Universals of Left-right Asymmetry And Their Relevance for Theories of Online Processing

Abstract

There are numerous patterns of left-right asymmetry in grammatical universals, i.e. systematic preferences for A before B rather than the reverse, which compete with the more symmetrical patterns of the Greenbergian VO vs OV universals, i.e. A+B and B+A. These asymmetries are of relevance for psycholinguists and computational linguists because language processing takes place word by word, and here the grammars of the world's languages are systematically placing some items before others regardless of general language type. Speakers and listeners are accordingly processing certain items first. We see left-right asymmetries particularly clearly whenever one linguistic category is strongly dependent on another, like a gap on a moved WH filler, or a reflexive anaphor on its antecedent, or a narrow-scope quantifier on a wide-scope one, or a predication on its topic. In all these cases the dependent category prefers a later positioning in the string and the one on which it depends an earlier one. In order to explain these asymmetries the present paper builds on the principle of Maximize Online Processing proposed by Hawkins (2002) and compares it with two other theories that have been developed for processing in general, including for linear ordering phenomena, Working Memory Reduction (Gibson 1998) and Surprisal Theory (Levy 2008). Maximize Online Processing claims that full and correct processing of each item as it is encountered is what drives the asymmetries. The other theories assume instead that online prediction is a, or the, major force in reducing processing load and determining efficiency among competing structures, though they differ over whether they prefer predicting before predicted words, or the reverse. The present paper shows that there is no consistent relationship between online prediction and asymmetrical orderings of A+B: sometimes A predicts certain aspects of B and not vice versa, sometimes B predicts A and not vice versa, sometimes neither is predictive. In this area of language, therefore, online prediction is not the major or all-important determinant of linear ordering that it has been claimed to be. The asymmetries are better accounted for by Maximize Online Processing, an integration-based rather than a prediction-based preference.

1. Introduction

There are numerous patterns of left-right asymmetry in grammatical universals that compete with the more symmetrical patterns of the Greenbergian VO vs OV universals (Greenberg 1963, Dryer 1992). We see asymmetrical ordering especially clearly whenever one linguistic category is asymmetrically dependent on another, e.g.:

(1) a gap on a moved WH filler:
Who$_i$ did you go to the movies with 0$_i$?
(2) a reflexive anaphor on its antecedent:
Mary$_i$ was very pleased with herself$_i$
(3) a narrow-scope quantifier on a wide-scope one:
Everyone was reading a linguistics book
(4) a predication on its topic:
As for John, things are not going well

In such cases the dependent category B generally, or sometimes invariably, stands to the right of the independent category A and is processed after it, i.e. A+B.

Left-right asymmetries have received less attention in psycholinguistic and computational research compared with the Greenbergian correlations. These latter have been explained in terms of efficiency and ease of processing, e.g. as a result of minimal domains of phrase structure or head-dependent processing (Hawkins 1994, 2004, 2014, Futrell et al. 2015, Futrell et al. 2020).[(1)] Minimal domains can be realized in both head-initial (VO) and head-final (OV) languages, with their mirror-image orderings, and the quantities of these two language types are roughly equal across the globe (Dryer 2013a, see fn.8 below). There is also evidence for domain minimization within the asymmetric dependencies as well, e.g. in shorter distances between filler and gap, and between anaphor and antecedent (Hawkins 2004). But domain minimization does not explain the asymmetry itself, i.e. it does not explain the preference for A before B in these structures.

The question to be addressed in this paper is: What do these universals of left-right asymmetry tell us about online processing? Processing takes place word-by-word, so if the world's languages are systematically placing some items before others and having speakers and listeners process them first, then prima facie there should be a good processing reason for it.

The answer given here builds on a principle of processing and grammar that was first proposed in Hawkins (2002, 2004): Maximize Online Processing. Grammatical and lexical properties are preferably assigned to each item X in online processing as completely and as accurately as possible as each X is encountered (see (5) below for the definition). The independent category A that is more completely and correctly processable online precedes the dependent one, B, in these left-right asymmetries, while B is positioned later where it too can be immediately and fully and correctly interpreted at the time of its processing, but only by looking back to A on which it is dependent for at least some of its property assignments. If the ordering is reversed, B+A, the processing of the dependent B will be delayed and incomplete and possibly incorrect at the time it is encountered (with so-called "unassignments" of properties and possibly "misassignments").

In other words, efficient processability at the time these items are encountered, both for the independently referring ones occurring early and for the dependent ones that follow, is what drives the asymmetry in ordering.

In what follows, I first explain (§2) why conventions of grammars should be relevant at all for considerations of processing ease and efficiency in language use. I then define Maximize Online Processing (§3), and give eight asymmetric orderings in grammars showing how this principle accounts for them (§4). The next section (§5) compares Maximize Online Processing with two other processing theories, both of which assume instead that online prediction is a, or the, major determinant of processing ease and efficiency. I show that these theories make mutually contradictory, and often incorrect, predictions for left-right asymmetries. Section 6 concludes with some suggestions regarding the role of online prediction in an overall theory of processing ease and efficiency and outlines key issues for further research.

2. The relevance of grammars for theories of processing and efficiency

Many early proposals for communicative efficiency were first formulated by linguists working in language typology, on the basis of patterns in the grammars of the world's languages. This inference from grammars to hypothesized preferences in performance was often tested on small samples of manually collected usage data. For example, Greenberg (1966) proposed morphological hierarchies such as Singular > Plural > Dual > Paucal for number marking on nouns, and Nominative > Accusative > Dative > Other for case marking, based on the cross-linguistic distribution of these morphemes and on patterns of allomorphy (with declining allomorphy down the hierarchies, and more zero expression for properties at the top end of these hierarchies and increasing phonological complexity in lower positions). In order to explain these patterns in grammars he pointed to correlations with (declining) frequencies of usage within languages (e.g. Singular nouns in a corpus of Sanskrit = 70.3%, Plural = 25.1%, Dual = 4.6%). These performance-grammar correspondences led to an efficiency principle of Minimize Forms in Hawkins (2004, 2014) and to a similar formulation with even more extensive cross-linguistic performance-grammar support in Haspelmath (2021).

Greenberg's (1963) word order universals and his proposed correlations between head orderings in grammars (whereby VO languages prefer Prepositions before NP, OV languages prefer Postpositions after NP, etc, cf. Dryer 1992) led to a processing efficiency explanation in Hawkins (1990), before performance data were actually collected involving word order choices among alternatives within typologically different languages and presented in Hawkins (1994). The preferred selections within languages corroborated the central idea derived from the variation patterns across grammars. This idea was called Early Immediate Constituents in Hawkins (1990, 1994) and was subsequently generalized into Minimize Domains in Hawkins (2004, 2014) (see again fn.1). This same minimal distance idea between words that are grammatically connected to one another has now been supported in large cross-linguistic electronic databases and captured using a Dependency Grammar framework as Dependency Length Minimization (Futrell et al. 2015, Futrell et al. 2020).

These examples show that it has long been recognized in language typology that considerations of processing ease and efficiency have had a profound impact on grammars and on their

evolution and are visible in the typological variation we see today. This was captured as the Performance-Grammar Correspondence Hypothesis in Hawkins (2004, 2014):[2]

> Performance-Grammar Correspondence Hypothesis (PGCH)
> Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by patterns of selection in corpora and by ease of processing in psycholinguistic experiments.

It is because of the PGCH that a comparison of many grammars can reveal the efficiencies and performance preferences that led to their conventionalization and suggest principles for actual testing on corpus data and in experiments on different languages. In the present context with its focus on asymmetric A+B patterns the correspondences between grammars and performance are particularly close since grammars generally make available only one order for speakers to choose from in performance. I will argue that these asymmetries suggest a principle of efficiency rather different from that which lies at the heart of some current processing theories. At issue will be the role and significance of online prediction in efficient processing.

3. Maximize Online Processing

This principle is defined in (5):[3]

(5) **Maximize On-line Processing (MaOP)**
> The human processor prefers to assign the full set of properties to each item X as X is processed and to minimize the number of properties that are unassigned or misassigned to it in the online parse. Alternative orderings for a given X in competing structures will be preferred that avoid these unassignments and misassignments.

Unassignments involve the number of words and phrases that undergo some temporary unassignment of properties online, the number of any mother-daughter attachments temporarily unassignable, and the number of relations of combination or dependency temporarily unassignable. For example, antecedents preceding reflexive anaphors, as in languages with basic word orders shown in (6), are preferred to reflexive before antecedent as in (7):

(6) John$_i$ washed himself$_i$; Washed John$_i$ himself$_i$
(7) Washed himself$_i$ John$_i$

In (6) the antecedent *John* occurs first and is fully processable as is, and the anaphor *himself* can then immediately receive its co-index and semantic coreference when it is encountered by looking back to *John$_i$*. In the reverse order (7) the anaphor occurs first and its complete processing is delayed until *John$_i$* occurs, whereupon a co-index and semantic coreference can be assigned retrospectively to *himself* by a look-back operation.

Hence two fewer properties (syntactic co-indexing and semantic coreference) are assignable immediately to word B as it is encountered in the order B+A, and two more to A, compared with the sequence A+B in which all of A's properties are assigned at A, while all of B's properties can also be assigned at B, in part by looking back at A.

