Argument!
Another look at the *X or no X* construction

Gregory M. Kobele
Humboldt Universität zu Berlin

Abstract

Pullum and Rawlins (2007) argue that deviations from absolute string identity in the English *X or no X* construction are either fatal for syntactic analyses thereof, or merely problematic. They suggest an alternative analysis, involving 'semantic identity', augmented with 'pragmatic considerations,' which, however, turns out to be rife with problems. The deviations from identity are certainly not fatal; in fact, widespread syntactic assumptions conspire to provide an elegant analysis for this construction type.

In a recent paper, Pullum and Rawlins (2007) (henceforth P&R) examine an argument made by Alexis Manaster-Ramer in the late 1980s (alluded to in (Manaster-Ramer, 1986)) which hoped to establish that English was not context-free. The argument revolved around a particular construction, called the *X or no X* construction, which is exemplified by the following sentences.

(1) Babysitter or no babysitter, the kid is staying home.

(2) Promise to resign or no promise to resign, I'm not quitting.

(3) Den of iniquity or no den of iniquity, I call it home.

Each of the above example sentences begin with a (for the sake of argument) noun phrase X, which is followed by the words *or no*, and then is followed by that same noun phrase, X. An utterance of an *X or no X* sentence conveys that the speaker explicitly discounts X from influencing his thoughts about the part following the *X or no X*. For example, sentences 1–3 can be paraphrased as in 4–6.

(4) Our child will stay home tonight, even if it means that he will be unsupervised.

(5) Even though I made a promise to do so, I will not resign.

(6) Although it may in fact be a den of iniquity, it feels like home to me.

Things get interesting once we try to delimit the range of sentences which participate in this construction. As example sentences 7–9 show, in an *X or no X* sentence, not just any two noun phrases may flank the *or no*.

(7) *Babysitter or no supervision, the kid is staying home.

(8) *Promise to resign or no oath to step down, I'm not quitting.

(9) *Den of iniquity or no house of ill repute, I call it home.

Indeed, although the NPs flanking the *or no* in examples 7–9 are near synonyms, and might therefore naively be thought to convey the intended meanings as given in 4–6, not even they are allowed to participate in the *X or no X* construction. A natural first approximation, and the one suggested by Manaster-Ramer, is that the NPs flanking the *or no* in the *X or no X* construction must be formally identical, in the sense that the spell-out of both NPs must yield the same string—the same sequence of words.

With this as their starting point, P&R undertake a very careful study of the *X or no X* construction. They find that this first approximation in fact erroneously excludes many actual *X or no X* sentences from membership in the construction thus described. They note two (they count four, but I will engage in some lumping together) differences that may obtain between the yields of the two NPs in this construction.

First, ellipsis of either a part (their difference one) or of the whole (their difference two) of the second NP may render their yields non-string identical, as in 11 and 12 (from 'source' 10).

(10) War with Iraq or no war with Iraq, innocent people are likely to die.

(11) War with Iraq or no war, innocent people are likely to die.

(12) War with Iraq or no, innocent people are likely to die.

Second, parentheticals (their difference three) and/or expressive expletives (their difference four) may appear in one of the NPs to the exclusion of the other, thereby making the respective strings obtained by spelling-out differ. Sentence 13 illustrates a parenthetical, and 14 expressive expletives, occuring in one but not the other NP of an *X or no X* sentence.

(13) War with Iraq (which, by the way, would be a disaster for the country) or no war with Iraq, innocent people are likely to die.

(14) War with Iraq or no fucking war with piece of shit Iraq, innocent people are likely to die.

In addition, they note a relationship between the *X or no X* construction, and another construction, the *not just* X_1 *but* X_2 construction. They observe that, given a broad range of X_1 and X_2, where X_1 and X_2 are nominal phrases, X_1 *or no* X_2 is a well-formed *X or no X* construction when and only when *not just* X_1 *but* X_2 is ill-formed. This leads them to make the obvious prediction. These are solid observations, and P&R should be lauded for advancing the state of our knowledge about this construction.

Unfortunately, as exceptional as their paper is when it comes to the linguistic data, it is at times unclear in its argumentation.

