order, we can measure the degrees of this variability by using quantitative measures, such as entropy, in a crosslinguistically comparable manner. Along with increasing descriptive adequacy, this allows us to move beyond categorical claims which stipulate that rigid word order provides a cue for assigning grammatical functions, such as Subject and Object, to noun phrases; instead, we can measure the reliability and strength of word order as a cue for argument structure based on behavioral data.¹

¹ Note that our claims apply to the order of words and larger-than-word units, as well, such as syntactic constituents and clauses; in this paper, we use “word” as a general term, and specify constituents/clauses when relevant.

A gradient approach to the description of a particular language can be illustrated by a simple analogy: Instead of using categorical color terms like “white”, “blue” or “orange”, we can encode them numerically by using different codes. For example, we can talk about a color for which there is a word in English, “orange”, in RGB terms as being 100% red, 64.7% green, and 0% blue. In CMYK color space, it is 0% cyan, 35.3% magenta, 100% yellow and 0% black, and it is assigned a hex code #ffa500. However, an additional advantage of this continuous measure is that we can also describe a color for which there is no standard label, such as #BA55D3, which is 73% red, 33% green, and 83% blue, and is given the label “medium orchid” in the X11 color names system. Thus, using various continuous measures, we can not only account for the underlyingly gradient properties of color, but we can also describe more and less prototypical colors on equal footing. Likewise, we can use cross-linguistically comparable gradient measures to describe a particular language, to describe how languages vary, and as a basis for understanding how linguistic variables may interact to produce typological patterns and motivate change.

1.2 Theoretical and practical concerns with gradience

Despite some previous research problematizing categorical distinctions in linguistics (e.g., Wälchli 2009), non-gradient approaches to word order have been traditionally dominant in cross-linguistic studies (e.g., Dryer 1992, 2013a, 2013b; Greenberg 1963; Lehmann 1973; Vennemann 1974). In contemporary quantitative typology and Bayesian phylogenetic models of language evolution, word order is still usually treated categorically (Cathcart et al. 2018; Dunn et al. 2011; Greenhill et al. 2017; Hammarström 2015; Jäger and Wahle 2021; Maurits and Griffiths 2014). Popular large-scale cross-linguistic databases, such as WALS (Dryer and Haspelmath 2013), AUTOTYP (Bickel et al. 2022), Grambank² and DiACL (Carling 2017), provide categorical word-order data.

There are both theoretical and practical reasons for this situation. From a theoretical perspective, a widespread view has been that grammar in general and word order in particular should be described by discrete features, categories, rules, or parameters (e.g., Guardiano and Longobardi 2005; Jackendoff 1977; Lightfoot 1982; Vennemann and Harlow 1977), although probabilistic approaches to grammar are becoming increasingly influential (Bod et al. 2003; Bresnan et al. 2007; Grafmiller et al. 2018; also see Section 1.2.2). A significant practical barrier to adopting a gradient perspective is that it is still difficult to obtain the types of behavioral or corpus data in many languages which would be necessary for descriptions using gradient measures. However, these barriers are rapidly falling away; we have access to new data

² www.glottobank.org (accessed 1 March 2022).

sources in the form of large corpora, as well as software for processing and analyzing them statistically. To take one example, the Universal Dependencies project (UD; Zeman et al. 2020) had 10 treebanks for 10 languages in 2015. Version 2.9, released in November 2021, contained 217 treebanks for 122 languages, with more in the works. Computational algorithms, such as multilingual word alignment applied to massively parallel corpora, can be used to scale up gradient approaches, providing word order information for almost a thousand languages (Östling 2015). The proliferation of quantitative approaches to different areas of linguistics (Janda 2013; Kortmann 2020), which has been possible thanks to the development of user-friendly statistical software and robust experimental methods, has made us better equipped than ever for investigating gradient phenomena, though many practical issues are far from being fully solved, as we show in Section 4.

1.2.1 Gradience across generative approaches

Even in approaches where, historically, gradience has not been a central concern, various strategies have emerged to account for variability in word and constituent order. With the recent diversification of methodologies in the generative paradigm, researchers are recognizing the need to account for gradience and variability in theory building.³ This section first discusses so-called ‘mainstream’ generative approaches (cf. Culicover and Jackendoff 2006) and accounts therein of gradience and variability in order. Then, we look to other major generative approaches, namely Lexical-Functional Grammar and Head-Driven Phrase Structure Grammar, in which accounting for variation in word/constituent order has played a central role in theory construction. We review these here as part of demonstrating why gradience might not have been as central to previous accounts of word order flexibility; the theoretical moves we advocate for in this article are certainly compatible with a range of approaches to syntax, though the details of implementation and, perhaps most crucially, how interested practitioners might be in the associated questions, will of course vary.

Historically, gradience in production or speaker judgments was explained as a difference among grammars: in short, crosslinguistic variation related to proposed parametric differences (e.g., Chomsky 1986). Although the Minimalist Program’s feature-checking model generates canonical word orders, subject position, as well as derived word orders in a step-wise, phase-based derivation (though cf. Fanselow 2009), optionality and gradience are problematic and often not assumed to exist.⁴ There is

³ For an experimental perspective, see, e.g., Leal and Gupton (2021). For a review of the micro-parametric approach and its application to diachronic studies, see, e.g., Roberts (2019).

⁴ For the Minimalist Program, see Chomsky (1995, 2001, 2008). See Preminger (2017) for arguments in favor of a *feature-valuation* model over a *feature-checking* model.

work that aims to account for insights related to frequency and innovation in inter-generational language transmission and acquisition (e.g., Biberauer 2019; Gravelly 2021; Yang 2002) or from variationist sociolinguistics (e.g., Adger and Smith 2005, 2010) by combining the assumptions and theoretical machinery of the Minimalist Program with probabilistic analyses of variable word order. In this type of work, variability can be accounted for by gradient or stochastic probabilistic constraints on lexical/feature selection, motivated by a variety of factors, including functional ones. However, because the organizing principle of this framework is to reduce the specificity of syntactic operations, the types of processes which are relevant to account for different orders within languages or varieties of languages, for example, are often explicitly constructed as being extra- or post-syntactic operations – a theoretical conundrum, given that such processes are reflected in the syntax itself. As such, analyses which include variation or gradience are necessarily about interfaces (as this is a modular framework) with other levels of linguistic analysis.

Syntactic variation related to discourse-information structure is one particularly relevant example, and it presents one of the biggest challenges for the Y-model of language (e.g., Irurtzun 2009). Although no crosslinguistic model accounting for variability due to information structure currently exists in this framework, numerous attempts have centered on particular phenomena or languages.⁵ Examples are Rizzi's (1997) Cartographic Program, Frascarelli and Hinterhölzl's (2007) isomorphism of syntax, information structure, and intonation, and Zubizarreta (1998), who elegantly accounts for syntax, prosody, and information structure in Germanic and Romance without the split-CP architecture. However, they are not unproblematic. In particular, recent experimental tests of Spanish varieties find variation not predicted by Zubizarreta's (1998) account, especially for subject information focus (i.e., rheme).⁶ In fact, Feldhausen and Vanrell (2015), which builds upon Zubizarreta's account, does so using a non-categorical approach, namely Stochastic Optimality Theory (Boersma and Hayes 2001).

Indeed, Optimality Theoretic approaches have been a productive space for generative syntacticians interested in making variation (be it gradient or stochastic) more central to syntactic analyses. Ortega-Santos (2016) is a compelling application of an Optimality Theoretic approach to focus-related word order variation in Spanish. Müller (2019) uses a Gradient Harmonic Grammar (Smolensky and Goldrick 2016) approach to shed light on extraposed infinitive constructions in German, showing

5 See, e.g., Culicover and Rochement (1983) on focus and stress in English, Horvath (1986) on focus in Hungarian, and López (2009) on a model for Catalan and Spanish which also integrates information structure, and Williams (1977) on VP anaphora in English.

6 See, e.g., Gabriel (2010) for Argentine Spanish; Hoot (2016) for Northern Mexican Spanish; Jiménez-Fernández (2015) for Andalusian Spanish; Leal et al. (2018) for Chilean and Mexican Spanish, as well as Spanish-English bilinguals living in the US; Muntendam (2013) for Andean Spanish.

how this approach can explain apparent variable strength in the CP realm. Since *Syntactic Structures* (Chomsky 1957), the vast majority of generative approaches have sought descriptive as well as explanatory adequacy; therefore, not assuming or accounting for gradience means not capturing an important portion of speaker competence when it comes to word order. The result is a theoretical model that is, at best, incomplete, and, at worst, incorrect.

The question of how to deal with languages with flexible word order has played an important role in how non-hegemonic generative approaches to syntax diverged from other generative approaches. For example, Lexical-Functional Grammar (LFG) is characterized by a division between constituent structure and functional structure, and this has been used specifically to deal with languages with highly flexible constituent order (Austin and Bresnan 1996; Dalrymple et al. 2019). In LFG, languages are not *a priori* required to be specified for particular orders, and languages such as Plains Cree or Malayalam are argued to be non-configurational (Dahlstrom 1986; Mohanan 1983; see also Nordlinger 1998 and Simpson 2012 for discussion of non-configurational Australian languages). Accounting for flexible constituent order languages was not as central to the development of Head-Driven Phrase Structure Grammar (HPSG), in which linearization of constituents is similarly separate from constituent structure (Wechsler and Asudeh 2021), but, as HPSG has been implemented in a variety of languages with flexible constituent order, the framework is flexible enough to account for the gradience seen in natural language (e.g., Fujinami 1996 for Japanese; Mahmud and Khan 2007 for Bangla; Müller 1999 for German; Simov et al. 2004 for Bulgarian). While gradience is not central to these approaches, the existence of flexibility in constituent order, and, relatedly, whether and where grammatical relations are specified in formalisms and/or speakers' knowledge,⁷ has not only been addressed in these frameworks, but dealing with word order flexibility has been the source of direct cross-framework comparison. What remains underdeveloped in these accounts is the ability to explain and measure degree of flexibility.

1.2.2 Gradience in usage-based linguistics

Though all theoretical approaches must deal with it in some way, word order gradience follows naturally from approaches that assume a dynamic usage-based view of grammar (e.g., Bybee 2010; Diessel 2019). This can be illustrated by grammaticalization processes. For example, English is mostly a prepositional language. Since many prepositions represent a result of the grammaticalization of verbs, and English has predominantly VO order, the normal outcome of this process is the development

7 For a more recent discussion of this general issue which compares formal to cognitive-functional perspectives, see Johnston (2019), who argues that grammatical relations are not specified in AusLan.

of prepositions. However, English also has a few postpositions, such as *ago* and *notwithstanding*. Nouns that are now used with those postpositions were originally the subjects of the source verbs (Dryer 2019). The bottom-up emergence of grammar from language use thus is very likely to result in variability/construction-specific idiosyncrasies, such that a categorical approach of applying labels to languages as a whole is less accurate than a gradient approach, which can quantify degrees of divergence from some canonical extreme.

The usage-based approach assumes that the user's knowledge of grammar is probabilistic (Bod et al. 2003; Grafmiller et al. 2018) because it is derived and updated based on exemplars of usage events stored in the memory (Bybee 2010), from which more abstract generalizations can be formed (Goldberg 2006). Individual language users implicitly learn statistical variability, or "soft constraints" (Bresnan et al. 2001), from the input (MacDonald 2013). This view is supported by the fact that language users are able to predict the likelihood of a variant in a specific context, closely matching predictions based on corpora (Bresnan and Ford 2010; Klavan and Divjak 2016). The probabilistic variation is captured by complex statistical models of language users' behavior (e.g., Bresnan et al. 2007; Gries 2003; Szmrecsanyi et al. 2016, to name just a few). Gradience thus forms a part of mental grammar, and categories are emergent or epiphenomenal.

Both word order variability and rigidity in such a framework result from competition between factors involved in language processing and learning, which can have different weights in one language and across languages (see Section 2.1). In this sense, the probabilistic view has much in common with Optimality Theory (see Lee 2003; Keller 2006, *et alia* for work which deals directly with word order).⁸ The difference is that probabilistic grammars do not assume a fixed set of innate constraints (Grafmiller et al. 2018). This variability is instead represented in the individual speaker's grammar in the form of sequential associations of different strength, depending on the degree of entrenchment of a sequence in an individual mind (Diessel 2019), which both depends on and determines the degree of conventionalization of this sequence in the community (Schmid 2020).

1.3 Moving from categories to gradience in language description

In this article, we argue that a gradient approach is more descriptively adequate than the approaches which rely on categorical labels like "VO" and "OV", or "rigid/fixed", "flexible" and "dominant". In particular, rigid order means that some orders are

⁸ See also the Competing Grammars Framework, Kroch (1989), formalized in Kauhanen and Walkden (2018).

“either ungrammatical or used relatively infrequently and only in special pragmatic contexts” (Dryer 2013b). If different possible word orders are grammatical, languages have flexible order. Many flexible-order languages are claimed to have a dominant word order, i.e., the more frequent one (e.g., the order that is used at least twice as frequently as any other order is considered dominant); in the absence of frequency data, or in languages where the dominant order might differ based on contexts such as register or genre (see Payne 1992), this could also be the order labeled as “basic”, “pragmatically neutral” or “canonical” in language descriptions.

However, as has been noted (e.g., Hale 1983; Salzmann 2004), a clear-cut distinction between rigid-order languages and flexible-order languages with a dominant order is problematic. First of all, pragmatic neutrality is a slippery notion that depends on the specific situation. In spontaneous informal conversations in Russian, for example, there is nothing pragmatically marked about putting a newsworthy object first. For example, (1a), which comes from a transcribed spoken text, has OSV. Compare it with (1b) taken from an online news report, which has SVO, or given – new order, which is pragmatically neutral here. See more on register and modality effects in Section 4.1.3.

(1) Russian

a. Spontaneous conversation

Čajnik ja postavi-l-a.
kettle.ACC 1SG.NOM set.PFV-PST-SG.F
'I set the kettle.'

(Zemskaja and Kapanadze 1978: “A day in a family”)

b. Online news

Ona postavi-l-a čajnik na plit-u...
3SG.F.ACC set.PFV-PST-SG.F kettle.ACC on stove-ACC
'She set the kettle on the stove... (and forgot about it)'
(mir24.tv)

The second criterion, text frequency, is also problematic. Converting a continuous measure to a categorical one means loss of data, and there is not always a clear cut-off point. Consider an illustration. Using corpora of online news in the Leipzig Corpora Collection (Goldhahn et al. 2012) in 31 languages, annotated with the Universal Dependencies (Zeman et al. 2020), we computed the proportions of Subject – Object order in the total number of clauses in which both arguments were expressed by common nouns. The sample sizes were very large, with a median number of about 138,000 clauses. The data are available online as *Dataset1.txt* in an OSF directory.⁹

The result is shown in Figure 1. In all 31 languages, Subject usually comes before Object. The distribution of the scores represents a continuum from flexible languages

⁹ https://osf.io/6m7ec/?view_only=1eb102ddd91d41dcaaea0f266e5eab79.

(Lithuanian, Hungarian, Tamil, Latvian, and Czech) to well-known rigid languages (e.g., Indonesian, French, English, Norwegian, Danish, and Swedish). This plot demonstrates that there is no clear boundary between the two types, so it is difficult to choose the most appropriate cut-off point.

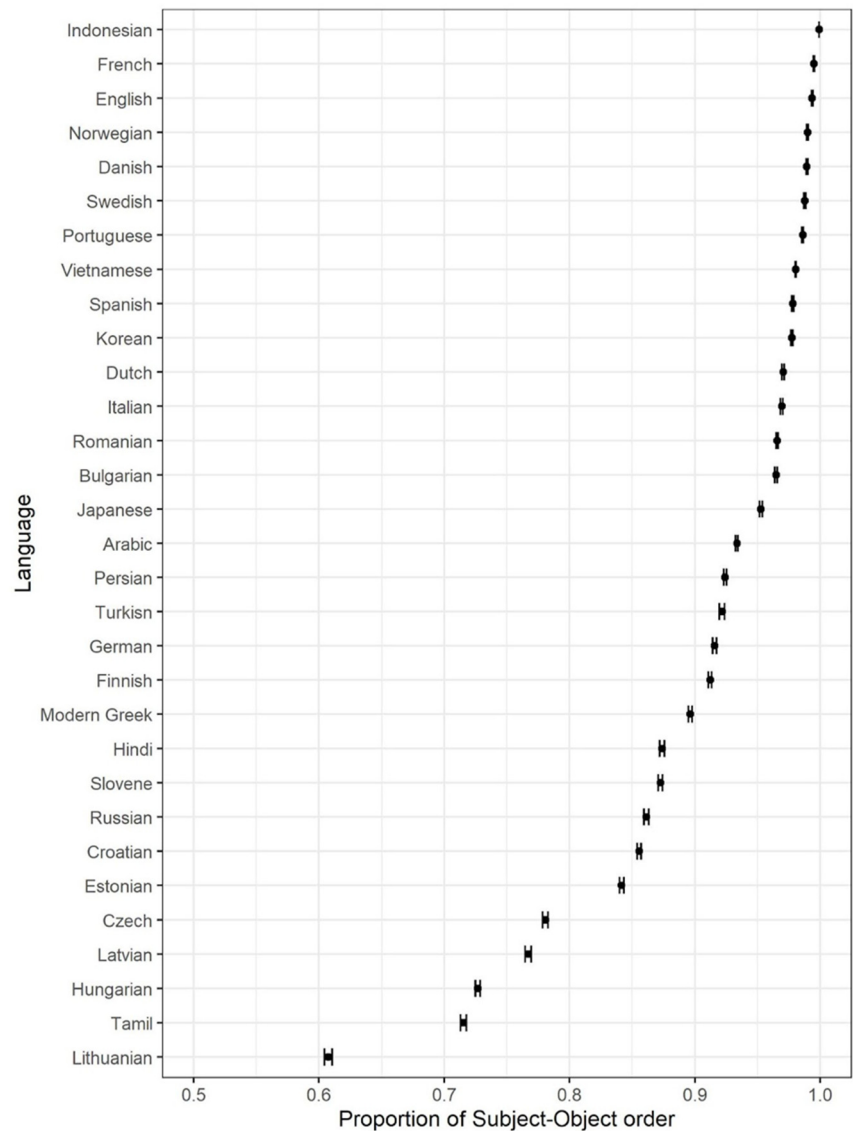


Figure 1: Proportions of the Subject first, Object second order in sentences with nominal Subject and Object in online news corpora. The lines delimit the 95% confidence intervals. Due to large samples, they are very narrow.

Approaches such as Dryer (2013a), which includes languages without a dominant order, resulting in a simple three-way scale (e.g., Adjective – Noun, no dominant order, Noun – Adjective), and Siewierska (1998), which determines relative flexibility based on the number of grammatical orders in a language, can be seen as precursors of gradient approaches. Here, we advocate for taking the next logical step and moving to more fully continuous variables, as these are not only descriptively more adequate, they also allow us to ask and answer more questions about word order, as discussed in more detail in Section 3. A gradient approach may not always lead to dramatically different results from a categorical one. For example, word order correlations, e.g., the correlation between the order of verb and object, and the order of adpositions and nouns, would still be correct under the gradient approach because both orders display low variability in many languages (Levshina 2019; see also Section 3.5). Yet, we can investigate more linguistic phenomena and languages if we characterize word order with the help of continuous measures.

