Syntactic change in the Parallel Architecture: The case of parasitic gaps*

Peter W. Culicover

1 Introduction

By way of introduction, I quote from the beginning of my overview chapter (Culicover 2001) for a volume of articles on parasitic gaps that Paul Postal and I edited (Culicover and Postal 2001). The notation E83 refers to Elisabet Engdahl's seminal 1983 article on the topic.

> The parasitic gap (henceforth P-gap) construction is exemplified by the following examples. The first gap, marked *t*, is called a 'true gap' because it is in a position that normally permits extraction; e.g., *which articles did John file t.* The second gap, marked *pg* for 'parasitic gap,' appears in a location that normally

*It is a tremendous privilege to contribute to this issue; my debt to Ray Jackendoff as a friend and collaborator is beyond estimation. I am grateful to Jefferson Barlew, Rebecca Morley, Heidi Quinn and Richard Samuels for discussions about the material discussed here. This paper was first presented at the Rumelhart Symposium in honor of Ray Jackendoff at the 2014 Cognitive Science Society Meeting in Québec City. I am grateful to the audience at that meeting and to the University of Canterbury Visiting Erskine Fellowship for the opportunity to make numerous improvements. All errors are of course my responsibility.

does not permit extraction.

(1) a. Which articles did John file *t* without reading *pg*? [E83:5,(1)]

b. This is the kind of food you must cook *t* before you eat *pg*. [E83:5,(2)]

The key property of the P-gap construction is that a single filler, e.g. *which articles*, is the antecedent of more than one gap. Engdahl (1983) notes the following examples in support of this point.

(2) a. Here is the paper that John read *t* before filing *pg*. [E83:14,(35a)]

b. ? Here is the paper that John read his mail before filing *t*. [E83:14,(35b)]

c. Here is the paper that John read *t* before filing his mail. [E83:14,(35c)]

(3) a. Who did John's talking to *pg* bother *t* most? [E83:14,(36a)]

b. ? Who did John's talking to *t* bother you most? [E83:14,(36b)]

c. Who did John's talking to Mary bother *t* most? [E83:14,(36c)]

The ungrammatical examples are typical violations of extraction constraints (Chomsky 1973, 1981, 1986).

A striking fact that the investigation of P-gaps in Culicover and Postal (2001) made very clear was that while some languages, like English and Swedish, allow P-gaps in a wide variety of syntactic contexts, some languages, like Spanish, allow P-gaps only in very restricted contexts, and some languages, like German and Dutch, do not allow P-gaps at all. This is the case in spite the fact that P-gaps appear to share many properties with across-the-board A′ extraction in coordination; see (4).

(4) Here is the paper that John read *t* and then filed *t*.

ATB extraction is found both in languages with P-gaps and languages without parasitic gaps.

Moreover, it appears that if a language allows P-gaps at all, it allows P-gaps in 'without'-clauses, e.g. Spanish *sin*-clauses; see (5).

(5)

los	articulos	que	archivaste	*t*	sin	leer	*pg*
the	articles	that	you.filed	*t*	without	to-read	*pg*

'the articles that you filed without reading (them).'

(Bordelois 1986)

Some, but not all, languages also permit P-gaps with other subordinating conjunctions like *after, before, although*, as in (6).

(6) We found out which talk$_i$ she criticized t_i after trying very hard to understand pg_j.

A few languages, such as English and Swedish, permit a more complex type of 'non-conjunctive' P-gap construction, e.g. (7).

(7) We found out which talk$_i$ everyone who listened to pg_i failed to understand t_i.

Thus, there appears to be a parasitic gap hierarchy, summarized in (8).

(8) *Parasitic Gap Hierarchy*

without > subordinate > non-conjunctive

These observations raise some fundamental questions:

- why do parasitic gaps exist at all?
- why are different syntactic configurations possible for P-gaps? and why just these configurations?
- why is there a Parasitic Gap Hierarchy?

The answers that I suggest here make crucial use of constructional overgeneralization, formulated straightforwardly in the framework of Jackendoff's Parallel Architecture (Jackendoff 1997, 2002).

There have of course been prior analyses of parasitic gaps in the literature. What is most striking is that for the most part they stipulate that P-gaps exist and encode this stipulation in some notational system. For example, in the *Barriers* framework, Chomsky (1986) accounts for the acceptability of a P-gap by assuming that it is the product of the movement of an invisible operator to the left edge of the island, which forms a chain with the gap, and a stipulated 'chain composition' mechanism that links this operator to the chain of the true gap.

The GPSG account of P-gaps in Sag (1983) differs from Chomsky's account in that it relates the existence of P-gaps to ATB extraction. Sag proposes a "parasitic gap metarule" that says that if a structure contains a constituent with a gap, then it may have another constituent with a gap as well. This is also a stipulation: it says that English has a P-gap construction. A language that lacks the metarule, lacks the construction.

Munn (1992, 2001) goes beyond the *Barriers* and GPSG accounts by observing some striking similarities between ATB extraction in coordinate structures and P-gaps. He argues on the basis of these similarities that coordinate structures are essentially the same as the structures in which P-gaps are licensed. In the canonical P-gap construction, the P-gap is in a subordinate clause, as in (1a). Munn's proposal is that the second conjunct in ATB extraction is in the same configuration with respect to the first clause, and undergoes *Barriers*-style chain-composition, as in (9).

(9) Who$_i$ did [[$_{\text{IP}}$ John see t_i] [$_{\text{BP}}$ Op$_i$ [$_{\text{B}'}$ and [$_{\text{IP}}$ Bill talk to e_i]]]]

My goal in this article is to explore the mechanisms behind the stipulations of earlier work on P-gaps, in order to find answers to the questions raised above. Sag's and Munn's association of parasitic gaps with ATB extraction is fundamental. However, I propose that in order to fully explain the connection between the two phenomena, and the hierarchy noted above, a constructional perspective is required, and particularly one that addresses constructional change. I assume Jackendoff's Parallel Architecture for this purpose. A formulation of the phenomena in terms of the correspondences of the Parallel Architecture provides a natural framework in which to articulate gradual transitions between grammars, each of which incorporates a more general version of ATB/P-gap structures than the one before. Through this approach it is possible to see not only the necessity of the relationship between ATB and P-gaps, but to explain what kinds of P-gap constructions are possible and to constrain the range of variation.

