# Simpler Syntax and explanation

Peter W. Culicover
The Ohio State University

Proceedings of the 20th International Conference on
Head-Driven Phrase Structure Grammar

Freie Universität Berlin

Stefan Müller (Editor)

2013

Stanford, CA: CSLI Publications

pages 263–282

Culicover, Peter W. 2013. Simpler Syntax and explanation. In Stefan Müller (ed.), *Proceedings of the 20th International Conference on Head-Driven Phrase Structure Grammar, Freie Universität Berlin*, 263–282. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2013.14.

**Abstract**

Simpler Syntax is an approach to grammar that calls for very restrictive limits on the notion of 'grammatical competence'. Specifically, it does not account for unacceptability judgments for sentences that are well-formed if they are fully licensed by the constructions of the language. SS leads us to seek accounts for such judgments in terms other than grammar per se, e.g., processing complexity, semantic or pragmatic well-formedness, discourse coherence, etc. I review several examples that suggest that the line that SS draws between competence on the one hand and performance and other mechanisms on the other is on the right track.

# 1 Introduction: What constitutes progress in linguistics?

What counts as progress in linguistics?[1] One way to gauge whether we are making progress, is to first be clear about what it is that we are trying to accomplish. In the case of syntax, we have essentially two options: (1) we can stick closely to the Chomskian program that has largely dominated the field since the 1960s, or (2) we can think 'outside of the box' and decide for ourselves what the goals of syntactic theory should be. Of course, following this second option does not preclude overlaps with the first, but it does mean that we may set goals for ourselves that may not always be widely shared.

The Chomskian premise is that there is an 'organ' of the mind that embodies the native speaker's knowledge of language. This knowledge is referred to as 'I-language', or competence (Chomsky 1986). Through interaction with various peripheral mechanisms, I-language accounts for the ability of the speaker to produce and understand sentences, the entire set of which comprises 'E-language'. Crucially, since this 'organ' is an expression of the human biological capacity for language, its architecture entails the existence of linguistic universals – those properties of languages that are not acquired through experience but imposed upon it by the structure and limitations of this 'organ'. The fundamental goal, as articulated in Chomsky 1965, is formulated narrowly in terms of linguistic competence: The theory makes available descriptively adequate grammars, and incorporates an evaluation metric that ranks competing descriptions of the same data; this

[†]I thank Ray Jackendoff for comments on an early version of this paper; the participants of the Workshop on Progress in Linguistics, and especially Gereon Müller, for helpful questions and comments, and Jefferson Barlew, Carmelo Bazaco, Lorena Sainz-Maza Lecanda, Susanne Winkler, Philip Hofmeister, William Schuler, Martin van Schijndel, Paij Lintz, Dave Howcroft and the Synners discussion group at OSU. Thanks to Micha Elsner and Stefan Müller for helping me with LaTeX.

[1]Although the title of this workshop is "Progress in Linguistics", my remarks are focused on progress in syntax.

latter function is characterized in terms of an abstract model of the language learner. Such a theory defines the notion 'possible language' in terms of 'possible grammar' (along the lines of I-Language) and explains what we find in language in terms of 'most highly ranked grammar'.

However, it is important to recognize that what such a theory seeks to explain – linguistic competence – is characterized in terms of grammaticality judgments. So at least on the more classical notion of the goals of a linguistic theory, the practical measure of success is progress in explaining these judgments. On the standard view, the grammar is responsible for judgments of grammaticality, while 'performance', that is, the real-time computation of the correspondence between sound and mean, is responsible for judgments of acceptability (Chomsky and Miller 1963; Chomsky 1965: Chapter 1).

One could go on at great length about what constitutes an explanation of grammaticality judgments in classical syntactic theory – see Chapters 2 and 3 of *Simpler Syntax* (SS; Culicover and Jackendoff 2005) for example. Rather, I focus here on where SS proposes that we draw the line between judgments that reflect grammar and judgments that reflect performance. Determining the location of this line constitutes progress, in that it makes it more likely that we will find explanations for the phenomena, rather than simply stating generalizations about the phenomena using the descriptive vocabulary of the grammatical theory. Along the way, SS also suggests a particular characterization of competence and performance that I think facilitates progress in finding explanations.

In §2 I summarize briefly the Simpler Syntax Manifesto and its implications for these issues. Then in §3 I discuss several cases where I think that drawing the line where SS proposes improves our chances of finding genuine explanations, rather than simply interesting statements of the problems. §4 draws a connection between the distinction drawn by SS and that of 'hard' and 'soft' constraints due to Sorace and Keller (2005). §5 is a brief conclusion.

## 2 The Simpler Syntax Manifesto

The Simpler Syntax Hypothesis (SSH) holds that syntactic descriptions should be no more complex than is necessary to account for interpretation, while still capturing all of the true syntactic and semantic generalizations. Reasons to believe that SSH is the right approach to syntactic analysis are given at some length in Culicover and Jackendoff 2005; 2006; I won't try to review them here.

SS adopts the Parallel Architecture perspective of Jackendoff 2002, in which phonological, syntactic and semantic representations have their own well-formedness conditions, and are related to one another through correspondence rules. We assume that the basic components of grammars (both

in the head and in the linguistic description) are constructions, defined in terms of these correspondences. Particular syntactic structures participate in such correspondences to the extent that there are generalizations to be captured that go beyond the association of a particular basic form (such as a word) with a particular meaning. Following SSH, we assume, among other things, no abstract syntactic structure (i.e. no functional heads, no massive binary branching, etc.); no movement; no invisible constituents (except maybe A′ trace); no UG syntactic constraints.

