Thomas Wasow (Stanford University)

Factors Influencing Word Ordering*

A number of factors have been shown to influence word order preferences, both across languages and within particular languages. These include preferences for: subject before object; verb adjacent to direct object; short phrases before long phrases or the reverse, depending on the headedness pattern of the language; old information before new information; and more accessible before less accessible information. Plausible explanations for many of such word order tendencies have been offered in terms of ease of processing, discourse coherence, or nuances of meaning to be expressed.

1 Introduction

Speech unfolds over time, necessitating that words in spoken languages appear in some sequential order. In principle, that ordering could be free; that is, it is logically possible that word order plays no role in the well-formedness or meaning of utterances. But that is not the case. Every natural language seems to have some constraints on possible word orders. Even in so-called free word order languages, some orderings are ungrammatical, and differences among the grammatical orders can serve communicative functions (see, for example, Simpson 2007).

Constraints on word ordering are to some extent conventionalized, as is evident from the fact that different languages have different patterns of word order. But there are some generalizations that hold across languages, both in terms of the dominant patterns of word order and in terms of the types of word order variation we find within and across languages.

The word-order generalizations discussed in this chapter are all probabilistic — that is, they have exceptions. There are no doubt some restrictions on word order in particular languages that hold without exception. I do not, however, know of any exceptionless universal constraints on word order.

A great deal of the literature on word order has been primarily descriptive. Two of the pioneers in this domain, Otto Behaghel and Joseph Greenberg, discovered some of the most important generalizations about word order but wrote little about why those generalizations hold. Starting with Hawkins (1994), however, much of the research on word order has looked for explanations as to why languages exhibit the word orders they do. The explanations are of two general types, which I will refer to as processing and pragmatic.[1] Processing explanations account for word order preferences on the basis of communicative efficiency. That is, some word orders are more easily produced or comprehended than others, owing to either general properties of human cognition or to even more general properties of information

* Comments on earlier versions of this chapter by John A. Hawkins, the editors of this volume, and an anonymous reviewer led to various improvements. I am grateful for their help.

[1] A third type of explanation can be found in the generative tradition, consisting of abstract representations and principles from which observable facts can be derived. I consider this sort of explanation incomplete without a grounding of the abstract representations and principles in human psychology, biology, or communicative needs. Paradigm examples of the generative approach applied to word order can be found in Chomsky (1970) and Kayne (1995).

transmission. Pragmatic explanations account for word order preferences by showing that they convey aspects of meaning that go beyond literal meaning (that is, truth conditions); these include conveying the relative importance of parts of a sentence or how the sentences in a discourse are connected.

This chapter reviews some of the main generalizations about word order discovered by Behaghel, Greenberg, and others, and discusses some of the work that has been done to try to explain those generalizations. This is a rich literature, and I do not pretend to be able to provide a comprehensive survey; rather, the chapter is a sampling of research on this topic, along with some reflections on the research described. For a more thorough survey, see Song (2012).

2 Cross-language Patterns

Greenberg's seminal 1963 paper "Some universals of grammar with particular reference to the order of meaningful elements" was the first of a very large number of typological studies of word order. One of its important contributions was the classification of languages on the basis of the relative positions of verb, subject, and object in simple transitive clauses. Greenberg's first word-order universal was that the basic word order in languages is "almost always" one in which the subject precedes the object. That is, of the six logically possible orderings among V(erb), S(ubject), and O(bject), only three, VSO, SVO, and SOV, occur with substantial frequency among the languages of the world. This has been confirmed in studies involving many more than Greenberg's original sample of 30 languages (see for example Hammarström (2016)).

To explain why natural languages overwhelmingly favor SO over OS, it is necessary first to consider how linguists distinguish grammatical subjects from objects. Keenan's (1976) detailed examination of the properties of subjects contains many insights on this question. Two that are of particular relevance in the present context are the following:[2]

(1) "…subjects normally express the agent of the action, if there is one." (p. 321)

(2) "…subjects are normally the topic of the …sentence, i.e., they identify what the speaker is talking about." (p. 318)

Taken together with the many grammatical properties Keenan identifies as characteristic of subjects, what (1) says is that noun phrases denoting agents characteristically will exhibit those properties. In clauses denoting actions involving two participants (arguably the canonical type of transitive clause), the agent, as the perpetrator of the action, is psychologically more salient than the other participant. Greenberg's universal then has a fairly straightforward explanation: all else being equal, speakers will mention the more salient participant before the less salient one. Support for this comes from the studies described by Goldin-Meadow et al. (2008), in which participants given the task of describing an event using only gestures consistently produced gestures to denote agents before those used to denote patients.