In order to make this parsing difference between (6) and (7) more explicit let us focus on just twelve of the grammatical and lexical properties that need to be assigned in each of these sentences, four to each word, from among their many syntactic, lexical, morphological and phonological properties. For the verb-initial ordering in (6) these properties would be assigned word by word in the manner shown in (6'):[4]

(6')	Washed	$John_i$	$himself_i$
	[Verb]	[Name]	[Reflexive Pronoun]
	[Construct Clause]	[SU of wash]	[DO of wash]
	[Activate Co-occurrences]	[Assign Index]	[Co-index to $John_i$]
	[Semantics for wash]	[Semantic Reference]	[Semantic co-reference to $John_i$]

When the ordering is reversed in (7) and himself is the second word, only two of the four properties assigned to it in (6') can be assigned at this stage in the parse, as seen in (7'):

(7')	Washed	himself
	[Verb]	[Reflexive Pronoun]
	[Construct Clause]	[DO of wash]
	[Activate Co-occurrences]	
	[Semantics for wash]	

When the third word is reached, $John_i$, the conditions will have been met for assigning the two remaining properties to himself by a look-back operation, shown in (7"):

(7")	Washed	$himself_i$	$John_i$
	[Verb]	[Reflexive Pronoun]	[Name]
	[Construct Clause]	[DO of wash]	[SU of wash]
	[Activate Co-occurrences]		[Assign Index]
	[Semantics for wash]		[Semantic Reference]
			[Co-index $himself_i$ to $John_i$]
			[Semantic co-reference of $himself_i$ to $John_i$]

In other words, two fewer properties are assigned to himself in the online parse of (7") compared with (6') where it occurs in the third position following its antecedent John. Property assignments to the first word washed, meanwhile, are identical in (6') and (7')/(7"), and property assignments to the third word $John_i$ in (7") contain the same properties as in (6'), plus the two extra ones assigned retrospectively to the anaphor. Between the two competing orders (6) and (7), therefore, one is characterized by incomplete property assignments to the dependent category himself at the point when it is parsed, (7'), the other (6') avoids these online unassignments and is accordingly preferred. Numerically, the online assignment of these properties to (6) and (7) can be shown in (8) and (9) respectively.

(8)	Word 1	Word 2 = A	Word 3 = B
Properties Assigned:	4	4	4

(9)	Word 1	Word 2 = B	Word 3 = A
Properties Assigned:	4	2	6

Misassigned properties mentioned in the definition for Maximize Online Processing in (5) are those that are both unassigned online and simultaneously a wrong property, such as a wrong mother-daughter attachment or relation of combination or dependency, is misassigned that needs to be corrected subsequently, i.e. there is what is traditionally called a garden path. Predictions for the severity of these can be made by the grammatical criteria given in Hawkins (2004: 51-61). For example, the famous garden-path sentence *The horse raced past the barn fell* involves far worse unassignments and misassignments by these criteria compared to *The horse that was raced past the barn fell*, than *I believe John is a fool* does compared to *I believe that John is a fool*.

The essential idea underlying Maximize Online Processing is that complete and correct processing of each item as soon as it is encountered is preferred over partial processing (with unassignments) or erroneous processing (with misassignments) requiring backtracking, reconstruction of the tree, and so on. When a particular category B can receive all of its properties completely and correctly at the time that it is encountered in one ordering but not in another, on account of its required access to A, the ordering with complete and correct property assignments will be preferred. In the next section we will see that this same preference for completeness and correctness in online parsing is what underlies a large number of left-right asymmetries, eight of which are illustrated here.

4. Eight Left-Right Asymmetries

4.1 WH-first before its Gap

English is one of a large number of languages that has a WH-fronting rule and in which a sentence like (10a) is grammatical while (10b), with WH-postposing, is not:

(10a) What$_i$ did you see a boy stealing 0_i in the grocery store last night?
(10b) *You saw a boy stealing 0_i in the grocery store last night what$_i$?

In the cross-linguistic database of the World Atlas of Language Structures (WALS) there are some 29% (212/730) of languages (Dryer 2005) that have a WH-fronting rule as in (10a). The remainder have either WH in situ, in which WH remains in the normal position that would be assigned to it by grammatical rules that apply to non-WH elements, or else there is some attachment or adjacency of WH to V, especially in SOV languages. This movement to the left, as opposed to the right, creates a "filler before gap" structure (Hawkins 2014: 172-176) and is almost exceptionless. Polinsky (2002) cites only American Sign Language and Tibeto-Burman Meithei as WH-last languages of which she is aware, i.e. just 2 versus the 212 with WH-first in WALS.

There are numerous processing considerations that have long been known to make gap-filling difficult (J.D. Fodor 1978, 1984, 1989, Frazier et al. 1989). The cross-linguistic asymmetry in WH+Gap vs Gap+WH ordering is striking and I attribute it to Maximize Online Processing. In (10a) the filler can be processed at the first word, and a gap is then postulated as soon as it can

be (by the Active Filler Hypothesis of Clifton & Frazier 1989), in a way compatible with the co-occurrence requirements of the grammar and lexicon, here right after *stealing* as its direct object. Both WH and the gap are processed as soon as they are encountered, therefore. In (10b), by contrast, where the gap precedes the filler, there is no filler to activate the gap, and the gap will be regularly unassigned, and possibly even misassigned online in languages that allow productive deletions of arguments of verbs and their recovery through contextual cues (cf. Gilligan 1987).(5)

We can illustrate the online processing advantages for (10a) over (10b) with a sample of property assignments to the key words $\underline{\text{what}_i}$ and stealing in the two orderings. For (10a) these are shown in (10a') (see again fn.3 for property labels shown here):

(10a')	$\underline{\text{What}_i}$	did you see a boy	stealing O_i ….
	[WH Pronoun]		[Verb + ing]
	[Assign Index]		[Construct VP]
	[WH-question Semantics]		[Activate Co-occurrences of Verb]
			[Assign SU to a boy]
			[Assign DO Gap after stealing]
			[Co-index Gap to $\underline{\text{what}_i}$]
			[Construe $\underline{\text{what}_i}$ as DO]

It is crucially the verb stealing that performs an integrating function, by activating its co-occurrence options from the lexicon (see Melinger, Pechmann & Pappert 2009 for the role of the verb in parsing), so assigning the preceding a boy to its subject (SU) slot, postulating a direct object (DO) gap site and filling that gap through co-indexation with the fronted $\underline{\text{what}_i}$, which is then interpreted as the DO of stealing. For (10b) fewer properties would be assignable to stealing at the point when it is parsed, as shown in (10b'):

(10b')	You	saw	a boy	stealing
				[Verb + ing]
				[Construct VP]
				[Activate Co-occurrences of Verb]
				[Assign SU to a boy]

The different co-occurrence possibilities for the verb steal can be activated at this point, but the parser does not yet know whether stealing is intransitive, in which case no further arguments will be assigned to it beyond the subject a boy which has already been parsed, or transitive, in which case the parser must look ahead for a DO. This uncertainty is not resolved until a plausible DO is reached, at what, and a gap site co-indexed to $\underline{\text{what}_i}$ can then be assigned to the immediate right of stealing, as shown in (10b"):

(10b")	You	saw	a boy	stealing O_i ….		$\underline{\text{what}_i}$
				[Verb + ing]		[WH Pronoun]
				[Construct VP]		[Assign Index]
				[Activate Co-occurrences of Verb]		[WH-question semantics]

[Assign SU to a boy] [Assign DO Gap after stealing]
[Co-index Gap to what$_i$ after stealing]
[Construe what$_i$ as DO of stealing]

There is a delay in the assignment of the full properties to stealing at the time that it is encountered, therefore, when what is in clause-final position. Misassignments of referents introduced in the previous discourse are also possible in languages with productive argument deletions (Gilligan op.cit.). The fronting of WH is accordingly preferred.

Notice that even if the verb in structures such as (10ab) is unambiguously transitive (e.g. buying in You saw a boy buying …) rather than ambiguous like stealing, the parser would still need to search after the verb for the direct object and could only co-index the DO gap to what$_i$ once it reached this word at the end of the clause. Either way there are significant delays in parsing assignments to the verb when it is encountered in structures like (10b).

4.2 Head Noun before Relative Clause

There is a similar ordering asymmetry of a head noun as filler before its gap (or subcategorizer) in relative clauses such as the English (11a), as opposed to the gap before filler in (11b):

(11a) the book$_i$ [that the professor wrote O$_i$]
(11b) [the professor O$_i$ wrote that] book$_i$

Almost all languages with VO or head-initial syntax have N+Rel, i.e. N+Gap (Dryer 1992).[6] Among OV languages Hawkins (2014:148-153) shows, using data from WALS (Dryer 2005, Dryer with Gensler 2005), that Rel+N is found only in rigid SOV languages like Japanese and not in non-rigid ones like Persian. In fact, rigid SOV languages are evenly split between those like Japanese that are Rel+N on the one hand and those on the other hand that have only N+Rel or both orders or certain other variants which avoid the Gap+N structure.[7] The distribution in SOV languages is shown in (12):

(12)

WALS	Rel+N	N+Rel/Both/Other
Rigid SOV	50% (17)	50% (17)
Non-rigid SOV	0%	100% (17)

Overall, therefore, of roughly half the world's languages with VO (Dryer 2013a) almost all have N+Rel and N+Gap. Plus two thirds of OV languages have N+Rel in (12), either as the exclusive order or in combination with Rel+N, or they have alternative structures that avoid Gap+N, which means that there is a clear asymmetrical preference for N+Gap across languages in relative clause structures.