On this view, 'grammatical rules' are maximally general constructions with non-idiosyncratic, compositional interpretations. A particular phonological form with a meaning is licensed if every part of the meaning corresponds to some part of the form as licensed by some construction in the grammar.[2]

Here are some simple concrete examples. I adapt the AVM notations used in HPSG and by Fillmore 1988 and Kay 2005 to the Parallel Architecture of Jackendoff (2002). Following the PA, I assume that a correspondence consists of three co-subscripted representations, PHON, SYN and CS. The representation for the lexical item *eat* is given in (1). **EAT** is the concept that corresponds to the act of eating. V abbreviates the syntactic information about the word; subcategorization information must be understood to be part of this information although it is not represented here. (I use boldface to identify elements of CS.)

(1) *eat*

$$\begin{bmatrix} \text{PHON} & [\text{it}]_1 \\ \text{SYN} & \text{V}_1 \\ \text{CS} & \lambda\text{y}.\lambda\text{x}.\mathbf{EAT}_1(\text{AGENT:x, PATIENT:y}) \end{bmatrix}$$

To keep things simple I ignore here the constructional details of inflected verbs.

Correspondences are licensed by constraint satisfaction, as in HPSG and Construction Grammar. For example, the construction for a non-idiomatic transitive VP is given in (2). The co-subscripting reflects the application of the correspondence rules that link components of PHON and SYN and components of SYN and CS. $\varphi$ is a variable phonetic string. '-' means 'immediately precedes'.

(2) TRANSITIVE VP

$$\begin{bmatrix} \text{PHON} & [\varphi]_1\text{-}[\varphi]_2 \\ \text{SYN} & [_{\text{VP}}\ \text{V}_1,\ \text{NP}_2] \\ \text{CS} & \mathbf{V}_1(\mathbf{NP}_2) \end{bmatrix}$$

[2]Hence meaning arrived at on line through metaphor cannot participate directly in such correspondences, but must be the consequence of inferential processes. Conventionalized coercion is, of course, part of meaning and therefore of correspondences.

This construction says that to form a VP, take a V and an NP that are sisters, linearize the form corresponding to V before that corresponding to NP, and apply the interpretation of V to the interpretation of NP. By assumption, this NP satisfies the subcategorization requirements of the verb.

A correspondence is licensed if there is a coindexing of its terms that satisfies the conditions imposed by the constructions of true language. This can be done by checking the PHON, SYN and CS of the particular correspondence against those of the constructions. For example, in (4) we check (3) against (1).

(3)
$$\begin{bmatrix} \text{PHON} & [\text{it}] \\ \text{SYN} & \text{V} \\ \text{CS} & \lambda\text{y}.\lambda\text{x}.\mathbf{EAT}(\text{AGENT:x, PATIENT:y}) \end{bmatrix}$$

(4)
$$\begin{bmatrix} \text{PHON} & [\varphi]_1 \subseteq [\text{it}] \\ \text{SYN} & \text{V}_1 \subseteq \text{V} \\ \text{CS} & \lambda\text{y}.\lambda\text{x}.\mathbf{EAT}_1(\text{AGENT:x, PATIENT:y}) \subseteq \\ & \lambda\text{y}.\lambda\text{x}.\mathbf{EAT}(\text{AGENT:x, PATIENT:y}) \end{bmatrix}$$

The properties of the construction constitute a subset of its properties. If the subset relation holds, then we coindex the correspondence so that it matches the construction, which licenses it as well-formed with respect to the construction.

To take a more complicated example, let the interpretation for *the pizza* be **PIZZA**[DEF]. Assuming the appropriate construction for the NP, the result of checking *eats the pizza* against the transitive VP construction (2) is (5).

(5) *eat the pizza*
$$\begin{bmatrix} \text{PHON} & [\ldots]_1\text{-}[\ldots]_2 \subseteq [[\text{it}]\text{-}[\eth\text{ə pitsə}]] \\ \text{SYN} & [_{\text{VP}}\ \text{V}_1,\ \text{NP}_2] \subseteq [_{\text{VP}}\ [_{\text{V}}\ \text{eat}],\ [_{\text{NP}}\ \text{the, pizza}]] \\ \text{CS} & \mathbf{V}_1(\mathbf{NP}_2) \subseteq \lambda\text{y}.\lambda\text{x}.\mathbf{EAT}(\text{AGENT:x, PATIENT:y})) \\ & ([\mathbf{PIZZA}[\text{DEF}]) \end{bmatrix}$$

The checking of PHON allows to assign subscripts to [it] and [ðə pitsə]. Checking of SYN goes through because the categories of the correspondence match those of the construction. The same holds for CS, assuming the appropriate semantic types and lambda-reduction.

Idioms and constructions with idiomatic properties take a similar form, where PHON specifies the linear order of elements, SYN describes the structure and CS the corresponding interpretation. Representations for *kick the bucket* and *sell NP down the river* are given in (6) and (7), respectively.