In their paper, P&R claim to be examining Manaster-Ramer's argument which purports to establish that English is not context-free. What they are really doing is not straightforward, as they explicitly acknowledge that the conclusions they draw about the argument examined[1] do not warrant the conclusion they come to as regards the context-freeness of English.[2] Instead, they seem to be making an argument that their characterization of the *X or no X* construction cannot be given a revealing syntactic description,[3] and offering a 'semantic' analysis of their own. The remainder of the paper is structured as follows. I begin by examining P&R's analysis of the *X or no X* construction. I show that it does not describe what they intend it to describe, and, that once you unpack the proposal, it presupposes something like the syntactic account of copying they are proposing it as an alternative to. I then turn to their argument against syntactic characterizations of the *X or no X* construction. I correct a possible misunderstanding of their argument, and then show that the tools for revealing syntactic analyses of the *X or no X* construction are already in existence. In an appendix, I give a careful reconstruction based on the ideas of Manaster-Ramer (1993) of the notion of a 'revealing analysis' that plays such a large role in P&R's argument.

[1] "The set of strings containing all and only the fully acceptable instances of the construction thus probably is not CF." (page 282)

[2] "What we are suggesting is that the *X or no X* construction offers no compelling reason for abandoning the position that English is one of the natural languages that does have a CF word-stringset." (page 285)

[3] "The many examples of this sort are perhaps the most convincing evidence that we are not dealing with syntactic reduplication, where word-for-word identity between strings is demanded on pain of ungrammaticality." (page 281)

1 P&R's analysis of the *X or no X* construction

P&R claim that what is required of the two disjuncts in the *X or no X* construction is simple denotational identity.[4] We might for the sake of concreteness formalize this in terms of the following construction-specific rule.

$$\begin{aligned}\textsc{x-or-no-x}(X_1, X_2) := \ & \textbf{if } [\![X_1]\!] = [\![X_2]\!] \textbf{ and } cat(X_i) = \texttt{N} \\ & \textbf{then } X_1^\frown \text{“ or no ”}^\frown X_2 \\ & \textbf{else undefined}\end{aligned}$$

What this predicts is that the following should be grammatical members of the *X or no X* family:

(15) gnu or no wildebeest

(16) pretty girls or no beautiful girls

(17) man that I hate or no man hated by me

(18) insane bailout or no bailout which is insane

(19) girl that it seems John has a crush on or no girl that John seems to have a crush on

(20) garden where three boys were standing around or no garden where there were three boys standing around

However, P&R, in taking the permissible deviations from string identity outlined in the introduction to this paper to be the only such permissible in this construction, must claim that the above phrases are not in the antecedently circumscribed class of *X or no X* phrases. They consider examples 15 and

[4] How fascinating would this be if it were true! Computational linguists aren't simply making weak generative capacity arguments because they have nothing better to do; they are attempting thereby to understand the kind of computational resources speakers of languages must be able to bring to bear when producing or understanding expressions of their language. From this perspective, telling them "oh, don't worry about X, it's a semantic thing" doesn't cut it—regardless of where the buck gets passed, it's gotta stop somewhere. Or as Chomsky (1981) puts it:

> [...]it is evident that a reduction in the variety of systems in one part of the grammar is no contribution to these ends if it is matched or exceeded by proliferation elsewhere.

16 in their paper, and judge them to be unacceptable. In order to account for this discrepancy between the predictions of the theory they have provided and their characterization of the phrases they want to describe, they postulate the following:

> This is explained by the fact that, as often noted, speakers tend to assume a distinction of meaning between any two expressions of different form. This tendency will lead to pressure to regard any two distinct nouns in close proximity to be understood as having different denotations. Thus, although all and only gnus are wildebeests, a use of the phrase *gnu or no wildebeest*, with an unmotivated choice of mismatched nouns when *wildebeest or no wildebeest* could have been used instead, would prompt the hearer to doubt the identity of denotation (via an inference that could doubtless be related to Grice's maxim of manner), and thus render the use of the phrase perplexing and unacceptable.

Let's think about this for a little bit. Essentially, they are looking for a way to reign in the over-generation of their theory. The mechanism they propose seems to work by comparing what was said to a set of alternatives. We are not told how to obtain this set of alternatives. How does this comparison work? It needs access to the syntactic structure of the sentence (otherwise it won't know that there is an *X or no X* phrase, nor would it be able to find the parts). Moreover, this syntactic structure must be in the form of a full syntactic parse tree, so as to make identity of lexical items (or structures, if one believes that words like *gnu* are syntactically complex, see e.g. (Marantz, 1997)) visible. Then, it must check whether the syntactic bits which make up the two halves of the *X or no X* phrase are identical. If they aren't, then if there is some other alternative parsed sentence where these two halves *are* identical, an 'inference' is triggered that 'prompts the hearer to doubt the identity of denotation' of the two halves of the *X or no X* phrase in the observed utterance.