I attribute this preference to Maximize Online Processing. In the English (11a) the filler head noun precedes the gap and can be processed as is, and the gap itself can then be immediately activated and processed by Clifton & Frazier's (1989) Active Filler Hypothesis once a relative clause structure has been activated and the subcategorizer wrote has been reached. This is illustrated in (11a') for a sample of property assignments to the filler and subcategorizer in (11a):

(11a')	the	book$_i$	that the professor	wrote O_i
		[Noun]		[Verb]
		[Construct NP]		[Construct VP]
		[Assign Index]		[Activate Co-occurrences of Verb]
				[Assign SU to the professor]
				[Assign DO Gap after wrote]
				[Co-index Gap to book$_i$]
				[Construe book$_i$ as DO]

In the reverse Gap+N relative clause of an SOV language given in (11b) the parser can activate the verb's cooccurrences at the point when wrote is encountered, it can assign *the professor* to this verb as a subject, and in the event that the verb is transitive it can even postulate a gap as direct object to the left of *wrote*, since verbs are final in the clause, as shown in (11b'):

(11b')	the professor	O	wrote
			[Verb]
			[Construct VP]
			[Activate Co-occurrences of Verb]
			[Assign SU to the professor]
			[Assign DO Gap before wrote]

But only when the filler ***book$_i$*** is reached can the remaining properties be assigned retrospectively to *wrote*, namely co-indexing the pre-verbal object gap to book$_i$ and construing it as the direct object of *wrote*, as shown in (11b"):

(11b")	the professor	O_i	wrote	that	book$_i$
			[Verb]		[Noun]
			[Construct VP]		[Construct NP]
			[Activate Co-occurrences		[Assign Index]
			of Verb]		[Co-index Gap before wrote to book$_i$]
			[Assign SU to the professor]		[Construe book$_i$ as DO of wrote]
			[Assign DO Gap before wrote]		

The result is a delay in assigning all the properties to the verb wrote when it precedes the filler in (11b"), compared to the ordering in (11a') when the filler occurs first and wrote can be assigned all of the information relevant for its argument structure interpretation at the time that it is actually encountered. (11a) is accordingly preferred over (11b).

Misassignments have also been shown to arise in rigid SOV languages as a result of the delayed access to the head noun filler at the time that a subordinate verb within a relative clause is processed. A gap that is activated but unfilled when the subordinate verb is processed can be erroneously filled by an earlier NP in the parse string, as was shown for structures such as (13) in Japanese by Clancy et al. (1986):

(13)	Zoo-ga	[[0$_i$	kirin-o	taoshi-ta]	shika-o$_i$]	nade-ta

elephant-NOM Gap giraffe-ACC knocked down deer-ACC patted
'The elephant patted the deer that knocked down the giraffe.'

Zoo-ga was regularly first misrecognized as the nominative subject of the subordinate verb *taoshi-ta* within a simpler main clause parsing of *Zoo-ga kirin-o taoshi-ta* (with the meaning 'the elephant knocked down the giraffe') whereupon the words and phrases of this sentence were subsequently reanalyzed in accordance with the Rel+N structure shown in brackets. The on-line misassignments are quite severe in this example, by the criteria in Hawkins (2004:51-61,205-210). In the event that the first two NPs encountered in the online parse do not match the co-occurrence requirements of the first verb, *taoshi-ta,* then the phrases and relations that involve misassignments in (13) will be temporarily unassigned instead. Either way, unassignments and/or misassignments are extensive for Rel+N structures, and it is to this that I attribute their limited distribution typologically.

Because of these parsing inefficiencies, Rel+N is predicted to occur in this processing approach to language typology built on the Performance-Grammar Correspondence Hypothesis (cf. §2) only in those languages in which there is a competing motivation for head-finality in the noun phrase. This competing motivation comes from the syntactic preference for Cross-Category Harmony (Hawkins 1983), which is ultimately explainable by Minimize Domains (Hawkins 2004) (recall fn.1): the more head-finality there is in a language, and specifically in phrases that contain noun phrases such as VP and PP whose processing can benefit from the adjacency of a head-final noun to head-final verbs and adpositions, the more a language will favor noun-finality and resist the independent, and overall empirically stronger, preference for N+Rel and N+Gap stemming from Maximize Online Processing.

4.3 Antecedent before Reflexive Anaphor

This asymmetry and the motivation for it in terms of Maximize Online Processing was exemplified for (6)-(9) above. Empirically the distribution of antecedents and reflexive anaphors is highly correlated with the asymmetric preference for subjects before objects across languages (see §4.4 below), since antecedent and reflexive anaphor are typically subject and object (or subject plus other non-subject) respectively, with the non-subject being asymmetrically dependent on the subject syntactically and semantically and c-commanded by it (Reinhart 1983). Hence this asymmetrical ordering is motivated both by general subject before object considerations and by the co-indexing and coreference property assignments mentioned above.

The quantities of languages exemplifying antecedent before reflexive anaphor will be at least those that have subject before object, therefore, and exceptions with object reflexives preceding subjects should generally be limited to languages like Malagasy that have a strong basic VOS word order (Keenan 1976a), see (14):

(14) *Manaja* *tena*$_i$ *Rabe*$_i$
respect self Rabe
'Rabe respects himself.'

4.4 Subjects before Objects

According to Tomlin (1986) some 96% of languages had subject before object in his 402 language sample, as shown in (15):[8]

(15)

SOV (168)				OVS (5)
	> VSO (37)	> VOS (12)	>	
SVO (180)				OSV (0)
87%	9%	3%		1%

This asymmetry becomes even more compelling when basic grammatical relations are broken down, more precisely, into their component properties of morphological marking, syntactic configuration, and semantics (theta-roles), as Primus (1999) does in her detailed and typologically well-informed study. She proposes the following hierarchies in which lower positions are asymmetrically dependent on each higher position:

(16) Case Morphology: Nominative > Accusative > Dative > Other
Absolutive > Ergative > Dative > Other

(17) Syntax: higher structural position (c-commanding) > lower position (c-commanded)

(18) Semantics (theta-roles): Agent > Recipient > Patient
Experiencer > Stimulus

Each of these hierarchies is then preferably linearized with higher positions occurring to the left of each lower position, with particularly strong linear asymmetries resulting when multiple hierarchies reinforce one another in a given language type, and with greater variation when hierarchies conflict. So a significant number of languages that have been (mis)classified as O+S, she argues, are those with ergative-absolutive case morphology, and these often position the absolutive first, in accordance with the case hierarchy (16). The preference for S+O is accordingly stronger in nominative-accusative languages in which all higher hierarchy positions can align, nominative, c-commanding configuration and agent, than in ergative languages in which absolutive is the highest case position in (16). This has been confirmed in the figures given in Hawkins (2022), taken from WALS (Comrie 2013) and shown in (19) and (20), which compare the orderings of subject and object in languages with nominative-accusative vs ergative-absolutive case marking. The nominative-accusative languages have a higher proportion of S+O basic orders (88%) as opposed to O+S and no dominant order, compared with ergative-absolutive languages (67%).

(19)	S_{Nom} + O_{Acc}	44/50 languages	(88%)
	O_{Acc} + S_{Nom}	2/50 (4%)	
	No dominant order	4/50 (8%)	
(20)	S_{Erg} + O_{Abs}	18/27 languages	(67%)
	O_{Abs} + S_{Erg}	2/27 (7%)	
	No dominant order	7/27 (26%)	

The asymmetric dependencies in (16)-(18) have different ultimate causalities underlying the asymmetry. In the morphology (16), for example, an accusative requires a co-occurring nominative but not necessarily vice versa, an ergative a co-occurring absolutive but not necessarily vice versa (see §5.1 and fnn.11&12 for exemplifying languages and data). In the syntax (17) a lower c-commanded position in the tree requires a higher c-commanding one (Reinhart 1983). For the

semantics (18), for example, you cannot have a patient without an agent semantically and conceptually (Jackendoff 1972, Dowty 1991).(9)

In all these cases the preferred linearizations position the A category first and the dependent B later. In this way, Primus (op. cit.) argues, precise details of B's interpretation and syntax can be assigned immediately by reference to a prior A, and without unassignments or misassignments, in accordance with Maximize Online Processing. There will be no attempt to assign a c-commanded object into a tree structure when the higher c-commanding subject has not yet been constructed online, and the theta-role of the second argument can be more precisely and immediately interpreted by looking back to the first. Interestingly, the asymmetry in ordering is least strong for the morphological hierarchies (16). In these hierarchies if the cases can be clearly recognized in their surface morphology and a unique value assigned to them at the time they are processed, then there will be no unassignments or misassignments in whatever order they occur in the parse string. This may explain why the linearization preferences of the syntactic (17) and semantic hierarchies (18) are stronger and generally override the case hierarchy, in favor of agent first and c-commanding position first, thus placing a morphologically ergative case first in 67% of languages in (20), even though absolutive is higher on the case hierarchy in (16) (see again §5.1 and fnn.11 and 12).

This preference for subjects before objects is is then further reinforced by the many additional properties and dependencies that correlate with subject and object in transitive clauses, and which provide additional online processing motivations for the A+B ordering through Maximize Online Processing. Keenan (1976b) mentions Independent Existence (the subject does not depend on a transitive verb for its reference), Autonomous Reference (the subject does not depend on other NPs in the clause, i.e. no **Heself loves John*), High Referentiality (the subject attracts highly referential NPs, like definite NPs only in Malagasy, Tagalog, and Bantu languages), Wide Scope (subjects have wider scope, see §4.6 below), and Topicality (subjects are regularly topics, see §4.5). In all these cases the transitive subject has referential properties that are quite independent of the object and can be fully and immediately processed at the outset, or indeed anywhere in the string, whereas in many of them the dependent object can only be fully processed by referring back to an already processed subject with the result that unassignments and misassignments would result in the reversed order.