(6) *kick the bucket*

$$\begin{bmatrix} \text{PHON} & [[\text{kɪk}_1\text{-}[\text{ðə}_2\ \text{bʌkɪt}_3]_4]_5 \\ \text{SYN} & [_{\text{VP}}[_{\text{V}}\text{kick}]_1,\ [_{\text{NP}}\ \text{the}_2,\ \text{bucket}_3]_4]_5 \\ \text{CS} & \lambda\text{x}.\mathbf{DIE}_5(\text{EXP:x}) \end{bmatrix}$$

(7) *sell NP down the river*

$$\begin{bmatrix} \text{PHON} & [\text{sɛl}]_1\text{-}[...]_2\text{-}[\text{dɑ}^{\text{u}}\text{n}_3\ \text{ðə}_4\ \text{rɪvɹ}_5]_6 \\ \text{SYN} & [_{\text{VP}}\ \text{sell}_1,\ \text{NP}_2,\ [_{\text{PP}}\ \text{down}_3\ [_{\text{NP}}\text{the}_4,\ \text{river}_5]]_6] \\ \text{CS} & \lambda\text{y}.\lambda\text{x}.\mathbf{BETRAY}_{6+1}(\text{AGENT:x},\ \text{PATIENT:y})(\mathbf{NP}_2) \end{bmatrix}$$

Because SS is a constructional theory, it strongly favors minimal syntactic structures to account for the correspondence between form and interpretation. For instance, given the sequence V-NP, if the corresponding interpretation is **V**(**NP**), it is simpler to state this directly in terms of the structure [$_{\text{VP}}$ V, NP], rather than posit a more abstract syntactic structure such as [$_{\text{vP}}$ $\text{V}_i$, $\text{NP}_j$ [$_{\text{VP}}$ $t_i$, $t_j$] or something even more complex. In other words, the constructional approach per se does not rule out complex structures in which there are filler-gap chains, but under the SSH such structures would have to be strongly motivated by the facts. So SS assumes a filler-gap chain in A′ constructions, for example, because doing so facilitates the interpretation, simplifies the grammatical description, and accounts for such things as unbounded dependencies and reconstruction effects.

Moreover, SS sharply restricts the scope of syntactic explanation to phenomena that have to do with the correspondence between syntactic structure and phonological form. It says for a given structure what the ordering possibilities of the constituents are, and what the morphological form of these constituents must or may be.

In other words, 'grammatical competence' in SS is a very restricted notion. Specifically, it does not account for unacceptability judgments for sentences that are well-formed if they are fully licensed by the constructions of the language, along the lines outlined above. Rather, it forces the explanation for such cases into other domains of explanation, e.g. semantic well-formedness, pragmatic coherence, processing, and so on.

# 3 Some cases

## 3.1 Island constraints and SS

It follows that one important consequence of SS for theories of the representation of language in the mind is that island constraints like subjacency, the complex NP constraint, the subject condition and the like are not grammatical phenomena in the narrow sense. There is in fact a growing literature that argues that they are the consequence of processing complexity arising

from particular configurations that are otherwise well-formed; see, for example, Kluender 1992; 1998; 2004; Hofmeister et al. submitted; Hofmeister et al. to appear; Hofmeister et al. 2007; Sag et al. 2007; Sag et al. 2008; Hofmeister and Sag 2010; Staum Casasanto et al. 2010; Hofmeister 2011; Hofmeister et al. 2012). Precisely **why** and **how** complexity gives rise to these effects is an intriguing question that I return to briefly below.

It is first instructive to reflect a little on why island constraints were considered to be part of the grammar, and in fact thought to be syntactic universals that constitute a part of the language faculty (Ross 1967; Chomsky 1977; Wexler and Culicover 1980). In early generative grammar, judgments of unacceptability were taken by default to be judgments of ungrammaticality, that is, to be accounted for by theories of competence. Except for obvious cases such as center-embedded relative clauses, there was no plausible account of such judgments in terms of performance. Since the island constraints applied not to particular constructions but to general configurations, it was reasonable to conclude that they were grammatical universals.

On the other hand, if these constraints are not a matter of grammar, they are universal only in the sense that they follow from properties of the universal processing mechanism. To show that this is plausible, I consider next an example where the data have very much the character of the kinds of data that motivated the island constraints, but the explanation is more likely to be external to grammar, as SS suggests. This conclusion supports in turn the view that the data that the island constraints account for should have a similar explanation. Then I consider a purported grammatical 'freezing' constraint that for which there is good empirical evidence that suggests that it is strictly the consequence of processing complexity.

### 3.2 Zero-relatives

In English it is possible to omit the relative marker *that* when a non-subject NP is relativized (Culicover in press).[3]

(8) (this is) a book $\left\{ \begin{array}{c} \text{which} \\ \text{that} \\ \emptyset \end{array} \right\}$ you should read

It is unacceptable in English to adjoin a constituent to the left periphery of a zero-relative clause, while similar adjunction to a marked relative following the relative marker is acceptable. A typical example is given in (9).

(9) (this is) a book $\left\{ \begin{array}{c} \text{which} \\ \text{that} \\ *\emptyset \end{array} \right\}$ if you have time you should read

[3]It is also possible to omit *that* when the subject is relativized in non-standard varieties.

A classical approach to such unacceptability would take the starred example to be ungrammatical, and formulate a grammatical account of this fact, either in terms of the configurations (i.e. what structures are possible or impossible, similar to the island constraints of Ross 1967), or a general constraint that blocks deletion of the complementizer (on analogy with application of the ECP to account for the *that-t* effect (Kayne 1984)).

I think that neither of these is the right type of solution. The evidence suggests that there is no simple syntactic characterization of the contexts that render zero-relatives unacceptable. Rather, the generalization appears to have to do with the identifiability of the relative clause on the basis of the sequence that marks its left periphery.

Here is the evidence. Note that it is possible, although somewhat complex, to position a non-subject constituent at the left periphery of a relative clause. This can be done in a number of ways, including topicalization, negative inversion, and stylistic inversion, illustrated in (10).