[2] Keenan's discussion refers to "b-subjects," where the b is short for "basic". That is, he is concerned with subjects of basic, unmarked clauses (which he calls "b-sentences"), excluding passives and other marked constructions.

Similarly, (2) helps to explain Greenberg's universal. Linguistic communication is not limited to isolated sentences. Rather, people usually speak in connected discourses, in which the connections between sentences can be as important to understanding as are the relations among the components of a sentence. It makes sense, therefore, to link an utterance to what preceded it early in each sentence. This promotes discourse coherence.

There is of course no guarantee that the semantic and discourse functions of subjects given in (1) and (2) will always converge. Hence, exceptions to both generalizations are not hard to find, and some marked constructions in languages serve to indicate such mismatches. Both, however, help to explain why the basic word order in almost all languages has subjects preceding objects.

Among the three orders in which the subject precedes the object, languages with SOV as their basic order are the most common, followed by SVO languages, with VSO being considerably less common. Intriguingly, in the Goldin-Meadow et al. (2008) studies cited above, participants overwhelmingly produced gestures denoting the action after having produced gestures denoting the agent and patient. That is, they used the analogue to SOV ordering over 80% of the time, even if that was not the dominant order of their native languages. This suggests that the high proportion of languages that are SOV might be attributable to some deeper cognitive preference. Exploring this speculation further goes beyond the scope of this chapter.

Why are VSO languages less frequent than SOV or SVO languages? One possible explanation is based on a generalization due to Behaghel (1932: 4), "*daß das geistig eng Zusammengehörige auch eng zusammengestellt wird*."[3] Keenan (1979) argues that the denotation of the object of a transitive verb restricts the meaning of the verb in a way that the subject of a transitive verb typically does not. He cites (along with several other examples) the different types of actions denoted by *cut* when its object is a body part (*cut a finger*), a substance (*cut heroin*), or an abstraction (*cut prices*). This close semantic association between transitive verbs and their objects, combined with Behaghel's generalization above suggests that word orders in which the verb and object are adjacent should be preferred. Of the three basic word orders observing the preference for subjects before objects, only VSO separates the verb and the object.

Among Greenberg's other observations about cross-language word order regularities is that the position of the verb with respect to the object in the basic word order tends to correlate strongly with other ordering preferences. For example, VO languages tend to have prepositions, while OV languages tend to have postpositions; and genitives follow nouns in VO languages, but precede them in OV languages. Languages that conform to such tendencies are often referred to as head-initial or head-final languages, and consistency of head position across phrases is sometime called harmonic word order.

[3] In Wasow (2002) I translated this as, 'that what belongs close together conceptually also gets placed close together.' Behaghel labels this *"[d]as oberste Gesetz*" ('the highest law') of word order.

3 Language-internal Order Variation

Some of the factors influencing word ordering across languages also play a role in word-order variation within languages. For example, in a normally SVO language like English, OSV sentences may be used to promote discourse coherence, as in the following:

(3) A: Do you know the Smiths?
B: Him I've met; her I haven't.

Here B uses two OSV clauses in order to highlight the link between the objects of those clauses and the object in A's question, and to emphasize the contrast between the two parts of the answer. In this case, the violation of the general preference for subjects before objects in English serves a discourse function, much as the canonical ordering does most of the time.

Differences in word order within a language very often serve to signal some semantic or pragmatic difference. If B in example (3) responded with the same words, but using the normal SVO order, the meaning conveyed would be subtly different, focusing attention more on the speaker and less on the two other individuals mentioned. Bolinger (1968: 127) went so far as to claim "a difference in syntactic form always spells a difference in meaning." I believe this is an overstatement (see Wasow 2009: 262), but it is true that different word orders are very often used to convey different nuances of meaning.