The organization of this paper is as follows. Section 2 sketches the basic features of the constructional framework. Section 3 defines the basic constructions used in the analysis. Section 4 focuses on the analysis of ATB within the constructional framework. Section 5 shows how a plausible mechanism of language change gives rise to the extensions of ATB that result in P-gaps. Section 6 details some of the empirical predictions of the constructional account and draws some general conclusions about the constructional approach to grammar.[1]

2 A constructional framework

Following the Parallel Architecture of Jackendoff (2002), I assume that there is a component of the grammar that generates all of the well-formed phonological strings, called PHON, another component of the grammar that generates all of the well-formed syntactic structures, called SYN, and a component of the grammar that generates all of the well-formed meanings (conceptual structures), called CS. These components are called TIERS.

A CONSTRUCT is the representation of an individual expression, while a CONSTRUCTION is a generalization over expressions. The description of a

[1]Still to be explained, and beyond the scope of this study, is why the ATB extension to parasitic gaps ameliorates a gap that is otherwise less than fully acceptable.

construction may contain variables, while that of a construct contains only constants. Constructs and constructions reflect correspondences between terms on different tiers. A construct is LICENSED if the correspondences between the tiers instantiate correspondences expressed by constructions of the language. (For formal definitions of these and related terms, see Barlew and Culicover 2015.)

Consider, for instance, the construction that licenses VP in English. It says that a VP may consist of an expression of category V followed by other, possibly null, material; it has the form in (10). 1 in PHON is the phonological string corresponding to V_1 and so on. The linear order is specified in PHON, and the hierarchical structure of the phrase is specified in SYN.

(10) V-INITIAL VP

$$\begin{bmatrix} \text{PHON} & 1 > 2 \\ \text{SYN} & [_{\text{VP}}\ \text{V}_1,\ \text{X}_2] \end{bmatrix}$$

A token that satisfies the conditions of this construction instantiates this construction. Such a token must have a phonological string correctly corresponding to a verb, it may have a phonological string corresponding to other constituents of the VP, and the strings must be in the order given.

In the Parallel Architecture, a word is also a construction. It specifies a correspondence between a string of sounds, syntactic properties, and a meaning. For example, (11) shows the lexical entry of the verb *kick*. For simplicity, I use the spelling of the word as the phonological representation of the word.[2]

[2]It can be seen that the entry in (11) is approximate, because it does not take into account the fact that *kick* may take different phonological forms depending on its inflection:

(11) $\begin{bmatrix} \textit{kick} & \\ \textsc{phon} & \text{kick}_1 \\ \textsc{syn} & [\text{V}_1] \\ \textsc{cs} & \lambda\text{y}.\lambda\text{x.kick}'_1(\textsc{agent:x,theme:y}) \end{bmatrix}$

Each term of each tier of a well-formed construct must instantiate in its correspondences with elements on other tiers some condition of at least one construction. For instance, in *John kicked Fido*, PHON must have the forms of the words in the appropriate order, and the meaning must be one in which the relation kick′ holds of the entities j, the CS representation of *John*, and f, the CS representation of *Fido*. The construct is given in (12).

(12) $\begin{bmatrix} \textit{John kicked Fido} & \\ \textsc{phon} & \text{john}_2\text{–kick}_1\text{–fido}_3 \\ \textsc{syn} & [_{\text{S}}\ \text{NP}_2,\ [_{\text{VP}}\ \text{V}_1,\ \text{NP}_3]] \\ \textsc{cs} & \lambda\text{y}.\lambda\text{x.kick}'_1[\textsc{agent:x,patient:y}](j_2)(f_3) \end{bmatrix}$

This representation assumes correspondence rules that link the subject to the x-argument and the object to the y-argument.

3 Some basic constructions

This section defines two basic constructions. One determines the distribution and interpretation of an A′ operator and a gap, and the other licenses

kick, kicks, kicked, kicking. For irregular verbs, the set of forms is larger, e.g. *go, goes, went, gone, going.* In fact, what must be included somehow in the lexical representation is the 'paradigm function' (Stump 2001), that is, the set of alternative forms for the word that are compatible with various syntactic and semantic contexts. I put this issue aside here.

coordination. They are combined in section 4 to form the ATB construction. That construction in turn is the foundation for the generalizations that give rise to the possibility of P-gap constructions in section 5.

The basic A′-extraction construction is (13). (There are other A′ constructions that do not conform exactly to (13), but this simplified variant will serve to illustrate how such constructions interact with coordination.) The null element in PHON corresponds to the λ-bound variable x in 5′, the CS representation of the clause XP_3. The extracted XP corresponds to an operator that binds the variable α in CS.

(13)

A′-EXTRACTION	
PHON	$1\text{-}2\text{-}\emptyset_3\text{–}4$
SYN	XP_1, $[_S\ Y_2, XP_3, Z_4]_5$
CS	$OP_1^{\alpha}[\lambda x.5'](\alpha_1)$

Recall that a construct is licensed if all of the correspondences between terms on the tiers of the construct instantiate conditions imposed by the constructions in the grammar. In this case, (13) licenses an example such as the question *(the people) who$_i$ we were talking to t_i at the party*, as shown in (14). The operator REL_1^{α} and the variable α[PERSON] corresponding to *who* that it binds wrap around the lambda expression that corresponds to the clause containing the gap.