(10) a. He is a man to whom$_j$ liberty$_i$, we will never grant $t_i$ $t_j$. [Baltin 1981]

b. This is a dog $\left\{ \begin{array}{c} \text{which} \\ \text{that} \end{array} \right\}_i$ under no circumstances should you ever try to feed $t_i$. [Culicover 1992]

c. Detroit is a town $\left\{ \begin{array}{c} \text{where} \\ \text{in which} \\ \text{that} \end{array} \right\}$ in almost every garage can be found a car manufactured by GM.

Omitting the complementizer makes these examples less acceptable, as seen in (11).

(11) a. * He is a man$_j$ liberty$_i$ we will never grant $t_i$ to $t_j$.[4]

b. * This is a dog$_i$ under no circumstances should you ever try to feed $t_i$.

c. * Detroit is a town in almost every garage can be found a car manufactured by GM.

An important property of these constructions is that when there is no complementizer, there is nothing that marks the relative clause as such. I return to this point below.

A classical approach to the *-ed examples in (9) and (11) would be to rule them out by invoking one or more grammatical principles, perhaps stated explicitly in terms of the observed structures. However, each of the structures in (11) is different. In the case of topicalization, the initial constituent is

[4]Note that (11a) is somewhat marginal even with *who* or *that*, because of the multiple extraction and stranded preposition.

adjoined to the highest clausal node and is arguably higher than [Spec,CP], as shown by cases in which the topic precedes an initial wh-phrase (12).

(12) To Mary$_j$, what$_i$ are you going to give $t_i$ $t_j$?

In the case of negative inversion, the initial constituent is in the position that fronted wh-phrases appear in, [Spec,CP], as evidenced by subject-aux inversion. And in the case of stylistic inversion, the initial constituent is in the subject position [Spec,IP] (see Culicover and Levine 2001). These configurations are summarized in (13).

(13) a. [$_{\text{CP}}$ XP [$_{\text{CP}}$...]]
b. [$_{\text{CP}}$ [$_{\text{Spec}}$ XP] [$_{\text{IP}}$...]]
c. [$_{\text{IP}}$ XP [$_{\text{I}'}$...]]

Since the structures are all different, there is no simple configurational generalization that can be used to rule out the unacceptable examples. We could of course stipulate an abstract analysis in which XP occupies the same position in all three constructions, but this would be ad hoc.

The simplest generalization that covers these three cases is that in each case there is no overt marker of the relative clause, and the subject is not in initial position in the clause. It is therefore plausible that the unacceptability is due to difficulty processing the relative clause when the two most common indicators that what is being processed is a relative clause are absent.

In fact there is independent evidence that suggests that this is the problem. In stacked relative clauses, a zero-relative is most acceptable when it is the first clause in the sequence, and hence immediately adjacent to the head N, as in (14a). As we move the zero-relative further from the head, as in (14b,c), acceptability decreases.

(14) a. (I'll tell you about) the actor$_i$ [∅ I interviewed $t_i$] [ $\left\{ \begin{array}{c} \text{who} \\ \text{that} \end{array} \right\}$ I didn't like $t_i$very much] [ $\left\{ \begin{array}{c} \text{who} \\ \text{that} \end{array} \right\}$ we just saw $t_i$ in a movie last week]

b. ? (I'll tell you about) the actor [ $\left\{ \begin{array}{c} \text{who} \\ \text{that} \end{array} \right\}$ I interviewed $t_i$] [∅ I didn't like $t_i$ very much] [ $\left\{ \begin{array}{c} \text{who} \\ \text{that} \end{array} \right\}$ we just saw $t_i$ in a movie last week]

c. ?? I'll tell you about) the actor [ $\left\{ \begin{array}{c} \text{[who} \\ \text{that} \end{array} \right\}$ I interviewed $t_i$] [ $\left\{ \begin{array}{c} \text{who} \\ \text{that} \end{array} \right\}$ I didn't like $t_i$ very much] [∅ we just saw $t_i$ in a movie last week]

The fact that (14b,c) are mildly unacceptable is consistent with the the idea that the source of unacceptability here is not a matter of syntactic configuration per se, but of processing complexity. While the left edge of the zero-relative is not marked, it is not obscured by an initial non-subject.

Finally, it is possible to have sentential subjects in relative clauses instead of NPs. These configurations are complex, but they are more or less acceptable when the relative clause is introduced by a relative pronoun. However, when there is no marker for the relative clause, complete unacceptability follows. The (a) examples in (15) show marked relative clauses with sentential subjects, and the (b) examples show unmarked relative clauses with the same sentential subjects.[5]

(15) *that*-clause

a. ? Otto appears to be a man [ $\left\{ \begin{array}{l} \text{who} \\ \text{that} \end{array} \right\}_i$ [S that it is snowing hard] apparently doesn't bother $t_i$].

b. * Otto appears to be a man$_i$ [ $\emptyset_i$ [S that it is snowing hard] apparently doesn't bother $t_i$].

(16) *for-to* infinitive

a. ? Colette is the kind of woman [ $\left\{ \begin{array}{l} \text{who} \\ \text{that} \end{array} \right\}_i$ [S for us to speak better French] would probably have pleased $t_i$].

b. * Colette is the kind of woman [ $\emptyset_i$ [S for us to speak better French] would probably have pleased $t_i$].

(17) embedded wh-question

a. ? We interviewed a candidate [ $\left\{ \begin{array}{l} \text{who} \\ \text{that} \end{array} \right\}_i$ [S whether it is polite to make eye contact] apparently was not obvious to $t_i$].

b. * We interviewed a candidate [$\emptyset_i$ [S whether it is polite to make eye contact] apparently was not obvious to $t_i$].