Note that we do not argue for completely banishing descriptive labels like “canonically SVO language.” Such labels can still be useful as a shortcut, especially in the absence of more precise measures. However, under a gradient approach, we make explicit that these labels reflect convenient simplifications rather than representing inherent properties of particular languages. Reiterating our two main arguments, presuming within-language variability and cross-linguistic gradience when it comes to typological categorization, we posit that gradience itself might be a good candidate for an inherent property of language.

In some cases, conventional categories can be misleading. For example, the plot in Figure 2 displays the distribution of proportions of head-final phrases in 123 corpora annotated with Surface-syntactic Universal Dependencies (Osborne and Gerdes 2019).¹⁰ The dataset is available as *Dataset2.txt* in the OSF directory, which also contains the Python script that was used for the data extraction. The plot shows that the corpora do not follow a distribution with two peaks at each end, confirming the results obtained by Liu (2010) based on a small sample of languages. In other words, the corpora are not strongly head-initial and only rarely strongly head-final (cf. Polinsky 2012). In fact, the main bulk of the corpora contains between 25% and 50% of head-final phrases. The conventional labels mask the asymmetric and gradient shape of the distribution, which is best represented by continuous measures.

Studying gradience requires interdisciplinarity. In order to understand why word order exhibits more or less variability, we need to consider diverse factors related to language acquisition, processing, language contact, language change, prosody, and many

¹⁰ The corpora can be downloaded from <https://grew.fr/download/sud-treebanks-v2.8.tgz>. Only the training corpora were taken. Some languages were represented by more than one corpus.

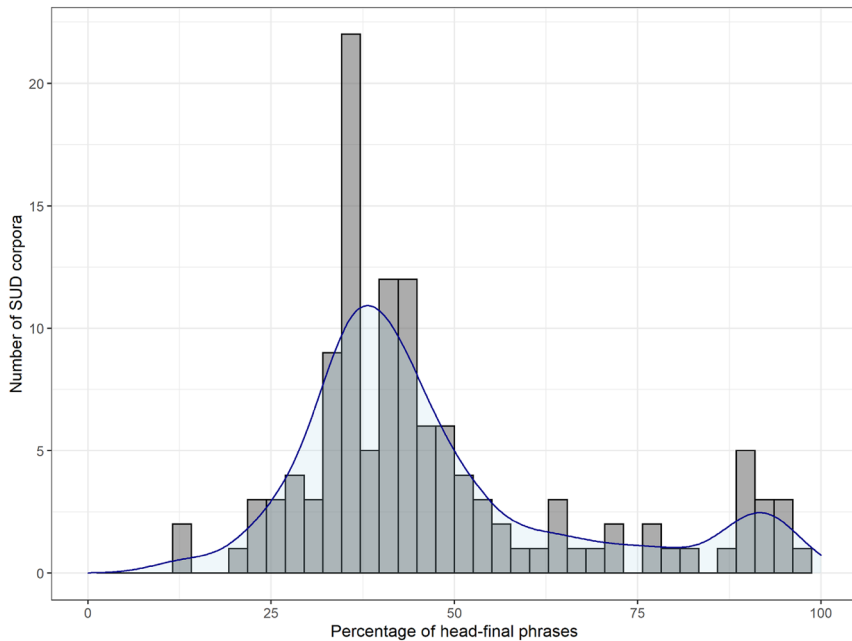


Figure 2: Distribution of head-finalness in SUD corpora.

others. These are discussed in Section 2. In Section 3, we discuss the research questions in different linguistic domains that cannot be asked and answered without taking a gradient approach. In order to study word order as a gradient phenomenon, we need quantitative measures of variability, such as probabilities or entropy, and empirical data from experiments or corpora. We suggest some methodological solutions and discuss the main challenges in Section 4. Finally, Section 5 provides conclusions and poses pertinent questions for future research across different linguistic disciplines.

2 What conditions word order gradience?

The aim of this section is to identify the main factors that contribute to the emergence of gradient word order patterns, and which make it possible for those patterns to remain in a language community. We begin with individual cognitive processes leading to gradient patterns, reviewing a large body of experimental and corpus-based work on language processing and acquisition (Sections 2.1 and 2.2). After that, we move on to word order gradience at the level of communities, focusing on

language change (Section 2.3) and the processes of language contact, which interact with language variation in vernaculars (Section 2.4).

2.1 Language processing

There is a large body of psycholinguistic work that has investigated what conditions word order variability at the level of an individual language user and learner. An important role is played by the accessibility of information expressed by constituents, depending on their semantics, information status, and formal weight. It is advantageous for the speaker/signer to place more accessible information first because it helps to save time for planning the less accessible parts. Specifically, there is a general preference to place human (and, more broadly, animate) referents in early appearing and/or prominent sentence positions (i.e., to realize them as grammatical subjects) (Branigan et al. 2008; see also Meir et al. 2017, who found a “me first” principle across three groups of sign languages and in elicited pantomime). The explanation for these effects on word order concern *conceptual accessibility*: humans find it easier to access concepts denoting animate entities along with their linguistic labels from memory (Bock and Warren 1985). Accordingly, Tanaka et al. (2011) reported that Japanese-speaking adults were more likely to recall OSV sentences as having SOV word order when this resulted in an animate entity appearing before an inanimate entity. Thus, sentence (2a) was often recalled as (2b).

(2) Japanese

- a. minato de, booto-o ryoshi-ga hakonda
harbor in boat-ACC fisherman-NOM carried
'In the harbor, the fisherman carried the boat.'
(Tanaka et al. 2011: 322)
- b. minato de, ryoshi-ga booto-o hakonda
harbor in fisherman-NOM boat-ACC carried
'In the harbor, the fisherman carried the boat.'
(Tanaka et al. 2011: 322)

This effect may ultimately have its origin in preferred modes of event construal. Studies of scene perception and sentence planning using eye-tracking show that speakers can rapidly extract information about participant roles in events, with speakers fixating on characters and entities that are informative about the scene as a whole (Konopka 2019).

Constituent order is also influenced by the discourse status of referents. The earliest proposal for this principle was perhaps by Behaghel (1932: 4): “es stehen die

alten Begriffe vor den neuen” [old concepts come before new ones]. This generalization was then later captured as “given before new” by Gundel (1988). The effects of discourse information status have been empirically attested in various contexts. For example, the production experiments of Bock and Irwin (1980) showed that, for English speakers, given information tends to be produced earlier. Similar results have been replicated in Ferreira and Yoshita (2003) with Japanese. Arnold et al. (2000) demonstrated that in both heavy NP shift and the dative alternation in English, speakers prefer to produce the relatively newer constituent later. The influence of information structure would in turn affect ordering flexibility, showing more fixed preference for given information to appear first (for example, Israeli Sign Language has been said to have a Topic-Comment order [Rosenstein 2001], see also McIntire [1982] on American Sign Language). Although most languages discussed in that regard have predominantly given-before-new order, some languages prefer to put new and/or newsworthy information first, e.g., Biblical Hebrew, Cayuga (Iroquoian, the USA), Ngandi (Gunwinyguan, Australia), and Uto (Uto-Aztecan, the USA) (Mithun 1992). This can be explained by the competing principle: “More important or urgent information tends to be placed first in the string” (Givón 1991: 972). But manifestations of this principle can also be found in languages with given-first order, e.g., the clause-initial placement of contrastive topic and focus, but also of full nominal phrases, as in (1a). Crucially, these principles interact, and they are often strong tendencies as opposed to absolute principles; competition between factors such as “given first” and “important first” can lead to gradient effects within and across languages.

Intonation plays an important role in these processes. Across languages, we see that less frequent orders are associated with particular intonational patterns (e.g., Downing et al. 2004; Patil et al. 2008; Vainio and Järviö 2006; see also Büring 2013), but also there seems to be a relationship between word order flexibility and the degree to which information structure is encoded prosodically, via word order, or both. Swerts et al. (2002) compared prominence and information status in Italian and Dutch. Italian has relatively flexible word order within noun phrases as compared to Dutch, and Swerts et al. found that, within noun phrases, Dutch speakers encoded information status prosodically in production, and took advantage of this information in perception. However, the connection between information status and prosodic prominence was much less straightforward in Italian production, and it was unclear whether Italian listeners were attending to prosodic prominence as a cue for information status. This work suggests a trade-off between word order flexibility and prosodic encoding of information structure.

A related factor influencing word order is “heaviness” (i.e., the length, usually in words, of a constituent, especially relative to another in the same utterance). In the already mentioned study, Arnold et al. (2000) also showed that heaviness is an

independent factor that determines constituent order in English, such that speakers prefer to produce short and given phrases before long and new ones (e.g., *I gave the Prime Minister a signed copy of my latest magnum opus* is preferred to *I gave a signed copy of my latest magnum opus to the Prime Minister*). In contrast, Yamashita and Chang (2001) showed that Japanese speakers prefer to order long phrases before short ones, suggesting that the typological differences between the languages lead to speakers weighting formal and conceptual cues in production differently (for a connectionist model, see Chang 2009).

These heaviness effects and their cross-linguistic variation have been explained by the pressure to minimize dependency lengths and similar processing principles, including Early Immediate Constituents (EIC) (Hawkins 1994), Minimize Domains (MiD) (Hawkins 2004), Dependency Locality Theory (DLT) (Gibson 1998) and Dependency Length Minimization (DLM) (Ferrer-i-Cancho 2004; Temperley and Gildea 2018).¹¹ These principles share a similar general prediction, which posits that words or phrases that bear syntactic and/or semantic dependency/relations with each other tend to occur closer to each other. For example, dependents of the verb move to a position close to the verb. Empirical support and motivation for these principles mainly come from language comprehension studies (Gibson 2000), to the effect that shorter dependencies are preferred in order to ease processing and efficient communication (Gibson et al. 2019; Hawkins 1994, 2004; Jaeger and Tily 2011). These principles explain the above-mentioned preference “light before heavy” in VO languages like English, and “heavy before light” in OV languages like Japanese.

As revealed by corpus-based investigations, the effect of these principles varies cross-linguistically. Previous work looking at syntactic dependencies has provided evidence that languages minimize the overall or average dependency length to a different extent (Futrell et al. 2015a, 2020; Gildea and Temperley 2010; Liu 2008; Liu 2020). The effects also depend on specific constructions. A study by Futrell et al. (2020) showed further that the average dependency length given fixed sentence length is longer in head-final contexts (see also Rajkumar et al. 2016). Other experiments have investigated DLM in syntactic constructions with flexible constituent orders (Gulordava et al. 2015; Liu 2019; Wasow and Arnold 2003; Yamashita 2002); with an examination of the double PP construction across 34 languages, Liu (2020) demonstrated a typological tendency for constituents of shorter length to appear closer to their syntactic heads. The results also indicated that the extent of DLM is weaker in preverbal (head-final) than postverbal (head-initial) domains, a contrast that is less

¹¹ While EIC, MiD and DLT were all built upon a phrase structure framework, DLM was formulated using the dependency grammar framework. Though EIC, MiD and DLT have not used the term *dependency length* directly, the notion can be easily derived.

constrained by traditionally defined language types (Hawkins 1994); this contrast also seems to hold when the effects of additional factors such as lexical frequency and contextual predictability are controlled for (Liu 2022). Note that the patterns in the preverbal context are in opposition to previous findings with transitive constructions in Japanese (Yamashita and Chang 2001) and Korean (Choi 2007).

Word order flexibility can be advantageous for language users by helping to maximize fluency, permitting users to articulate easily retrievable elements early, and leaving time to plan the more cognitively demanding elements. At the same time, language users are inclined to re-use more global phrasal and sentence plans. This is observed in recycling of highly accessible abstract schemas known as syntactic priming (Ferreira and Bock 2006). Rigid word order, which allows language users to recycle the same routinized articulation plan, can also facilitate language production (MacDonald 2013). The re-use of highly entrenched production plans in conceptually similar constituents can also account for analogy effects in word order (for instance, the order of genitive and noun matches that of adjective and noun, cf. Diessel 2019). This kind of recycling is also important for frequently co-occurring lexemes (e.g., Sasano and Okumura 2016). Since the sequential associations between linguistic units can be of different strength (Diessel 2019), we observe gradience in the individual speaker/signer's grammar knowledge and in their interaction with other users. Highly frequent sequences have strong sequential associations due to automatization and chunking (Bybee 2010; Diessel 2019). Frequency-driven grammaticalization is usually accompanied by a loss of syntagmatic freedom (Lehmann 1995 [1982]). We can speak therefore of a trade-off between the ease of production of individual words and their sequences. Different languages and varieties have different points of balance, leading to cross-linguistic gradient patterns. The differences between languages can be explained by particular features of different grammars: cross-linguistic studies reveal that grammatical features of individual languages (e.g., verb agreement or aspectual marking) constrain language production (e.g., Kimmelman 2012; Norcliffe et al. 2015; Nordlinger et al. 2022; Proske 2022; Sauppe et al. 2013), even at the earliest stages of sentence planning. Taking a gradient approach to how these factors interact is crucial for capturing both the distribution of cross-linguistic variation and the dynamics of within-language variation.

Although research on comprehension and production regularly uses different methods and tests mode-specific hypotheses, it is likely that each makes use of the same underlying representational vocabulary and processes (Momma and Phillips 2018). Thus, all things being equal, we should expect many of the constraints on word order gradience in production to determine expectations about word order variability in comprehension (see MacDonald 2013, and commentaries, for discussion of

one proposal along these lines), although *exactly* what those constraints are, how they are implemented on parsing and production architectures, is still an open question.

2.2 Language acquisition

The outcomes of learning vary at different points across the lifespan. By and large, children faithfully replicate the stable distributional properties of their input (Ambridge et al. 2015), thus maintaining the variability observed cross-linguistically in word order patterns.¹² That is, an English-acquiring child will acquire the variability exhibited in English, and a Latvian-acquiring child will acquire the variability exhibited in Latvian. Word order flexibility is in principle not problematic for acquisition since children acquire so-called free-word order languages such as Warlpiri without problem (e.g., see Bavin and Shopen 1989). All this means that the processes of language development allow for the learning of gradient patterns and support the maintenance of word order variability if it exists in the ambient language.

As for adult L2 learners, experimental evidence shows that they have a strong bias towards regularity during language diffusion (Smith and Wonnacott 2010). In these experiments, input languages exhibiting free variation become increasingly regular (Wonnacott and Newport 2005).¹³ Also, it seems that adults are good at learning abstract patterns that can be captured by a few simple rules, while children are better at memorizing strings without any underlying rules (Nowak and Baggio 2017). Based on this, we can hypothesize that a high number of L2 adult learners may result in lower variability in word order in the target language, although this needs to be further investigated. Taking a gradient approach to word order flexibility would allow for a more nuanced picture which could potentially incorporate more complex contexts of language contact (see Section 2.4).¹⁴

¹² Although in the case of variable or incomplete input, word order preferences may emerge spontaneously, such as in children learning homesign (Goldin-Meadow and Mylander 1983).

¹³ At the same time, adults tend to probability-match free variation in an input language under certain conditions more than children do (Hudson Kam and Newport 2005), but this effect is restricted by different factors. In particular, it is observed when adults reproduce already familiar input and when the free variation is between only two alternatives.

¹⁴ Cf. related work by Bentz and Winter (2013) which shows a statistical tendency for reduced case marking in languages with more L2 users.

2.3 Language change

While word order change may be caused by language contact (see Section 2.4, and also Bradshaw 1982; Doğruöz and Backus 2007; Heine 2008, etc.), including substrate influence, this section focuses on internal change, which is influenced by the cognitive factors discussed above.¹⁵ These changes are transferred and/or amplified via language acquisition, which leads to language- or lineage-specific pathways in word order change (cf. Dunn et al. 2011).

Hawkins (1983: 213) states that word order change takes place through “doubling”: a new order comes in, co-exists with the old one, and replaces the old one through increase in frequency of occurrence and grammaticalization. We can represent this scenario informally as follows: $WO_A > WO_A \& WO_B > WO_B$. This may not be true for all word order change, but this process is a basic benchmark. Hence, change in word order is inherently tied up with variation. For example, Bauer (2009) describes word order change in Latin as part of a general tendency or “drift” (Sapir 1921) in Indo-European, with change away from rigid left-branching (OV) order in Proto-Indo-European (Bauer 2009: Sect. 2.1), towards flexible right-branching (VO) order in Latin, to more rigid right-branching order in Romance.

Importantly, this change happens first at the level of specific constructions, not sweepingly across all constituents at once.¹⁶ The doubling and word order variability are a result of these local changes. Bauer (2009: Section 3) describes in detail the constructions that changed from left-branching to right-branching order in their consecutive order, with verb-complement order changing after right-branching relative clauses, prepositions, and noun-adjective order emerged. Variation arose through long-term development of moving away from left-branching structure, where postpositions were archaic and prepositions arose, both were in non-free variation (i.e., some lemmas were postpositional, others prepositional) for an extended period of time. In addition, stylistic and pragmatic word-order variation arose, such as fronting of the subject and the verb. Several of these pragmatic word orders have been grammaticalized in Romance languages, most importantly using the cleft construction for emphasis. All of these changes are interdependent and dependent on the previous state of the language, i.e., Bauer (2009: 306) emphasizes that variation is not arbitrary, but “in fact [is] rooted in the basic characteristics of the language and [is] connected to many other linguistic features”.

¹⁵ Substrate influence versus internal change have been suggested for change to VSO in Celtic; Pokorny (1964) argues for a substrate influence account, Watkins (1963) and Eska (1994) for a language-internal account.

¹⁶ Though cf. Kroch (1989) and Taylor and Pintzuk (2012) who argue that English OV to VO change happened in a categorical manner.

However, variation in word order does not always imply (ongoing) language change. Word order variability can be stable for centuries. Examples are adjective-noun and noun-adjective order in Romance and many other languages, such as in Greater Philippine (Austronesian), where four of the six languages in Dryer (2013a) show variable order and at least seven more languages have both orders (Santo Domingo Bikol in Fincke [2002: 93–94]; Mandaya in Estrera [2020: 26–27]; Mansaka in Svelmoe and Svelmoe [1974: 51]; Kalagan in Collins [1970: 66–67]; Cebuano in Tanangkingsing [2009: 134]; Romblomanon in Law [1997: 19–20]; Inonhan in van den Heuvel [1997: 39]). Other examples are verb-object order in main/subordinate clauses in Germanic and many other languages, such as Western Nilotic (for Dinka and Nuer see Salaberri [2017], other relevant Western Nilotic languages are Jumjum in Fadul et al. [2016: 39] and Reel in Cien et al. [2016: 88]), and co-existence of prepositions and postpositions in Gbe and Kwa languages (Aboh 2010; Ameka 2003; see in general Hagège 2010: 110–114). There are usually different diachronic sources for each part of the doublet (Hagège 2010; Hawkins 1983), but this does not necessarily imply that one word order will outcompete the other over the course of generations. While word order change implies variability, variability does not necessarily imply change.

On the other hand, language change can involve rigidification and loss of flexibility of frequently used constructions. Croft (2003: 257–258), citing Lehmann (1995 [1982]) and Heine and Reh (1984), calls the grammaticalization of word order rigidification, “the fixing of the position of an element which formerly was free” (see also Lehmann 1992). See Hawkins (2019) for the relation between rigidification and Sapir’s (1921) notion of drift, and Harris and Campbell (1995: Ch. 8) for a summary of word order change.