4.5 Topic before Comment

In languages with clear topic positions, especially those that have explicit topic marking and "sentence-external topics" and are topic-prominent like Chinese (Li & Thompson 1981), there is a strong preference for the topic to precede the comment. This can sometimes be reversed into Comment + Topic, and there are numerous grammatical and pragmatic intricacies that are relevant here, as has been shown by Lambrecht (1994), Polinsky (2002), and others. There are also correlations between topic status and givenness, definiteness and genericness. This is illustrated for Japanese in (21) (from Kuno 1973: 59), in which an indefinite topic is ungrammatical and must instead be either definite or generic:

(21) *Oozei no hito wa party ni kimasita
many people TOP party to came
'Many people came to the party'

Primus (1999) and Hawkins (2004:235-240) have argued that there is no purely pragmatic and information structure explanation that is convincing for why topics can be generic, but not indefinite, and for the preferred ordering. Instead they argue that there are numerous asymmetrical dependencies between them, and it is these and the unassignments and misassignments of properties in the reverse ordering that explain the left-right asymmetry. Gundel (1988:210) captures this asymmetry in dependency when she states that the Comment predication is "assessed relative to the topic". Reinhart (1982) talks of its "aboutness". Hawkins (op.cit.) gives numerous examples from Mandarin Chinese, taken from Tsao (1978), showing that the Comment depends asymmetrically on the Topic for argument assignments and for argument and predicate enrichments in examples such as (22)-(24), given here in their literal renderings into English:

(22) Jang San (Topic Particle), yesterday came see me. I.e yesterday he came to see me.
(23) This man (Topic Particle), mind simple. I.e. this man's mind is simple.
(24) Fish (Topic particle), tuna is now the most expensive. I.e. tuna is now the most expensive thing relative to fish.

If the Comment occurred first and the Topic last, there would be regular unassignments and misassignments to the Comment. The Topic can be fully processed in initial, or indeed in any position, as long as it is definite and mutually identifiable by speaker and hearer and cognitively accessible (Levelt 1989:260), or else generic and universal when the speaker refers to all potential referents of a noun as opposed to some unspecified and undetermined subset that requires contextual delimitation. This explains why Topics regularly have these properties across languages. The Comment needs access to the Topic for its full online processing, however, and this explains why it regularly follows the Topic.

Compare (22a) and (22b):

(22a) Jang San (Topic Particle), yesterday came see me.
(22b) Yesterday came see me, Jang San (Topic Particle).

A subset of relevant property assignments made to key words in the Topic-first ordering is shown in (22a'):

(22a') Jang San$_i$ (Topic Particle) yesterday O_i came see me

Jang San$_i$	yesterday O_i	came	see	me
[Name]		[Verb]		
[Assign Index]		[Activate Co-occurrences of Verb]		
[Semantic Reference]		[Assign SU Gap to left of came]		
		[Co-index Gap to Jang San$_i$]		
		[Construe Jang San$_i$ as SU]		

In this ordering the verb *came* can be fully processed as soon as it is encountered. The parser looks back to the Topic *Jang San$_i$,* assigns a subject gap to the left of came, co-indexes this gap to the topic (or co-indexes the subcategorizing verb directly to it in theories that dispense with gaps, cf. fn.5), and construes the topic as its subject. In the reverse ordering of (22b), a subject

gap can be assigned to the left of came, but no co-indexation or subject construal can be made to the topic since this has not yet been encountered. In other words, the hearer does not know at this point who the subject is.

(22b') yesterday O came
[Verb]
[Activate Co-occurrences of Verb]
[Assign SU Gap to left]

The subsequent Topic provides the subject argument and assigns it retrospectively to *came*, through co-indexation and construal, as shown in (22b"):

(22b") yesterday O_i came see me Jang San_i (Topic Particle)
[Verb] [Name]
[Activate Co-occurrences] [Assign Index]
[Assign SU Gap to left] [Semantic Reference]
[Co-index Jang San_i to Gap before came]
[Construe Jang San_i as SU of came]

For (23), in which the Topic serves to enrich the subject argument in the Comment, if the ordering in (23a) were reversed to (23b) it would be unknown when the Comment was being processed which person and whose mind the speaker was talking about, and this aspect of the argument in the Comment would be unprocessable until this was clarified.

(23a) This man (Topic Particle), mind simple.
(23b) Mind simple, this man (Topic Particle).

(23a') shows that the property 'mind of Gap' (i.e. the person whose mind is being referred to) can be activated when *mind* is processed, and this gap is immediately co-indexed to man_i and a possession semantics is assigned to *mind of* man_i.

(23a') This man_i (Topic Particle) mind of Oi simple
[Noun] [Noun]
[Assign Index] [Semantic Reference]
[Semantic Reference] [Activate mind of Gap]
[Co-index Gap to man_i]
[Construe possession semantics for mind of man_i]

In the reverse ordering 'mind of Gap' is activated first in (23b'), and the co-indexation to man_i and construal of the possession semantics for *mind of* man_i must then wait until the Topic is reached, as shown in (23b"):

(23b') mind of O simple
[Noun]
[Semantic Reference]
Activate mind of Gap]

(23b") mind of Oi simple this man$_i$ (Topic Particle)
[Noun] [Noun]
[Semantic Reference] [Assign Index]
[Activate mind of Gap] [Semantic Reference]
[Co-index Gap to man$_i$]
[Construe possession semantics for mind of man$_i$]

For the example in (24) it is the superlative predicate in the Comment *most expensive* that is enriched by referring back to the Topic. In the Topic-first ordering (24a) the semantics of the superlative 'tuna is more expensive than all other X' is appropriately constrained and relativized to the semantic content of the Topic, *fish*. In the reverse ordering (24b) the superlative semantics is not so constrained and an overly general interpretation can be assigned online, 'tuna is more expensive than everything else'. Only later in the parse is this misassignment recognizable and the comparison set 'all other X' limited to *fish.*

(24a) Fish (Topic Particle), tuna is now the most expensive.
(24b) Tuna is now the most expensive, fish (Topic Particle)

The assignment of key properties in the parsing sequence for (24a) is shown in (24a'):

(24a') Fish (Topic Particle) tuna is now the most expensive
[Noun] [Adjective Phrase]
[Semantic Reference] [Semantics of Superlative]
[Construe: Tuna is more expensive than all other X]
[Assign Topic fish to X]

For (24b) the online assignments are as shown in (24b') and (24b"):

(24b') tuna is now the most expensive
[Adjective Phrase]
[Semantics of Superlative]
[Construe: Tuna is more expensive than all other X]

(24b") tuna is now the most expensive fish (Topic Particle)
[Adjective Phrase] [Noun]
[Semantics of Superlative] [Semantic Reference]
[Construe: Tuna is more expensive than all other X [Assign Topic fish to X in superlative semantics for most expensive]
[Recognize earlier misassignment and reduce reference to X through Topic]

One particularly strong cross-linguistic tendency for topic positions, noted first by Gundel (1988) and by Lambrecht (1994) and Polinsky (2002), is the preference for topics that mark a new topic or that signal a topic shift to require initial position. This makes sense by Maximize Online Processing and can be explained by it. If there is a new Topic, and that Topic follows the Comment, then the dependencies inherent in the Comment will be regularly unassignable to it at the time the Comment is processed, and they can be readily misassigned to a previous Topic. In other words, the unassignments and misassignments would be particularly severe in the case of new topics, resulting in an even stricter Topic before Comment ordering.

4.6 Wide Scope before Narrow Scope Quantifiers & Operators

Across languages there appears to be a correlation, which has yet to be quantified but which is frequently observed both in the semantics literature (e.g. Horn 1989) and in cross-linguistic comparison (e.g. Kiss 2002), between semantic scope and linear ordering: the first quantifier/operator encountered in a clause generally has wide scope and the second narrow scope. Some languages like English permit more variability and allow interpretations in which the second quantifier/operator has wide scope (as in *All that glitters is not gold*, which is preferably interpreted as 'Not all that glitters is gold' despite the surface structure ordering). Other languages (e.g. Hungarian, cf. Kiss op.cit.) are much less tolerant of such reversals. Even in English, with its permitted variability, there is a strong tendency for the relative ordering of quantifiers/operators to iconically match their relative scope, as one major factor that interacts with others in determining semantic interpretations. These additional factors determining relative scope include the configurational position of a scope-bearing item in a clause, i.e. whether it is c-commanding or c-commanded (Reinhart 1983), and the inherent strength of a quantifier/operator (cf. Ioup 1975, Horn op.cit.). A sentence like (23), repeated from (3) above, is accordingly preferably interpreted as referring to a possibly different book for different people, with the indefinite NP *a linguistics book* having narrow scope and looking back to the initial *everyone*.

(23) Everyone was reading a linguistics book.

Passivization to *A linguistics book was read by everyone* is preferred in the event that the scope is to be reversed and reference made to one and the same book. This preference for scope to mirror the order of presentation is predicted by Maximize Online Processing since there can be frequent unassignments and semantic misassignments to the narrow scope element in the event that it is presented prior to the wide scope item on which the narrow scope interpretation ultimately depends, as in *All that glitters is not gold* where *not* has wide scope over the quantifier *all* in initial position.