It is important to stress that the judgments shown here are qualitatively no different from those that have been encountered in many other contexts in the course of syntactic theorizing over the past fifty-plus years. Some examples are strongly unacceptable, so that we might want to say that they are "ungrammatical". Others are mildly unacceptable. In the case of phenomena such as extraction from islands it was possible to identify a relatively simple syntactic configuration that could be held responsible for the unacceptability. The same can be said for many other analyses, while

[5]These and other unacceptable examples can be rendered more acceptable by using intonation to signal the clause, which is another indication that what we are dealing with here is processing and not grammar.

recognizing that in many cases only the strongest unacceptability judgments could be accounted for in grammatical terms, while exceptions and gradient judgments were left out of the account. In the present case, we have seen that zero-relative clauses are unacceptable in cases of (i) topicalization, (ii) negative inversion, (iii) stylistic inversion, (iv) stacking and (v) sentential subjects. The diversity of syntactic configurations responsible for the unacceptability suggests that there is in fact no grammatical explanation for these judgments about zero-relatives.

This state of affairs is precisely what SS predicts. SS rules out (other things being equal), grammatical constraints that rule out otherwise well-formed syntactic configurations. SS forces us to seek explanations for such phenomena outside of syntax proper, e.g. in terms of processing, semantic well-formedness, pragmatics, discourse coherence, etc. As suggested above, a plausible place to look in the case of zero-relatives would be sentence processing. The reduction in acceptability occurs for reasons having to do with the particular linear string of elements, i.e. when there is no overt marker of a relative clause and the subject of the relative clause is not immediately adjacent to the head. My proposal is that in such a case, it is more difficult for the processor to recognize that there is a relative clause and correctly project the appropriate structure for further processing.

Why should difficulty of recognition lead to judgments of unacceptability? Assume, following Jackendoff (2002; 2007), that pieces of syntactic structure and their corresponding interpretations are "pieces of structure stored in memory". Assume as well that the processing of a sentence proceeds from the beginning of the sentence by projecting possible continuations of the string in the form of hypothesized projected structure. For example, if the sequence is *the man that*, the complementizer *that* triggers a rule in the processor that projects the structure [$_{\text{DP}}$ *the man* [$_{\text{S-REL}}$ *that* ...]]. If the sequence is *the statement that*, there are (at least) two rules triggered: [$_{\text{DP}}$ *the fact* [$_{\text{S}}$ *that* ...]], as in *the fact that I disputed* and [[$_{\text{DP}}$ *the fact* [$_{\text{S-REL}}$ *that* ...]], as in *the fact that I disputed the result.*

Since at most if not all points in the processing there is typically more than one possible continuation, a plausible theory of sentence processing may take the perspective of a probabilistic phrase structure grammar, in which the probability of each expansion of a phrase at any point in the processing of the string is determined at least in part by the relative frequency of its structure in the corpus on which the parser has been trained (e.g. Nguyen et al. 2012).[6] An additional assumption is that the processing complexity of a sequence and, in extreme cases, its acceptability, is determined in part by the correspondence between the projected structure in the course of processing,

[6]The low or zero probability of a string of words with familiar structure but unfamiliar lexical items is not predicted to be unacceptable on this formulation. But a construction specified in terms of specific lexical items is predicted to be unacceptable when it contains a lexical item that is not typically used in this construction.

and the actual structure settled on (Hale 2001; 2003; Levy 2005; 2008). In the case of *the fact that*, for example, the probability of the sentential complement analysis is greater than that of the relative clause analysis. In the relative clause case there is a significant slowdown in reading times, suggesting that there is greater processing complexity at the gap, when the processor realizes that the correct structure is that of a relative clause (Chen et al. 2005). Ungrammatical sentences of course have virtually zero probability of occurrence, so they will be judged highly unacceptable.

Consider now the case of the zero-relative. When there is no explicit marker at the left edge of the relative clause, the processor must depend on other familiar evidence to project the structure, i.e. an initial subject DP. Topicalization in relative clauses even with overt markers appears to be very rare. Since zero-relatives are frequent in English, a sequence such as *a man we (could never grant liberty to)* can be reliably assigned the structure [$_{\text{DP}}$ *a man* [$_{\text{S-REL}}$ *we* ...]], where *we* is clearly the subject of the relative clause. In example (15) the initial sequence is *a man liberty*, which can be reliably assigned the structure [[$_{\text{DP}}$ [$_{\text{S-REL}}$ *liberty* ...], where *liberty* is the subject of the relative clause. But in the sequence *a man liberty we*, the subject of the relative clause is actually *we*. There is no basis for treating *liberty* as a topicalized DP with a following subject DP, presumably because the sequence DP-DP-DP does not occur in the corpus. The absence of a rule for processing this sequence leads the process to engage in some type of repair strategy, with a corresponding reduction in acceptability.

Finally, the less than fully acceptable zero-relatives appear to be very rare or non-existent in the corpus. The interaction between probability and judgments in extreme cases appears to account for the judgments in examples (9)-(17).

### 3.3 Freezing

Consider in this regard the phenomenon of 'freezing'. Ross (1967:305) observed that extraction from a PP that has been extraposed is reduced in acceptability, as shown by (18b).

(18) a. You saw [a picture] yesterday [$_{\text{PP}}$ of Thomas Jefferson].
b. ? Who$_i$ did you see [a picture $t_j$] yesterday [$_{\text{PP}}$ of $t_i$]$_j$?

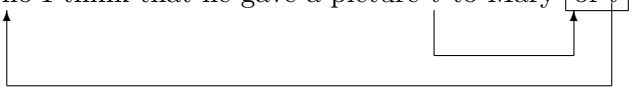
Ross's (1967) formulation of the Frozen Structure Constraint in (15) deals specifically with such examples.