2.4 Language contact and vernaculars

While borrowing of orders wholesale from one language to another is one potential outcome, language contact can also have a non-obvious effect on word order variability. For example, Heine (2008) discusses change to flexibility as one potential outcome of language contact. Contact may lead to increased or decreased variability. This has been documented in a number of Australian Aboriginal languages, which have famously flexible word order, after they came under pressure from the relatively more rigid English or Kriol (e.g., Lee 1987; Meakins 2014; Meakins and O’Shannessy 2010; Richards 2001; Schmidt 1985a, 1985b). In this situation, the frequency of SVO orders may increase, and the degree of flexibility (entropy) may also decrease. This is the main conclusion of Namboodiripad et al. (2018), who found that English-dominant Korean-speakers showed a greater relative preference for the canonical

SOV order than Korean-dominant speakers, that is, they rated non-canonical orders relatively low as compared to the canonical SOV order. Thus, we see a significantly lower constituent order flexibility correlating with more English dominance, but not overt borrowing of the dominant English order into Korean. A greater preference for canonical order corresponding to increased language contact was also found within a community of Malayalam speakers by Namboodiripad (2017).

Depending on the social context, language contact within a community can often correspond to intergenerational variation. An example can be found in Pitjantjatjara, a Pama-Nyungan language of Central Australia, which has traditionally been described as having a dominant SOV order, with a great degree of flexibility. Bowe (1990) finds SOV order to be approximately twice as frequent as SVO order (although clauses with at least one argument omitted are much more frequent in her corpus). Langlois (2004), in her study of teenage girls' Pitjantjatjara, finds that this is reversed, and SVO order is 50% more frequent than SOV in her sample. An experimental study of word order in Pitjantjatjara (Nordlinger et al. 2020; Wilmoth et al. Forthcoming) also substantiated this general trend; the younger generations had a higher proportion of SVO order, presumably as a result of more intense language contact and more English-medium education. However, while the distribution of different orders appears to be changing, the overall degree of flexibility as measured by entropy (see Section 4.1.1) remained approximately stable among all generations of speakers. There was some effect of older female participants, many of whom had worked as teachers or translators, being more rigidly SOV. This may be a result of their prescriptive attitudes interacting with the experimental setting.

The presence or absence of pressure of prescriptive norms in a particular language variety is an important and relatively under-investigated factor in the dynamics of word order. Specifically, vernacular varieties are relatively free from the pressure of prescriptive norms and are more inclined to word order variation compared to “standard” varieties. An example is reported in the study by Naccarato et al. (2020) on genitive noun phrases in several spoken varieties of Russian. In Standard Russian, the neutral and most frequent word order in genitive noun phrases is Noun followed by Genitive modifier, while two types of vernacular Russian – dialects and contact-influenced varieties – often show the alternative order Genitive modifier followed by Noun. Interestingly, kinship semantics, which was the strongest factor affecting the choice of a specific word order, turns out to be the same for both dialectal and contact-influenced Russian (irrespective of the area and the indigenous languages spoken there). This situation resembles instances of so-called “vernacular universals”, that is, features that are common to different spoken vernaculars (Filppula et al. 2009; see also Röthlisberger and Szmrecsanyi 2020), which are sometimes interpreted as language “natural tendencies” in the sense of Chambers (2004). If this is the case, we could argue that

word order rigidity is imposed through standardization, and variation might be the result of the natural development of language systems that, due to geographical and historical reasons, are less tightly connected to a particular standard, or which have resisted or rejected standardization.

3 What we (would) miss out without gradient approaches

The previous section demonstrated that word order variability is supported by diverse cognitive and social factors. How all these factors interact with each other is not yet fully understood. The case studies presented above have also shown that investigating gradient patterns and their causes, is a fruitful endeavor. In this section, we strengthen this argument, presenting a series of research questions that could not be asked or answered without applying a gradient approach systematically.

3.1 Language processing

A focus on word order gradience is very much the bread and butter of psycho- and neurolinguistic studies of grammar, where substantial effort has been invested in explaining ordering constraints on production and the implication of word order variability (and its correlation with grammatical structure) for comprehension. As an illustration, consider corpus-based research on dependency length minimization (DLM), which was discussed in Section 2.1. Some studies suggest that the crosslinguistic variation in the effect of dependency length is due to the fact that certain languages have more word order freedom (Futrell et al. 2015a; Futrell et al. 2020; Gildea and Temperley 2010; Yamashita and Chang 2001). The argument goes that with more ordering variability, the ordering preferences might be less subject to DLM and possibly abide more by other constraints such as information structure (Christianson and Ferreira 2005). Nevertheless, recent findings from Liu (2021) showed that is not necessarily the case. Using syntactic constructions in which the head verb has one direct object noun phrase dependent and exactly one adpositional phrase dependent adjacent to each other on the same side (e.g., *Kobe praised his oldest daughter from the stands*), the results indicated that there is no consistent relationship between overall flexibility and DLM. On the other hand, when looking at specific ordering domains (e.g., preverbal vs. postverbal), on average there is a very weak correlation between DLM and word order variability at a constructional level in the preverbal

contexts (e.g., preverbal orders in Czech); while no correlation, either positive or negative, seems to exist between the two in postverbal constructions (e.g., postverbal orders in Hebrew).

Another illustration is the connection between prosody and syntax, which is an active area of investigation (Eckstein and Friedericil 2006; Franz et al. 2020; Kreiner and Eviatar 2014; Luchkina and Cole 2021; Nicodemus 2009; Vaissière and Michaud 2006). The work which has investigated the connection between prosody and word order from a phonetic perspective does take phonetic gradience into account and adding a gradient approach to word order further enriches this area of research. For example, Šimík and Wierzbica (2017) combine gradient constraints in prosody and syntax to model flexible word order across three Slavic languages. They combine this with a gradient measure of flexibility in the form of acceptability judgments and find complex interactions between prosody and information structure (see also Gupton 2021). Asking and answering these questions about language production and comprehension would be impossible without taking a gradient approach.

3.2 Language acquisition

Conceptualizing word order in terms of gradience naturally aligns with the large focus on language acquisition as a cue-based process, which is most notable in probabilistic functionalist approaches such as the Competition Model (Bates and MacWhinney 1982) or constraint-satisfaction (Seidenberg and MacDonald 1999) but is also a feature of formal approaches such as Optimality Theory (Prince and Smolensky 2004). The challenge for acquisition researchers is to determine cue-based hierarchies across languages and how they interact; in addition, they must determine whether children bring pre-existing biases to this problem, such as the oft-cited preference to produce agents before lower-ranking thematic roles (Jackendoff and Wittenberg 2014), and to interpret early appearing nouns as agents (e.g., Bornkessel-Schlesewsky and Schlesewsky 2009).

For instance, while both German and Turkish allow for flexibility in the order of the main constituents with thematic roles marked via case, the transparency of the Turkish case system relative to German means that Turkish-acquiring children find variability in word order less problematic than their German-acquiring counterparts, who tend to prefer to attend to word order over case until they are much older (Dittmar et al. 2008; Slobin and Bever 1982; Özge et al. 2019). In the case of German, it appears that children settle on word order as a “good enough” cue to interpretation while they acquire the nuances of the case system (for a similar effect in Tagalog, see Garcia et al. 2020), although they are also sensitive to deviations in word order marked by cues such as animacy (Kidd et al. 2007).

A focus on cue-weightings (or constraints) has the potential to both inform the creation of more dynamic models of acquisition (thereby linking them to theories of adult language processing and production), and also to force questions common to acquisition concerning representation. Namely, are the cues coded in the structure, as might be argued in construction/usage-based approaches (e.g., Ambridge and Lieven 2015), or do they guide, but are independent of the sequencing choices of the processing system? For instance, do children acquiring European languages like English or German learn non-canonical structures such as object relative clauses as constructions by encoding common properties such as [-animate Head Noun] as part of a prototype structure (e.g., Diessel 2009), or do these distributional properties provide cues to a more abstract structure building mechanism? Understanding what causes word order gradience in the target languages and how children identify these variables is key to answering these questions.

3.3 Language change

Word order has typically been measured in a categorical fashion in typology, making gradient phenomena somewhat marginal (see Section 3.5). We can observe the same in historical linguistics: Harris and Campbell (1995: 198) point out that word order reconstructions have ignored frequently attested alternative word orders. Barðdal et al. (2020: 6) hold that under a constructional view, this should not be allowed, while “variation in word order represents different constructions with different information-structural properties, as such constituting form-meaning pairings of their own, which are by definition the comparanda of the Comparative Method”. Modeling sentential word order change in terms of categorical bins such as SVO, VSO, etc. does not allow for the inclusion of doubling, the intermediary step of an alternative word order arising; the process of rigidification is glossed over completely (see Section 2.3). Excluding variation from reconstruction makes reconstructions incomplete at best, and wrong at worst. See Section 4.4 for a discussion of how to incorporate gradient measures in a phylogenetic analysis, and which caveats can arise when doing so.

3.4 Language contact

A gradient approach to word order appears to be useful in studies in language contact. Although syntactic changes usually start at an intense stage of contact, word order is one of the linguistic features that are more easily borrowed (cf. Thomason 2001: 70–71; Thomason and Kaufman 1988: 74–76). Investigations on contact-induced word order variation in the world’s languages include studies focusing on word order at the sentence level (cf. numerous examples listed in Heine 2008: 34–35 and

Thomason and Kaufman 1988: 55), and studies devoted to variation within the noun phrase, e.g., focusing on the order of the head noun and its genitive modifier. See, for instance, Leisiö (2000) on Finnish Russian, and some observations in Sussex (1993) on émigré Polish, as well as Friedman (2003) on Macedonian Turkish. Most of these studies take a more categorical approach, focusing on the addition or loss of particular orders.

However, word order variability has been (indirectly) implicated as a potential source of the facility of contact-induced change in this domain, via convergence (e.g., Aikhenvald 2003): More flexible languages are more likely to have more frequent alternative orders, and thus are also more likely to have convergent structures with their contact languages. As such, contact could boost the frequency or change the status of a previously less frequent or non-canonical but still attested order (e.g., Manhardt et al. 2023). So “borrowing” of an order in this case could simply be a change in status of a previously derived or less frequent order to becoming a canonical or more frequent order, as in the case of Pitjantjatjara discussed in Section 2.4. Measuring the degree of flexibility, and accounting for the relative status of all grammatical orders in a language, allows us to see these types of contact effects, which might not be otherwise detectable.

An illustration of the usefulness of gradient approaches is a study of English-dominant and Korean-dominant Korean speakers by Namboodiripad et al. (2018), which was mentioned in Section 2.4. They found that English-dominant Korean speakers rated canonical SOV order the highest, followed by OSV, the verb-medial orders the next highest, and verb-initial orders the lowest of all. This is the same pattern shown by age-matched Korean speakers who grew up and live in Korea. However, there was a significant quantitative difference based on the degree of contact with English, as discussed above: more precisely, the English-dominant Korean speakers showed a greater relative preference for the canonical SOV order than Korean-dominant speakers. In taking a gradient approach, we are able to see contact effects that might not otherwise be visible. That is, even when we do not see an outright change to the basic constituent order in a language, or even if we do not see the addition of a constituent order due to contact, we might see a decrease in the number of possible orders, or a decrease in the relative frequency or acceptability of some orders (see Section 4.2 for more details on this method).

Moving to another example, what at first glance looks like the result of L1 calquing is actually the by-product of a more complex interaction of different factors. In previous research it has been frequently pointed out that contact in fact reinforces some existing language-internal tendencies (cf. Poplack 2020; Poplack and Levey 2010; Thomason and Kaufman 1988: 58–59). A similar conclusion is drawn in a recent study by Naccarato et al. (2021) devoted to word order variation in noun phrases with a genitive modifier in the variety of Russian spoken in Daghestan. In this variety of L2

Russian the Genitive - Noun order in noun phrases is unexpectedly frequent, whereas in Standard Russian the Noun - Genitive order is the neutral and by far most frequent option (see also Section 2.4). A quantitative analysis of the Daghestanian data and the comparison with monolinguals' varieties of spoken Russian suggest that contact is not the only factor involved in word order variation in Daghestanian Russian. Rather, L1 influence in bilinguals' speech (here, Nakh-Daghestanian or Turkic languages) seems to interact with some language-internal tendencies in Russian that, to a certain extent, are observed for non-bilinguals' varieties too. Without a gradient approach, it would be impossible to test this interaction.

3.5 Linguistic typology

Due to theoretical and practical reasons (e.g., the over-emphasis on basic word order in syntax, and the level of granularity available in grammar descriptions), typologists have mostly coded word order using categorical variables. The famous Greenbergian “correlations” (Greenberg 1963), for instance, represent associations between categorical variables, e.g., VO/OV and pre- or postpositions. This practice is responsible for two types of data reduction in typology (cf. Wälchli 2009): first, languages with orders with low word order entropy are more likely to be described accurately and systematically than languages with high entropy; second, word order variables studied by typologists are biased towards those that represent bimodal distributions. This can have consequences for characterizing, documenting or revitalizing the language.

Using a gradient approach can help us deal with these limitations, exploring the points between the extremes. Moreover, it allows us to add more different constructions. For example, oblique nominal phrases and adverbials are highly variable positionally with regard to their heads, which is why they are less frequently considered in word order correlations than the “usual suspects”, such as the order of Verb and Object (Levshina 2019). A gradient approach also allows us to formulate more precise predictions for typological correlations and implications, minimizing the space of possible languages, and making the universals more complete (see Gerdes et al. 2021; Levshina 2022; Naranjo and Becker 2018). For example, instead of including only predominantly VO or OV languages, we can also make predictions for more flexible languages which may not fit neatly into one category or another.

Consider an illustration. It is well known that there is a negative correlation between word order rigidity and case marking (Sapir 1921; Sinnemäki 2008). This categorical information can be represented by continuous variables, as shown in Figure 3, which is based on data from 30 online news corpora. Word order variability

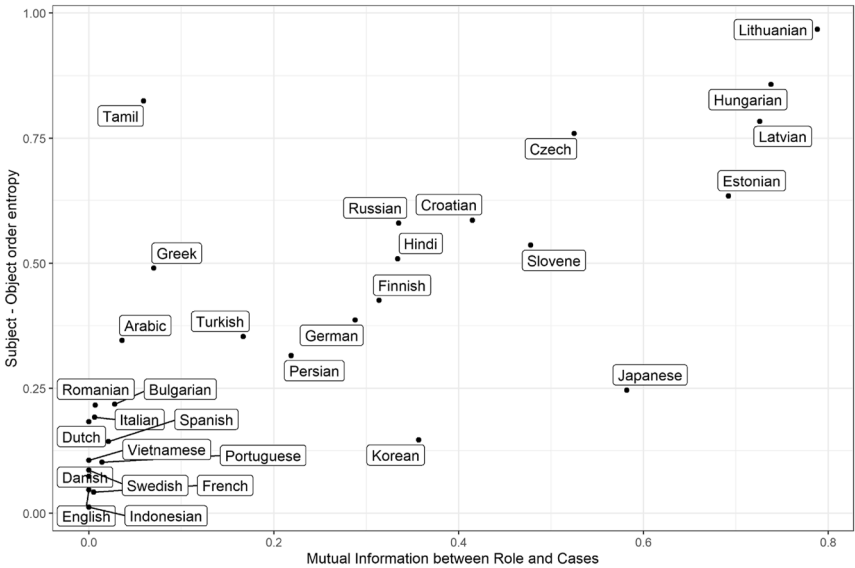


Figure 3: Mutual information between syntactic role (Subject or Object) and formally marked case plotted against Subject – Object order entropy based on online news corpora.

is measured as entropy of Subject and Object order (see Section 4.1.1), while case marking is represented as Mutual Information between case markers and the syntactic roles of Subject and Object. It shows how much information about the syntactic role of a noun we have if we know the case form, or, in other words, how systematically cases can actually help one guess if the noun is Subject or Object in the corpus (see Levshina 2021b for more details). The dataset is available in the OSF directory as *Dataset3.txt*. The plot shows clearly that the correlation is observed in the entire range of values of both variables. In particular, languages with differential marking tend to occupy the mid-range of word order variability, while languages with very little overlap between Subject and Object case forms like Lithuanian and Hungarian, allow for maximal freedom (Levshina 2021a). This supports the functional-adaptive view of language, showing that case and word order are not inherent built-in properties, but efficiently used communicative cues (cf. Koplenig et al. 2017). Word order gradience plays thus an important role in functional explanations of cross-linguistic generalizations (cf. Schmidtke-Bode et al. 2019).

3.6 Summary

As demonstrated in this section, gradient approaches have already allowed researchers to discover many important facts about word order across different linguistic domains – from language contact to prosody. Without a gradient perspective, these phenomena could not be investigated. Using a gradient approach also allows us to include more constructions and languages when testing theoretical hypotheses, avoiding data reduction. At the same time, the potential of gradient approaches has not been fully tapped. An important reason is the numerous methodological challenges and caveats. In the next section, we discuss them and propose some solutions.

4 How to investigate gradience? Challenges and some solutions

In this section, we delve into the *how* of implementing our gradient approach. As we have discussed, methodological limitations have been one barrier to more widespread adoption of gradient approaches to word order. Here, we discuss these challenges along with current solutions, covering corpora (Section 4.1), experimental methods (Section 4.2), fieldwork practices (Section 4.3), and phylogenetic comparative methods (Section 4.4). Along the way, we present some novel analyses of corpora and experimental results using these approaches.

4.1 Measuring gradience in corpora: methods and challenges

4.1.1 Gradient measures and sample size

Recent work on word order variability tends to use a gradient approach to provide a more accurate picture of word order patterns in language (Ferrer-i-Cancho 2017; Futrell et al. 2015b; Hammarström 2016; Levshina 2019; Östling 2015). Most of these studies are based on large-scale annotated corpora such as the Universal Dependencies (Zeman et al. 2020; also see Croft et al. [2017] for a discussion of using UD corpora to address questions in linguistic typology). Two common gradient measures are the proportion of a particular word order and the Shannon entropy of a particular word order. As an example, let us imagine that we are interested in the order of the object and the verb in a language. One straightforward way to measure the stability of such order in a language is to consider the proportions of OV and VO orders. Another measure frequently used is Shannon's entropy (Shannon 1948),

which represents the uncertainty of the data. The higher the entropy (which ranges between 0 and 1 in the case of two possible outcomes), the more variable the word order. For instance, a distribution of 500 sentences with OV and 500 sentences with VO would result in an entropy of 1.¹⁷ If the word order is always OV or always VO, the entropy will be 0. If OV is observed 95% of the time, the entropy will be equal to $-(0.05 \cdot \log_2(0.05)) + (0.95 \cdot \log_2(0.95))$, which is 0.2864.¹⁸ Generally speaking, the choice between entropy or simple proportions depends on what is more important for the researcher: reliability of word order as a cue for interpreting the speaker's message, or the preference for a specific word order. That said, entropy is particularly useful when considering word order variability in case of more than two possibilities, e.g., the order of S, V, and O.

While these measures are relatively established, the effect of corpus size is frequently mentioned as a potential source of error. As an example, "the corpora available vary hugely in their sample size, from 1,017 sentences of Irish to 82,451 sentences of Czech. An entropy difference between one language and another might be the result of sample size differences, rather than a real linguistic difference" (Futrell et al. 2015b: 93–94). The question of how large a corpus should be to provide robust measures of word order is very difficult because the answer depends on the frequency of the construction we investigate, which can also vary substantially across languages and genres. Here, we will focus on how many instances of a construction are necessary to measure entropy for the relative ordering of Subject and Verb (S&V) and Verb and Object (V&O). We aim to show what the appropriate sample size might be.