3.1 Weight

Another important factor influencing word order variation is ease of processing. In three books (Hawkins 1994, 2004, 2014) and numerous papers, Hawkins posited several principles of efficient processing that he claims influence the grammatical forms speakers use. The one of most relevance to word order is what he calls Minimize Domains (MiD). Hawkins's formulation of MiD is fairly complex, involving supporting definitions, but the concept is quite intuitive: if there is some syntactic and/or semantic dependency between two elements, processing that dependency will be facilitated when they are in close proximity. In most sentences, there are multiple pairs of elements with such dependencies, and, according to MiD, word orders will be favored to the extent that they minimize the average distance between dependent pairs of elements.

Notice that this is an extension of Behaghel's generalization above. Behaghel's observation was evidently limited to semantic dependencies ("*geistig eng Zusammengehörige*"), but Hawkins extends this to syntactic dependencies, such as agreement and case-marking, which do not always have a clear semantic basis.

One application of MiD is as an explanation and extension of another of Behaghel's generalizations. He dubbed this one, first published in Behaghel (1909/10), "*das Gesetz der wachsenden Glieder*" ('the law of the growing constituents'). Based on his examination of word order patterns in a number of European languages, he concluded (Behaghel 1932: 6) "*daß von zwei Gliedern,*

soweit möglich, das kürzere vorausgeht, das längere nachsteht."[4] This generalization, sometimes known as "end weight" or "short-before-long" has been widely discussed in the literature. Hawkins (1994) sought to explain it in terms of processing, using a predecessor to (and special case of) MiD that he called Early Immediate Constituents (EIC). The idea was that processing efficiency would be maximized by minimizing the number of words that needed to be processed for the determination of immediate constituent structure.

As an illustration, in the two examples in (4), taken from Hawkins (2003), the underlined portion indicates the words that must be processed to determine that the verb phrase contains a verb and two prepositional phrases.

(4) a. The gamekeeper [$_{VP}$ looked [$_{PP}$ through his binoculars] [$_{PP}$ into the blue but slightly overcast sky]]
b. The gamekeeper [$_{VP}$ looked [$_{PP}$ into the blue but slightly overcast sky] [$_{PP}$ through his binoculars]]

The underlined portion in (4a) is five words long, whereas in (4b) it is nine words long. Hence, according to Hawkins, structures like (4a) should occur more frequently and be processed more easily than structures like (4b). Notice that the VP in (4a) obeys Behaghel's *Gesetz der Wachsenden Glieder*, whereas the VP in (4b) does not.

Hawkins (1994) deduced predictions from EIC and then checked corpus data from ten languages to test their accuracy. One novel prediction was that in consistently head-final languages, longer constituents should precede rather than follow shorter ones. To see why that is the case, consider the analogue to (4) in an SOV language with postpositions, rather than prepositions. The two VPs would look schematically like (5).

(5) a. [$_{VP}$ [$_{PP}$ X X P] [$_{PP}$ X X X X X X P] V]]
b. [$_{VP}$ [$_{PP}$ X X X X X X P] [$_{PP}$ X X P] V]]

Here the Xs stand for arbitrary words, the Ps are postpositions, and V is a verb. As in (4), the underlined regions represent the portion of the sentences that must be processed to determine the top-level constituent structure of the VP. But in this case, the underlined region is shorter in the (b) example. Thus, EIC predicts that consistently head-final languages should show a preference for long constituents before short ones, contrary to Behaghel's generalization. Two of the languages Hawkins (1994) applied EIC to were Japanese and Korean, which are consistently verb-final and postpositional; they did indeed show the long-before-short pattern. Subsequent corpus and psycholinguistic studies (e.g., Yamashita & Chang 2001) have supported this finding.

Gibson (1998, 2000) put forward a similar processing-based account of such phenomena (which I will henceforth refer to as 'weight effects'), grounding his

[4] 'that of two constituents, to the extent possible, the shorter one precedes, the longer one follows.'

account on properties of short-term memory. There are subtle but significant differences between Hawkins's and Gibson's proposals for how to compute dependency length, but they are conceptually very similar.

Faghiri and Samvelian (in press) argue that these differences lead to different predictions regarding the word order preferences in a verb-final but prepositional language (a combination that Greenberg (1963) noted was relatively uncommon). The reason for this can be seen by considering the two schematic structures in (6).