(19) The Frozen Structure Constraint (FSC): If a constituent C, where C is a clause or a prepositional phrase, has been extraposed from a noun phrase whose head is lexical, this noun phrase may not be moved, nor may any element of C be moved out of C (pp. 160, 165).

Subsequently, Wexler and Culicover (1980) proposed the Freezing Principle, based on considerations of language learnability. The basic idea was that a structure that is created transformationally that is not compatible with the base phrase structure rules of a language is frozen. Such a derivation is non-structure-preserving, in the sense of Emonds (1970; 1976). However, note that an extraposed PP is in the position of an argument or adjunct PP in the VP, and hence should not be frozen on the Wexler-Culicover definition of freezing.

Culicover and Winkler (2013) in fact propose that the unacceptability of extraction from an extraposed PP depends in large part, if not entirely, on the fact that it demands that the gap corresponding to the extracted filler is inside of a constituent whose connection to a preceding head is unanticipated, which results in processing complexity. The structure is given in (20).

(20) the person who I think that he gave a picture $t$ to Mary $\boxed{\text{of } t}$



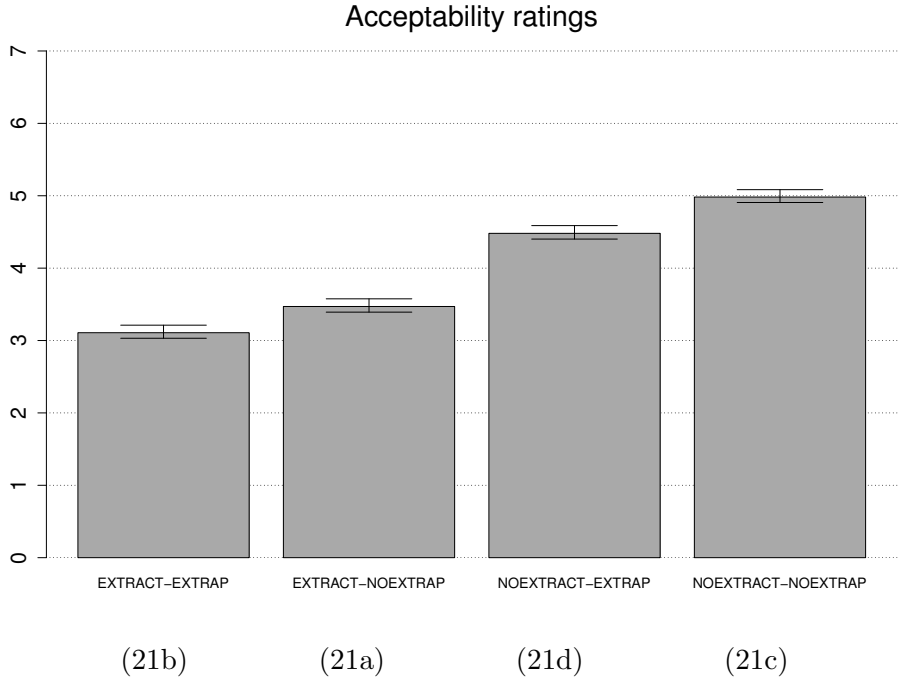
As before, if processing complexity leads to the avoidance of certain configurations, such configurations will have lower probability and therefore produce judgments of lower acceptability. With this in mind, Hofmeister et al. (2012) did several experiments to confirm that distance-based effects on acceptability judgments occur in the case of extraction and in the case of extraposition. The results strongly resemble findings from the psycholinguistics literature on effects of dependency locality (Gibson 1998; 2000; Grodner and Gibson 2005): in general, the longer the dependency, the lower the acceptability judgment. Hofmeister et al. constructed an experiment in order to determine if the acceptability judgments due to extraposition and extraction in combination are in some way dependent on the two factors occurring together in the same examples, that is, if there is a measurable freezing effect. A sample of the examples used in this experiment is given in (21).

(21) a. Tell me which actor your friend read a story about twice while having breakfast.

b. Tell me which actor your friend read a story twice about while having breakfast.

c. You told me your friend read a story about an actor twice while having breakfast.

d. You told me your friend read a story twice about an actor while having breakfast.




(21b) (21a) (21d) (21c)

Figure 1: Mean acceptability judgments from Experiment. Error bars show ± 1 standard error.

This design allows us to determine how much extraposition and extraction independently lower judgments, and whether combining the two lowers judgments beyond what is expected on the basis of each independent source of unacceptability.

The results of this experiment are summarized in Figure 1. The figure shows that there is no interaction: extraposition is no worse in contexts with extraction, (21b), than in contexts without, (21d). The freezing violations in (21b) have an average rating that is predictable on the basis of the independent average penalties for extraposition and extraction. The data suggest that the low ratings for freezing violations are in fact attributable solely to the combined penalties resulting from extraction and extraposition. This experiment appears to eliminate the need for any constraint specific to the freezing configuration. Rather, this freezing effect appears to be due entirely to processing complexity.