4.7 Complementizers before Subordinate Clauses

The distribution of free-standing complementizers before or after a subordinate S has a very similar distribution to that of head noun and relative clause (see Hawkins 2014:153-158) and can be similarly explained by an interaction of Maximize Online Processing and Minimize Domains (recall fn.1). Dryer (2009) gives the following figures:

(25)	VO languages	Comp + S	74% (140)
		S + Comp	0% (0)

OV languages	Comp + S	12% (22)
	S + Comp	14% (27)

In a VO language an initial Comp before a subordinate clause provides a minimal domain for processing a verb and sentential object combination like *I believe that John is a smart guy* (the sentential object being projected from the complementizer by construction principles that are discussed at length in Hawkins 1994). The initial Comp also avoids unassignments and misassignments that arise in the absence of the complementizer (as in *I believe John is a smart guy*). A final Comp after S does not satisfy these desiderata in VO languages: it does not construct the subordinate S on its left; it does not bring about a minimal domain for phrase structure processing; and it does not avoid online unassignments and misassignments. It is striking that all of the VO languages in (25) have Comp + S, and none has S + Comp.

In OV languages the morphemes that correspond to free-standing complementizers in VO languages are regularly affixes (mostly suffixes) on verbs, and these affixes signal and construct the subordinate status of the clause, most often in clause-final position. For this reason there are far fewer free-standing complementizers in OV languages in (25) compared with those in VO (49 vs 140), just over half of the OV total (27) having harmonic final position, while the remainder (22) have initial position in OV languages and are predicted to occur in mainly non-rigid OV languages with numerous head-initial structures co-existing with OV (cf. Hawkins 2014: op.cit.), just as NRel occurs mainly in non-rigid SOV languages as shown in (12).

4.8 Restrictive precedes Appositive modifiers of N

Consider the relative ordering of restrictive and appositive relative clauses in a head-initial VO language like English:

(26) a. Students that major in mathematics, who must work very hard (R+A)
b.*Students, who must work very hard, that major in mathematics (A+R)

If (26b) were grammatical there would be regular semantic misassignments online with the appositive claim being predicated first of all students, and only later being reduced to the restricted subset of students that major in mathematics, which goes against Maximize Online Processing.[(10)]

5. Some Alternative Processing Ideas for Asymmetries

The principle of Maximize Online Processing, with its emphasis on completeness, correctness and more generally on speed and avoiding delays in online processing, was originally proposed in Hawkins (2002) in response to the basic insights and findings from psycholinguistic studies on parsing and production captured in foundational works such as Fodor, Bever & Garrett (1974) and Levelt (1989). Since then, with the advent of big data and computational techniques for measuring the frequencies of occurrence and co-occurrence for linguistic items, we have seen an increasing emphasis in psycholinguistics on the role of online prediction in measuring processing ease and difficulty. The key theoretical issue motivating these frequency-based studies is: to what extent does the occurrence of one item A lead to the expectation that another item B will follow later in the string, and how does this impact online processing load? Current

theories differ in the way they implement this idea, in the significance they attach to it within an overall theory of processing, and in the precise predictions they make. But it is safe to say that online prediction has now become one of the central determinants of processing ease and difficulty, and for some its central and even unique determinant (see §5.2). Universals of left-right asymmetry provide a revealing set of data in this context since the world's languages, and their users, are systematically placing some items before others.

The author of the present paper is not convinced that online prediction has this centrality or uniqueness. Instead, assigning the full set of properties to items as soon as they are encountered, assigning them correctly and integrating them with already processed items, appears to be more important. Online prediction about future items is certainly one contributor to overall processing load. There are numerous constructions and sequences of items that provide evidence for it, such as the binomials of Morgan & Levy (2016) (*bread and …*, *salt and …*, etc). Wasow, Jaeger & Orr (2011) provide evidence for degrees of online predictability correlating with the omission of the relativizer in (non-subject) English relative clauses: the more predictable a relative clause was following a prenominal determiner, adjective or head noun in their corpora, the more often the relativizer was omitted. Different language types can now also be usefully compared from the perspective of online prediction, e.g. in Engelhardt, Filipovic & Hawkins (in press) where these authors test whether the earlier position of the verb in English versus Japanese makes processing of the verb's argument structure faster and easier (they find that it does). And countless other studies, too numerous to mention, provide evidence for online prediction.

Online prediction clearly does have an important role to play in language processing, therefore, but the precise manner in which it is being implemented in different theories is not uniform and this has led to different preferences being defined. In the remainder of this section I compare two different implementations of online prediction for one of the data sets summarized in section 4, one that has been widely discussed and that both theories make clear predictions for, involving subject before object ordering (§4.4): Gibson's (1998) Working Memory Reduction proposal (§5.1) and Levy's (2008) Surprisal Theory (§5.2). I also make brief reference (in §5.1) to a third theory, Jaeger's (2010) Uniform Information Density theory, in which prediction plays a role in accounting for the relative ordering of subject and object. This comparison of theories with respect to one of the data sets considered here gives us some common ground for assessing how they deal, or would deal, with left-right asymmetries in general. We can then determine how online prediction fares in comparison with Maximize Online Processing (§5.3).

5.1 Working Memory Reduction

The Dependency Locality Theory of Gibson (1998, 2000) is a systematic attempt to quantify the complexity of online parsing operations and their demands on working memory. It comprises both an integration component and online prediction (see Ferreira & Chantavari 2018 for a more recent formulation of the respective roles of these two components). Gibson discusses the alternation between SVO and OVS orders in languages like German and Finnish and he argues that the clear preference for SVO can be attributed to online prediction and to the extra demands that OVS places on working memory. He writes (Gibson 1998: 59):

"OVS … orders are more complex at the initial nouns because it is necessary to retain the prediction of a subject noun at this location. … SVO sentences are expected to be more frequent than OVS …, because they require less memory to produce and comprehend."

Paraphrasing, language users avoid structures like OVS in which a category B, a direct object, predicts another A, a subject, and precedes it, thereby adding to working memory load. More generally, online prediction adds to the amount of material that needs to be held and processed simultaneously in working memory and is undesirable, Gibson argues.

The factual basis for this explanation is complicated by typological differences across languages in case marking. It is certainly true that an accusative predicts the co-occurrence of a nominative in the transitive clauses of German and Finnish, in accordance with the case morphology hierarchy of (16) above, and hence that an initial accusative would predict a following nominative. But working memory reduction is not a plausible general explanation for the nominative before accusative preference in German and Finnish language usage, and more generally in the 88% of grammars with nominative before accusative orderings shown in (19).

The reason is that languages with ergative-absolutive morphology also have a significant preference for ergative subjects before absolutive objects (67% in (20)), and yet in these languages ergative case is now the predictive category which predicts a following absolutive in the clause (see the case morphology hierarchy of (16)).[(11)] In other words, the working memory reduction theory, based on online prediction, lacks generality across languages.[(12)]

An alternative explanation to Gibson's for why accusative-first structures are dispreferred is Jaeger's Uniform Information Density theory, see Jaeger (2010), Maurits et al. (2010) and Clark et al. (2023). This theory claims that the predictiveness of the direct object for a subject results in an uneven spike in information at the beginning of the clause, which makes it non-uniform with the rest of the clause compared with other orders, especially SVO and VSO.

Again, the 67% of ergative subjects in (20) go counter to this. Plus, SOV is counterfactually predicted in this approach to be a dispreferred order! In the most recent WALS data (Dryer 2013a) SOV is actually the most common single type (SOV=48%, SVO=41%, VSO&VOS=10%, see fn.8).

5.2 Surprisal Theory

Levy (2008) sets out a general theory of syntactic comprehension according to which online prediction is the most important determinant of processing ease or difficulty. His theory is called Surprisal Theory, and its basic premise is summarized in his recent co-authored paper (Futrell, Levy & Gibson 2020) as follows:

"… the results from psycholinguistics indicat[e] that the bulk of language processing load comes from the degree to which linguistic elements such as words are unexpected in context. Surprisal theory … formalizes this idea and claims that *all processing difficulty* [italics, JAH] results from the extent to which elements are unexpected in context ..."

Paraphrasing, the less expected and predicted a word is, the larger its surprisal and the greater the processing load, while the more predicted it is, the less its surprisal, and the less processing load there is. More generally, online prediction is good for processing, and in fact all processing difficulty is claimed to correlate with the extent to which items are not predicted online.

It follows that if one category systematically predicts another in the left-right asymmetries of this paper, then placing the predicting one first should reduce the surprisal and processing load for the second and should be preferred. But this contradicts Gibson (1998): objects should now precede subjects by this logic in nominative-accusative languages since the subject in the OVS order is predicted and expected in context! But only 12% of the nominative-accusative languages in (19) have productive (preferred or unique) accusative object before nominative subject order, the vast majority (88%) have subject before object, with the non-predicting nominative first.

Empirically, Surprisal Theory would work well for the 67% of ergative subjects in (20) that have the ergative first (see fn.11), since the ergative then predicts the upcoming absolutive, but not for the nominative-accusative languages of (19) in which nominative-first is significantly preferred.

5.3 There is No Consistent Relation Between Online Prediction and Asymmetric Ordering

The theories we have considered assume that online prediction is **a**, or **the**, major force in reducing processing load and in favoring certain sequences of words, whether it be the predicting before the predicted or predicted before predicting. But left-right asymmetries of A+B cannot in general be explained by online prediction in either of these variants.