(14) *who we were talking to t at the party*

$$\left[\begin{array}{ll} \textsc{phon} & \text{who}_1\text{-[we-were-talking-to]}_2\text{-}\emptyset_3\text{-[at-the-party]}_4 \\ \textsc{syn} & \text{NP}_1,\ [_\text{S}\ \text{Y}_2,\ \text{NP}_3,\ \text{Z}_4]_5 \\ \textsc{cs} & \text{REL}_1^{\alpha}[\lambda \text{x}.[\underbrace{\text{we-talk-to}}_{2}\text{-x}_3\text{-}\underbrace{\text{at-the-party}'}_{4}]_5](\alpha[\textsc{person}]_1) \end{array}\right]$$

By lambda-reduction, the argument containing α[PERSON] is substituted for the lambda-bound variable x_3.[3]

(15) $\text{REL}_1^{\alpha}[\lambda \text{x}.[\underbrace{\text{we-talk-to}}_{2}\text{-x}_3\text{-}\underbrace{\text{at-the-party}'}_{4}]_5](\alpha[\textsc{person}]_1)$
$\Rightarrow \text{REL}_1^{\alpha}[\underbrace{\text{we-talk-to}}_{2}\text{-}\alpha[\textsc{person}]_1\text{-}\underbrace{\text{at-the-party}'}_{4}]_5$

We turning now to the construction for coordination in English. It states that the conjuncts are linearly ordered, with the conjunction preceding the last one.

(16)
$$\left[\begin{array}{ll} \textsc{phon} & \text{1-}\ldots\text{-(n-1)-0-n} \\ \textsc{syn} & [\left\{\begin{array}{c}\text{and}\\ \text{or}\end{array}\right\}_0,\ \text{XP}_1,\ \ldots,\ \text{XP}_n\ ] \\ \textsc{cs} & \left\{\begin{array}{c}\wedge\\ \vee\end{array}\right\}'_0(1',\ldots,\text{n}') \end{array}\right]$$

I assume that *and* and *or* are sui generis grammatical features, each of which corresponds to a CS logical connective.[4]

---

[3] This is equivalent to 'reconstruction' (Culicover 2013a). If the A′ constituent contains the interrogative phrase, the gap in PHON does not correspond to the variable α. But reduction produces a CS representation in which the operator binds the variable. E.g.,

(i) *to whom we were talking t at the party*
$\text{REL}^{\alpha}[\lambda\text{x.[we-talk-x-at-the-party}'](\text{[to-}\alpha]) \Rightarrow \text{REL}^{\alpha}\text{[we-talk-to-}\alpha\text{-at-the-party}']$

[4] The question of what properties of the coordinated XPs make them amenable to coor-

4 The ATB construction

Consider now how across the board, or ATB extraction, may emerge as an individuated construction that can be further generalized to yield P-gaps. The characteristic of ATB extraction is that there is a gap in every conjunct of a coordinate structure. ATB A′-extraction is possible even when there is no constituent in overt A′ position, as in (17).

(17) a. a person (that) [$_S$ Sandy talked to *t* but everyone else ignored *t*]

b. a person (that) Sandy [$_{VP}$ talked to *t* and didn't convince *t*]

c. a person (that) Sandy talked [$_{PP}$ to *t* and about *t*]

d. a person (that) Sandy was [$_{AP}$ impressed by *t* and angry at *t*]

In order to figure out how to state the correspondence between such a structure and its interpretation, consider (17c). The representation of the construct is essentially (18).

$$(18)\quad \begin{bmatrix} \text{PHON} & \text{person}_1\text{-Sandy}_2\text{-talk}_4\text{ed}_3\text{-to}_5\text{-}\emptyset_6\text{-and}_7\text{-about}_8\text{-}\emptyset_9 \\ \text{SYN} & [\text{N}_1, [_{\text{S}}\ \text{NP}_2, \text{AUX}_3, [_{\text{VP}}\ \text{V}_4, [_{\text{PP}}\ [_{\text{PP}}\ \text{P}_5, \text{NP}_6], \text{and}_7, \\ & [_{\text{PP}}\ \text{P}_8, \text{NP}_9]]]]_{10}] \\ \text{CS} & 1'^{\alpha}[\text{REL}^{\alpha}[10'](\alpha)] \end{bmatrix}$$

Each of the conjuncts contains a gap, so each corresponding part of the interpretation $10'$ contains the λ-bound variable x. For convenience, I

dination is an independent question. It has been known since Ross (1967) that asymmetric extraction from coordinate structures is in general problematic, although there are some notable exceptional cases (Deane 1991).

assume that the operator in this case is an abstract REL$^{\alpha}$, meaning 'such that', that is bound by the head of the relative clause.[5] The CS interpretation in (18) is thus essentially (19).

(19) $\text{person}'^{\alpha}[\text{REL}^{\alpha}[\lambda \text{x.talk}'(\textsc{agent}{:}s,\ (\textsc{goal}{:}\text{x}\wedge\textsc{theme}{:}\text{x}))](\alpha)]$

Assuming distributivity of conjunction and lambda-reduction, we get the appropriate interpretation.

(20) $\Rightarrow \text{person}'^{\alpha}[\text{REL}^{\alpha}[\text{talk}'(\textsc{agent}{:}s,\ (\textsc{goal}{:}\alpha\wedge\textsc{theme}{:}\alpha))]]$

It appears, then, that nothing special needs to be said to license ATB extraction that does not follow directly from A′ extraction and coordination. That is, if A′ extraction and coordination exist in a language, sentences with ATB A′ extraction will be licensed.

However, just as learners may generalize beyond the evidence presented to them, so learners may hypothesize individuated constructions for constructs that are themselves licensed by combinations of other constructions. Müller (2006) argues against the inclusion in the grammar of such combined constructions for arbitrary constructs, e.g. passives combining with datives, or relative clauses combining with resultative predication. These combinations of constructions are accidental and do not reflect significant patterns of the language. But I assume that in principle two constructions can combine to form an independent individuated construction, but only if learners encounter appropriate constructs sufficiently often that the cooccurrence of properties has a reasonable likelihood of being non-accidental.