This analysis of freezing phenomena illustrates once again the implications of SS for explanation of judgments. These acceptability judgments cannot be accounted for in the grammar per se, because the configurations that produce them cannot be formulated in terms of constructions. The sentences are strictly speaking well-formed, in that every local configuration conforms to the requirements of the grammar. Therefore, SS suggests that

the judgements fall outside of grammar. Processing complexity may not be the correct account in every case, but at least in the cases I have reviewed here it appears to offer a plausible account.[7]

## 4 Hard and soft constraints

The line drawn by SS brings to mind a distinction made by Sorace and Keller (2005), as follows: "We assume a fundamental dichotomy between hard constraints (that trigger categorical linguistic judgments) and soft constraints (that trigger gradient judgments)." Since SS distinguishes sharply between those phenomena that are the province of well-formedness and those that are not, the distinction drawn by SS may provide a basis for the hard/soft distinction. Sorace and Keller categorize as 'hard' constraints cases such as the following:

(22) Hard constraints on extraction
   a. INVERSION (INV): subject and auxiliary have to be inverted.
   b. AGREEMENT (AGR): subject and verb have to agree in number.
   c. PRESUMPTIVE (RES): resumptive pronouns are disallowed in wh-questions.

These are all what we would take to be constructional well-formedness conditions and therefore a matter of grammar – by definition, they admit of no exception. On the other hand, 'soft' constraints are those that yield gradient judgments. SS requires that gradient phenomena fall outside of the grammar – they must be the result of variable processing complexity, or perhaps reflect aspects of discourse complexity and coherence that are sensitive to contextual effects. Some examples of 'soft' constraints cited by Sorace and Keller with respect to extraction from NP are the following:

(23) Soft constraints on extraction
   a. DEFINITENESS (DEF): a picture NP is marked [-DEFINITE].

[7] When it comes to irregularities and idiosyncrasies, processing complexity certainly is not the only plausible story. Low or zero frequency of occurrence in the corpus can occur for a variety of reasons that have nothing to do with complexity. For instance, there are collocations, such as *(to look) high and low* but **(to look) low and high*. In the sluice-stranding construction (Culicover 1999), certain combinations of wh-phrase and preposition are possible while others are not, for no apparent reason, e.g. *who with, what about* but **who about, *who next to, *what under*. Regardless of the source, low frequency appears to produce judgments of unacceptability, however. This said, frequency must be measured against a background of expectation, since individual words, phrases and especially sentences may be of very low frequency and yet be perfectly acceptable, as Chomsky (1965) noted.

b. VERBCLASS (VERB): a verb subcategorizing for a picture NP has to be marked [-EXISTENCE].

c. REFERENTIALITY (REF): an NP extracted from a picture NP has to be marked [+REFERENTIAL].

These constraints all arguably reflect complexities in the construction of coherent discourse representations.

The differences between these two types of constraints are precisely what SS predicts. Sorace and Keller's hard constraints are those that produce strong unacceptability, are minimally sensitive to context and show no developmental optionality, that is, they are faithfully observed in development. Soft constraints, on the other hand, produce mild unacceptability, are sensitive to context and show developmental optionality. In present terms, the 'hard' constraints produce judgments of ungrammaticality, while the 'soft' constraints produce extragrammatical judgments of unacceptability on well-formed structures.

## 5 Conclusions

SS is a constructional theory that makes very restrictive assumptions about what falls within the domain of syntactic competence. Competence is limited to well-formedness as defined by constructions, as exemplified by (2).

(2) TRANSITIVE VP

$$\begin{bmatrix} \text{PHON} & [\varphi]_1\text{-}[\varphi]_2 \\ \text{SYN} & [_{\text{VP}}\ \text{V}_1,\ \text{NP}_2] \\ \text{CS} & \mathbf{V}_1(\mathbf{NP}_2) \end{bmatrix}$$

So, if a VP is not properly linearized, or if its interpretation does not conform to the CS as given here or to any idiomatic CS, the sentence that contains it is ungrammatical. Any sentence that is fully licensed by conforming to the set of constructions of the language is grammatical in the strict sense. If such a sentence is judged to be unacceptable in some way, the unacceptability judgment is not a matter of grammar, but something else. For the cases that I have discussed, I suggest that processing is responsible.[8]

In sum, SS draws the line between competence and performance so that everything that does not have to do with satisfying the conditions of constructions in a grammar is not competence. All judgments that cannot be accounted for in constructional terms must be explained in other ways. Of course, SS does not automatically provide explanations for such judgments

[8] I suspect that this is the case for a substantial number of judgments cited in the literature as "ungrammatical" that have formed the basis of proposals that have found their way into the theoretical literature, but this is far too big an issue to take on here.

– theories of processing, discourse structure and so on have to be independently formulated and verified. But it does make clear claims about where to find explanations, and, to the extent that the explanations are there, one can argue that progress is being made.

## References


Baltin, Mark. 1981. Strict bounding. The logical problem of language acquisition, ed. by Lee Baker and John Mccarthy, 247-95. Cambridge, MA: MIT Press.

Chen, Evan, Edward Gibson, and Florian Wolf. 2005. Online syntactic storage costs in sentence comprehension. Journal of Memory and Language 52.144-69.

Chomsky, Noam. 1965. Aspects of the theory of syntax. Cambridge, MA: MIT Press.

Chomsky, Noam. 1977. On wh movement. Formal syntax, ed. by Peter W. Culicover, Thomas Wasow, and Adrian Akmajian, 71-132. New York: Academic Press.

Chomsky, Noam. 1986. Knowledge of language: Its nature, origin, and use. New York: Praeger.

Chomsky, Noam and George A Miller. 1963. Introduction to the formal analysis of natural languages. Handbook of Mathematical Psychology, vol. 2, ed. by R.D. Luce, R.R. Bush, and E. Galanter, New York: Wiley.

Culicover, Peter W. 1992. Topicalization, inversion, and complementizers in English. OTS Working Papers. University of Utrecht, Utrecht. Going Romance and Beyond pp 1-43. Reprinted in Peter W. Culicover. 2013. Explaining Syntax. Oxford: Oxford University Press.