Sometimes A predicts certain aspects of B and not vice versa, as shown in (27), sometimes B predicts A and not vice versa, see (28), and sometimes neither is predictive, see (29):

(27) A predicts (certain aspects of) B and not vice versa
WH predicts there will be a gap in §4.1, but not necessarily where (Fodor 1984), cf. the Active Filler Hypothesis of Clifton & Frazier (1989)
Ergative subjects predict an absolutive object in §4.4 and fn.7
Topics (if clearly topic-marked) predict some upcoming Comment in §4.5
A complementizer predicts a subordinate clause in §4.7

(28) B predicts (certain aspects of) A and not vice versa
Reflexive anaphor predicts there will be an Antecedent in §4.3
Accusative objects predict a nominative subject in §4.4

(29) Neither A nor B predicts the other
A noun does not predict that it is going to be modified by a relative clause and contain a gap, nor is a co-indexed gap even recognizable in advance of its head noun filler, §4.2
A wide scope quantifier does not predict that a narrow scope one will follow, nor is a narrow scope quantifier recognizable as such and predictive in advance of a wide scope one, §4.6
A restrictive relative does not predict that an appositive will follow, nor does the latter predict the former, §4.8

There appears to be no consistent and general relationship between online prediction and asymmetrical left-right ordering, therefore. For the patterns in (27) online prediction may indeed be contributing to the preference for the favored orders from the perspective of Surprisal Theory. We referred in §4.1, for example, to the Active Filler Hypothesis of Clifton & Frazier (1989) whereby an initial fronted WH-item activates the expectation that there will a gap later in the string and so predicts it (although it does not predict exactly where or what it will be). Similarly a clearly-marked topic phrase in leftmost position predicts an upcoming comment. But at the same time the grammatically more specific and more constrained prediction made by accusative case marking in (28) to the effect that there will be a co-occurring nominative (as in Gibson's theory) is not matched by asymmetrical ordering in favor of the predicting item first (Gibson argues for the reverse, predicted first, as we have seen). Nor is there asymmetrical ordering for a predicting reflexive anaphor before its antecedent. For the pairs of categories in (29) no online predictions are made by either one for the co-occurrence of the other.

We can conclude that online prediction lacks generality as an explanation for cross-linguistic asymmetries in ordering, whether in the form of predicting before predicted or the reverse. By contrast, the fundamental determinants of processing ease and difficulty that underlie Maximize Online Processing, in terms of the completeness and correctness of property assignments online and the speed and efficiency of integrating items with the material that precedes them, do provide a consistent and general explanation. In each of (27)-(29), if the preferred A+B is reversed to B+A there will be incomplete or incorrect property assignments online in the form of unassignments or misassignments (recall §3).[(13)] For (27) online prediction can be claimed to contribute to processing ease according to Surprisal Theory, and for (28) Working Memory Reduction could be claimed to be relevant, but even here these theories cannot be held to be uniquely responsible for the data since unassignment and misassignment avoidance are also operative and neither accounts for (29) where online prediction is irrelevant. Whatever the precise role of online prediction is in preferring certain linearizations, therefore, it appears to be a weaker contributor to overall processing ease.[(14)]

The contradictions between the theories in §5.1 and §5.2 also raise some issues that have yet to be resolved in the theoretical psycholinguistic literature. For example, if predictability can aid processing, according to Surprisal Theory, why does it not also have the reverse effect, of greatly increasing the number of items held in working memory, so making processing more difficult, as Gibson (1998) originally argued? If a verb can activate numerous possible co-occurrence frames for its various arguments, why isn't the working memory of speakers of verb-early languages so clogged up with alternative syntactic and semantic possibilities in online processing, most of which will be discarded later in the parse string, that they make verb-argument processing domains very complex? Perhaps they do, and this is one of the benefits of having verb-last in SOV languages! But then why should predictability be such an advantage and how does it offset this putative disadvantage? More generally, how do we reconcile the claimed advantages or disadvantages of online predictability with the roughly equal distribution of VO to OV languages across the world, given that verbs are far more predictive of nouns than nouns are of verbs?

Hawkins (1994, 2004, 2014) argued that this distribution is expected when phrase structure and dependency domains are calculated in a more traditional way by looking at actually occurring items in the parse string and calculating the distances between their various heads of phrases,

without looking forward to activated and predicted co-occurrences of these heads, many or most of which will eventually be dismissed in the online parse. Putting this another way, the cross-linguistic distribution of VO to OV is expected when we calculate integration domains in processing in the sense of Gibson (1998, 2000) and Ferreira & Chantavari (2018), but not when we look at online predictions, according to which verb-early languages with their rich online predictions about verbal co-occurrences should either be favored (per Levy) or disfavored (for working memory load reduction, per Gibson). If integration and not prediction holds the key to the Greenbergian VO/OV symmetries (cf. Dryer 1992, Hawkins 1990, 1994), this would provide further support for the conclusion we have derived from left-right asymmetries that online prediction is not a strong determinant of linear ordering.

6. Conclusions

Asymmetric orderings in cross-linguistic grammatical conventions point systematically to an efficiency principle whereby each item is processed as completely and as correctly as it can be and as soon as it is encountered. This is why certain categories are regularly ordered A+B, especially those in which B is asymmetrically dependent on A. Reversing them to B+A would make the processing of B incomplete and possibly erroneous prior to A.

It is the Maximize Online Processing principle in (5) that receives general support from left-right asymmetries, not the other theories in §§5.1-2. The preferred A+B may or may not reduce Working Memory Load and it may or may not reduce Surprisal. Likewise the dispreferred B+A may or may not conform to these theories. They both assume a major role for online prediction in processing, yet we have seen that this results in contradictory preferences and provides no consistent and general explanation for left-right asymmetries. I conclude from this that online prediction is not the unique or even central factor that determines ease of processing and usage preferences, as revealed by these data in which the world's languages are systematically placing certain items A before others B in both grammatical conventions and in the language usage of speakers.

Much more important are: completeness (avoid look-ahead); correctness (avoid garden paths); and integration of items with preceding items at the time they are encountered in the parse string so as to achieve this completeness and correctness. On the other hand, there can be a role for online prediction in certain data sets that have been argued to support it (recall §5). It can play an additional role, per Surprisal Theory, in the asymmetries of (27), e.g. in predicting that there will be a gap downstream after a WH element. It may also play a role in those of (28) in the event that predicted items before predicting are preferred, as in Working Memory Reduction (though see fn.14). But whatever preferences are defined by these different online prediction theories, they do not appear to be strong enough to oppose the general benefits of Maximize Online Processing in favor of A+B whenever there is a conflict with this latter, and they lack the generality to extend to (29) in which no prediction is involved. Future research will need to address where and how exactly online prediction plays an additional role in left-right asymmetries, and how it interacts with the more traditional factors of completeness, correctness and backward-looking integration that have been considered and supported here.

Footnotes

(1) Minimize Domains is defined in Hawkins (2004:31 and 2014:11) as follows:
The human processor prefers to minimize the connected sequences of linguistic forms and their conventionally associated syntactic and semantic properties in which relations of combination and/or dependency are processed. The degree of this preference is proportional to the number of relations whose domains can be minimized in competing sequences or structures, and to the extent of the minimization difference in each domain.
By positioning heads consistently to the left or right of their non-heads in different language types (e.g. VO co-occurs with initial prepositions in PP, and with initial or early nouns in NP, etc, and OV co-occurs with final postpositions in PP, and final nouns in NP) heads of phrases will be consistently adjacent or close to one another and the processing of phrase structure and head-dependent structure can proceed efficiently within minimal domains for processing.

(2)
Many further examples and theoretical proposals from several areas of the language sciences support the role of performance in shaping grammatical conventions across languages in this way. Examples include:

- the Keenan-Comrie (1977) Accessibility Hierarchy (SU>DO>IO/OBL>GEN) whereby relative clause cut-off points across languages match declining ease of processing in English and other languages (Keenan & S. Hawkins 1987, Hawkins 1999, Lau & Tanaka 2021);
- a similar Accessibility Hierarchy effect is found across languages that conventionalize gaps in simpler relativization domains higher on the hierarchy and resumptive pronouns in more complex domains lower on the hierarchy, thereby matching the performance distribution of gaps to pronouns in languages such as Hebrew and Cantonese in which both are grammatical (in some syntactic positions), gaps being preferred in the simpler and pronouns in the more complex relatives (Hawkins 2004, Matthews & Yip 2003, Ariel 1999);
- more generally there are filler gap hierarchies for increasingly complex clause-embedding environments in which the cut-off points across grammars again correspond to declining processing ease in languages with numerous gap-containing environments (including subjacency-violating languages like Akan, Saah & Goodluck 1995, Hawkins 1999);
- performance preferences for subjects that obey the Person Hierarchy (1st,2nd > 3rd) in English (whereby the bus hit me is preferably passivized to I was hit by the bus) have been conventionalized into a grammatical/ungrammatical distinction in languages such as Lummi (Bresnan et al. 2001) - more generally "hard constraints" (in e.g. Lummi) are claimed to correlate with "soft" constraints (in e.g. English).

Hawkins' (1994, 2004, 2014) books are devoted to a full discussion and exemplification of such correspondences between performance preferences and grammatical conventions.

