Constantin Freitag, Humboldt-Universität zu Berlin, constantin.freitag@gmx.de

From local preferences to global configurations

In this contribution I will argue for a view that considers language output as an output of a self-organizing system. The formation of patterns is upward bound by constraints on linear processing and linguistic form emerges as a consequence of the interplay of some basic processing factors. No matter if all languages are equally complex or not, we can state that the overall complexity of all languages is equally constrained.
Hawkins (2004) argued that the parser prefers efficient structures. This is the reason why we can observe coss-linguistically the combination of certain patterns, e. g. the correlation between verb initial order (VO) and prepositions on the one hand and between verb final order (OV) and postpositions at the other hand. Hawkins (2004) shows that the more dependency relations are established between two linguistic elements the closer they will appear in the surface structure to facilitate processing. If we assume that processing capacities are upwards bound and also sum up, it follows logically that only certain pattern combinations are allowed. In order to make these assumptions concrete we have to formulate a set of relevant factors and their impact on processing. There are relevant proposals in Gibson (1998) and Hawkins (2004). In order to illustrate the idea I want to have a look at center embeddings (CE) and syntactic islands.
Karlsson (2007) showed that CEs are upper bound. In written language one finds a deeper level of embedding than in spoken language which correlates with lower processing cost in written language. Additionally Karlsson (2007) states that deep CEs correlate with grammatical factors which are known to facilitate processing of linguistic structure like position of embedding (SUB>OBJ), NP-type (full NP < pronoun < $\emptyset$)and type of the embedded clause. Similar observations are made for syntactic islands (SI). As Kluender (2004) and Sag et al. (2007) show, there are many factors that influence the acceptability/grammaticality of SIs such as finiteness, NP-type, among others. Ungrammaticality of Island configurations can be gradually improved by changing these factors (2). This in turn provides evidence for the assumption that SIs are the result of processing constraints rather than grammatical constraints. What these factors, discussed for CEs and SIs, have in common is that they directly influence linear processing. Assuming a strict left-to-right parser, I propose to deduce them from at least two underlying factors which describe dependency relations between linguistic elements: number of actual dependencies and number of interveners (in words). For example the NP-type: full NPs are usually longer than pronouns which in turn are longer than anything non-existing ($\emptyset$). With decreasing length the element in question becomes a shorter intervener and consequently the distance between an encompassing dependency becomes shorter (table 1 and 2). Furthermore with decreasing number of elements the number of dependencies decreases. While no element ($\emptyset$) establishes no syntactic dependency relation (table 1c and 2c), a pronoun establishes at least one (table 1b and 2b) and full NPs normally also establish dependency relations between determiner, modifier and noun (table 1a and 2a).
As exemplified, these underlying processing factors act locally and also interact globally. The respective NPs establish only local dependency relations. No matter how costly the encompassing dependencies are, by using a full NP the parser will have more dependencies to keep track of. Additionally it must hold the early elements longer active in short term memory. Therefore the longer sentences become the more efficient structures must be used to maintain processability. I don't want to propose a metric here but rather argue for an approach that searches the reasons for the exclusion of certain configurations in the interactions of local preferences of the parser. The approaches to CEs and SIs show that grammaticality/acceptability is of gradual nature. This graduality can be explained by local processing peaks and their influence on long-distance dependencies. In this view linguistic structures are the output of a self-organizing system and surface syntax is a pattern that emerges rather than is generated.

(1) based on (Karlsson, 2007, ex. 7a)

a. [$_M$ The girl [$_{C1}$ who was clothed in the tightest-fitting pair of slacks [$_{C2}$ the salesman had ever seen on a woman] and a sweater] wanted to be sociable.]

b. [$_M$ The girl [$_{C1}$ who was clothed in the tightest-fitting pair of slacks [$_{C2}$ I had ever seen on a woman] and a sweater] wanted to be sociable.]

c. [$_M$ The girl [$_{C1}$ who was clothed in the tightest-fitting pair of slacks [$_{C2}$ ever seen on a woman] and a sweater] wanted to be sociable.]

a.		M	C1	the	salesman	had	ever	seen	on	a	woman	C1	M
	Dep-M	⊢	-	-	-	-	-	-	-	-	-	-	⊣
	Dep-C1		⊢	-	-	-	-	-	-	-	-	⊣	
	Dep-NP			[+	+]								
	Dep-C2				[+	+]							
b.		M	C1		I	had	ever	seen	on	a	woman	C1	M
	Dep-M	⊢	-		-	-	-	-	-	-	-	-	⊣
	Dep-C1		⊢		-	-	-	-	-	-	-	⊣	
	Dep-C2				[+	+]							
c.		M	C1			having	ever	seen	on	a	woman	C1	M
	Dep-M	⊢	-			-	-	-	-	-	-	-	⊣
	Dep-C1		⊢			-	-	-	-	-	-	⊣	

Table 1: Influence of NP type on processing center embeddings.

(2) (Kluender, 2004, ex. 4)

a. That's the campaign [that I finally thought of the aide [$_{S3}$ who could spearhead _]].

b. That's the campaign [that I finally thought of someone [$_{S3}$ who could spearhead _]].

c. That's the campaign [that I finally thought of someone [$_{S3}$ to spearhead _]].

a.		That is the	campaign	that	I	finally	thought	of	the	aide	who	could	spearhead	GAP
	Extraction		⊢	-	-	-	-	-	-	-	-	-	-	⊣
	PP							[+	+	+]				
	NP								[+	+]				
	REL-PRO									[+	+]			
	S3										[+	+]		
b.		That is the	campaign	that	I	finally	thought	of		someone	who	could	spearhead	GAP
	Extraction		⊢	-	-	-	-	-		-	-	-	-	⊣
	PP							[+		+]				
	REL-PRO									[+	+]			
	S3										[+	+]		
c.		That is the	campaign	that	I	finally	thought	of		someone		to	spearhead	GAP
	Extraction		⊢	-	-	-	-	-		-		-	-	⊣
	PP							[+		+]				

Table 2: Influence of NP type on processing syntactic islands

References

Gibson, Edward (1998) Lingusitic Complexity: Locality of Syntactic Dependencies. *Cognition* **68**: 1–76.

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

Karlsson, Fred (2007) Constraints on Multiple Center-Embedding of Clauses. *Journal of Linguistics* **43**(2): 365–392.

Kluender, Robert (2004) Are Subject Islands Subject to a Processing Account. In *Proceedings of the 23rd West Coast Conference on Formal Linguistics (WCCFL 23)*, Vineeta Chand, Ann Kelleher, Angelo J. Rodrìguez, & Benjamin Schmeiser, eds., Somerville: Cascadilla Press, 475–499.

Sag, Ivan A., Philip Hofmeister, & Neal Snider (2007) Processing Complexity in Subjacency Violations: The Complex Noun Phrase Constraint. In *Proceedings of the 43rd Annual Meeting of the Chicago Linguistic Society*, vol. 1, Malcolm Elliott, James Kirby, Osamu Sawada, Eleni Staraki, & Suwon Yoon, eds., Chicago: CLS, 215–229.