Of course, this is predicated on the assumption that we want somehow to exclude examples 15–20 from being 'correct' *X or no X* phrases, an assumption which P&R seem to (at least implicitly[5]) be making. Is this assumption warranted? P&R claim that examples 15 and 16 are unacceptable, and I agree. As far as the other examples go, I don't find 19 or 20 nearly as grat-

[5] "This new body of data does not in fact alter the situation as regards the stringset mathematics. Our examples show that the construction may sometimes take the form 'X_1 *or no* X_2', where X_2 is only a (possibly null) initial part of X_1, or where either X_1 or X_2 contains extra interpolated material." (page 282)

ing as 15 and 16.[6] So, while we probably can agree that we don't want examples 15 and 16, we might want to let the theory decide the other cases. How does P&R's theory decide these other cases? They don't say. The procedure I sketched above has lots of leeway in what counts as 'identical' enough to avoid triggering hearer doubt.[7]

2 On the Nature and Limitations of Syntactic Identity

P&R are not very careful with their argument against syntactic accounts of the *X or no X* construction. I believe the argument they intend to make is based on simplicity, and might run as follows:

> although a syntactic account of the *X or no X* construction may be technically possible, there is currently no natural, elegant, yet restrictive syntactic account which allows for the kinds of deviations from identity which obtain.[8]

This is exactly the right kind of argument to make, and I will return to it in section 2.2. However, P&R might easily be misconstrued as trying to argue against *any* possible syntactic account of the identity effects found in the *X or no X* construction. This erroneous interpretation of their position could be forgiven, as P&R hint repeatedly at the premise that syntactic accounts of identity effects in grammar cannot deal with deviations from perfect string identity.[9] This is simply a false premise. In section 2.1 I show how such almost-identity is easily describable syntactically by example.

[6] However, these are hardly the clear cases we ought to take as the foundation for linguistic theories.

[7] A condition which has been proposed before in a different context (Chung, 2006) is that the two constituents in question have identical numerations (which, in the context of minimalist grammars (Stabler, 1997), we can understand in terms of identity of the derivation tree, as proven by (Hale and Stabler, 2005)). This would then correctly rule out 15 and 16, while possibly allowing 19 and 20 (depending on your analysis of expletives).

[8] "We are not aware of a framework for syntactic description that would be able to describe an infinite set of phrases with string identity but structure of the sort needed here." (page 285)

[9] "The many examples of this sort are perhaps the most convincing evidence that we are not dealing with syntactic reduplication, where word-for-word identity between strings is demanded on pain of ungrammaticality." (page 281)

2.1 Almost-Identity, Syntactically

The languages L_1 and L_2 below, given by P&R as rough characterizations of the kinds of deviations from exact identity that exist in the *X or no X* construction, have any number (literally) of purely syntactic descriptions.[10]

$$L_1 = \{u\mathsf{c}uv : u, v \in \{\mathsf{a}, \mathsf{b}\}^*\}$$
$$L_2 = \{u\mathsf{c}v : \exists w.\ u, v, w \in \{\mathsf{a}, \mathsf{b}\}^* \wedge v \in u \sqcup\!\sqcup w\}$$

The language L_1 is a gross approximation of the first deviation from perfect string identity talked about in the introduction, namely, that ellipsis can apply to one of the Xs in the *X or no X* construction. L_1 models this by allowing the first 'X' (there, u) to be a copy of just the beginning part of the second 'X' (there, uv).

The language L_2 approximates the addition of parenthetical and expressive material to one of the two Xs. In the definition of L_2, v is the result of adding parentheticals and expressives (represented by w) *ad libitum* throughout u.