Let us consider the order of S&V and V&O within a sample of 30 languages extracted from the Universal Dependencies: Arabic, Basque, Bulgarian, Catalan, Chinese, Croatian, Danish, Dutch, English, Estonian, Finnish, French, German, Hebrew, Hindi, Indonesian, Italian, Japanese, Korean, Latvian, Norwegian, Persian, Polish, Portuguese, Romanian, Russian, Slovenian, Spanish, Swedish, and Ukrainian. This sample is obviously biased toward Indo-European languages, but it serves here solely for illustrating the methodology. To visualize the effect of corpus size, we randomly sample clauses containing a verb with a dependent nominal subject/object from the existing corpora, with an increasing range of 20 clauses. That is to say, we first sample 20 clauses, then 40, 60, and so on until we reach a size of 2,000 clauses. For each sample, we measure the proportion of the S&V and V&O orders. The data are

¹⁷ It is worth pointing out that different studies may use variants to calculate the entropy, and these variants may be affected in different ways by corpus size. For instance, Futrell et al. (2015b) have a more complex definition of word order entropy (as opposed to the coarse-grained entropy we use in our examples), which is conditional on several factors and may be more unstable in small corpora.

¹⁸ See www.shannonentropy.net/mark.pl for a live calculator of the entropy.

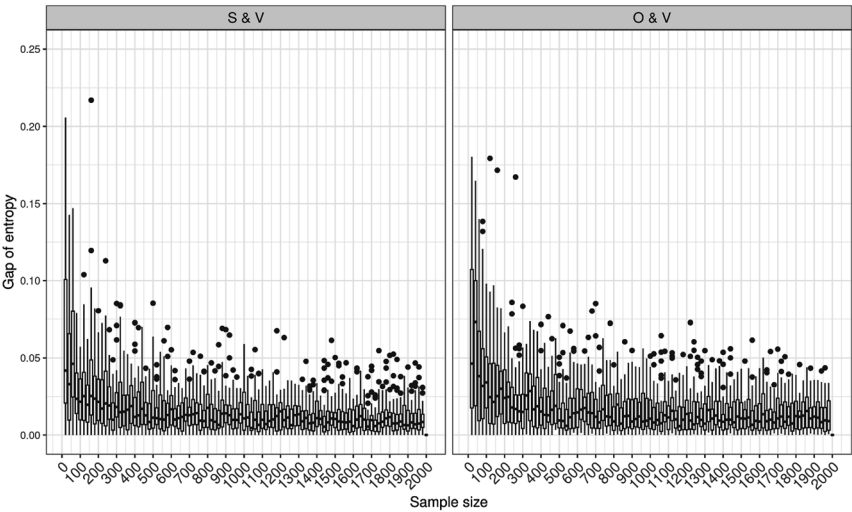


Figure 4: The stabilization of the word order entropy across different sample sizes. Each point indicates the gap of entropy between a given sample size and the sample size set as 2,000 instances.

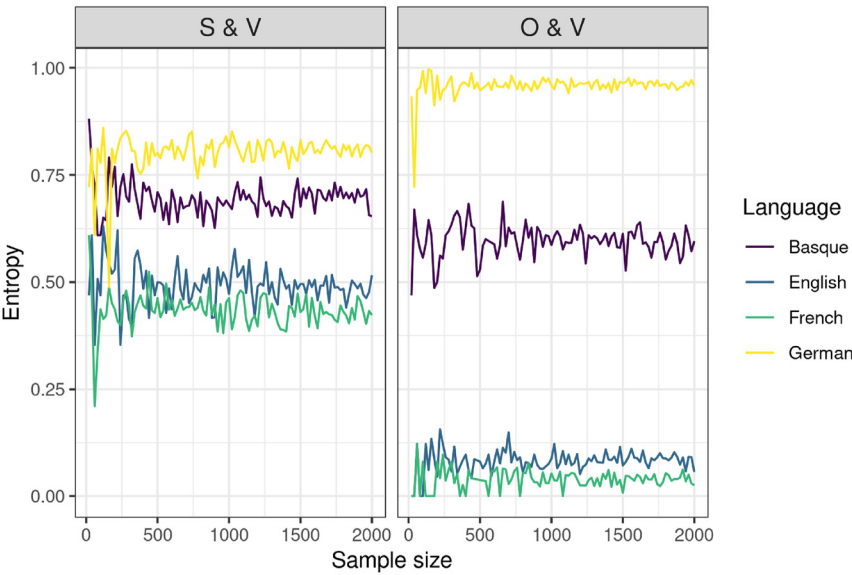


Figure 5: The variation of entropy across different sample sizes for Basque, English, French, and German.

available as *Dataset4.txt* in the OSF directory. Next, we calculate the entropy as described above and also the difference as compared to the entropy measured with the corpus size set as 2,000 clauses. This difference is called here “gap of entropy”. The results averaged across all languages are plotted in Figure 4.

The results indicate that extremely small sample sizes result in unstable results, as the entropy is likely to be biased by the sampled sentences, but sample sizes of more than 300 clauses produce entropy gaps less than 0.1. If we zoom in on a few individual languages, such as Basque, English, French, and German (see Figure 5) and plot the entropy measures, we observe that the entropy varies a lot within approximately the first 500 clauses but stabilizes after that threshold. This means that using a sample size of 500 or even larger is likely to give similar results in terms of S&V and O&V order.

Additional computational measures can be used to quantify where exactly the entropy converges for each language (see Janssen 2018: 79 and Josserand et al. 2021: 10 for an example of how to calculate where a curve stabilizes). Nevertheless, this example provides a quick-and-dirty demonstration of how sample size could be considered in studies related to word order. We recommend that such analyses should be performed on different word order categories and language families to investigate to what extent the entropy measure depends on the sample size cross-linguistically and/or across different constructions.

4.1.2 Potential biases in corpora: annotation methods

Parallel and comparable corpora of reasonable size usually rely on automatic annotation, thus introducing a potential bias depending on the quality of annotation. The quality is strongly influenced by the size and text types of the treebanks available for training the software performing the automatic annotation (the “parser”).

If we take a look at the current version (at the time of writing: 2.7) of the Universal Dependency Treebanks, we see that just 21 languages out of 111 have more than 3 treebanks. These languages are mostly national languages of Asian and European countries, and among them, only Arabic, Czech, French, German, Japanese, Russian and Spanish have treebanks with a total number of more than 1M tokens. In order to quantify the annotation quality bias both within and across languages, we run a comparison between the automatic annotation of the parallel corpus of Indo-European Prose and More (CIEP+: Talamo and Verkerk 2022) and the corresponding UD treebanks used for training the parser.¹⁹

Our comparison focuses on the Shannon entropy of the relative order of nominal heads and modifiers in eleven languages from five different genera of the Indo-

¹⁹ UD v.2.5, except for Welsh: v.2.6.

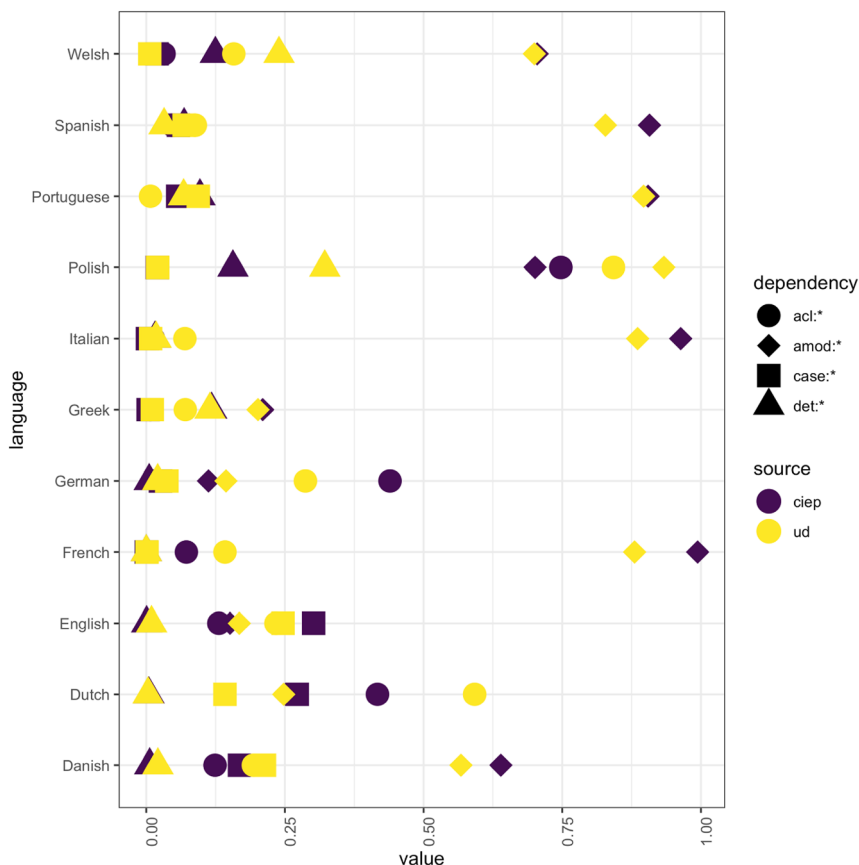


Figure 6: A comparison between the entropy of four nominal dependencies in eleven languages from two data sources: an automatically-parsed corpus (CIEP+) and the UD treebanks that have been employed for training the parser.

European family. Furthermore, we limit our analyses to the following syntactic relations/UD relations: adpositions (*case*), determiners (*det*),²⁰ adjectival modification (*amod*) and relative clauses (*acl:relcl*).²¹ Results are plotted in Figure 6. The dataset is available as *Dataset5.txt* in the OSF directory, which also contains the code for the

²⁰ Italian, Polish and Portuguese have specific UD relations for possessive pronouns and/or quantifiers; for data consistency, we have added the entropy of these relations to the entropy of the *det* relation.

²¹ The UD German parser (model: GSD) has a very limited support of the *acl:relcl*; accordingly, we do not have data for CIEP+ relative clauses in German.

statistical analysis. Assuming that UD treebanks are manually annotated and/or corrected data sources, thus serving as gold standards, the different – and, in many cases, higher – rates of entropy in the CIEP+ corpus are likely due to wrong annotations, or “noise”. Yet, despite some outliers such as *amod* and *det* in Polish and *case* in Dutch, the amount of noise in automatic annotations is overall low. A regression model with absolute differences between entropies estimated based on CIEP+ and UD, and languages and dependencies as random intercepts shows that the expected absolute difference (represented by the intercept) is 0.05. It is statistically significant ($t = 3.429$).²² Although the acceptability of this amount of noise depends on the level of precision required for a specific research question, this value can be seen as relatively low.

Note that source languages may influence target languages in translation (Johansson and Hofland 1994; Levshina 2017), such that the distributions of words or grammatical constructions may differ in translated versus spontaneously-generated texts. However, our results suggest that this is not the case in most languages and dependencies. Needless to say, both sources of bias (i.e., annotation and translationese) require further investigation, especially for low-resource languages, but we hope that our case study gives us some reasons for cautious optimism with regard to the practical feasibility of the gradient approach.

From a more theoretical perspective, we should add that the UD do not reflect categorial universalism in the sense that they do not presuppose any innate set of categories (cf. Haspelmath 2010). The UD were developed based on several desiderata, which are often in conflict: they should be suitable for both language description and language comparison, and for both rapid, consistent annotation by a human annotator and high-quality computational annotation. They should also be accessible to non-linguists, which explains why traditional Eurocentric terms are used.²³ According to Croft et al. (2017), most of these desiderata match the goals of a typologist. Importantly, the UD represent an inventory of universal constructions that serve some communicative functions. For example, the dependency ‘case’ used for adpositions reflects the fact that they are analogous to morphological marking. This allows (together with the annotation of morphological case in a separate annotation layer of the UD) to extract all cases of the universal construction whose function is to relate an argument dependent to its head (Croft et al. 2017: 65–66).

If annotators believe that it is important to reflect some important idiosyncrasies in a particular language, they can add extensions of the annotation tag to make it more fine-grained. For example, for some Slavic languages, the UD include ‘det:nummod’,

²² The model structure was $\text{abs_diff} \sim (1|\text{Language}) + (1|\text{Dependency})$. The standard deviations of the random intercepts were as follows: 0.022 (Language) and 0.026 (Dependency).

²³ See the full list of criteria at <http://www.universaldependencies.org/introduction> (last access 02.03.2022).

which marks a special type of quantifiers that do not agree with the quantified noun in case and require the counted noun to be in its genitive form. Another option is to introduce a special tag, e.g., classifiers for Chinese. Importantly, these differences should be well documented. A linguist planning to use the UD corpora and tools should first check how the categories of interest are encoded in the corpora. But such a check is necessary, in principle, for all kinds of cross-linguistic data.

We would like to acknowledge that in some cases, the UD framework is not yet directly applicable. For instance, the UD approach is based on annotation of words, which will not be meaningful for analysis of polysynthetic languages. At the same time, the UD annotation guidelines allow flexible modifications and extensions. For example, we could treat each individual morpheme within a polysynthetic word as the alternative of ‘one word’ in a sentence in English, then design additional dependency relations to show their grammatical properties in relation to the root of the word. For examples of work in that direction, see Spence et al. (2018) on Hupa, a Dene language of Northwestern California, and Park et al. (2021) on St. Lawrence Island Yupik, an endangered language spoken in the Bering Strait region.

4.1.3 The influence of register, text type, and modality

In addition to the problems of corpus size, annotation quality and translationese, another potential source of bias is different linguistic registers and text types, which can have different word order distributions (Batova 2020; Brunato and Dell’Orletta 2017; Panhuis 1981). The extent of this variability is an open question: some evidence suggests, in particular, that register and text type does not have a significant impact on overall patterns of dependency direction (Liu 2008, 2010). For many languages and constructions, however, we lack sufficient evidence. The most obvious problem is the bias towards written texts. In particular, the UD corpora are almost universally written, rather than spoken or signed. In fact, out of the 111 varieties represented in UD 2.7, only 22 include spoken or signed data, and this data is not always clearly separated from written data within a given corpus (Zeman et al. 2020). For many languages, available corpora are restricted to news and web-crawled texts (e.g., the Leipzig Corpora Collection, Goldhahn et al. 2012). Collection of texts containing the same content in different languages, or parallel corpora (Cysouw and Wälchli 2007), may help to mitigate this problem, providing new text types (e.g., fiction or film subtitles).

The importance of modality for estimation of word order is obvious from the following case study. We took spontaneous conversations, fiction and online news in Russian (from the 1980s to our days), focusing on the order of finite verbs and nominal objects in declarative sentences. Each text type was represented by 100 observations. The data are available as *Dataset6.txt* in the OSF directory. There was

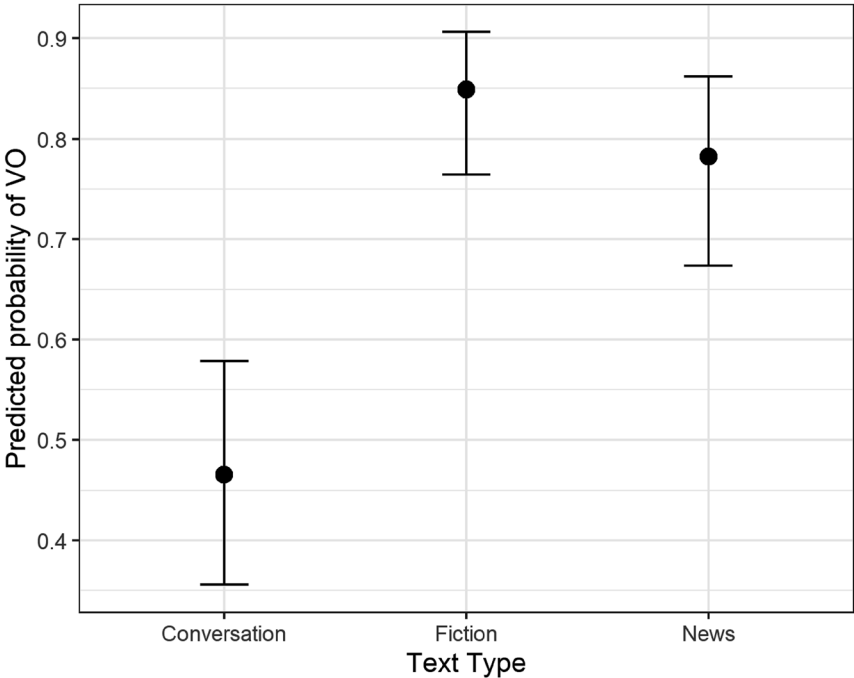


Figure 7: Predicted probabilities of VO order in different text types in Russian.

an obvious difference between the conversations and the written texts: the conversations contained 61 examples of OV, and only 39 examples of VO, whereas both the fiction and news contained 17 examples of OV and 83 examples of VO each. In order to control for potential accessibility effects, associated with the length and animacy of the nominal objects, a multiple logistic regression was fitted, with the word order as the response variable (OV or VO), and text type, animacy, and log-transformed length of the object as the predictors. As was shown in Section 2.1, it is advantageous for language users to produce more accessible (short and animate) constituents first. This is why we could not exclude that the preferences for OV in the conversations could be explained by a preference for short and animate objects. The modeling results indicate, however, that modality has a significant effect on the use of word order. Both of the written text types had higher chances of VO in comparison with the conversations (Fiction: $b = 1.86$, $p < 0.001$; News: $b = 1.42$, $p < 0.001$). The predicted probabilities of VO are shown in Figure 7. In addition, we found that longer objects had higher chances to be placed after the verb, as we would expect ($b = 0.71$, $p = 0.004$). Animacy did not have a significant effect ($p = 0.443$), probably, due to the low frequency of animate objects in the data (only 13%).

In addition to differences in word order variability, modalities and text types can also differ with regard to other features, such as argument drop (e.g., Ueno and Polinsky 2009; Zdorenko 2010). These differences can have implications for other popular continuous measures related to word order, such as dependency lengths (see Section 2.1). A study by Kramer (2021) using dependency-parsed YouTube captions to generate naturalistic spoken corpora in seven languages (English, Russian, Japanese, Korean, French, Italian, and Turkish) found that while dependency length minimization in speech patterned similarly, at a broad level, to dependency length minimization in written UD corpora, the difference in minimization across speech and writing varied in a consistent way. In particular, “flexible”, argument-dropping languages that were more head-final (Japanese, Korean, Turkish) showed more minimization in speech than in writing, whereas languages with similar properties that were less head-final (Russian, Italian, French) showed less minimization. Finally, English did not differ significantly across modalities (Kramer 2021).

Based on these case studies, we can make some preliminary conclusions. First, we do not need very large samples to obtain relatively stable estimates of word order entropy. Second, we have not found strong indications against using automatically annotated corpora and translated texts. Third, we need to be careful about using comparable genres and modalities (cf. Levshina 2022). Although more research is needed to test these preliminary conclusions and formulate best practices, it seems that computing gradient measures of word order from corpora is a realistic task if sufficient care is taken (see also Schnell and Schiborr [2022] for a thorough review of practical and conceptual concerns associated with using corpora for studying typology).

Note that the same limitations apply in principle to the traditional sources of typological information, such as reference grammars. The ‘dominant word order’ of a given language could be determined after seeing a couple of sentences from a specific modality, a specific topic and/or a small hand-collected corpus. One of the advantages of using corpus data is that we are more aware of these limitations.