(6) a. [$_{VP}$ [$_{PP}$ P X] [$_{PP}$ P X X X X X] V]]
b. [$_{VP}$ [$_{PP}$ P X X X X X] [$_{PP}$ P X] V]]

In cases like this, where recognition of the three immediate constituents of VP requires the full length of the VP, irrespective or order, Hawkins suggests that, at each word, the ratio of immediate constituents recognized to words processed should be computed, and then averaged over the whole structure. He calls this the IC-to-Word ratio. In (6), the calculations would be as follows:

(7) a. [$_{VP}$ [$_{PP}$ P X] [$_{PP}$ P X X X X] V]]
1/1 1/2 2/3 2/4 2/5 2/6 2/7 2/8 3/9 IC-to-Word ratio = 47.4%
b. [$_{VP}$ [$_{PP}$ P X X X X X] [$_{PP}$ P X] V]]
1/1 1/2 1/3 1/4 1/5 1/6 2/7 2/8 3/9 IC-to-Word ratio = 36.8%

Hawkins's theory says that the higher IC-to-Word ratio (that is, (6a)) should be preferred. Gibson's account, on the other hand, simply compares the sum of the dependency distances in the two structures, with the lower total being preferred. In this case, that would be 8+6=14 in (6a) and 8+2=10 in (6b). Thus, Hawkins predicts that a language with structures like those in (6) should prefer the short-before-long (a) order, whereas Gibson predicts it should prefer the long-before-short (b) order. Faghiri and Samvelian offer Persian as a language of this type and provide evidence that the long-before-short order is in fact preferred.

In the last decade, a number of linguists have applied computational tools to test the hypothesis that minimization of dependency length is an important factor in word ordering across languages (e.g., Gildea and Temperley (2010), Futrell et al. (2015), Futrell et al. (2020)). This research has taken advantage of the availability of corpora from an increasing number of languages that have been parsed for syntactic dependencies. In particular, the Universal Dependencies project (see https://universaldependencies.org/) has provided dependency-parsed corpora for dozens of languages. In each language, then, the dependency lengths in the corpus can be compared to the dependency lengths that would result if the ordering of the words in the sentences of the corpus is scrambled. Multiple such randomly scrambled alternatives are constructed for each corpus and compared to the original. In the languages that have been tested in this way, the total dependency lengths in the scrambled versions are consistently greater than in the original corpus, providing support for Hawkins's and Gibson's insight that dependency minimization is an

important factor in determining word ordering. For a survey of a variety of types of research on dependency minimization, see Liu et al. (2017).

A fascinating next step in this research program is described by Hahn et al. (2020). They explored the hypothesis that some word-order generalizations could be explained in terms of abstract properties of communication systems, independent of who or what is communicating. While earlier research advocating some version of dependency minimization as an explanation of word order facts cited human short-term memory limitations as the motivation for dependency minimization, Hahn et al.'s results do not rest on any assumptions about human psychology. Rather, they define a mathematically precise notion of communicative efficiency and argue that the actual word orders in the languages whose corpora they tested scored better on their metric of communicative efficiency than scrambled versions of the corpora. Communicative efficiency in their models correlated strongly (and inversely) with dependency length, suggesting that dependency minimization may be derivable from very general considerations about efficient communication. Moreover, in their models, languages with harmonic (that is, consistently head-initial or head-final) word orders scored higher on their metric of communicative efficiency than those whose basic word order is not harmonic.

(4) and (5) illustrate the effects of dependency minimization in the relative ordering of two PPs. Similar effects can be observed in the ordering of constituents of different types. Three alternations in English that have received considerable attention with regard to the effects of grammatical weight are so-called heavy NP shift (HNPS), the verb-particle construction, and the dative alternation. Examples are given in (8).

(8) a. I put the keys to my house and car on the table. ~ I put on the table the keys to my house and car.
b. I picked my car keys up. ~ I picked up my car keys.
c. I handed my car keys to the valet. ~ I handed the valet my car keys.