[5]A slightly different notation is used in Culicover (2013a).

In the present case, then, an individuated construction may be hypothesized that combines the conditions of (13) and (16), yielding (21). What corresponds in PHON to the conjoined YPs as 2 are strings containing gaps, and the interpretations of the individual conjuncts are connected by the appropriate logical connective.

(21)
$$\begin{bmatrix} \text{ATB CORE} & \\ \text{PHON} & \text{1-0-2} \\ \text{SYN} & [_{YP}\ [\begin{Bmatrix}\text{and}\\ \text{or}\end{Bmatrix}_0,\ [_{YP}\ldots XP_3\ldots]_1,\ [_{YP}\ldots XP_4\ldots]_2]] \\ \text{CS} & \begin{Bmatrix}\wedge\\ \vee\end{Bmatrix}_0[\lambda x.[[\ldots x_3\ldots]_1,[\ldots x_4\ldots]]_2] \end{bmatrix}$$

The individuated construction ATB Core is composed with ATB extraction (13) to produce (22).

(22)
$$\begin{bmatrix} \text{ATB+OPERATOR} & \\ \text{PHON} & 1-6 \\ \text{SYN} & XP_1,[_{XP}\ [\begin{Bmatrix}\text{and}\\ \text{or}\end{Bmatrix}_0,\ [\ldots XP_4\ldots]_2,\ [\ldots XP_5\ldots.]_3]_6] \\ \text{CS} & OP_1^{\alpha}[(6')](\alpha_1) = \\ & OP_1^{\alpha}[\begin{Bmatrix}\wedge\\ \vee\end{Bmatrix}_0(\lambda x[\ldots x_4\ldots]_2,[\ldots.x_5\ldots]_3)](\alpha_1) \end{bmatrix}$$

5 ATB extensions to parasitic gaps

5.1 Learning constructions

Let us now apply the idea that language acquisition accounts for language change to the analysis of P-gaps. The general idea is of course is not new. A persuasive case has been made by Lightfoot (1991, 1995), and it is an idea that is often found in discussions of creole formation and the effects of language contact (for example, Bickerton 1988, Trudgill 1983, 1986). Moreover, it is central to the computational modeling of language change and evolution of Niyogi and Berwick (1997) and Niyogi (2002, 2006).

My goal is to describe a plausible generalization path that ATB+Operator may have taken on the way to licensing P-gaps. The key to identifying this path is that in the course of acquisition of a grammar in a population, certain terms in a construction may change or even drop out of the construction.

Suppose a population is divided into Target and Learner.[6] This is a radical idealization, but it is helpful in modeling the interaction between acquisition and change, as Niyogi demonstrates.

Here, we are concerned with the precise formulation of some construction C^T. This construction is shared by the members of Target, and it is the job of the members of Learner to figure out what it is, based on the linguistic evidence.

The linguistic evidence consists of a sequence of constructs $\mathbf{c}$={c_1, c_2, ...

[6] In a more realistic model, members of Target may function as Learners and vice versa over the course of time. Especially in cases of complex constructions such as P-gaps, it is plausible that adult speakers influence one another's grammars and their judgments. Thanks to Maryellen Macdonald for raising this point.

} produced by Target's grammar over time. Each construction c_i is licensed by C^T, in the sense of section 2. On the basis of this evidence, at time t Learner posits a set of constructions $\mathbf{C}_t^L$={C_1^L, C_2^L, ... }. Some may be novel, and others may already have been hypothesized on the basis of prior experience.

A minimum criterion of success for Learner is that at some point it gets the construction 'right'. What 'right' means depends of course to a considerable extent on what we think learning a construction (or some other representation of a grammar) amounts to. One view is that learning a construction means arriving at a correct hypothesis that is stable at some point in the course of acquisition. That is, there is a set of constructions $\mathbf{C}_n^L$ at some time n that accounts for all and only the constructs produced by the Target.[7]

However, the key to language change is that the Learner does not get the grammar exactly 'right'. Niyogi (2006) suggests the Learner would get the grammar right in the limit, but stops learning at some point, due to maturation. Alternatively, we could require that even if the Learner never converges on the Target's grammar, not even in the limit, at minimum the Learner's grammar eventually accounts for all of the Target's productions. I adopt the second approach here.

In order to guarantee that the Learner arrives at a construction that

[7] I do not consider here how other outcomes might occur as a consequence of language acquisition. For example, constructions may become more restricted over time – the transition from Germanic V2 to English subject AUX inversion is a good example (see Culicover (2008) for a constructional analysis.) An account of how constructions become 'repurposed' requires a more elaborated model of the interactions between components of a grammar and the dynamics of acquisition in a social network than can be explored here.

at least includes the constructs licensed by the Target's construction, we have to posit a learning algorithm that specifies precisely what Learner does when its current hypothesis fails to account for a particular datum. Such algorithms exist for modeling the acquisition of phrase structure grammars (Wexler and Culicover 1980) and parameter-based grammars (Niyogi 2006), but none for the acquisition of construction grammars. Formulating such an algorithm is non-trivial and requires formalization that would take us far beyond the scope of this study. So the scenario in the remainder of this section is necessarily informal and intuitive.

Consider how a learner could formulate a construction C_i^L that is an overgeneralization of C^T. By assumption, the Target is producing a sequence of constructs that are all fully licensed by C^T. If the Learner's hypothesis is completely conservative, C_i^L would converge to C^T in the limit, other things being equal. We would be assuming, on such a scenario, that there is only one construction compatible with **c**, and generalizing to multiple constructions, only one set of constructions, that is, only one grammar, that is compatible with the Learner's experience.[8]

So in order to allow for change internal to the population consisting only of Target and Learner (that is, without external contact), we must suppose that the Learner guesses which features and feature values in the construct are relevant. That is, the Learner is probabilistic.[9]

[8]Complications arise in the case of multiple constructions. For example, the linguistic evidence may serve as evidence for two or more competing constructions (Gibson and Wexler 1992, Berwick and Niyogi 1996). I leave aside such complications here.