(3) The definition of Maximize Online Processing given in (5) differs from that in Hawkins (2002, 2004). This latter employed a metric for measuring it using "online to ultimate property ratios". Generally when unassignments and misassignments are avoided online, as in (6) versus (7) in the main text, more of the ultimate properties to be assigned to these sentences can be

assigned sooner, as seen in (6') versus (7") and in (8) versus (9), resulting in higher online to ultimate property ratios for the sentence as a whole. The definition in (5) does not use this metric. There are always online benefits for the key categories in a left-right asymmetry when there are unassignments or misassignments to one of its categories, as shown and quantified in the examples given in sections 3 and 4 of this paper. But the benefits for the sentence as a whole can sometimes be lost depending on extraneous categories and properties present in the remainder of a sentence. The definition of Maximize Online Processing given in (5) accordingly focuses on these local processing advantages for the categories A and B that figure in these left-right asymmetries and no longer employs the metric of online to ultimate property ratios for the whole sentence.

(4) In (6') and subsequent examples SU refers to grammatical subject, DO to direct object, Co-occurrences refers to the various strict subcategorization and selectional restriction options listed in the lexical entry for the verb, cf. Chomsky (1965), and Construct is a parsing operation assigning higher phrasal nodes based on lower grammatical categories that project to them, cf. Kimball (1973), Hawkins (1994).

(5) In "direct association theories", which link fillers directly to their subcategorizers and do not postulate gaps (Pickering et al. 1994), there would be a similar delay in linking *what*$_i$ to *stealing*$_i$ in (10b) as there is in linking *what*$_i$ to a gap that has been assigned to the immediate right of *stealing*.

(6) An apparent anomaly in this regard is Chinese, in which RelN occurs in a language with basic SVO. A recent paper by Song (2023), using a scalar approach to typology, shows that there are also significant head-final features in Chinese with which RelN is harmonic. It is still the case, however, that RelN is normally found across languages only in combination with rigid SOV, not with SVO, see (12) in the main text.

(7) These other relative clause variants that avoid the Gap+N structure of Rel+N and that are strongly characteristic of OV and not VO languages are the "head-internal" strategy of languages like Diegueño, Yavapai and Bambara (Cole 1987, Basilico 1996, Lehmann 1984) and the "correlative" construction of Hindi (Keenan 1985), see Hawkins (2014: 150-153) for examples of all these strategies and further discussion.

(8) The same figure for SO vs OS languages (96.6% or 1017 vs 40) has been given in the larger and more recent WALS database of Dryer (2013a), for languages having a dominant order in the sense of Dryer (2013b). The full figures in Dryer's (2013a) sample are: SOV (564), SVO (488), VSO (95), VOS (25), OVS (11), OSV (4), no dominant order (189).

(9) These semantic and conceptual dependencies shown in (18) hold at the level of semantic representations. In the grammars and surface forms of languages certain higher thematic roles can be deleted and understood, of course, as when an agent is removed in passive structures and only the patient is expressed. (Thanks to Tom Wasow for pointing out the need for this clarification.) It is still the case, in structures like *John was killed*, that some agentive force is understood semantically, however unspecified or uncertain the identity of that agent is.

(10) See Lehmann (1984) for the generality of this R+A patterning in NRel structures in VO languages, and for a discussion of interesting differences between VO and OV languages with respect to whether appositive interpretations occur at all in the RelN order. It is a promising area of further research to investigate whether this resistance to appositive interpretations in RelN can be linked to the unavailability of the head noun at the time the relative is interpreted, i.e. through unassignments or misassignments.

(11) The ergative-absolutive case marking pattern is found in languages such as Avar (North-East Caucasian), cf. Comrie (1978):

(i) a. Vas-as: jas j-ec:ula. (Avar)
boy-ERG girl-ABS SG.FEM.ABS-praise
'the boy praises the girl'
b. Jas j-ekerula.
girl-ABS SG.FEM.ABS-run
'the girl runs'

The agent subject receives distinctive (ergative) coding (-as) in the transitive (ia) while the patient object has (zero-marked) absolutive case, like that of the intransitive subject in (ib). The significance of this coding in the present context is that ergative is now the predicting category. Ergatives are unique to transitive clauses and they predict the co-occurrence of an absolutive. So when ergative precedes absolutive, as in (ia), it will add to working memory load. The reverse absolutive + ergative should therefore be preferred, according to Gibson's theory, just as nominative before a predicting accusative is preferred in German and Finnish.

Empirically, the majority (67%) of ergative-absolutive grammars in the WALS sample of (20) in the main text prefer ergative + absolutive, as in Avar (Primus, 1999). But interestingly for Gibson's theory, the remaining 33% do have either a preference for the absolutive + ergative order or no dominant order and hence productive absolutive + ergative orders. A language with such a preference for absolutive before ergative is the Australian language Dyirbal (Comrie, 1989 p.106, Dixon, 1972).

(ii) a. Balan dʸugumbil baŋgul yaṛaŋgu balgan (Dyirbal)
CLASSIF-ABS woman-ABS CLASSIF-ERG man-ERG hit
'The man hit the woman'
b. Balan dʸugumbil baninʸu
CLASSIF-ABS woman-ABS came-here
'The woman came here'

Grammars with ergative-absolutive morphology like Dyirbal comprise most of the languages that have been classified as object before subject in the typological literature (Tomlin's 4%), according to Primus (1999).

(12) The frequency of ergative subjects before absolutives (67%) is less than that of nominative subjects before accusatives (88%) in (19) and (20). Recall that the orderings preferred by all three sets of hierarchies in (16) – (18) are generally supported in nominative-accusative languages, whereas only two of them are supported in ergative-absolutive languages. The relative weakness of the case hierarchy in (16) in imposing an ordering preference compared with the syntactic and semantic hierarchies in (17)-(18) is plausibly because explicit surface case marking creates less of an asymmetric dependency between one case on the other in terms of actually recognizing which

case is which, and hence less pressure for asymmetric ordering. I.e. cases can generally be interpreted regardless of their ordering, unless they are ambiguous in surface form.

(13) Some further examples of left-right asymmetry that have been discussed in the typological literature and that merit systematic consideration from the perspective of this paper comparing Maximize Online Processing with online prediction are the following: SOV languages, in which nouns precede the verb, typically have "rich" case-marking (Tallerman 1998, Dryer 2002) and so provide immediate information about the thematic role and grammatical case of an argument as soon as it is encountered prior to the verb (see Bornkessel 2002 for neurocognitive confirmation of the activation of this information prior to the verb); SOV languages also have a higher ratio of lexical nouns to verbs compared with SVO and V-initial languages, and so provide richer lexical content and semantic information about the event being described at the time that nouns are processed prior to verbs, many of which are grammatically "light" and lexically uninformative in SOV languages (see Polinsky & Magyar 2020 for quantitative support of this); Verb-initial languages, by contrast, have rich verb agreement which provides immediate access to a verb's co-occurrence structure at the time the verb is encountered (Dryer 2002, Hawkins 2002) and other verb-coding strategies as well, e.g. rich passives (Keenan 1976a) which also signal argument structure information immediately at the initial verb; and the processing of main and subordinate clause orderings, e.g. in conditional sentences and temporal sequences, appears to optimize the causality and temporal relations between A+B for processing, such that B can be more readily and fully interpreted when it follows A rather than preceding it (see Diessel 2001 for relevant cross-linguistic data).

(14) Overall predicting before predicted, i.e. Surprisal Theory, appears to be the stronger of the alternative prediction-based theories considered here, based on the number of studies that provide support for it.

References

Antinucci, F., A. Duranti & L. Gebert (1979) 'Relative clause structure, relative clause perception and the change from SOV to SVO', Cognition 7:145-76.

Ariel, M. (1999) 'Cognitive universals and linguistic conventions: The case of resumptive pronouns', Studies in Language 23: 217-269.

Basilico, D. (1996) 'Head position and internally headed relative clauses', Language 72: 498-532.

Bornkessel, I. (2002) 'The argument dependency model: A neurocognitive approach to incremental interpretation', Ph.D. dissertation, University of Potsdam.

Bresnan, J., S. Dingare, & C.D. Manning (2001) 'Soft constraints mirror hard constraints: Voice and person in English and Lummi', in M. Butt & T.H. King (eds.) Proceedings of the LFG 01 Conference. Stanford: CSLI Publications, 13-32.

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

Clancy, P.M., H. Lee & M. Zoh (1986) 'Processing strategies in the acquisition of relative clauses', Cognition 14: 225-62.

Clark, T.H., C. Meister, T. Pimentel, M. Hahn, R. Cotterell, R. Futrell, & R. Levy (2023) 'A cross-linguistic pressure for uniform information density in word order', arXiv:2306.03734

Clifton, C. & L. Frazier (1989) 'Comprehending sentences with long-distance dependencies', in