Each of these has idiosyncracies, and Melnick (2017) has shown that they differ in their sensitivity to grammatical weight. In the case of HNPS, there is strong bias against the shifted ordering — that is, with the direct object non-adjacent to the verb. Even when the object is longer than the other postverbal constituent by as many as 3 or 4 words, corpora show a preference for the canonical ordering of V-NP-X (where X is often a PP, but may be of some other type). The canonical ordering is favored by the fact, discussed above, that verbs and their direct objects have a close semantic relationship. Hence, the shifted ordering is relatively rare. Indeed, in experimental studies in which participants are asked for acceptability judgments, the shifted order never gets higher average scores than the canonical order, no matter how large the length difference between the postverbal constituents (see Medeiros, Mains, & McGowan 2020).

The verb-particle construction differs from most other weight-sensitive phenomena in that one of the two elements whose order can be reversed consists of

a single (usually monosyllabic) word. Consequently, according to EIC, the ordering in which the particle is adjacent to the verb should be preferred in all cases except when the object NP is a single word. However, while corpus studies have shown a powerful weight effect, examples with the object NP intervening between the verb and the particle are by no means rare, provided that the object is under five words long (see, e.g., Lohse et al. 2004). In this construction, as in HNPS, the tendency for the object NP to appear adjacent to the verb even when considerations of weight would lead one to expect it to occur later in the verb phrase can plausibly be attributed to the close semantic connection between verbs and their objects. Lohse et al. pointed out that, in this construction, there is sometimes also a semantic dependency between the verb and its particle. To illustrate this, consider examples like *lift the glass up*, in which the glass is lifted and goes up, with *figure the problem out*, in which the problem cannot be said to be figured or to go out. Dependency minimization predicts that verb-particle pairs exhibiting such a dependency should appear adjacent (with the object following) at higher rates than verb-particle pairs without such a semantic dependency. The corpus study of Lohse et al. confirmed this prediction.

The dative alternation has two idiosyncracies. The first is simply that the two variants differ not just in word order, but also in the presence or absence of a preposition. The second is that there are subtle meaning differences between the variants, as noted by Green (1974) and Oehrle (1976). For example, the contrast between (9) and (10) shows that the object of *to* in the prepositional variant may be either a recipient or a location, whereas the first object in the double object variant must be a recipient.

(9) a. The pitcher threw the ball to Jones.
    b. The pitcher threw Jones the ball.
(10) a. The pitcher threw the ball to first base.
    b.*The pitcher threw first base the ball.

In spite of these complicating features of the dative alternation, it exhibits clear weight effects in corpus data. This was noted by Hawkins (1994), and corroborated by Wasow (2002) and Bresnan et al. (2007), among many others. The latter paper provides an especially careful simultaneous examination of multiple factors that have been claimed in the literature to influence whether a speaker uses the prepositional or double object dative construction. It revealed not only that weight plays an important role in this alternation, but also that many other factors do too.

3.2 Information Structure

One of these other factors is what is sometimes called "information structure" — that is, the distinction between information already available in the discourse and information that is newly introduced. There is a large literature on how word order is influenced by information structure, most of which claims a universal tendency to

order old (also known as "given") information before new information.[5] This ordering seems natural in terms of discourse coherence, since the old information in a sentence provides a link to what was said earlier. Bresnan et al.'s (2007) multivariate models of the English dative alternation in the Switchboard corpus showed a strong effect of information structure, so that sentences with new theme arguments tended to occur in the double-object construction and those with new recipient arguments tended to occur in the prepositional construction.

A natural question that arises is whether the influence of information structure and grammatical weight are in fact distinct phenomena. Since old information can, in general, be comprehensibly expressed with fewer words than are required for new information, could it be that weight effects can be reduced to a preference for ordering old information before new? Arnold et al. (2000) investigated that question via a corpus study and an experiment, finding that the two influences on word order could in fact be teased apart. Bresnan et al.'s (2007) study reinforces this conclusion, since both information structure and weight emerge as significant factors in their model.

This finding should not have been surprising, for the two factors (weight and information structure) appear to differ in their cross-linguistic patterning. As noted above, the influence of weight on word order depends on the basic word order of the language: head-final languages prefer long-before-short ordering, whereas other types of languages seem to prefer short-before-long. But none of the literature on information structure claims any such bifurcation in the effects of information newness. The tendency to place old information before new information does not appear to depend on what the basic word order of the language is. And if weight effects are best explained in terms of processing while information structural effects are best explained in terms of discourse coherence, there is no reason to expect that their influence will always converge.