[9]For an extensive treatment of the probabilistic approach to language learnability, see Niyogi (2006), who uses "the probabilistic characterization of learnability ... to quantify the degree to which the grammar of children might differ from that of their parents – thereby opening the door to the study of language change."

To put this more explicitly, let φ_k be a term on a particular tier in a construct that is licensed by C^T. Suppose that there is a greater than zero probability that the Learner decides incorrectly that φ_k is irrelevant to the licensing of the construct by the construction to be learned. For example, the actual condition might be a particular lexical form, but the learner guesses that it is a category, or the learner fails to observe a semantic condition.[10] Failure to include φ_k in C_i^L will not be corrected, because the more general construction will be consistent with all subsequent constructs that Target produces using C^T. But the Learner's C^L will be different than the Target's C^T, and the Learner will overgenerate.

Overgeneralization is subject to subsequent correction, in principle – learners do retreat from overgeneralization (Ambridge et al. 2013).[11] However, correction of an error depends on sufficiently frequent experience with the correct form in comparison with the frequency of forms consistent with the error. Such experience, in turn, depends at least in part on the configuration of the social network in which communication between agents occurs. Computational modeling of social interaction in networks suggests that subgroups of agents can sustain features that are quite different from

[10] For concreteness, we could also assume that all of the features of the category *a* are ordered in terms of the probability of irrelevance, so that as we go up the hierarchy, the probability goes down. For example, [GENDER] would have a higher probability than [NUMBER], which would have a higher probability than N. This hierarchy would have to be wired into the Learner or somehow induced from experience, neither of which are unproblematic.

[11] There is a substantial literature on this topic in the area of Bayesian learning. Much of it has to do with general issues of statistical learning, and some has to do with language acquisition per se. See, for example, Gopnik and Tenenbaum (2007); Griffiths et al. (2008); Kemp et al. (2007); Tenenbaum et al. (2006); Xu and Tenenbaum (2007), and for application specifically to grammar acquisition, Pearl (2005, 2009); Pearl and Mis (2011); Pearl (2011); Pearl and Sprouse (2012).

the larger surrounding environment – the key is how the agents interact in the network and the amount of interaction among the agents (Latané et al. 1995; Vallacher and Nowak 1994; Liebrand et al. 1998; Nowak and Vallacher 1998; Culicover and Nowak 2003). So if the network is such that innovators of some construction have sufficiently strong influence, or interact only with those who have not yet learned the construction, it may come to pass that there is a distinct subgroup of speakers with the more general version of the construction.

If the broader community is subsequently exposed to the more general construction as well as the original version, the situation becomes one of constructional competition. Under certain circumstances the original construction may reassert itself and the innovation will die out. But in other situations, the innovation wins. A relevant factor may be the relative complexity of the alternatives (Culicover 2013a). Again, see Niyogi (2006), whose computational modeling of population dynamics shows how divergence of grammars can lead to variation, and in the limit, change.

5.2 Generalizing Basic ATB

Assuming the foregoing as a plausible framework for tracking constructional change, I propose that the Learner's failure to accurately assess the precise conditions on ATB extraction leads to the development and extension of P-gap constructions. This failure must be probabilistic, so that it is not inevitable – if it was, every language that has ATB extraction would have the full range of P-gaps.

Consider first the specification of the conjunction in ATB+Operator.

The corresponding SYN and CS terms of (22) are shown as boxed in (23).

$$(23)\quad \begin{bmatrix} \textsc{phon} & 1-6 \\ \textsc{syn} & \mathrm{XP}_1,[_{\mathrm{XP}}\ [\ \boxed{\left\{\begin{array}{c}\text{and}\\ \text{or}\end{array}\right\}}_0,\ [\ldots\mathrm{XP}_4\ldots]_2,\ [\ldots\mathrm{XP}_5\ldots]_3]_6] \\ \textsc{cs} & \mathrm{OP}_1^{\alpha}[\boxed{\left\{\begin{array}{c}\wedge\\ \vee\end{array}\right\}}_0\ (\lambda x[\ldots x_4\ldots]_2,[\ldots x_5\ldots]_3)](\alpha_1) \end{bmatrix}$$

Recall now our earlier observation that the canonical P-gap construction is one with *without*, e.g. *which books did Terry read without understanding.* Such cases are licensed if the SYN-condition $\left\{\begin{array}{c}\text{and}\\ \text{or}\end{array}\right\}$, is generalized to CONJ, since *without* has a coordinate interpretation, i.e. 'and not'.

(24) a person ($\left\{\begin{array}{c}\text{that}\\ \text{who}\end{array}\right\}$) we talked to *t* for a long time

a. and didn't find anything out about *t*

b. without finding anything out about *pg*

Extension of the ATB construction to semantically non-coordinate subordination structures is straightforward and minimal – it is a generalization from coordinating conjunctions to CONJ, the set of coordinating and subordinating conjunctions.

(25) $$\begin{bmatrix} \text{PHON} & 1 > 6 \\ \text{SYN} & \text{XP}_1,[_{\text{XP}}\ [\ \boxed{\text{CONJ}_0}, [\ldots \text{XP}_4 \ldots]_2, [\ldots \text{XP}_5 \ldots]_3]_6] \\ \text{CS} & \text{OP}_1^{\alpha}[\boxed{\begin{Bmatrix} \wedge \\ \vee \end{Bmatrix}}_0 (\lambda x[\ldots x_4 \ldots]_2,[\ldots x_5 \ldots]_3)](\alpha_1) \end{bmatrix}$$

This construction will license semantically coordinate P-gap configurations such as (24b) along with standard ATB extraction such as (24a).