4.2 Capturing gradience in experiments: methods and challenges

Psycholinguistic studies of syntactic comprehension typically manipulate properties of word order in order to test models of structure building, measuring those processes by indexing processing difficulty via reaction times, reading and looking times, pupil dilations, neurophysiological data (e.g., event-related potentials), or offline accuracy. These measures are well suited for capturing gradient effects. Production studies reveal those word orders that are preferred given those properties of

the message that need to be sequenced into an utterance (for examples from a language with several word order options see Christianson and Ferreira 2005; Norcliffe et al. 2015), which allows us to investigate the weights of relevant factors (see Section 2.1).

This section focuses on acceptability judgment experiments, which can be used to capture gradient patterns of variability. While the measure itself is reflective of potentially many factors (familiarity, conventionality, language ideology, processing effort), there is a robust literature on analyzing and interpreting acceptability ratings, as well as comparing these ratings to other psycholinguistic measures (see, among others, Dillon and Wagers 2021; Goodall 2021; Langsford et al. 2019; and, notably, Francis 2021 on how theoretical assumptions affect the interpretation of gradient results in acceptability judgment research). Because this method potentially yields gradient results (given that a gradient response scale is used, such as a 1–7 Likert scale; see Langsford et al. 2019 for discussion of various scales), it is suited to approaches which take gradience as a default assumption. Here, we discuss one particular way that acceptability judgment experiments have been used as a measure of constituent order flexibility (as seen in Namboodiripad 2017 and Namboodiripad et al. 2018; see also another example in Section 3.4).

In this method, participants hear all six logical orderings of subject, object, and verb, pseudo-randomly distributed amongst many other sentences of varying structure. Audio stimuli are used, in order to ensure that the proper intonation associated with each order is held constant, and to ensure that participants do not posit their own silent prosody which might affect how the sentences are parsed or interpreted (see Sedarous and Namboodiripad 2020). While the particular experiments referenced here do not include discourse context, instead holding “neutral context” constant across items, participants can be asked to rate sentences given particular discourse contexts.

For languages with canonical orders, flexibility is operationalized as the difference in acceptability between canonical and non-canonical orders; a bigger difference in acceptability between canonical and non-canonical order is interpreted as an increased preference for canonical order, and therefore decreased flexibility. As such, this measure both captures intuitions about constituent order flexibility (constituent orders feel more interchangeable or similar in flexible languages), and it is analogous to the entropy-based measures discussed elsewhere in this article (e.g., Levshina 2019). These “flexibility scores” can be calculated for individuals, or, when averaged across individuals, for particular varieties or languages. This allows for cross-variety or cross-linguistic comparison, as well as individual differences analyses (e.g., Dąbrowska 2013).

Figure 8 shows flexibility scores for several languages using this method. The data are available as *Dataset7.txt* in the OSF directory. Of these, Avar, Malayalam,

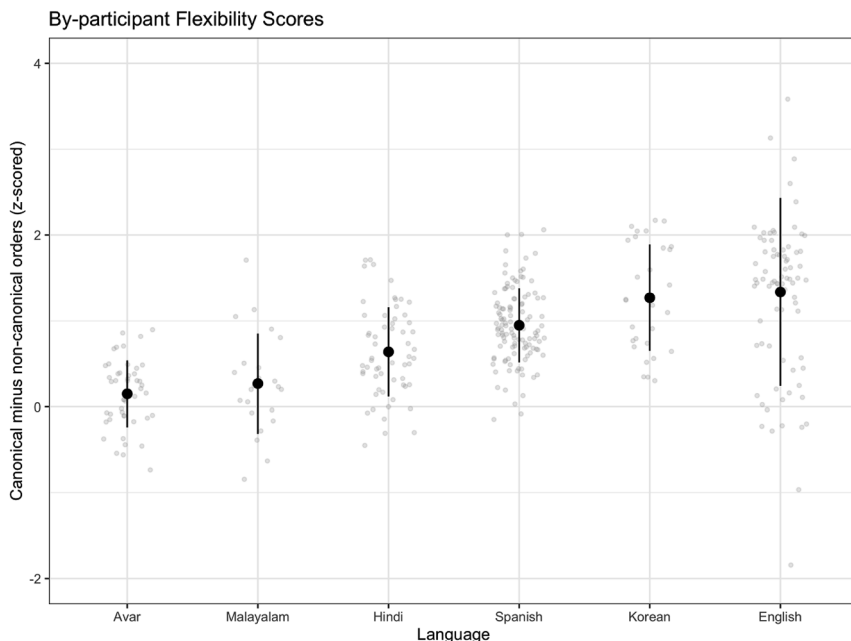


Figure 8: Flexibility scores for varieties of Avar, Malayalam, Hindi, Spanish, Korean, and English. Greater scores represent lower flexibility. Each light gray dot represents the difference between the mean rating for canonical order in each language minus the mean ratings for non-canonical orders for each participant; 1–7 ratings have been transformed into z-scores to account for individual differences in how the scale is used and to increase cross-experiment comparability. The languages are in order of mean flexibility, which is the mean flexibility score across participants, represented by the black dots. The bars represent two standard deviations from the mean.

Hindi, and Korean would all be labeled as “SOV flexible” languages, were we to take a categorical approach. However, taking a gradient approach allows us to see meaningful heterogeneity within that group, as well as similarity across languages which might be categorized differently – we see continuity between the “SOV flexible” languages and English, for example. Of course, more languages are required to make stronger cross-linguistic claims, but this nonetheless shows that asking gradient questions can yield patterns which we might not otherwise have seen.

These experiments are quite portable, robust to environmental noise (especially compared to other psycholinguistic measures), and relatively straightforward to explain. Therefore, this method is suited to a wide range of contexts (various field-work contexts, online, and in the lab), which is ideal for cross-linguistic comparison. However, careful experimental design is required; we refer the reader to other sources for more information on experimental design (e.g., Gibson et al. 2011; Sprouse and Almeida 2017), but considerations include counterbalancing, creation of

lexicalization sets, and controlling items for plausibility, familiarity, cultural appropriateness, length, and other relevant factors for the language. Non-canonical orders often have special intonational contours; sentences should be recorded to have as natural-as-possible intonation (see Sedarous and Namboodiripad 2020 for tips). As such, a crucial aspect of conducting these experiments is familiarity with the language; either working on languages one uses, in close collaboration with the language community, or both, is key.

Acceptability judgment experiments on their own provide partial information; they can measure individuals' relative preference for various orders in a highly controlled context. This method shares a common problem with more traditional psycholinguistic methods: it tests hypotheses about language in a restricted and somewhat artificial manner (Liu et al. 2017). While experiments are indispensable because they allow causal inference, the criticism is not without merit. It will be important to both test more naturalistic materials in ecologically valid contexts (see Kandylaki and Bornkessel-Schlesewsky 2019) and move beyond primarily studying psycholinguistic processes via human-computer interaction to study conversational interaction, since it is these contexts for which both language evolved and is acquired (Levinson 2016). It is also fruitful to combine corpus-based and experimental approaches, as it was done, for example, in Arnold et al. (2000).

4.3 Gradience and fieldwork practices

A gradient approach to word order faces some natural restrictions when we deal with field data, especially with hardly accessible data of minor, under-investigated and endangered languages. Word order variation is hard to investigate by means of the most simplistic elicitation tasks alone (translating sentences, grammaticality judgments), which are often used by field linguists. At the same time, experiments with a complex design involving a large number of speakers are not always possible in the field (though see Christianson and Ferreira 2005; Norcliffe et al. 2015; Nordlinger et al. 2022). In this case, textual collections remain the main source of data. They are usually much more restricted than for better studied languages. However, unlike many corpora of major world languages, small field corpora consist mostly of oral spontaneous texts, which represent a better source of data for studies on word order variation with respect to normalized written ones. Data from spontaneous texts such as these can then be used to inform targeted elicitation to test the grammaticality of particular structures and combinations.

The current state of affairs can be illustrated by the data of Turkic, Tungusic and Uralic indigenous languages of Russia. Outside the gradient approach, they are usually described as OV-languages. However, many of them feature VO word order

as a more marginal strategy or an ongoing change from OV to VO, which needs a separate explanation. VO word order might be motivated both by language-internal factors such as information structure (see Section 2.1) and by contact with Russian, which has predominantly verb-medial order (approximately 90% in online news corpora, according to Levshina 2021b). This puzzle is being discussed actively for different languages and based on different types of field data. The few experimental studies available concern major languages of the area: for instance, Grenoble et al. (2019) investigated word order variation in Sakha (Turkic) based on data from several picture description experiments. Most recent studies focus on textual data, especially data from small field corpora, cf., e.g., Rudnitskaya (2018) on Evenki (Tungusic), Däbritz (2020) on Enets, Nganasan (Uralic, Samoyedic), and Dolgan (Turkic). A hybrid technique is the use of semi-spontaneous texts, such as Pear-stories or Frog-stories, which provide more comparable data than spontaneous ones. Stapert (2013: 241–265) shows for spontaneous narratives and Pear-stories in Dolgan and Sakha (Turkic) that the frequency distribution of competing word orders is pretty much the same in both types of texts. Finally, some studies combine the analysis of textual data with data from questionnaires, e.g., Asztalos (2018) on word order variation in Udmurt (Uralic, Permic). Such a hybrid, pluralistic approach represents a viable strategy for research on word order variability in lesser-studied languages and non-standard varieties.

4.4 Gradience and phylogenetic comparative methods

Using a gradient measure of word order allows us to model cross-linguistic variation and word order change more accurately because of the continuous nature of the resulting variables. From a statistical perspective, this is not problematic at all; rather than using the chi-squared test, logistic regression, or discrete phylogenetic comparative methods, we can use (generalized) linear mixed-effect models and continuous phylogenetic comparative methods.

The following presents a case study on the phylogenetic signal of entropy of 24 different word orders taken from Levshina (2019) in Indo-European. Levshina (2019) calculated entropy using Universal Dependency treebanks (Zeman et al. 2020). The dataset is available as *Dataset8.txt* in the associated OSF directory. The code can be found there, as well. Phylogenetic signal is the notion that related languages are more similar to each other than languages that are less closely related. Here we use the measure λ (see Pagel 1999), which ranges from 0 (no evidence for historical signal) to 1 or slightly above (tree-like evidence for historical signal, the maximum value of λ depends on the phylogenetic tree). Figure 9 shows a high phylogenetic signal for some word order entropies (top) but not for all (bottom), showing that

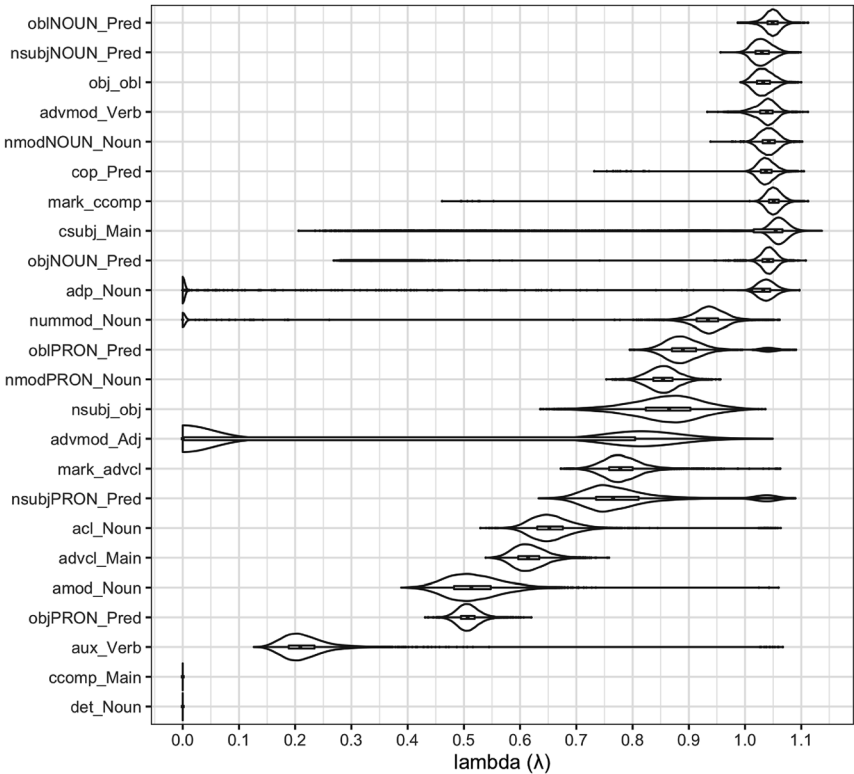


Figure 9: Phylogenetic signal (lambda, λ) on the entropy of 24 word orders (see labels on y axis) from Levshina (2019) in 24–36 Indo-European languages. Analyses conducted with trees from Dunn et al. (2011), using *geiger* in R (Harmon et al. 2008; R Core Team 2020).

variation or non-variation in word order can be related to genealogical inheritance of patterns for many but not all word orders.

At the top, we find *oblNOUN_Pred*, the variable dealing with the position of oblique noun phrases with regard to the predicate, as in *Jane looks_{PRED} [at the stars]_{OBL}* (Levshina 2019: 538), with very high phylogenetic signal. The reason for this is that 11/14 Balto-Slavic languages have an entropy above 0.96 (high variation) for oblique noun phrase–predicate word order; in addition, three out of four Indo-Aryan languages have an entropy less than 0.15 (low variation). Hence, attested variation in the order of oblique noun phrase and predicate is for a large part determined by genealogy: closely related languages have similar entropies, hence resulting in the high phylogenetic signal.

Most word orders tested by Levshina's (2019) have a high phylogenetic signal in Indo-European, but for others the effect is less strong. We can consider the difference between auxiliary-verb (around 0.2) and copula-verb order (around 1.0), and observe that for auxiliary-verb order, about half of the entropies are low, less than 0.14, and while Romance languages tend to have low entropies and Slavic languages high entropies, the pattern is not clear-cut at all. For copula-verb order, the same pattern exists but is much stronger: six out of eight Romance languages have entropies between 0.27 and 0.39; all Balto-Slavic languages (with the exception of Old Church Slavonic) have entropies between 0.47 and 0.84; all Indo-Aryan languages have entropies below 0.27, etc. Lastly, adjective-noun order, which has a bimodal distribution for phylogenetic signal, has high entropies for Romance, and low for Balto-Slavic and Germanic, the pattern is not very pronounced but is picked up in part of the phylogenetic tree sample (one part of the tree sample returns lambda approximately zero; another part returns lambda around 0.8). Phylogenetic signal measures of word order entropy will vary between families and between the word orders that are being studied.

This case study demonstrates that investigating gradience from an explicit historical perspective is not a moot point.²⁴ We can detect some meaningful patterns in the data. All noun-predicate dependencies (order of nominal subject, object, and oblique with the predicate) have higher phylogenetic signal than their pronominal counterparts, which may be explained by larger cross-linguistic differences between pronoun placement as opposed to noun placement; or to differences in word order in transitive and intransitive clauses that may exist even in closely related languages. Clausal word orders (*ccomp_Main*, *advcl_Main*, *acl_Noun*) tend to have a lower phylogenetic signal. Even closely related languages may have quite different entropies for clausal word order. A relevant factor for *ccomp_Main* could also be noisy annotation of direct speech across the language corpora, however. The lack of signal for the order of determiner and noun (*det_noun*) can be explained by the fact that determiners are a problematic category, which is lacking in some grammars (e.g., in Czech and Polish, see Talamo and Verkerk 2022). This is why there are many inconsistencies in the UD annotation. For example, possessive pronouns are annotated as determiner dependencies in some languages and as nominal modifiers in others.²⁵ Another issue of concern is the different frequencies of individual constructions in

²⁴ We already know, of course, that categorical analysis of word order carries phylogenetic signal. The stability of word order in four families investigated by Dunn et al. (2011) is clear; so is the analysis by Hammarström (2015), who states that the major "cause" of any language A to have any sentential word order is that the immediate ancestor of language A had that order.

²⁵ See <https://universaldependencies.org/u/dep/det.html>.

individual corpora. A safer approach could be using parallel corpora with more similarity between the texts in different languages.

Despite these caveats, the current results clearly show that closely related languages tend to have similar amounts of word order variation for at least half of the word orders investigated here. This case study demonstrates for the first time that continuous word order data, which reflect gradience, can be used for capturing phylogenetic signal.

4.5 Summary

In this section we discussed some methods and data for applying a gradient approach to word order in the lab and in the field. We have addressed some pressing issues, such as ecological validity, sample size, stylistic variation, and reliability of annotation, and suggested some recommendations. With the development of new corpora and cross-linguistic annotation tools, such as the Universal Dependencies pipeline (see Levshina 2022, for an overview), we can hope that gradient approaches will become increasingly accessible to researchers.

5 Conclusions and perspectives

In this article, we advocated for a gradient approach to word order which (a) assumes word order variation within languages, and (b) assumes that all languages vary in degree, not kind, when it comes to word order flexibility. Section 2 summarized the main factors that are responsible for emergence and maintenance of gradience. In Section 3, we highlighted how gradient approaches have been and could be relevant for a variety of fields of inquiry within linguistics. Finally, Section 4 dealt with some practical methodological issues when implementing gradient approaches. While we advocate for starting from the assumption of within-language variability and cross-linguistic continuity, we want to emphasize that categorical approaches could be fruitful for some languages, empirical domains, or more coarse-grained research questions. However, by taking non-categorical assumptions as the null hypothesis, we allow for the possibility for categoricity to emerge – the reverse cannot be true. Gradient approaches follow naturally from the emergentist usage-based view of languages and allow us to obtain a deeper understanding of the underlying cognitive and social processes that determine language structure and use. At the same time, we have argued that many theoretical frameworks and empirical domains can benefit

from taking into account word order variability more systematically. One useful “side-effect” of gradient approaches is the greater transparency about the source of data (register, modality) and annotation decisions, which is crucial for cross-linguistic comparison.

At this point we can formulate some pertinent questions, challenges and perspectives for investigating word order gradience across linguistic disciplines:

- **Language processing:** Psycholinguistic research has been conducted on a very small set of languages – less than 1% of the world’s languages at last count (Anand et al. 2011; Blasi et al. 2022; Jaeger and Norcliffe 2009; with a similar problem occurring in language acquisition, Kidd and Garcia 2022). Perhaps more worrying is that this sample is biased towards spoken European languages, and as such the influence that typologically diverse languages have on word order variability within individual speakers, be that as learners or mature language users, constitutes a notable gap in the literature (although see Garrido Rodriguez et al. [2023], Norcliffe et al. [2015], Nordlinger et al. [2022]; Sauppe [2017] and Sauppe et al. [2013] for recent work which addresses this gap). Also, explaining how the human language processing system produces word order variability and negotiates variability in comprehension requires explicit models of how these variables (and others, see Section 2.1) are weighted and implemented. Along with developing more robust experimental paradigms, an important method that will help achieve this goal is computational modeling, since it requires the postulation of explicit assumptions about how gradience is instantiated representationally.
- **Language acquisition:** We know that gradience in word order is not in principle a problem for the language learner, but how children deal with gradience as they develop towards the adult target requires further attention. Concerted efforts to study this across a wide range of languages representing a continuum of word order variability, from relatively low entropy languages like English to highly flexible languages like Warlpiri, would be a welcome addition to the literature.
- **Language evolution and change:** The prevailing norm in this area seems to be to treat word order variability as an unstable and transient state. This seems to be motivated by how constituent order change processes are dealt with in historical linguistics²⁶ (Croft 2003; Kroch et al. 2000; see also Section 2.3), but this potentially presumes a less flexible end-state, which is not the case for the many highly flexible languages. Further, Newmeyer (2000) and Givón (1979) have posited SOV as the potential constituent order of protolanguage, and subsequent empirical work on gestural communication has supported a bias for SOV, at least

²⁶ As discussed further in Section 2.3, reconstructing flexible constituent order is an especial challenge for the comparative method (e.g., Harris and Campbell 1995).

in some cases (Goldin-Meadow et al. 2008; Hall et al. 2013). However, these approaches do not engage directly with variability or variation in order, although they sometimes compare the communicative functions of word order variants (Gibson et al. 2013; Schouwstra and de Swart 2014). In short, taking the assumptions of gradient approaches presented in this article could lead to stronger models of language evolution; in particular, the idea that rigidity is associated with stability should be taken critically.