Here I provide two multiple context-free grammars (MCFGs, Seki et al., 1991), G_1 and G_2, for the languages L_1 and L_2 respectively.[11] Both G_1 and G_2 contain the following productions, which by themselves define the language $L_{w\mathsf{c}w} = \{w\mathsf{c}w : w \in \{\mathsf{a}, \mathsf{b}\}^*\}$:

$$A(\epsilon, \epsilon). \tag{i}$$
$$A(\mathsf{a}x, \mathsf{a}y) \rightarrow A(x, y). \tag{ii}$$
$$A(\mathsf{b}x, \mathsf{b}y) \rightarrow A(x, y). \tag{iii}$$
$$S(x\mathsf{c}y) \rightarrow A(x, y). \tag{iv}$$

It is natural to conduct derivations in this grammar 'bottom-up':

$$A(\epsilon, \epsilon) \Rightarrow_{iii} A(\mathsf{b}, \mathsf{b}) \Rightarrow_{ii} A(\mathsf{ab}, \mathsf{ab}) \Rightarrow_{iv} S(\mathsf{abcab})$$

To define G_1, we add the following productions, which allow the second 'copy' to grow independently to the right:

$$A(x, y\mathsf{a}) \rightarrow A(x, y). \tag{v}$$
$$A(x, y\mathsf{b}) \rightarrow A(x, y). \tag{vi}$$

[10] The operation of *shuffling* non-deterministically interleaves strings, preserving the order of elements within them, and is defined as per the following:

$$\epsilon \sqcup\!\sqcup x = x \sqcup\!\sqcup \epsilon = \{x\}$$
$$(\mathsf{a}x \sqcup\!\sqcup \mathsf{b}y) = \mathsf{a}(x \sqcup\!\sqcup \mathsf{b}y) \cup \mathsf{b}(\mathsf{a}x \sqcup\!\sqcup y)$$

[11] For readability, the grammars have been written using the notation of literal movement grammars (Groenink, 1997).

$$A(\epsilon, \epsilon) \Rightarrow_{iii} A(\mathsf{b}, \mathsf{b}) \Rightarrow_{ii} A(\mathsf{ab}, \mathsf{ab}) \Rightarrow_{v} A(\mathsf{ab}, \mathsf{aba}) \Rightarrow_{iv} S(\mathsf{abcaba})$$

To define G_2, we add instead the following productions, which allow the second 'copy' to grow independently to the left:

$$A(x, \mathsf{a}y) \rightarrow A(x, y). \tag{v'}$$
$$A(x, \mathsf{b}y) \rightarrow A(x, y). \tag{vi'}$$

$$A(\epsilon, \epsilon) \Rightarrow_{iii} A(\mathsf{b}, \mathsf{b}) \Rightarrow_{v'} A(\mathsf{b}, \mathsf{ab}) \Rightarrow_{iii} A(\mathsf{bb}, \mathsf{bab}) \Rightarrow_{iv} S(\mathsf{bbcbab})$$

2.2 The Argument from Elegance

Above, we saw that there exist simple grammars which derive string languages approximating in various respects what P&R have argued to be the case in the *X or no X* construction. Their major argument against potential syntactic accounts is that those frameworks which we view as plausible from a linguistic perspective are not able to derive *X or no X* phrases elegantly. They cite the frameworks of Tree Adjoining Grammars (TAGs (Joshi, 1987)) and Minimalist Grammars (MGs (Stabler, 1997)), which, although perfectly capable of defining copies (even with the kinds of deviations from perfect identity noticed by P&R), cannot do this in a natural way.[12] However, adding a primitive operation of copying to either of these grammar formalisms would presumably change the naturalness of descriptions of copy constructions. Kobele (2006), doing just this, introduces minimalist grammars with copy movement (CMGs) as a formalization of the minimalist's copy-theory of movement (Chomsky, 1995). The nature of the copying mechanism in CMGs is such that both 'copies' are assigned the same derived structure. Furthermore, the membership problem for CMGs remains in **PTIME**, despite their departure from mild context-sensitivity.[13] So, what CMGs provide us with is a restrictive syntactic framework in which copies are assigned the 'right kind of' structures. Thus, CMGs are a good choice of formalism for dealing with the *X or no X* construction, *as described before P&R*. One of the main contributions of P&R's paper lies in their improvement of the description of the *X or no X* construction. Now, a formalism

[12]It is no trivial matter to specify objectively what counts as a 'natural' way, here. As seems often to be the case, the objective measure of size correlates rather well with our intuitions regarding naturalness. P&R, in previous incarnations of their paper, consider the number of lexical items needed by TAGs and by MGs to define the copy language L_{wcw}. Section A gives a formal reconstruction of our intuitions regarding naturalness in terms of the rate of growth of TAG and MG lexica as new open-class items are added.