The next generalization step is the one that distinguishes English from many other languages that have only (24a). In this step, the pairing of CONJ with the coordinate connectives is extended to the pairing of CONJ with non-coordinate connectives, such as *before, after, when, while, since, because, although*. Most of these conjunctions implicate coordination, in that they presuppose as given the truth of the subordinate proposition while asserting the truth of the main proposition and the relationship between the propositions – the exceptions are *if, unless*. For example, expressions of the form P CONJ Q in (26) are false if Q is false, even when P is true.

(26) $$\text{Sandy baked the cake} \begin{Bmatrix} \text{before} \\ \text{after} \\ \text{when} \\ \text{while} \\ \text{since} \\ \text{because} \\ \text{although} \\ \text{until} \end{Bmatrix} \text{Terry gave away the cookies.}$$

If Terry did not give away the cookies, then we would say that the entire sentence is false, even though Sandy baked the cake.

The next plausible step, then, is generalization of (25) to connectives that in some way imply coordinate semantics.

(27)
$$\begin{bmatrix} \text{PHON} & 1 > 6 \\ \text{SYN} & \text{XP}_1,[_{\text{XP}}\ [\ \boxed{\text{CONJ}_0}, [\ldots\text{XP}_4\ldots]_2], [\ldots\text{XP}_5\ldots]_3]\ _6] \\ \text{CS} & \text{OP}_1^{\alpha}[\boxed{0'\ [\rightarrow \begin{Bmatrix} \wedge \\ \vee \end{Bmatrix}]}\ \lambda\text{x}.([\ldots.x_4\ldots]_2, [\ldots.x_5\ldots]_3)](\alpha_1) \end{bmatrix}$$

This extended construction licenses sentences such as those in (28).[12]

(28) the person $(\begin{Bmatrix} \text{that} \\ \text{who} \end{Bmatrix})$ Sandy talked to t several times

a. ? $\begin{Bmatrix} \text{before} \\ \text{after} \\ \text{when} \\ \text{while} \\ \text{since} \\ \text{because} \\ \text{although} \\ \text{until} \end{Bmatrix}$ Terry got angry at *pg*

[12] The difference in acceptability between (28a,b) is arguably due to the sensitivity of processing to the presence or absence of finite tense and an overt subject. For related discussion, see Gibson (1998, 12).

b. $\left\{\begin{array}{l}\text{before}\\ \text{after}\\ \text{while}\\ \text{since}\end{array}\right\}$ getting angry at *pg*

This construction also licenses constructs in which the subordinate clause precedes the main clause.[13]

(29) the person $\left\{\begin{array}{l}\text{that}\\ \text{who}\end{array}\right\}$

a. ? $\left\{\begin{array}{l}\text{before}\\ \text{after}\\ \text{when}\\ \text{while}\\ \text{since}\\ \text{because}\\ \text{although}\\ \text{until}\end{array}\right\}$ Terry got angry at *pg*, Sandy talked to *t* several times

b. $\left\{\begin{array}{l}\text{before}\\ \text{after}\\ \text{while}\\ \text{since}\end{array}\right\}$ getting angry at *pg*, Sandy talked to *t* several times

The next step in the generalization is to drop the requirement that the

[13]In Culicover (2013b) I provide evidence that the unacceptability of examples like these but without an overt relative marker is due to complexity in identifying the relative clause at the point at which the left-adjoined constituent is processed.

conjunction implicates coordination. This permits P-gaps in *if*- and *unless*-clauses.

(30) $\begin{bmatrix} \text{PHON} & 1-6 \\ \text{SYN} & \text{XP}_1,[_{\text{XP}}\ [\ \boxed{\text{CONJ}_0}, [...\text{XP}_4...]_2], [...\text{XP}_5...]_3]\ _6] \\ \text{CS} & \text{OP}_1^{\alpha}[\boxed{0'}(\lambda\text{x}.[[....x_4...]_2,[....x_5...]_3])](\alpha_1) \end{bmatrix}$

(31) a person who I will talk to *t*

a. if I can get in touch with *pg*

b. unless I can't meet personally with *pg*

6 Predictions and conclusions

The final step in the generalization trajectory is to drop the requirement that there should be a conjunction in a phrase containing one of the XPs.

(32) $\begin{bmatrix} \text{PHON} & 1-6 \\ \text{SYN} & \text{XP}_1,[_{\text{XP}}, [_{\text{YP}}\ ...\text{XP}_4...]_2], [_{\text{YP}}\ ...\text{XP}_5...]_3]\ _6] \\ \text{CS} & \text{OP}_1^{\alpha}[\lambda\text{x}[[....x_4...]_2,[....x_5...]_3]](\alpha_1) \end{bmatrix}$

The construction in (32) characterizes the very broad non-conjunctive P-gap construction of English. It predicts that where there is more than one XP, there can be multiple gaps and in fact P-gaps. It does not require that the XPs be conjoined, or even that they be in separate clauses. Such extraction need not be ATB. This variant of the construction predicts that there can be gaps in any two constituents in a sentence, regardless of their syntactic function or relationship to one another (subject of course to other factors constraining the distribution of gaps).

The examples in (33) show that the construction is in fact fully general.[14]

(33) a. (a person who) everyone who talks to *pg* likes *t* [P-gap in subject NP]

b. (a person who) I introduced some friends of *t* to some enemies of *t* [gaps in non-conjoined arguments]

c. (a person who) some friends of *t* introduced me to some enemies of *t* [gaps in non-conjoined arguments]

d. (a person who) you talk to *t/pg*, you immediately get angry at *t* [gaps in paratactic clauses (Culicover and Jackendoff 1997)]

e. (a person who) you talk to *t/pg*, some very large friend of *t* will come around to talk to you [gaps in paratactic clauses and subject NP]

f. (a person who) Sandy was kind to *t* although impatient with *pg* [P-gap in subordinate AP]

g. (a person who) Sandy was kind to *t* although not particularly a very good friend of *pg* [P-gap in subordinate NP]

h. (a person who) if angry at *pg*, Sandy would simply not talk to *t* [P-gap in preposed subordinate AP]

i. (a person who) Sandy swore to *t* that only very good friends of *pg* would be allowed to attend the party [P-gap in subject NP of sentential complement]

[14] All of these examples are well-formed in a narrow sense, given that the gaps are all in positions that locally permit gaps. However, they are quite complex and therefore somewhat less than fully natural, and for some speakers some may be unacceptable. For discussion of the distinction between well-formed and acceptable, see Culicover (2013a).