- **Language contact and emergence:** Studies of emergent languages (pidgins and creoles) can benefit from a gradient approach, as well. Though the Atlas of Pidgin and Creole Language Structures (Michaelis et al. 2013) does an excellent job in highlighting non-canonical orders and flexibility in word order, it is still a common misconception that creoles have rigid word order.²⁷ This is often part of the host of misguided and problematic claims about “simpler” grammars in creoles (DeGraff 2005; Mufwene 2000). In discussions of different definitions of simplicity and complexity as they relate to creoles (e.g., Good 2012, discussing morphological complexity), the role of gradience in word order must also be taken into account.
- **Linguistic typology:** The main challenge and task for the future is finding more comparable cross-linguistic data. The main bottleneck remains coverage of less-studied languages, including vernacular varieties and signed languages (Coons 2022). However, there are many ongoing projects whose goal is to provide annotated corpora for diverse samples of languages, including less well-described ones.²⁸ Also, new tools for online experiments should help us gather more data from language users in different places of the world. We expect a gradient approach to be particularly fruitful and pertinent in the study of languages which have highly flexible constituent order, and multiple canonical orders (e.g., Sign Language of the Netherlands [Klomp 2021]). This is where new cross-linguistic correlations and implications are likely to be found.
- **Language policy:** We can expect an important role of language policy in increasing or decreasing gradience. For example, a prescriptivist grammar may characterize a language as having a rigid or dominant word order, trying to fit the language in the Procrustean bed of popular word order types, while the actual behavior of speakers can be different. More research on a wider variety of languages *as they are used* is necessary.

²⁷ This misconception can be found in various introductory linguistics texts, such as Fromkin et al. 2011's *Introduction to Language*, which states that “Word order becomes more rigid in stabilized pidgins and creoles” (Fromkin et al. 2011: 457).

²⁸ See <http://doreco.info/>, <https://www.spur.uzh.ch/en/departments/research/textgroup/MorphDiv.html> and <https://multicast.aspra.uni-bamberg.de/>.

The co-authors of this article are researchers working across disciplines and across the world, all of whom have found taking a gradient approach to be crucial for the particular questions they are asking about word order. By bringing together diverse examples of gradient approaches, we have demonstrated a growing consensus for this type of theoretical move, and we have shown that it is both possible and necessary to be enacted. We hope that the theoretical ideas and practical advice in this article will encourage the reader to embark on their own investigations using gradient approaches, increasing the number and types of languages which are included in such research.

Acknowledgements: The authors would like to thank the participants and audience at the 2020 SLE workshop “New Approaches to Word Order”, where this collaboration was first conceived. In addition, the authors are grateful to Calle Börstell, Josef Fruehwald, Ingo Feldhausen, the two anonymous reviewers, and our editor Volker Gast for their helpful feedback on previous versions of this paper. Natalia Levshina’s work on this project was funded by the Netherlands Organisation for Scientific Research (NWO) under Gravitation grant Language in Interaction (024.001.006). Marc Allasonnière-Tang is thankful for the support of the French National Research Agency (ANR-20-CE27-0021). The contribution by Chiara Naccarato, Anastasia Panova and Natalia Stoyanova was supported by the Ministry of Science and Higher Education of the Russian Federation within Agreement No 075-15-2020-793.

Data availability statement: The datasets and code used for the quantitative case studies can be found in the OSF directory https://osf.io/6m7ec/?view_only=1eb102ddd91d41dcacaa0f266e5eab79.

References

- Aboh, Enoch O. 2010. The Morphosyntax of the Noun Phrase. In Enoch O. Aboh & Essegbey James (eds.), *Topics in Kwa syntax* (Studies in Natural Language and Linguistic Theory 78), 11–37. Dordrecht: Springer.
- Adger, David & Jennifer Smith. 2005. Variation and the Minimalist Program. In Leonie Cornips & Karen Corrigan (eds.), *Syntax and variation: Reconciling the biological and the social*, 149–178. Amsterdam & Philadelphia: John Benjamins.
- Adger, David & Jennifer Smith. 2010. Variation in agreement: A lexical featurebased approach. *Lingua* 120. 1109–1134.
- Ameka, Felix K. 2003. Prepositions and postpositions in Ewe: Empirical and theoretical considerations. In Anne Zribi-Hertz & Patrick Sauzet (eds.), *Typologie des langues d’Afrique et universaux de la grammaire*, 43–66. Paris: L’Harmattan.
- Aikhenvald, Alexandra. 2003. Mechanisms of change in areal diffusion: new morphology and language contact. *Journal of Linguistics* 39(1). 1–29.

- Ambridge, Ben, Evan Kidd, Anna Theakston & Caroline Rowland. 2015. The ubiquity of frequency effects in first language acquisition. *Journal of Child Language* 42(2). 239–273.
- Ambridge, Ben & Elena V. M. Lieven. 2015. A constructivist account of child language acquisition. In Brian MacWhinney & William O'Grady (eds.), *The handbook of language emergence*, 478–510. Malden, MA: Wiley Blackwell.
- Anand, Pranav, Sandra Chung & Matthew Wagers. 2011. *Widening the net: Challenges for gathering linguistic data in the digital age*. NSF SBE 2020. Rebuilding the mosaic: Future research in the social, behavioral and economic sciences at the National Science Foundation in the next decade.
- Arnold, Jennifer, Thomas Wasow, Anthony Losongco & Ryan Ginstrom. 2000. Heaviness vs. Newness: The effects of structural complexity and discourse status on constituent ordering. *Language* 76. 28–55.
- Asztalos, Erika. 2018. *Szórendi típusváltás az udmurt nyelvben* [Word order change in Udmurt]. Budapest: Eötvös Loránd University dissertation.
- Austin, Peter & Joan Bresnan. 1996. Non-configurationality in Australian aboriginal languages. *Natural Language & Linguistic Theory* 14. 215–268.
- Barðdal, Jóhanna, Spike Gildea & Eugenio R. Luján. 2020. *Reconstructing syntax*. Leiden: Brill.
- Bates, Elizabeth & Brian MacWhinney. 1982. Functionalist approaches to grammar. In Eric Wanner & Lila Gleitman (eds.), *Language acquisition: The state of the art*, 173–218. New York, NY: Cambridge University Press.
- Batova, Natalia. 2020. *Russian word order: The pragmatics of unconventional verb-initial types*. Melbourne: The University of Melbourne dissertation.
- Bauer, Brigitte M. 2009. Word order. In Philip Baldi & Pierluigi Cuzzolin (eds.), *New perspectives on historical Latin syntax, vol. 1: Syntax of the sentence*, 241–316. Berlin & New York: Mouton de Gruyter.
- Bavin, Edith L. & Timothy Shopen. 1989. Warlpiri children's processing of transitive sentences. In Brian MacWhinney & Elizabeth Bates (eds.), *The crosslinguistic study of sentence processing*, 185–205. New York: Cambridge University Press.
- Behaghel, Otto. 1932. *Deutsche Syntax: eine geschichtliche Darstellung*. Heidelberg: Carl Winter.
- Bentz, Christian & Bodo Winter. 2013. Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change* 3. 1–27.
- Biberauer, Theresa. 2019. Children always go beyond the input: The Maximise Minimal Means perspective. *Theoretical Linguistics* 45(3–4). 211–224.
- Bickel, Balthasar, Johanna Nichols, Taras Zakharko, Alena Witzlack-Makarevich, Kristine Hildebrandt, Michael Rießler, Lennart Bierkandt, Fernando Zúñiga & John Lowe. 2022. The AUTOTYP database (v1.0.1). *Zenodo*. <https://doi.org/10.5281/zenodo.6255206>.
- Blasi, Damián E., Joseph Henrich, Evangelia Adamou, David Kemmerer & Asifa Majid. 2022. Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences* 26(2). 1153–1170.
- Bock, Kathryn J. & David E. Irwin. 1980. Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior* 19. 467–484.
- Bock, Kathryn J. & Richard K. Warren. 1985. Conceptual accessibility and syntactic structure in sentence formulation. *Cognition* 21(1). 47–67.
- Bod, Rens, Jennifer Hay & Stefanie Jannedy (eds.). 2003. *Probabilistic linguistics*. Cambridge, MA: MIT Press.
- Boersma, Paul & Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32. 45–86.
- Bornkessel-Schlesewsky, Ina & Matthias Schlewsky. 2009. The role of prominence information in the real-time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass* 3(1). 19–58.
- Bowe, Heather J. 1990. *Categories, constituents, and constituent order in Pitjantjatjara: An aboriginal language of Australia*. London, New York: Routledge.

- Bradshaw, Joel. 1982. *Word order change in Papua New Guinea Austronesian languages*. Honolulu: University of Hawaii dissertation.
- Branigan, Holly P., Martin J. Pickering & Mikihiro Tanaka. 2008. Contributions of animacy to grammatical function assignment and word order during production. *Lingua* 118(2). 172–189.
- Bresnan, Joan, Shipra Dingare & Christopher D. Manning. 2001. Soft constraints mirror hard constraints: Voice and person in English and Lummi. In Miriam Butt, & Tracy H. King (eds.), *Proceedings of the LFG '01 Conference*, University of Hong Kong. Stanford: CSLI Publications. Available at: <http://csli-publications.stanford.edu/LFG/6/lfg01.html>.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina & Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Bouma, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.
- Bresnan, Joan & Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language* 86(1). 168–213.
- Brunato, Dominique & Felice Dell'Orletta. 2017. On the order of words in Italian: A study on genre vs complexity. In Simonetta Montemagni & Joakim Nivre (eds.), *Proceedings of the Fourth International Conference on Dependency Linguistics* (Depling 2017), 25–31. Linköping University Electronic Press. Available at: <https://aclanthology.org/W17-6505>.
- Büring, Daniel. 2013. Syntax, information structure and prosody. In Marcel den Dikken (ed.), *The Cambridge handbook of generative syntax*, 860–895. Cambridge: Cambridge University Press.
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Carling, Gerd (ed.). 2017. *Diachronic atlas of comparative linguistics online*. Lund: Lund University. <https://diac.ht.lu.se/> (accessed 01 March 2022).
- Cathcart, Chundra, Gerd Carling, Filip Larsson, Niklas Johansson & Erich Round. 2018. Areal pressure in diachronic evolution: An Indo-European case study. *Diachronica* 35(1). 1–34.
- Chambers, Jack K. 2004. Dynamic typology and vernacular universals. In Bernd Kortmann (ed.), *Dialectology meets typology: Dialect grammar from a cross-linguistic perspective*, 127–145. Berlin & New York: Mouton de Gruyter.
- Chang, Franklin. 2009. Learning to order words: A connectionist model of heavy NP shift and accessibility effects in Japanese and English. *Journal of Memory and Language* 61. 374–397.
- Choi, Hye-Won. 2007. Length and order: A corpus study of Korean dative-accusative construction. *Discourse and Cognition* 14(3). 207–227.
- Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton & Co.
- Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use*. Westport, CT & London: Greenwood Publishing Group.
- Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A life in language*, 1–52. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos Peregrín Otero & Maria Luisa Zubizarreta (eds.), *Foundational issues in linguistic theory. Essays in honor of Jean-Roger Vergnaud*, 133–166. Cambridge, MA: MIT Press.
- Christianson, Kiel & Fernanda Ferreira. 2005. Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition* 98. 105–135.
- Cien, Mark Akec, Andrew Agok Igai Piiny, David Mabor Makuei Angong, John Malak Balang Kerjok, Daniel Majier Johou & David Kuacrengh Nhial. 2016. *Reel grammar book*. Juba: SIL South Sudan.
- Collins, Grace. 1970. *Two views of Kalagan grammar*. Bloomington, IN: Indiana University dissertation.
- Coons, Caitie. 2022. Nominal word order typology in signed languages. *Frontiers in Communication* 6. 802596.

- Croft, William. 2003. *Typology and universals*. Cambridge: Cambridge University Press.
- Croft, William, Dawn Nordquist, Katherine Looney & Michael Regan. 2017. Linguistic typology meets Universal Dependencies. In Markus Dickinson, Hajic Jan, Sandra Kübler & Adam Przepiórkowski (eds.), *Proceedings of the 15th International Workshop on Treebanks and Linguistic Theories (TLT15)*, 63–75. Aachen: CEUR Workshop Proceedings. Available at: <http://ceur-ws.org/Vol-1779/05croft.pdf>.
- Culicover, Paul & Ray Jackendoff. 2006. Turn over control to the semantics. *Syntax* 9(2). 131–152.
- Culicover, Paul & Michael Rochemont. 1983. Stress and focus in English. *Language* 59(1). 123–165.
- Cysouw, Michael & Bernhard Wälchli. 2007. Parallel texts: Using translational equivalents in linguistic typology. *STUF – Sprachtypologie und Universalienforschung* 60(2). 95–99.
- Däbritz, Chris Lasse. 2020. Focus position in SOV ~ SVO-varying languages – evidence from Enets, Nganasan and Dolgan. *Journal of Estonian and Finno-Ugric Linguistics* 11(2). 99–118.
- Dąbrowska, Ewa. 2013. Functional constraints, usage, and mental grammars: A study of speakers' intuitions about questions with long-distance dependencies. *Cognitive Linguistics* 24(4). 633–665.
- Dalrymple, Mary, John J. Lowe & Louise Mycock. 2019. *The Oxford reference guide to Lexical Functional Grammar*. USA: Oxford University Press.
- Dahlstrom, Amy. 1986. *Plains Cree morphosyntax*. Berkeley, CA: University of California, Berkeley dissertation.
- DeGraff, Michel. 2005. Linguists' most dangerous myth: The fallacy of Creole Exceptionalism. *Language in society* 34(4). 533–591.
- Diessel, Holger. 2009. On the role of frequency and similarity in the acquisition of subject and non-subject relative clauses. In Talmi Givón & Masayoshi Shibatani (eds.), *Syntactic complexity*, 251–276. Amsterdam & Philadelphia: John Benjamins.
- Diessel, Holger. 2019. *The grammar network: How linguistic structure is shaped by language use*. Cambridge: Cambridge University Press.
- Dillon, Brian & Matt Wagers. 2021. Approaching gradience in acceptability with the tools of signal detection theory. In Goodall Grant (ed.), *The Cambridge handbook of experimental syntax*. Cambridge: Cambridge University Press.
- Dittmar, Miriam, Kirsten Abbot-Smith, Elena Lieven & Michael Tomasello. 2008. German children's comprehension of word order and case marking in causative sentences. *Child Development* 79. 1152–1167.
- Doğruöz, A. Seza & Ad Backus. 2007. Postverbal elements in immigrant Turkish: Evidence of change? *International Journal of Bilingualism* 11(2). 185–220.
- Downing, Laura J., Al Mtenje & Bernd Pompino-Marschall. 2004. Prosody and information structure in Chichewa. *ZAS Papers in Linguistics* 37. 167–186.
- Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68. 81–138.
- Dryer, Matthew S. 2013a. Order of adjective and noun. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: <http://wals.info/chapter/87>.
- Dryer, Matthew S. 2013b. Order of subject, object and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: <http://wals.info/chapter/81>.
- Dryer, Matthew S. 2019. Grammaticalization accounts of word order correlations. In Karsten Schmidtke-Bode, Natalia Levshina, Susanne Maria Michaelis & Ilja A. Seržant (eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*, 63–95. Berlin: Language Science Press.
- Dryer, Matthew S. & Martin Haspelmath. 2013. *WALS Online (v2020.3) [Data set]*. Leipzig: Zenodo.
- Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson & Russell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473. 79–82.

- Eska, Joe F. 1994. On the crossroads of phonology and syntax: Remarks on the origins of Vendryes' restriction and related matters. *Studia Celtica* 28. 39–62.
- Estrera, Edward G. 2020. *A grammatical sketch of Mandaya*. Diliman: University of the Philippines Unpublished manuscript.
- Eckstein, Korinna & Angela D. Friederici. 2006. It's early: Event-related potential evidence for initial interaction of syntax and prosody in speech comprehension. *Journal of Cognitive Neuroscience* 18(10). 1696–1711.
- Fadul, Yousif Aljuzuli, Terafi Mohadin Chol, Muusa Nuer Teebu, Bal Douwash Yousif, Abrahama Kidir Blang & Abdalmajid Juma Anur. 2016. *Jumjum grammar book*. Juba: Jumjum Language Committee & Literacy Partners International & Sudan Workshop Programme.
- Fanselow, Gisbert. 2009. Free constituent order: A minimalist interface account. *Folia Linguistica* 37(1/2). 405–437.
- Feldhausen, Ingo & Maria del Mar Vanrell. 2015. Oraciones hendidas y marcación del foco estrecho en español: Una aproximación desde la teoría de la optimidad estocástica. *Revista Internacional de Lingüística Iberoamericana* 13(2). 39–59.
- Ferreira, Victor S. & Kathryn Bock. 2006. The functions of structural priming. *Language & Cognitive Processes* 21. 1011–1029.
- Ferreira, Victor S. & Hiromi Yoshita. 2003. Given-new ordering effects on the production of scrambled sentences in Japanese. *Journal of Psycholinguistic Research* 32. 669–692.
- Ferrer-i-Cancho, Ramon. 2004. Euclidean distance between syntactically linked words. *Physical Review E* 70(5). 056135.
- Ferrer-i-Cancho, Ramon. 2017. The placement of the head that maximizes predictability: An information theoretic approach. *Glottometrics* 39. 38–71.
- Filppula, Markku, Juhani Klemola & Heli Paulasto (eds.). 2009. *Vernacular universals and language contacts: Evidence from varieties of English and beyond*. London: Routledge.
- Fincke, Steven C. 2002. *Sto. Domingo Bikol verbal portmanteau morphemes: Transitivity in conversation*. Santa Barbara, CA: University of California at Santa Barbara dissertation.
- Francis, Elaine J. 2021. *Gradient acceptability and linguistic theory*. Oxford: Oxford University Press.
- Franz, Isabelle, Markus Bader, Frank Domahs & Gerrit Kentner. 2020. Influences of rhythm on word order in German. In *Proceedings of 10th International Conference on Speech Prosody 2020*, 385–388. Tokyo: ISCA.
- Frascarelli, Mara & Roland Hinterhölzl. 2007. Types of topics in German and Italian. In Susanne Winkler & Kerstin Schwabe (eds.), *On information structure, meaning and form*, 87–116. Amsterdam & Philadelphia: John Benjamins.
- Friedman, Victor A. 2003. *Turkish in Macedonia and beyond: Studies in contact, typology and other phenomena in the Balkans and the Caucasus*. Wiesbaden: Harrassowitz.
- Fromkin, Victoria, Robert Rodman & Nina Hyams. 2011. *An introduction to language*, 9th edn. Boston, MA: Wadsworth, Cengage Learning.
- Fujinami, Yoshiko. 1996. *An implementation of Japanese grammar based on HPSG*. Edinburgh: Department of Artificial Intelligence, University of Edinburgh MA thesis.
- Futrell, Richard, Kyle Mahowald & Edward Gibson. 2015a. Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences* 112(33). 10336–10341.
- Futrell, Richard, Kyle Mahowald & Edward Gibson. 2015b. Quantifying word order freedom in dependency corpora. In Joakim Nivre & Eva Hajičová (eds.), *Proceedings of the Third International Conference on Dependency Linguistics* (Depling 2015), 91–100. Uppsala: Uppsala University.
- Futrell, Richard, Roger P. Levy & Edward Gibson. 2020. Dependency locality as an explanatory principle for word order. *Language* 96(2). 371–412.