3.3 Accessibility

Another factor that has been found to influence word order is animacy. Corpus and experimental studies in a variety of languages have shown that there is tendency for expressions denoting animate entities to precede those denoting inanimates (see, e.g., Branigan et al. (2008)). Indeed, Tomlin (1986: 102) put forward "The Animated First Principle," which he summarizes as: "In a transitive clause, all other things being equal, there is a tendency for the most 'animated' NP to precede other NPs."

This generalization overlaps to a significant extent with the observation, noted above, that agents tend to precede patients, for agents are typically (though not always) animate. Animate entities are psychologically highly salient. Hence, the tendency for expressions denoting animates to occur early seems natural from a production perspective. Spoken communication is highly time-constrained; if a

[5] As I noted in Wasow (2002: 62-65), this literature can be quite confusing, owing to a plethora of terms used for information structural distinctions, uncertainty about whether different terminologies denote exactly the same distinction, and occasional contradictory claims about how information structure correlates with word order.

person speaks too slowly or hesitantly, someone else will begin to speak. It makes sense, then, for speakers to utter what is readily accessible early, providing extra time to access and articulate what is less accessible. Branigan et al. (2008: 177), after reviewing a considerable body of literature on the subject, concluded that "animacy exerts its influence because it correlates with conceptual accessibility, an index of how easily a concept is retrieved from memory."

Notice that this idea also can be invoked as at least part of the explanation for the old-before-new preference, since old information is generally more accessible than new information. One might be tempted also to try to treat the short-before-long preference in many languages as another manifestation of accessibility, but this runs afoul of the fact that some languages exhibit a long-before-short preference. Yamashita and Chang (2001) investigated this issue and came to the conclusion that both conceptual accessibility and formal factors like length play a role in word order preferences.

4 Conclusion

Languages are primarily vehicles of spoken communication among humans. That fact alone goes a long way towards explaining many properties of language, including many of the regularities concerning word order that linguists have discovered. A variety of factors influence the sequencing of words in an utterance, including properties of the language being spoken, properties of the situations being referred to, the linguistic forms used in referring to those situations, what was said earlier in the discourse, and the state of mind of the speaker. There remain many puzzles about word order in particular languages and across languages, but great progress has been made in the last few decades both in documenting what word orders occur and explaining their distributions.

References

Arnold, Jennifer, Anthony Losongco, Thomas Wasow, & Ryan Ginstrom. 2000. "Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering". *Language* 76(1). 28-55.

Behaghel, Otto. 1909/10. "Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern". *Indogermanische Forschungen* 25. 110-142.

Behaghel, Otto. 1932. *Deutsche Syntax: Eine Geschichtleiche Darstellung. Band IV: Wortstellung. Periodenbau*. Heidelberg. Carl Winters Universitätsbuchhandlung.

Bolinger, Dwight. 1968. "Entailment and the Meaning of Structures". *Glossa* 2. 119-127.

Branigan Holly P., Martin J. Pickering, & Mikihiro Tanaka. 2008. "Contributions of animacy to grammatical function assignment and word order during production". *Lingua* 118. 172–189.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina, & Harald Baayen. 2007. "Predicting the Dative Alternation." In G. Boume, I. Kraemer, and J. Zwarts (eds) *Cognitive Foundations of Interpretation*, pp. 69-94. Amsterdam. Royal Netherlands Academy of Science.

Chomsky, Noam. 1970. "Remarks on Nominalization." In R. Jacobs and P. Rosenbaum (eds) *Readings in English Transformational Grammar*. Waltham, MA. Ginn & Co.

Faghiri, Pegah & Pollet Samvelian. 2020. "Word order preferences and the effect of phrasal length in SOV languages: evidence from sentence production in Persian." *Glossa.*

Futrell, Richard, Roger Levy, & Edward Gibson. 2020. "Dependency locality as an explanatory principle for word order" *Language* 96(2). 71-412.