Some of these examples were developed in conversation with Robert Levine.

j. (a person who) Sandy swore to *t* that [only very good friends of *pg*] would we permit *t* to attend the party [P-gap in topicalized non-subject NP of sentential complement]

k. (a person who) the parents of *pg/t* guaranteed that only very good friends of *pg* would be permitted to attend the party [P-gap in topicalized non-subject NP of sentential complement]

In fact, only contexts from which extraction *per se* is independently problematic, for reasons of processing complexity or grammatical constraints, are not possible P-gap configurations (Culicover 2013a).

An ATB account predicts as well that a context that absolutely prohibits extraction would also prohibit P-gaps. An especially dramatic case is documented by Dubinsky (2007). Dubinsky contrasts appositives with non-appositive, adjunct constructions. (34a) illustrates a non-appositive relative while (34b) is appositive. Similarly, (35a) is a non-appositive adjunct, while (35b) is the corresponding appositive.

(34) a. A few people who dislike John refuse to talk to him.

b. A few people, who dislike John, refuse to talk to him.

(35) a. We talked to that woman while smiling sweetly at her.

b. We talked to that woman, smiling sweetly at her.

Dubinsky goes on to observe that appositive constructions in English do not permit P-gaps. To illustrate this point, I construct minimal pairs of relative clauses based on (34)-(35) that will produce non-appositive and appositive P-gaps. The results are shown in (36)-(37).

(36) a. a man who a few people who dislike *pg* refuse to talk to *t*

b. a man who a few people, who dislike $\left\{ \begin{array}{c} *pg \\ \text{him} \end{array} \right\}$, refuse to talk to *t*

(37) a. a woman who we talked to *t* while smiling sweetly at *pg*

b. a woman who we talked to *t*, smiling sweetly at $\left\{ \begin{array}{c} *pg \\ \text{her} \end{array} \right\}$

It is significant that simple extraction from an appositive is itself quite unacceptable, while extraction from a non-appositive, although often quite degraded, is nevertheless at least marginally acceptable. The pairs in (38)-(39) illustrate.

(38) a. ? John is a person who a few people who dislike *t* refused to attend the lecture.

b. * John is a person who a few people, who dislike *t*, refused to attend the lecture.

(39) a. ? Mary is the person who we got very upset while smiling sweetly at *t*.

b. * Mary is the person who we got very upset, smiling sweetly at *t*.

This pattern is precisely what we predict on a strict ATB analysis of P-gaps. A P-gap must be in a position from which extraction is in principle possible, although possibly degraded. The presence of the true gap ameliorates the extraction at the site of the P-gap, because it is licensed by the ATB+Operator construction. But extraction from an appositive alone is not possible, for reasons that remain to be investigated. In any case, it does not fall under an extension of ATB extraction.

I conclude that Sag's (1983) and Munn's (1992, 2001) conflation of P-gaps and ATB extraction were on the right track. But we do not need to stipulate such an analysis; we can derive it as a natural extension of ATB extraction in coordinate structures. But crucially, the possibility of P-gaps is not a theorem, in the sense that it can be predicted from other grammatical properties of the language. A language may permit ATB extraction but not allow P-gaps.

As developed here, the explanation for the occurrence of P-gaps rests on a constructional formulation of ATB and the assumption that errors may occur in learners' identification of the conditions under which constructions license constructs. The errors become significant only if the population dynamics permit them to take hold. Assuming this, we are able to explain not only why English has a very broad range of P-gap configurations, but why other languages have narrower P-gaps and some even do not permit them, while at the same time they permit ATB extraction in coordinate structures. The Parasitic Gap Hierarchy follows from the progressive generalization of the ATB construction.

As far as I know, there is no account of P-gaps in non-constructional theories that goes beyond stipulating that P-gaps exist in a given language and that they are subject to particular conditions. It is a virtue of the constructional account, formulated in terms of Jackendoff's Parallel Architecture, that it offers a natural account of the the possibility of P-gaps, the observed variation, and the Parasitic Gap Hierarchy, all in terms of constructional change.

References

Ambridge, Ben, Julian M Pine, Caroline F Rowland, Franklin Chang, and Amy Bidgood. 2013. The retreat from overgeneralization in child language acquisition: Word learning, morphology, and verb argument structure. *Wiley Interdisciplinary Reviews: Cognitive Science* 4:47–62.

Barlew, Jefferson, and Peter W. Culicover. 2015. Minimal constructions Under review.

Berwick, Robert C., and Partha Niyogi. 1996. Learning from triggers. *Linguistic Inquiry* 27:605–622.

Bickerton, Derek. 1988. Creole languages and the bioprogram. In *Linguistics: The Cambridge survey(ii)*, ed. Frederick J. Newmeyer, 268–284. Cambridge: Cambridge University Press.

Bordelois, Ivonne. 1986. Parasitic gaps: Extensions of restructuring. *Generative Studies in Spanish Syntax, Foris, Dordrecht* 1–24.

Chomsky, Noam. 1973. Conditions on transformations. In *A Festschrift for Morris Halle*, ed. Stephen Anderson and Paul Kiparsky, 232–286. New York: Holt, Reinhart & Winston.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht, The Netherlands: Foris Publications.

Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.

Culicover, Peter W. 2001. Parasitic gaps: a history. In *Parasitic gaps*, ed. Peter Culicover and Paul Postal, 3–68. Cambridge, MA: MIT Press.