- Gabriel, Christoph. 2010. On focus, prosody, and word order in Argentinean Spanish: A minimalist OT account. *Revista Virtual de Estudos da Linguagem* 4. 183–222.
- Garcia, Rowena, Jens Roeser & Barbara Höhle. 2020. Children's online use of word order and morphosyntactic markers in Tagalog thematic role assignment: An eye-tracking study. *Journal of Child Language* 47(3). 533–555.
- Garrido Rodriguez, Gabriela, Elisabeth Norcliffe, Penelope Brown, Falk Huettig & Stephen C. Levinson. 2023. Anticipatory processing in a verb-initial Mayan language: Eye-tracking evidence during sentence comprehension in Tzeltal. *Cognitive Science* 47. E13292.
- Gerdes, Kim, Silvain Kahane & Xinying Chen. 2021. Typometrics: From implicational to quantitative universals in word order typology. *Glossa: A Journal of General Linguistics* 6(1). 17.
- Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68(1). 1–76.
- Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In Alec Marantz, Yasushi Miyashita & Wayne O'Neil (eds.), *Image, language, brain*, 95–126. Cambridge, MA: MIT Press.
- Gibson, Edward, Richard Futrell, Steven T. Piantadosi, Isabelle Dautriche, Kyle Mahowald, Leon Bergen & Roger Levy. 2019. How efficiency shapes human language. *Trends in Cognitive Sciences* 23(5). 389–407.
- Gibson, Edward, Steven T. Piantadosi, Kimberly Brink, Leon Bergen, Eunice Lim & Rebecca Saxe. 2013. A noisy-channel account of crosslinguistic word-order variation. *Psychological Science* 24(7). 1079–1088.
- Gibson, Edward, Steven T. Piantadosi & Kristina Fedorenko. 2011. Using Mechanical Turk to obtain and analyze English acceptability judgments. *Language and Linguistics Compass* 5(8). 509–524.
- Gildea, Daniel & David Temperley. 2010. Do grammars minimize dependency length? *Cognitive Science* 34(2). 286–310.
- Givón, Talmy. 1979. From discourse to syntax: Grammar as a processing strategy. In Talmy Givón (ed.), *Discourse and syntax*, 81–112. Leiden: Brill.
- Givón, Talmy. 1991. Isomorphism in the grammatical code: Cognitive and biological considerations. *Studies in Language* 15(1). 85–114.
- Goldberg, Adele. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Goldin-Meadow, Susan & Carolyn Mylander. 1983. Gestural communication in deaf children: Noneffect of parental input on language development. *Science* 221(4608). 372–374.
- Goldin-Meadow, Susan, Wing Chee So, Aslı Özyürek & Carolyn Mylander. 2008. The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences* 105(27). 9163–9168.
- Goldhahn, Dirk, Thomas Eckart & Uwe Quasthoff. 2012. Building large monolingual dictionaries at the Leipzig Corpora Collection: From 100 to 200 languages. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, Asuncion Moreno & Stelios Piperidis (eds.), *Proceedings of the 8th International Language Resources and Evaluation (LREC'12)*, 759–765. Istanbul: ELRA.
- Good, Jeff. 2012. Typologizing grammatical complexities: Or why creoles may be paradigmatically simple but syntagmatically average. *Journal of Pidgin and Creole Languages* 27(1). 731–747.
- Goodall, Grant (ed.). 2021. *The Cambridge handbook of experimental syntax*. Cambridge: Cambridge University Press.
- Grafmiller, Jason, Benedikt Szendrői, Melanie Röthlisberger & Benedikt Heller. 2018. General introduction: A comparative perspective on probabilistic variation in grammar. *Glossa: A Journal of General Linguistics* 3(1). 94.

- Gravely, Brian. 2021. *Language acquisition and endogenous grammar change: The rise of Galician complementizer agreement*. Athens, GA: University of Georgia dissertation.
- Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of human language*, 73–113. Cambridge, MA: MIT Press.
- Greenhill, Simon, Chieh-Hsi Wu, Xia Hua, Michael Dunne, Stephen C. Levinson & Russell D. Gray. 2017. Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences of the United States of America* 114(42). E8822–E8829.
- Grenoble, Lenore A., Jessica Kantarovitch, Irena Khokhlova & Liudmila Zamorshchikova. 2019. Evidence of syntactic convergence among Russian-Sakha bilinguals. *Suvremena Lingvistika* 45(7). 41–57.
- Gries, Stefan Th. 2003. *Multifactorial analysis in corpus linguistics: A Study of particle placement*. New York: Continuum Press.
- Guardiano, Christina & Giuseppe Longobardi. 2005. Parametric comparison and language taxonomy. In Montserrat Batllori, Maria-Lluïsa Hernanz, Carme Picallo & Francesc Roca (eds.), *Grammaticalization and parametric variation*, 149–174. Oxford: Oxford University Press.
- Gulordava, Kristina, Paola Merlo & Benoit Crabbé. 2015. Dependency length minimisation effects in short spans: A large-scale analysis of adjective placement in complex noun phrases. In Chengqing Zong & Michael Strube (eds.), *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, vol. 2: *Short papers*, 477–482. Beijing: ACL.
- Gundel, Jeanette K. 1988. Universals of topic-comment structure. *Studies in Syntactic Typology* 17(1). 209–239.
- Gupton, Timothy. 2021. Aligning syntax and prosody in Galician: Against a prosodic isomorphism account. In Timothy Gupton & Elizabeth Gielau (eds.), *East and west of the Pentacrest: Linguistic studies in honor of Paula Kempchinsky*, 41–67. Amsterdam & Philadelphia: John Benjamins.
- Hagège, Claude. 2010. *Adpositions*. Oxford: Oxford University Press.
- Hale, Kenneth. 1983. Walpiri and the grammar of nonconfigurational languages. *Natural Language & Linguistic Theory* 1. 5–47.
- Hall, Matthew L., Rachel I. Mayberry & Victor S. Ferreira. 2013. Cognitive constraints on constituent order: Evidence from elicited pantomime. *Cognition* 129(1). 1–17.
- Hammarström, Harald. 2015. The basic word order typology: An exhaustive study. Paper presented at the closing conference of the Department of Linguistics at the Max Planck Institute for Evolutionary Anthropology, Leipzig. Available at: https://www.eva.mpg.de/fileadmin/content_files/linguistics/conferences/2015-diversitylinguistics/Hammarstroem_slides.pdf.
- Hammarström, Harald. 2016. Linguistic diversity and language evolution. *Journal of Language Evolution* 1(1). 19–29.
- Harmon, Luke J., Jason T. Weir, Chad D. Brock, Richard E. Glor & Wendell Challenger. 2008. GEIGER: Investigating evolutionary radiations. *Bioinformatics* 24(1). 129–131.
- Harris, Alice C. & Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.
- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687.
- Hawkins, John A. 1983. *Word order universals*. San Diego, CA: Academic Press.
- Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

- Hawkins, John A. 2019. Word-external properties in a typology of Modern English: A comparison with German. *English Language and Linguistics* 23(3). 701–727.
- Heine, Bernd. 2008. Contact-induced word order change without word order change. In Peter Siemund & Noemi Kintana (eds.), *Language contact and contact languages*, 33–60. Amsterdam & Philadelphia: John Benjamins.
- Heine, Bernd & Mechthild Reh. 1984. *Grammaticalization and reanalysis in African languages*. Hamburg: Helmut Buske Verlag.
- van den Heuvel, Wilco. 1997. *Het Inunhan: verkenning van een Bisayaanse taal uit de Filipijnen*. Amsterdam: Vrije Universiteit Amsterdam MA thesis.
- Hoot, Brad. 2016. Narrow presentational focus in Mexican Spanish: Experimental evidence. *Probus* 28(2). 335–365.
- Horvath, Julia. 1986. *FOCUS in the theory of grammar and the syntax of Hungarian*. Dordrecht: Foris.
- Hudson Kam, Carla L & Elissa L. Newport. 2005. Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development* 1(2). 151–195.
- Irurtzun, Aritz. 2009. Why Y: On the centrality of syntax in the architecture of grammar. *Catalan Journal of Linguistics* 8. 141–160.
- Jackendoff, Ray. 1977. *X-bar syntax: A study in phrase structure*. Cambridge, MA: MIT Press.
- Jackendoff, Ray & Eva Wittenberg. 2014. What you can say without syntax: A hierarchy of grammatical complexity. In Frederick Newmeyer & Laurel Preston (eds.), *Measuring linguistic complexity*, 65–82. Oxford: Oxford University Press.
- Jaeger, T. Florian & Elisabeth J. Norcliffe. 2009. The cross-linguistic study of sentence production. *Language and Linguistics Compass* 3(4). 866–887.
- Jaeger, T. Florian & Harry Tily. 2011. On language ‘utility’: Processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews: Cognitive Science* 2(3). 323–335.
- Jäger, Gerhard & Johannes Wahle. 2021. Phylogenetic typology. *Frontiers in Psychology* 12. 682132.
- Janda, Laura A. (ed.). 2013. *Cognitive linguistics: The quantitative turn*. Berlin & Boston: De Gruyter Mouton.
- Janssen, Rick. 2018. *Let the agents do the talking: On the influence of vocal tract anatomy on speech during ontogeny and glossogeny*. Nijmegen: Radboud University & Max Planck Institute for Psycholinguistics dissertation.
- Jiménez-Fernández, Ángel. 2015. Towards a typology of focus: Subject position and microvariation at the discourse-syntax interface. *Ampersand* 2. 49–60.
- Johansson, Stig & Knut Hofland. 1994. Towards an English-Norwegian parallel corpus. In Udo Fries, Gunnel Tottie & Peter Schneider (eds.), *Creating and using English language corpora*, 25–37. Amsterdam: Rodopi.
- Johnston, Trevor A. 2019. Clause constituents, arguments and the question of grammatical relations in Auslan (Australian Sign Language): A corpus-based study. *Studies in Language* 43(4). 941–996.
- Josserand, Mathilde, Marc Allassonnière-Tang, François Pellegrino & Dan Dediú. 2021. Interindividual variation refuses to go away: A Bayesian computer model of language change in communicative networks. *Frontiers in Psychology* 12. 626118.
- Kandylaki, Katarina D. & Ina Bornkessel-Schlesewsky. 2019. From story comprehension to the neurobiology of language. *Language, Cognition and Neuroscience* 34(4). 405–410.
- Kauhanen, Henri & George Walkden. 2018. Deriving the constant rate effect. *Natural Language & Linguistic Theory* 36(2). 483–521.
- Keller, Frank. 2006. Linear Optimality Theory as a model of gradience in grammar. In Gisbert Fanselow, Caroline Féry, Matthias Schlewsky & Ralf Vogel (eds.), *Gradience in grammar: Generative perspectives*, 270–287. Oxford: Oxford University Press.

- Kidd, Evan, Silke Brandt, Elena Lieven & Michael Tomasello. 2007. Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Language & Cognitive Processes* 22(6). 860–897.
- Kidd, Evan & Rowena Garcia. 2022. How diverse is child language acquisition research? *First Language* 42(6). 703–735.
- Kimmelman, Vadim. 2012. Word order in Russian Sign Language. *Sign Language Studies* 12(3). 414–445.
- Klavan, Jane & Dagmar Divjak. 2016. The cognitive plausibility of statistical classification models: Comparing textual and behavioral evidence. *Folia Linguistica* 50(2). 355–384.
- Klomp, Ulrika. 2021. *A descriptive grammar of Sign Language of the Netherlands*. Amsterdam: University of Amsterdam dissertation.
- Konopka, Agnieszka E. 2019. Encoding actions and verbs: Tracking the time-course of relational encoding during message and sentence formulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 45(8). 1486–1510.
- Koplenig, Alexander, Peter Meyer, Sascha Wolfer & Carolin Müller-Spitzer. 2017. The statistical trade-off between word order and word structure – Large-scale evidence for the principle of least effort. *PLoS One* 12(3). e0173614.
- Kortmann, Bernd. 2020. Turns and trends in 21st-century linguistics. In Bernd Kortmann (ed.), *English linguistics*, 241–286. Stuttgart: J.B. Metzler.
- Kramer, Mathew Alex. 2021. Dependency lengths in speech and writing: A cross-linguistic comparison via YouDePP, a pipeline for scraping and parsing YouTube captions. *Proceedings of the Society for Computation in Linguistics* 4(1). 359–365.
- Kreiner, Hamutal & Zohar Eviatar. 2014. The missing link in the embodiment of syntax: Prosody. *Brain and Language* 137. 91–102.
- Kroch, Anthony S. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1(3). 199–244.
- Kroch, Anthony S., Ann Taylor & Donald Ringe. 2000. The Middle English verb-second constraint: A case study in language contact and language change. In Susan Herring, Pieter van Reenen & Lene Schøsler (eds.), *Textual parameters in older language*, 353–391. Amsterdam & Philadelphia: John Benjamins.
- Langlois, Annie. 2004. *Alive and kicking: Areyonga Teenage Pitjantjatjara*. Canberra, ACT: Pacific Linguistics.
- Langsford, Steven, Rachel G. Stephens, John C. Dunn & Richard L. Lewis. 2019. In search of the factors behind naive sentence judgments: A state trace analysis of grammaticality and acceptability ratings. *Frontiers in Psychology* 10. 2886.
- Law, Rence S. 1997. *The Romblomanon grammar essentials sketch*. Available at: <https://romblomanon.net/en/study-of-language/grammar>.
- Leal, Tania, Emilie Destruel & Brad Hoot. 2018. The realization of information focus in monolingual and bilingual native Spanish. *Linguistic Approaches to Bilingualism* 8(2). 217–251.
- Leal, Tania & Timothy Gupton. 2021. Acceptability judgments in Romance languages. In Goodall Grant (ed.), *The Cambridge handbook of experimental syntax*, 448–476. Cambridge: Cambridge University Press.
- Lee, Hanjung. 2003. Prominence mismatch and markedness reduction in word order. *Natural Language & Linguistic Theory* 21. 617–680.
- Lee, Jennifer R. 1987. *Tiwi today: A study of language change in a contact situation*. Canberra, ACT: Pacific Linguistics.
- Lehmann, Christian. 1995 [1982]. *Thoughts on grammaticalization: A programmatic sketch, vol. I*. Munich: LINCOM Europa.

- Lehmann, Christian. 1992. Word order change by grammaticalization. In Marinel Gerritsen & Dieter Stein (eds.), *Internal and external factors in syntactic change*, 395–416. Berlin & New York: Mouton de Gruyter.
- Lehmann, Winfred P. 1973. A structural principle of language and its implications. *Language* 9. 47–66.
- Leisiö, Larisa. 2000. The word order in genitive constructions in a diaspora Russian. *International Journal of Bilingualism* 4(3). 301–325.
- Levinson, Stephen. 2016. Turn-taking in human communication – origins and implications for language processing. *Trends in Cognitive Sciences* 20. 6–14.
- Levshina, Natalia. 2017. Film subtitles as a corpus: An n-gram approach. *Corpora* 12(3). 311–338.
- Levshina, Natalia. 2019. Token-based typology and word order entropy. *Linguistic Typology* 23(3). 533–572.
- Levshina, Natalia. 2021a. Communicative efficiency and differential case marking: A reverse-engineering approach. *Linguistics Vanguard* 7(s3). 20190087.
- Levshina, Natalia. 2021b. Cross-linguistic trade-offs and causal relationships between cues to grammatical subject and object, and the problem of efficiency-related explanations. *Frontiers in Psychology* 12. 648200.
- Levshina, Natalia. 2022. Corpus-based typology: Applications, challenges and some solutions. *Linguistic Typology* 26(1). 129–160.
- Lightfoot, David W. 1982. *The language lottery: Towards a biology of grammars*. Cambridge, MA: MIT Press.
- Liu, Haitao. 2008. Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science* 9(2). 159–191.
- Liu, Haitao. 2010. Dependency direction as a means of word-order typology: A method based on dependency treebanks. *Lingua* 120(6). 1567–1578.
- Liu, Haitao, Chunshan Xu & Junying Liang. 2017. Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews* 21. 171–193.
- Liu, Zoey. 2019. A comparative corpus analysis of PP ordering in English and Chinese. In Xinying Chen & Ramon Ferrer-i-Cancho (eds.), *Proceedings of the First Workshop on Quantitative Syntax (Quasy, Syntaxfest 2019)*, 33–45. Paris: ACL. Available at: <https://www.aclweb.org/anthology/W19-7905>.
- Liu, Zoey. 2020. Mixed evidence for crosslinguistic dependency length minimization. *STUF – Language Typology and Universals* 73(4). 605–633.
- Liu, Zoey. 2021. The crosslinguistic relationship between ordering flexibility and dependency length minimization: A data-driven approach. *Proceedings of the Society for Computation in Linguistics* 4(1). 264–274.
- Liu, Zoey. 2022. A multifactorial approach to crosslinguistic constituent orderings. *Linguistics Vanguard*. <https://doi.org/10.1515/lingvan-2020-0139>, In press.
- López, Luis. 2009. *A derivational syntax for information structure*. Oxford: Oxford University Press.
- Luchkina, Tatiana & Jennifer S. Cole. 2021. Perception of word-level prominence in free word order language discourse. *Language and Speech* 64(2). 381–412.
- McIntire, Marina L. 1982. Constituent order & location in American Sign Language. *Sign Language Studies* 37. 345–386.
- MacDonald, Maryellen C. 2013. How language production shapes language form and comprehension. *Frontiers in Psychology* 4. 226.
- Mahmud, Altaf & Mumit Khan. 2007. Building a foundation of HPSG-based treebank on Bangla language. In *Proceedings of 10th International Conference on Computer and Information Technology, Dhaka, Bangladesh*, 1–6. Piscataway, NJ: IEEE.
- Manhardt, Francie, Susanne Brouwer, Eveline Van Wijk & Aslı Özyürek. 2023. Word order preference in sign influences speech in hearing bimodal bilinguals but not vice versa: Evidence from behavior and eye-gaze. *Bilingualism: Language and Cognition* 26(1). 48–61.