Futrell, Richard, Kyle Mahowald & Edward Gibson. 2015. “Large-scale evidence of dependency length minimization in 37 languages”. *Proceedings of the National Academy of Sciences* 112(33).
Gibson, Edward. 1998. “Linguistic complexity: Locality of syntactic dependencies.” *Cognition* 68. 1-76.
Gibson, Edward. 2000. “The dependency locality theory: A distance-based theory of linguistic complexity.” *Image, Language, Brain*, pp. 95-126. Cambridge, MA. MIT Press.
Gildea, Daniel & David Temperley. 2010. “Do grammars minimize dependency length?” *Cognitive Science* 34(2). 286–310.
Goldin-Meadow, Susan., Wing Chee So, Asli 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. 9163–9168.
Green, Georgia. 1974. *Semantics and Syntactic Regularity*. Bloomington. Indiana University Press.
Greenberg, Joseph. 1963. “Some universals of grammar with particular reference to the order of meaningful elements” in J. Greenberg (ed) *Universals of Language*, pp. 73-113. Cambridge, MA. MIT Press.
Hammarström, Harald. 2016. “Linguistic diversity and language evolution.” *Journal of Language Evolution.* 1(1). 19–29.
Hahn, Michael, Dan Jurafsky, & Richard Futrell. 2020. “Universals of word order reflect optimization of grammars for efficient communication.” *Proceedings of the National Academy of Sciences* 117(5).
Hawkins John A. 1994. *A Performance Theory of Order and Constituency*. Cambridge, UK. Cambridge University Press.
Hawkins, John A. 2003. “Efficiency and Complexity in Grammars: Three General Principles.” In J. Moore and M. Polinsky (eds) *The Nature of Explanation in Linguistic Theory*, pp. 121-152. Stanford. CSLI Publications
Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford. Oxford University Press.
Hawkins, John A. 2014. *Cross-Linguistic Variation and Efficiency*. Oxford. Oxford University Press.
Kayne, Richard. 1995. *The Antisymmetry of Syntax*. Cambridge MA. MIT Press.
Keenan, Edward. 1976. “Towards a universal definition of "subject". In Charles Li (ed) *Subject and Topic*, pp. 303–333. Cambridge, MA. Academic Press.
Keenan, Edward. 1979. “On surface form and logical form.” *Studies in the Linguistic Sciences* 8(2). 163–203.
Liu Haitao, Xu Chunshan, & Liang Junying. 2017. “Dependency distance: a new perspective on syntactic patterns in natural languages” *Physics of Life Reviews* 21. 171-193.
Lohse, Barbara, John Hawkins, & Thomas Wasow. 2004. “Domain Minimization in English Verb-Particle Constructions” *Language* 80(2). 238-261.
Medeiros, David J., Paul Mains, & Kevin B. McGowan. 2020. “Ceiling Effects on Weight in Heavy NP Shift” accepted for publication in *Linguistic Inquiry*, available online at https://doi.org/10.1162/ling_a_00382
Melnick, Robin. 2017. *Consistency in variation: On the provenance of end-weight*. Stanford University Ph.D. thesis. https://www.academia.edu/39400150/Consistency_in_variation_On_the_provenance_of_end-weight
Oehrle, Richard T. 1976. *The Grammatical Status of the English Dative Alternation*. MIT Ph.D. thesis. https://pure.mpg.de/rest/items/item_405081/component/file_405080/content
Simpson, Jane. 2007. “Expressing pragmatic constraints on word order in Warlpiri.” In: A. Zaenen, J. Simpson, C. Manning, J. Grimshaw (eds.) *Festschrift for Joan Bresnan*, pp. 403–427. Stanford. CSLI Publications.
Song, Jae Jung. 2012. *Word Order*. Cambridge, UK. Cambridge University Press.
Tomlin, Russell. 1986. *Basic Word Order*. London. Routledge.
Wasow, Thomas. 2002. *Postverbal Behavior*. Stanford. CSLI Publications.
Wasow, Thomas. 2009. “Gradient Data and Gradient Grammars” *Proceedings of the 43rd Annual Meeting of the Chicago Linguistics Society*, pp. 255-271.

Yamashita, Hiroko & Franklin Chang. 2001. "'Long before short' preference in the production of a head-final language". *Cognition* 81(2). B45-B55

Thomas Wasow
wasow@stanford.edu
Department of Linguistics
Stanford University
Stanford, CA 94305-2150
USA