Culicover, Peter W. 2008. The rise and fall of constructions and the history of English do-support. *Journal of Germanic Linguistics* 20:1–52.

Culicover, Peter W. 2013a. *Grammar and complexity: Language at the intersection of competence and performance*. Oxford: Oxford University Press.

Culicover, Peter W. 2013b. Problems with English zero relatives. *International Review of Pragmatics* 5:253–270.

Culicover, Peter W., and Ray Jackendoff. 1997. Semantic subordination despite syntactic coordination. *Linguistic Inquiry* 28:195–217.

Culicover, Peter W., and Andrzej Nowak. 2003. *Dynamical grammar: Minimalism, acquisition and change*. Oxford: Oxford University Press.

Culicover, Peter W., and Paul Postal, ed. 2001. *Parasitic gaps*. Cambridge, MA: MIT Press.

Deane, Paul. 1991. Limits to attention: a cognitive theory of island phenomena. *Cognitive Linguistics* 2:1–63.

Dubinsky, Stanley. 2007. Parasitic gaps in restrictive and appositive clauses. In *Proceedings of Israel Association for Theoretical Linguistics (IATL)*, volume 22, 1–19. Jerusalem: The Hebrew University of Jerusalem: Israel Association for Theoretical Linguistics.

Engdahl, E. 1983. Parasitic gaps. *Linguistics and Philosophy* 6:5–34.

Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68:1–76.

Gibson, Edward, and Kenneth Wexler. 1992. Parameter setting, triggers and V2. *GLOW Newsletter* 28:16–17.

Gopnik, Alison, and Joshua B Tenenbaum. 2007. Bayesian networks, Bayesian learning and cognitive development. *Developmental Science* 10:281–287.

Griffiths, Thomas L, Charles Kemp, and Joshua B Tenenbaum. 2008. Bayesian models of cognition. *Cambridge handbook of computational cognitive modeling* 59–100.

Jackendoff, Ray. 2002. *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Jackendoff, Ray S. 1997. *The Architecture of the Language Faculty*. Cambridge, MA: The MIT Press.

Kemp, Charles, Amy Perfors, and Joshua B Tenenbaum. 2007. Learning overhypotheses with hierarchical Bayesian models. *Developmental science* 10:307–321.

Latané, Bibb, James H Liu, Andrzej Nowak, Michael Bonevento, and Long Zheng. 1995. Distance matters: Physical space and social impact. *Personality and Social Psychology Bulletin* 21:795–805.

Liebrand, Wim B.G., Andrzej Nowak, and Rainer Hegselmann. 1998. *Computer modelling of social processes*. London: Sage Publications.

Lightfoot, David. 1991. *How to set parameters: Arguments from language change*. Cambridge: Cambridge University Press.

Lightfoot, David. 1995. Why UG needs a learning theory: Triggering verb movement. In *Clause structure and language change*, ed. Adrian Battye and Ian Roberts, 31–52. Oxford: Oxford University Press.

Müller, Stefan. 2006. Phrasal or lexical constructions. *Language* 82:850–883.

Munn, Alan. 1992. A null operator analysis of ATB gaps. *The Linguistic Review* 9:1–26.

Munn, Alan. 2001. Explaining parasitic gap restrictions. In *Parasitic gaps*, ed. Peter Culicover and Paul Postal, 369–392. Cambridge, MA: MIT Press.

Niyogi, Partha. 2002. The computational study of diachronic linguistics. In *Syntactic Effects of Morphological Change*, ed. David Lightfoot. Oxford: Oxford University Press.

Niyogi, Partha. 2006. *The computational nature of language learning and evolution*. Cambridge, MA: MIT Press.

Niyogi, Partha, and Robert C Berwick. 1997. A dynamical systems model for language change. *Complex Systems* 11:161–204.

Nowak, Andrzej S, and Robin R Vallacher. 1998. *Dynamical social psychology*. New York: Guilford Press.

Pearl, Lisa. 2005. The input for syntactic acquisition: Solutions from language change modeling. In *Proceedings of the Workshop on Psychocomputational Models of Human Language Acquisition*, 1–9. Association for Computational Linguistics.

Pearl, Lisa. 2009. Acquiring complex linguistic systems from natural language data: What selective learning biases can do. Unpublished manuscript, University of California, Irvine.

Pearl, Lisa, and Benjamin Mis. 2011. How far can indirect evidence take us? anaphoric one revisited. In *Proceedings of the 33rd annual Conference of the Cognitive Science Society*, 879–884.

Pearl, Lisa, and Jon Sprouse. 2012. Computational models of acquisition for islands. In *Experimental syntax and the islands debate*, ed. Jon Sprouse and Norbert Hornstein. Cambridge, MA: Cambridge Univeristy Press.

Pearl, Lisa S. 2011. When unbiased probabilistic learning is not enough: Acquiring parametric system of metrical phonology. *Language Acquisition* 18:87–120.

Ross, John. 1967. Constraints on variables in syntax. Doctoral Dissertation, Massachusetts Institute of Technology.

Sag, Ivan A. 1983. On parasitic gaps. *Linguistics and Philosophy* 6:35–45.

Stump, Gregory T. 2001. *A theory of paradigm structure*. Cambridge: Cambridge Univeristy Press.

Tenenbaum, Joshua B, Thomas L Griffiths, and Charles Kemp. 2006. Theory-based Bayesian models of inductive learning and reasoning. *Trends in cognitive sciences* 10:309–318.

Trudgill, Peter. 1983. *On dialect: Social and geographical perspectives*. Oxford: Basil Blackwell.

Trudgill, Peter. 1986. *Dialects in contact*. Oxford: Blackwell.

Vallacher, Robin R, and Andrzej Ed Nowak. 1994. *Dynamical systems in social psychology.*. New York: Academic Press.

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

Xu, Fei, and Joshua B Tenenbaum. 2007. Word learning as Bayesian inference. *Psychological review* 114:245–272.