C.N. Carlson & M.K. Tannenhaus (eds.) Linguistic Structure in Language Processing. Dordrecht: Kluwer.
Cole, P. (1987) 'The structure of internally headed relative clauses', Natural Language and Linguistic Theory 5: 277-302.
Comrie, B. (1978) 'Ergativity', in W.P. Lehmann (ed.) Syntactic Typology: Studies in the Phenomenology of Language. Austin: University of Texas Press.
Comrie, B. (1989) Language Universals and Linguistic Typology, 2nd edn.. Chicago: University of Chicago Press.
Comrie, B. (2013) 'Alignment of case marking of full noun phrases', in M.S. Dryer & M. Haspelmath (eds.), ch.98.
Diessel, H. (2001) 'The ordering of main and adverbial clauses: A typological study', Language 77(3): 433-455.
Dixon, R.M.W. (1972) The Dyirbal Language of North Queensland. Cambridge: Cambridge University Press.
Dowty, D.R. (1991) 'Thematic proto-roles and argument selection', Language 67: 547-619.
Dryer, M.S. (1992) 'The Greenbergian word order correlations', Language 68: 81-138.
Dryer, M.S. (2022) 'Case distinctions, rich verb agreement, and word order type', Theoretical Linguistics 28: 151-157.
Dryer, M.S. (2005) 'Position of interrogative phrases in content questions', in M. Haspelmath et al. (eds.), 378-381.
Dryer, M.S. (2009) 'The branching direction theory of word order correlations revisited', in S. Scalise, E. Magni, & A. Bisetto (eds.), Universals of Language Today. Berlin: Springer.
Dryer, M.S. (2013a) 'Order of subject, object and verb', in M.S. Dryer & M. Haspelmath (eds.), ch.81.
Dryer, M.S. (2013b) 'Determining dominant word order', in M.S. Dryer & M. Haspelmath (eds.).
Dryer, M.S. & M. Haspelmath (eds.) (2013) The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at: http://wals.info)
Dryer, M.S. with O.D. Gensler (2005) 'Order of object, oblique and verb', in M. Haspelmath et al. (eds.), 342-345.
Engelhardt, P.E., L. Filipovic & J.A. Hawkins (in press) 'Prediction in SVO and SOV languages: Processing and typological considerations', Linguistics
Ferreira, F. & S. Chantavari (2018) 'Integration and prediction in language processing: A synthesis of old and new', Current Directions in Psychological Science 27: 443-448.
Fodor, J.A., T.G. Bever & M.F. Garrett (1974) The Psychology of Language. New York: McGraw-Hill
Fodor, J.D. (1978) 'Parsing strategies and constraints on transformations', Linguistic Inquiry 8: 425-504.
Fodor, J.D. (1984) 'Constraints on gaps: is the parser a significant influence?' in B. Butterworth, B. Comrie & Ö. Dahl (eds.) Explanations for Language Universals. Berlin: Mouton, 9-34.
Fodor, J.D. (1989) 'Empty categories in sentence processing', Language and Cognitive Processes 3:155-209.
Frazier, L. & G.B. Flores d'Arcais (1989) 'Filler driven parsing: A study of gap filling in Dutch', Journal of Memory and Language 28: 331-344.
Futrell, R., K. Mahowald & E. Gibson (2015) 'Large-scale evidence of dependency length

minimization in 37 languages', Proceedings of the National Academy of Sciences 112(33): 10336-41
Futrell, R., R.P. Levy & E. Gibson (2020) 'Dependency locality as an explanatory principle for word order', Language 96(2): 371-412.
Gibson, E. (1998) 'Linguistic complexity: Locality of syntactic dependencies', Cognition 68: 1-76.
Gibson, E. (2000) 'The Dependency Locality Theory: A distance-based theory of linguistic complexity', in Y. Miyashita P. Marantz & W. O'Neil (eds.), Image, Language, Brain. Cambridge, Mass.: MIT Press, 95-112.
Gilligan, G.M. (1987) 'A cross-linguistic approach to the pro-drop parameter', Ph.D. dissertation, University of Southern California.
Greenberg, J.H. (1963) 'Some universals of grammar with particular reference to the order of meaningful elements', in J.H. Greenberg (ed.), Universals of Language. Cambridge, Mass.: MIT Press, 73-113.
Greenberg, J.H. (1966) Language Universals with Special Reference to Feature Hierarchies. The Hague: Mouton.
Gundel, J.K. (1988) 'Universals of topic-comment structure', in M. Hammond, E. Moravcsik & J. Wirth (eds.) Studies in Syntactic Typology. Amsterdam/Philadelphia: John Benjamins, 209-239.
Haspelmath, M. (2021) 'Explaining grammatical coding asymmetries: Form-frequency correspondences and predictability', Journal of Linguistics (2021): 1-29
Haspelmath, M., M.S. Dryer, D. Gil & B. Comrie (eds.) (2005) The World Atlas of Language Structures. Oxford: Oxford University Press.
Hawkins, J.A. (1983) Word Order Universals. New York: Academic Press.
Hawkins, J.A. (1990) 'A parsing theory of word order universals', Linguistic Inquiry 21: 223-261.
Hawkins, J.A. (1994) A Performance Theory of Order and Constituency. Cambridge: Cambridge University Press.
Hawkins, J.A. (1999) 'Processing complexity and filler-gap dependencies across grammars', Language 75: 244-285.
Hawkins, J.A. (2002) 'Symmetries and asymmetries: their grammar, typology and parsing', Theoretical Linguistics 28(2):95-149.
Hawkins, J.A. (2004) Efficiency and Complexity in Grammars. Oxford: Oxford University Press.
Hawkins, J.A. (2014) Cross-linguistic Variation and Efficiency. Oxford: Oxford University Press.
Horn, L.R. (1989) A Natural History of Negation. Chicago: University of Chicago Press.
Ioup, G. (1975) 'Some universals for quantifier scope', in J.P. Kimball (ed.) Syntax and Semantics, Volume 4. Leiden: Brill, 37-58.
Jackendoff, R.S. (1972) Semantic Interpretation in Generative Grammar. Cambridge, Mass.: MIT Press.
Jaeger, T.F. (2010) 'Redundancy and reduction: Speakers manage syntactic information density', Cognitive Psychology 61:23-62.
Keenan, E.L. (1976a) 'Remarkable subjects in Malagasy', in C.N. Li (ed.) Subject and Topic. New York: Academic Press.
Keenan, E.L. (1976b) 'Towards a universal definition of "subject of"', in C.N. Li (ed.) Subject

and Topic. New York: Academic Press.
Keenan, E.L. (1985) 'Relative clauses', in T. Shopen (ed.) Language Typology and Syntactic Description, vol.II: Complex Constructions. Cambridge: Cambridge University Press, 141-170.
Keenan, E.L. & B. Comrie (1977) 'Noun phrase accessibility and Universal Grammar', Linguistic Inquiry 8: 63-99.
Keenan, E.L. & S. Hawkins (1987) 'The psychological validity of the Accessibility Hierarchy', in E.L. Keenan, Universal Grammar. London: Croom Helm, 60-85.
Kimball, J. (1973) 'Seven principles of surface structure parsing in natural language', Cognition 2: 15-47.
Kiss, K.E. (2002) The Syntax of Hungarian. Cambridge: Cambridge University Press.
Kuno, S. (1973) The Structure of the Japanese Language. Cambridge, Mass.: MIT Press.
Lambrecht, K. (1994) Information Structure and Sentence Form: Topic, Focus and the Mental Representation of Discourse Referents. Cambridge: Cambridge University Press.
Lau, E. & N. Tanaka (2021) 'The subject advantage in relative clauses: A review', Glossa 6(1): 1-34.
Lehmann, C. (1984) Der Relativsatz. Tuebingen: Gunter Narr Verlag.
Levelt, W.J.M. (1989) Speaking: From Intention to Articulation. Cambridge, Mass.: MIT Press.
Levy, R. (2008) 'Expectation-based syntactic comprehension', Cognition 106:1126-77.
Li, C.N. & S.A. Thompson (1981) A Functional Reference Grammar of Mandarin Chinese. Berkeley: University of California Press.
Matthews, S. & V. Yip (2003) 'Relative clauses in early bilingual development: Transfer and universals', in A.G. Ramat (ed.) Typology and Second Language Acquisition. Berlin: Mouton de Gruyter.
Maurits, L., A. Perfors & D. Navarro (2010) 'Why are some word orders more common than others? A uniform information density account', Proceedings of the 24th Annual Conference on Neural Information Processing, Vancouver, Canada.
Melinger, A., T. Pechmann, & S. Pappert (2009) 'Case in language production', in A. Malchukov & A. Spencer (eds.), Oxford Handbook of Case. Oxford: Oxford University Press, 384-401.
Morgan, E. & R. Levy (2016) 'Abstract knowledge versus direct experience in processing of binomial expressions', Cognition 157(3): 384-402.
Pickering, M., S. Barton & R. Shillcock (1994) 'Unbounded dependencies, island constraints and processing complexity', in C. Clifton, L. Frazier & K. Rayner (eds.), Perspectives on Sentence Processing. Hillsdale, NJ: Erlbaum, 199-224.
Polinsky, M. (2002) 'Efficiency preferences: refinements, rankings and unresolved questions', Theoretical Linguistics 28(2): 177-202.
Polinsky, M. & L. Magyar (2020) 'Headedness and the lexicon: The case of verb-to-noun ratios', Languages 5(1): 9.
Primus, B. (1999) Cases and Thematic Roles: Ergative, Accusative and Active. Tübingen: Niemeyer.
Reinhart, T. (1982) 'Pragmatics and linguistics: An analysis of sentence topics', Philosophica 27: 53-94.
Reinhart, T. (1983) Anaphora and Semantic Interpretation. London: Croom Helm.
Saah, K.K. & H. Goodluck (1995) 'Island effects in parsing and grammar: Evidence from Akan', Linguistic Review 12: 381-409.

Song, Kyung-An (2023) ‘Chinese word order and the two typological “exceptions”: A scalar approach’, elanguageresearch.org, 145-162.
Tallerman, M. (1998) Understanding Syntax. London: Arnold.
Tomlin, R.S. (1986) Basic Word Order: Functional Principles. London: Croom Helm.
Tsao, F.-F. (1978) A Functional Study of Topic in Chinese. Taipei: Student Book Company.
Wasow, T., T.F. Jaeger & D.M. Orr (2011) ‘Lexical variation in relativizer frequency’, in H. Simon & H. Wiese (eds.) Expecting the Unexpected: Exceptions in Grammar. Berlin: Mouton de Gruyter, 175-195.