Culicover, Peter W. in press. English zero-relatives and the competence-performance distinction. Papers dedicated to Mike Harnish. International Review of Pragmatics.

Culicover, Peter W. and Ray Jackendoff. 2005. Simpler syntax. Oxford: Oxford University Press.

Culicover, Peter W. and Ray Jackendoff. 2006. The simpler syntax hypothesis. Trends in Cognitive Sciences 10.413-18.

Culicover, Peter W., Levine, Robert. D. 2001. Stylistic inversion in English: A reconsideration. Natural Language & Linguistic Theory 19, 283-310. Reprinted in Peter W. Culicover. 2013. Explaining Syntax. Oxford: Oxford University Press.

Culicover, Peter W. and Susanne Winkler. 2013. Freezing: A conspiracy. Unpublished ms. Ohio State University and University of Tuebingen.

Emonds, Joseph. 1970. Root and structure preserving transformations. Bloomington, Indiana: Indiana University Linguistics Club.

Emonds, Joseph. 1976. A transformational approach to English syntax. New York: Academic Press.

Fillmore, Charles J. 1988. The mechanisms of construction grammar. BLS 14.35-55.

Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. Cognition 68.1-76.

Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. Image, language, brain, ed. by Yasushi Miyashita, Alec Marantz, and Wayne O'Neil, 95-126. Cambridge, MA: MIT Press.

Grodner, Daniel J. and Edward A. F. Gibson. 2005. Consequences of the serial nature of linguistic input for sentential complexity. Cognitive Science 29.261-91.

Hale, John T. 2001. A probabilistic Earley parser as a psycholinguistic model. Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies. 1-8.

Hale, John T. 2003. The information conveyed by words in sentences. Journal of Psycholinguistic Research 32.101-23.

Hofmeister, Philip. 2011. Representational complexity and memory retrieval in language comprehension. Language and Cognitive Processes 26.376-405.

Hofmeister, Philip, Inbal Arnon, T. Florian Jaeger, Ivan A. Sag, and N. Snider. submitted. The source ambiguity problem: Distinguishing the effects of grammar and processing on acceptability judgments. Language and Cognitive Processes

Hofmeister, Philip, Peter W. Culicover, and Susanne Winkler. 2012. Effects of processing on the acceptability of 'frozen' extraposed constituents. To appear, Syntax.

Hofmeister, Philip, T. Florian Jaeger, Ivan A. Sag, Inbal Arnon, and Neal Snider. 2007. Locality and accessibility in wh-questions. Roots: Linguistics in search of its evidential base, ed. by Sam Featherston and Wolfgang Sternefeld, 185-206. Berlin: de Gruyter.

Hofmeister, Philip and Ivan A. Sag. 2010. Cognitive constraints and island effects. Language 86.366-415.

Hofmeister, Philip, Laura Staum Casasanto, and Ivan A. Sag. to appear. Islands in the grammar? Standards of evidence. Experimental syntax and island effects, ed. by Jon Sprouse and Norbert Hornstein, Cambridge: Cambridge University Press.

Jackendoff, Ray. 2002. Foundations of language. Oxford: Oxford University Press.

Jackendoff, Ray. 2007. A parallel architecture perspective on language processing. Brain Research 1146.2-22.

Kay, Paul. 2005. Argument structure constructions and the argument-adjunct distinction. Grammatical constructions: Back to the roots, ed. by Mirjam Fried, 71-98. Amsterdam: John Benjamins.

Kayne, Richard S. 1981. ECP extensions. Linguistic Inquiry 22.93-133.

Kluender, Robert. 1992. Deriving island constraints from principles of predication. Island constraints: Theory, acquisition and processing, ed. by Helen Goodluck and Michael Rochemont, 223-58. Dordrecht: Kluwer.

Kluender, Robert. 1998. On the distinction between strong and weak islands: A processing perspective. The limits of syntax, ed. by Peter W. Culicover and Louise Mcnally, 241-79. New York: Academic Press.

Kluender, Robert. 2004. Are subject islands subject to a processing account? Proceedings of wccfl 23, ed. by Benjamin Schmeiser, Vineeta Chand, Ann Kelleher, and Angelo Rodriguez, 101-25. Somerville, MA: Cascadilla Press.

Levy, Roger. 2005. Probabilistic models of word order and syntactic discontinuity. Department of Linguistics: Stanford University dissertation.

Levy, Roger. 2008. Expectation-based syntactic comprehension. Cognition 106.1126-77.

Nguyen, Luan, Marten Van Schijndel, and William Schuler. 2012. Accurate unbounded dependency recovery using generalized categorial grammars. Proceedings of the 24th International Conference on Computational Linguistics (COLING '12).

Ross, John R. 1967. Constraints on variables in syntax. Cambridge, MA: MIT dissertation.

Sag, Ivan A., Philip Hofmeister, Inbal Arnon, Neal Snider, and Florian Jaeger. 2008. Processing accounts for superiority effects. Stanford.

Sag, Ivan A., Philip Hofmeister, and Neal Snider. 2007. Processing complexity in subjacency violations: The complex noun phrase constraint. Proceedings of the 43rd Annual Meeting of the Chicago Linguistic Society. Chicago.

Sorace, Antonella and Frank Keller. 2005. Gradience in linguistic data. Lingua 115.1497-524.

Staum Casasanto, Laura, Hofmeister, Philip, and Ivan A. Sag. 2010. Understanding acceptability judgments: Additivity and working memory effects. Proceedings of the 32nd Annual Conference of the Cognitive Science Society. Austin, TX.

Wexler, Kenneth and Peter W. Culicover. 1980. Formal principles of language acquisition. Cambridge, MA: MIT Press.