- Maurits, Luke & Thomas L. Griffiths. 2014. Tracing the roots of syntax with Bayesian phylogenetics. *Proceedings of the National Academy of Sciences* 111(37). 13576–13581.
- Meakins, Felicity. 2014. Language contact varieties. In Harold Koch & Rachel Nordlinger (eds.), *The languages and linguistics of Australia: A comprehensive guide*, vol. 3, 365–416. Berlin & Boston: De Gruyter Mouton.
- Meakins, Felicity & Carmel O'Shannessy. 2010. Ordering arguments about: Word order and discourse motivations in the development and use of the ergative marker in two Australian mixed languages. *Lingua* 120(7). 1693–1713.
- Meir, Irit, Mark Aronoff, Carl Börstell, So-One Hwang, Deniz Ilkbasaran, Itamar Kastner, Ryan Lepic, Adi Lifshitz Ben-Basat, Carol Padden & Wendy Sandler. 2017. The effect of being human and the basis of grammatical word order: Insights from novel communication systems and young sign languages. *Cognition* 158. 189–207.
- Michaelis, Susanne M., Philippe Maurer, Martin Haspelmath & Magnus Huber. 2013. *Atlas of pidgin and creole language structures online*. Max Planck Institute for Evolutionary Anthropology. Available at: <http://apics-online.info>.
- Mithun, Marianne. 1992. Is basic word order universal? In Doris L. Payne (ed.), *Pragmatics of word order flexibility*, 15–61. Amsterdam & Philadelphia: John Benjamins.
- Momma, Shota & Colin Phillips. 2018. The relationship between parsing and generation. *Annual Review of Linguistics* 4. 233–254.
- Mohanan, K. P. 1983. Lexical and configurational structures. *The Linguistics Review* 3. 113–139.
- Müller, Gereon. 2019. The third construction and strength of C: A gradient harmonic grammar approach. *Linguistische Arbeitsberichte* 94. 419–446.
- Mufwene, Salikoko. 2000. Creolization is a social, not a structural, process. *Creole Language Library* 22. 65–84.
- Müller, Stefan. 1999. *An extended and revised HPSG-Analysis for free relative clauses in German*. (Verbmobil-Report 225). Stuttgart: University of Stuttgart. Available at: <https://www.ims.uni-stuttgart.de/en/research/publications/verbmobil/>.
- Muntendam, Antje. 2013. On the nature of cross-linguistic transfer: A case study of Andean Spanish. *Bilingualism: Language and Cognition* 16(1). 111–131.
- Naccarato, Chiara, Anastasia Panova & Natalia Stoyanova. 2020. Word-order flexibility in genitive noun phrases: A corpus-based investigation of contact varieties of Russian. Paper presented at the SLE 2020 conference, 26th August 2020.
- Naccarato, Chiara, Anastasia Panova & Natalia Stoyanova. 2021. Word-order variation in a contact setting: A corpus-based investigation of Russian spoken in Daghestan. *Language Variation and Change* 33(3). 387–411.
- Namboodiripad, Savithry. 2017. *An experimental approach to variation and variability in constituent order*. La Jolla, CA: University of California San Diego dissertation.
- Namboodiripad, Savithry, Dayoung Kim & Gyeongnam Kim. 2018. English-dominant Korean speakers show reduced flexibility in constituent order. In Daniel Edmiston, Marina Ermolaeva, Emre Hakgüder, Jackie Lai, Kathryn Montemurro, Brandon Rhodes, Sankhagowit Amara & Michael Tabatowski (eds.), *Proceedings of the Fifty-Third Annual Meeting of the Chicago Linguistic Society*, 247–260. Chicago: Chicago Linguistic Society.
- Naranjo, Matías Guzmán & Laura Becker. 2018. Quantitative word order typology with UD. In Dag Haug, Stephan Oepen, Lilja Øvrelid, Marie Candito & Jan Hajic (eds.), *Proceedings of the 17th International Workshop on Treebanks and Linguistic Theories (TLT 2018)*, 91–104. Linköping, Norway: Linköping University Electronic Press.

- Newmeyer, Frederick J. 2000. On the reconstruction of “Proto-World” word order. In Chris Knight, Michael Studdert-Kennedy & James R. Hurford (eds.), *The evolutionary emergence of language: Social function and the origins of linguistic form*, 372–390. Cambridge: Cambridge University Press.
- Nicodemus, Brenda. 2009. *Prosodic markers and utterance boundaries in American Sign Language interpretation*. Washington: Gallaudet University Press.
- Norcliffe, Elisabeth, Agnieszka E. Konopka, Penelope Brown & Stephen C. Levinson. 2015. Word order affects the time course of sentence formulation in Tzeltal. *Language, Cognition and Neuroscience* 30(9). 1187–1208.
- Nordlinger, Rachel. 1998. *Constructive case: Evidence from Australian languages*. Stanford: CSLI Publications.
- Nordlinger, Rachel, Gabriela Garrido Rodriguez & Evan Kidd. 2022. Sentence planning and production in Murrinhpatha, an Australian ‘free word order’ language. *Language* 98(2). 187–220.
- Nordlinger, Rachel, Gabriela Garrido Rodriguez, Sasha Wilmoth & Evan Kidd. 2020. What drives word order flexibility? Evidence from sentence production experiments in two Australian Indigenous languages. Paper presented at the SLE 2020 conference, 26th August 2020, online.
- Nowak, Iga & Giosuè Baggio. 2017. Developmental constraints on learning artificial grammars with fixed, flexible and free word order. *Frontiers in Psychology* 8. 1816.
- Ortega-Santos, Iván. 2016. *Focus-related operations at the right edge in Spanish*. Amsterdam & Philadelphia: John Benjamins.
- Osborne, Timothy & Kim Gerdes. 2019. The status of function words in dependency grammar: A critique of Universal Dependencies (UD). *Glossa: A Journal of General Linguistics* 4(1). 17.
- Östling, Robert. 2015. Word order typology through multilingual word alignment. In Chengqing Zong & Michael Strube (eds.), *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, 205–211. Beijing: ACL.
- Özge, Duygu, Aylin Kuntay & Jesse Snedeker. 2019. Why wait for the verb? Turkish speaking children use case markers for incremental language comprehension. *Cognition* 183. 152–180.
- Pagel, Mark. 1999. Inferring the historical patterns of biological evolution. *Nature* 401. 877–884.
- Panhuis, Dirk. 1981. Word order, genre, adstratum: The place of the verb in Caesar’s topographical excursus. *Glotta* 59(3/4). 295–308.
- Park, Hyunji, Lane Schwartz & Francis Tyers. 2021. Expanding Universal Dependencies for polysynthetic languages: A case of St. Lawrence Island Yupik. In Manuel Mager, Arturo Oncevay, Annette Rios, Ivan Vladimir, Meza Ruiz, Alexis Palmer, Graham Neubig & Katharina Kann (eds.), *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*, 131–142. Zurich: ACL. <https://aclanthology.org/2021.americasnlp-1.14/> (accessed 1 February 2022).
- Patil, Umesh, Gerrit Kentner, Anja Gollrad, Frank Kügler, Caroline Féry & Shravan Vasishth. 2008. Focus, word order and intonation in Hindi. *Journal of South Asian Linguistics* 1. 53–70.
- Payne, Doris L. (ed.). 1992. *Pragmatics of word order flexibility*. Amsterdam/Philadelphia: John Benjamins.
- Pokorny, Julius. 1964. Zur Anfangstellung des inselkeltischen Verbum. *Münchener Studien zur Sprachwissenschaft* 16. 75–80.
- Polinsky, Maria. 2012. Headedness, again. In Thomas Graf, Denis Paperno, Anna Szabolcsi & Jos Tellings (eds.), *Theories of everything: In honor of Ed Keenan* (UCLA Working Papers in Linguistics 17), 348–359. Los Angeles, CA: UCLA Department of Linguistics.
- Poplack, Shana. 2020. A variationist perspective on language contact. In Evangelia Adamou & Yaron Matras (eds.), *The Routledge handbook of language contact*, 46–62. New York & London: Routledge.

- Poplack, Shana & Stephen Levey. 2010. Contact-induced grammatical change: A cautionary tale. In Peter Auer & Jürgen E. Schmidt (eds.), *Language and space: An international handbook of linguistic variation*, vol. 1: *Theories and methods*, 391–419. Berlin & New York: De Gruyter Mouton.
- Preminger, Omer. 2017. How can feature-sharing be asymmetric? Valuation as UNION over geometric feature structures. In Claire Halpert, Hadas Kotek & Coppe van Urk (eds.), *A pesky set: Papers for David Pesetsky*, 493–502. Cambridge, MA: MITWPL.
- Prince, Alan & Paul Smolensky. 2004. *Optimality theory: Constraint interaction in generative grammar*. Hoboken, NJ: Wiley Blackwell.
- Proske, Sina. 2022. *The impact of verb type on word order in German Sign Language*. Göttingen: Georg-August-Universität Göttingen dissertation.
- R Core Team. 2020. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: <https://www.R-project.org/>.
- Rajkumar, Rajakrishnan, Marten van Schijndel, Michael White & William Schuler. 2016. Investigating locality effects and surprisal in written English syntactic choice phenomena. *Cognition* 155. 204–232.
- Richards, Norvin. 2001. Leerdil yuujmen bana yanangarr (Old and new Lardil). In Rob Pensalfini & Norvin Richards (eds.), *MIT Working Papers on Endangered and Less Familiar Languages*, 2. Cambridge, MA: MIT.
- Rizzi, Luigi. 1997. On the fine structure of the left periphery. In Adriana Belletti & Luigi Rizzi (eds.), *Elements of grammar*, 281–337. Dordrecht: Kluwer.
- Roberts, Ian. 2019. *Parameter hierarchies and universal grammar*. Oxford: Oxford University Press.
- Rosenstein, Ofra. 2001. *Israeli Sign Language: A topic-prominent language*. Haifa: Haifa University MA thesis.
- Röthlisberger, Melanie & Benedikt Szmrecsanyi. 2020. Dialect typology: Recent advances. In Stanley D. Brunn & Roland Kehrein (eds.), *The handbook of the changing world language map*, 131–156. Cham: Springer.
- Rudnitskaya, Elena L. 2018. Porjadok slov (glagol i prjamoj objekt) v ustnyh rasskazah na évenkijskom jazyke [Word order (V and O) in spoken narratives in Evenki]. *Sibirskij filologičeskij žurnal* 1. 219–234.
- Salaberri, Iker. 2017. *Word order in subordinate clauses: Innovative or conservative? A typology of word order change*. Vitoria-Gasteiz: University of the Basque Country BA thesis.
- Salzmann, Martin David. 2004. *Theoretical approaches to locative inversion*. Zurich: University of Zurich dissertation. Available at: <http://www.martinsalzmann.com/files/salzmann-loc-inv-distr04-01.pdf>.
- Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt.
- Sasano, Ryohei & Manabu Okumura. 2016. A corpus-based analysis of canonical word order of Japanese double object constructions. In Katrin Erk & Noah A. Smith (eds.), *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 2236–2244. Berlin: ACL.
- Sauppe, Sebastian. 2017. Symmetrical and asymmetrical voice systems and processing load: Pupillometric evidence from sentence production in Tagalog and German. *Language* 93(2). 288–313.
- Sauppe, Sebastian, Elisabeth Norcliffe, Agnieszka E. Konopka, Robert D. Van Valin, Jr. & Stephen C. Levinson. 2013. Dependencies first: Eye tracking evidence from sentence production in Tagalog. In Markus Knauff, Michael Pauen, Natalie Sebanz & Ipke Wachsmuth (eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, 1265–1270. Austin, TX: Cognitive Science Society.
- Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system: Usage, conventionalization, and entrenchment*. Oxford: Oxford University Press.
- Schmidt, Annette. 1985a. The fate of ergativity in dying Dyirbal. *Language* 61(2). 378–396.
- Schmidt, Annette. 1985b. *Young People's Dyirbal: An example of language death from Australia*. Cambridge: Cambridge University Press.

- Schmidtke-Bode, Karsten, Natalia Levshina, Susanne Maria Michaelis & Ilija A. Seržant (eds.). 2019. *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*. Berlin: Language Science Press.
- Schnell, Stefan & Nils Norman Schiborr. 2022. Crosslinguistic corpus studies in linguistic typology. *Annual Review of Linguistics* 8. 171–191.
- Schouwstra, Marieke & Henriëtte de Swart. 2014. The semantic origins of work order. *Cognition* 131(3). 431–436.
- Sedarous, Yourdanis & Savithry Nambodiripad. 2020. Using audio stimuli in acceptability judgment experiments. *Language and Linguistics Compass* 14(8). e12377.
- Seidenberg, Mark S. & Maryellen C. MacDonald. 1999. A probabilistic constraints approach to language acquisition and processing. *Cognitive Science* 23. 569–588.
- Shannon, Claude. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27(3). 379–423.
- Siewierska, Anna. 1998. Variation in major constituent order; a global and European perspective. In Siewierska Anna (ed.), *Constituent order in the languages of Europe*, 475–552. Berlin & New York: Mouton de Gruyter.
- Šimík, Radek & Marta Wierzba. 2017. Expression of information structure in West Slavic: Modeling the impact of prosodic and word-order factors. *Language* 93(3). 671–709.
- Simov, Kiril, Petya Osenova, Alexander Simov & Milen Kouylekov. 2004. Design and implementation of the Bulgarian HPSG-based treebank. *Research on Language and Computation* 2(4). 495–522.
- Simpson, Jane. 2012. *Warlpiri morpho-syntax: A lexicalist approach*, vol. 23. Dordrecht: Springer.
- Sinnemäki, Kaius. 2008. Complexity trade-offs in core argument marking. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 67–88. Amsterdam & Philadelphia: John Benjamins.
- Slobin, Daniel I. & Thomas G. Bever. 1982. Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition* 12(3). 229–265.
- Smith, Kenny & Elizabeth Wonnacott. 2010. Eliminating unpredictable variation through iterated learning. *Cognition* 116(3). 444–449.
- Smolensky, Paul & Matthew Goldrick. 2016. Gradient symbolic representations in grammar: The case of French liaison. *Rutgers Optimality Archive* 1552. 1–37.
- Spence, Justin, Zoey Liu, Kayla Palakurthy & Tyler Lee-Wynant. 2018. Syntactic annotation of a Hupa text corpus. In Willem de Reuse & Siri Tuttle (eds.), *Working papers in Athabaskan languages 2017*, 37–53. Fairbanks, AK: Alaska Native Language Center.
- Sprouse, Jon & Diogo Almeida. 2017. Design sensitivity and statistical power in acceptability judgment experiments. *Glossa: A Journal of General Linguistics* 2(1). 1–32.
- Stapert, Eugénie. 2013. *Contact-induced change in Dolgan: An investigation into the role of linguistic data for the reconstruction of a people's (pre)history*. Utrecht: LOT dissertation.
- Sussex, Ronald. 1993. Slavonic languages in emigration. In Comrie Bernard & Greville G. Corbett (eds.), *The Slavonic languages*, 999–1035. London: Routledge.
- Svelmoe, Thelma & Gordon Svelmoe. 1974. *Notes on Mansaka grammar*. Huntington Beach, CA: Summer Institute of Linguistics.
- Swerts, Marc, Krahmer Emiel & Cinzia Avesani. 2002. Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics* 30. 629–654.
- Szmrecsanyi, Benedikt, Jason Graffmiller, Benedikt Heller & Melanie Röthlisberger. 2016. Around the world in three alternations: Modeling syntactic variation in varieties of English. *English World-Wide* 37(2). 109–137.

- Talamo, Luigi & Annemarie Verkerk. 2022. A new methodology for an old problem: A corpus-based typology of adnominal word order in European languages. *Italian Journal of Linguistics* 34(2). 171–226.
- Tanangkingsing, Michael. 2009. *A functional reference grammar of Cebuano*. Taipei: National Taiwan University dissertation.
- Tanaka, Mikihiro N., Holly P. Branigan, Janet F. McLean & Martin J. Pickering. 2011. Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language* 65. 318–330.
- Taylor, Ann & Susan Pintzuk. 2012. Rethinking the OV/VO alternation in Old English: The effect of complexity, grammatical weight, and information status. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English, 835–837*. Oxford: Oxford University Press.
- Temperley, David & Daniel Gildea. 2018. Minimizing syntactic dependency lengths: Typological/cognitive universal? *Annual Review of Linguistics* 4(1). 67–80.
- Thomason, Sarah G. 2001. *Language contact*. Edinburgh: Edinburgh University Press.
- Thomason, Sarah G. & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley, CA: University of California Press.
- Ueno, Mieko & Maria Polinsky. 2009. Does headedness affect processing? A new look at the VO-OV contrast. *Journal of Linguistics* 45. 675–710.
- Vainio, Martti & Juhani Järviö. 2006. Tonal features, intensity, and word order in the perception of prominence. *Journal of Phonetics* 34(3). 319–342.
- Vaissière, Jacqueline & Alexis Michaud. 2006. Prosodic constituents in French: A data-driven approach. In Yuji Kawaguchi, Ivan Fónagy & Tsunekazu Moriguchi (eds.), *Prosody and syntax: Cross-linguistic perspectives*, 47–64. Amsterdam & Philadelphia: John Benjamins.
- Vennemann, Theo. 1974. Theoretical word order studies: Results and problems. *Papere zur Linguistik* 7. 5–25.
- Vennemann, Theo & Ray Harlow. 1977. Categorical grammar and consistent basic VX serialization. *Theoretical Linguistics* 4. 227–254.
- Wälchli, Bernhard. 2009. Data reduction typology and the bimodal distribution bias. *Linguistic Typology* 13(1). 77–94.
- Wasow, Thomas & Jennifer Arnold. 2003. Post-verbal constituent ordering in English. *Topics in English Linguistics* 43. 119–154.
- Watkins, Calvert. 1963. Preliminaries to a historical and comparative analysis of the syntax of the Old Irish verb. *Celtica* 6. 1–49.
- Wechsler, Stephen & Ash Asudeh. 2021. HPSG and lexical functional grammar. In Stefan Müller, Anne Abeillé, Robert Borsley & Jean-Pierre Koenig (eds.), *Head-driven phrase structure grammar*. Berlin: Language Science Press.
- Williams, Edwin. 1977. Discourse and logical form. *Linguistic Inquiry* 8(1). 101–139.
- Wilmoth, Sasha, Rachel Nordlinger, Gabriela Garrido Rodriguez & Evan Kidd. Forthcoming. Word order flexibility in Pitjantjatjara, across genres and generations. In Carmel O'Shannessy, James Gray & Denise Angelo (eds.), *Projecting voices: Studies in language and linguistics in honour of Jane Simpson*. Canberra: Asia-Pacific Linguistics.
- Wonnacott, Elizabeth & Elissa L. Newport. 2005. Novelty and regularization: The effect of novel instances on rule formation. In Alejna Brugos, Manuela R. Clark-Cotton & Seungwan Ha (eds.), *BUCLD 29: Proceedings of the 29th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadia Press.
- Yamashita, Hiroko. 2002. Scrambled sentences in Japanese: Linguistic properties and motivations for production. *Text – Interdisciplinary Journal for the Study of Discourse* 22(4). 597–634.

- Yamashita, Hiroko & Franklin Chang. 2001. “Long before short” preference in the production of a head-final language. *Cognition* 81(2). B45–B55.
- Yang, Charles. 2002. *Knowledge and learning in natural language*. Oxford: Oxford University Press.
- Zdorenko, Tatiana. 2010. Subject omission in Russian: A study of the Russian National Corpus. In Stefan Th. Gries, Stefanie Wulff & Mark Davies (eds.), *Corpus-linguistic applications*, 119–133. Leiden: Brill.
- Zeman, Daniel, Joakim Nivre, Mitchell Abrams, et al. 2020. Universal dependencies 2.6. In *LINDAT/CLARIAH-CZ Digital library at the Institute of formal and applied linguistics (ÚFAL)*. Prague: Faculty of Mathematics and Physics, Charles University. Available at: <http://hdl.handle.net/11234/1-3226>, <http://universaldependencies.org>.
- Zemskaja, Elena L. & Lara A. Kapanadze (eds.). 1978. *Russkaja Razgovornaja Reč. Teksty* [Russian colloquial speech. Texts]. Moscow: Nauka.
- Zubizarreta, María Luisa. 1998. *Prosody, focus, and word order*. Cambridge, MA: MIT Press.