[13]CMGs are straightforwardly seen to be weakly equivalent to *parallel* multiple context-free grammars (Seki et al., 1991).

for linguistic description needs not just to be able to elegantly account for perfect string identity, as before, but also to account for the various kinds of deviations from perfect string identity given in the introduction.

The question that we turn to now is whether CMGs are expressive enough to elegantly account not just for cases of perfect copying, but also for cases of 'imperfect' copying. The copying mechanism in CMGs is integrated into the derivational system, which means that copying happens *at a certain point* in the derivation of an object. At that point, a full, exact, copy is made. As the derivation continues, the exactness of the copy relation might be obscured by further operations.[14] The range of operations in the grammar, therefore, is the limiting factor on what kind of deviations from exact string identity may obtain. With only phrasal movement (i.e. without head movement, as proposed in a more general minimalist setting by Mahajan (2000)), CMGs can only generate surface identical copies. Adding head movement, either a variant of the standard version (Baker, 1988) (as in Stabler (1997)) or à la Brody (2000) (as in Kobele (2002)), allows for a limited kind of non-identity, in which a single (complex) word is positioned in a different position in the copy than in the 'source'. The Yoruba cases described in Kobele (2006) are of this kind. Very roughly, the Yoruba sentence 21 below has the three different verbal relative clause possibilities 22–24.

(21) *Jimọ ọ ra adiẹ se*
Jimọ HTS buy chicken cook
"Jimọ bought the chicken to cook."

(22) *Rira ti Jimọ ọ ra adiẹ se*
buying TI Jimọ HTS buy chicken cook
"the fact/way that Jimọ bought the chicken to cook"

(23) *Rira adiẹ se ti Jimọ ọ ra adiẹ se*
buying chicken cook TI Jimọ HTS buy chicken cook

(24) *Rira se ti Jimọ ọ ra adiẹ se*
buying chicken TI Jimọ HTS buy chicken cook

Based on 22 and 23, the generalization we would like to pursue is that the 'head' of the verbal relative clause is a (nominalized) copy of a verbal con-

[14]This is exactly the situation in rule-based theories of phonological reduplication, where after an exact copy of the base is made, further rule applications may render the base-reduplicant identity relation opaque. In phonology, it has long been of interest how to characterize the base-reduplicant relation, especially given its opaque relation to string identity. CMGs offer syntacticians the possibility of investigating the limits of opacity in *syntactic* copying constructions.

stituent (the main verb ra in 22, and the entire VP ra adiẹ se in 23).[15] However, expression 24 requires special attention—it seems clear that the head rira se bears some copy-like relation to the lower clause, but what? Adopting a derivation broadly along the lines of the below for the sentence in 21, we see that ra se is, although not a surface one, an underlying constituent, as are ra and ra adiẹ se.

1. [ra se] (merge ra and se)
2. [adiẹ [ra se]] (merge adiẹ)
3. [ra [adiẹ [se]]] (head move ra)
4. [Jimọ [ra [adiẹ [se]]]] (merge Jimọ)

In order to describe along these lines the kind of deviations from exact identity that are attested in the *X or no X* construction, we need to be able both to insert expressive adjectives and to elide material *after* copying has taken place. Interestingly, mechanisms which will do just this are widely discussed in the minimalist program. Late adjunction (of relative clauses) has been proposed by Lebeaux (1988), and is discussed in the context of reconstruction (and has been adapted to the minimalist grammar framework in Gärtner and Michaelis (2003)). Ellipsis has, since Merchant (2001), been commonly assumed in minimalist circles to be late deletion under some form of identity. Adopting both mechanisms uncritically for the moment, *X or no X* phrases like 25 below can be derived by first constructing the phrase war with Iraq, copying it, deleting with Iraq from the second disjunct, and then late adjoining the expressive adjective fucking.

(25) war with Iraq or no fucking war

3 Conclusions

I hope to have convinced you by now of two things. First, that it is possible to give a straightforward, syntactic, account of the *X or no X* construction, despite the fact that the two disjuncts stand in a relation more abstract than that of simple surface identity. Second, that even though P&R's account might sound simple at first (denotational identity plus 'pragmatics'), upon closer inspection the kinds of things they require their 'pragmatics' to do involve comparing syntactic objects in terms of their internal structure. This is, of course, no argument *against* their theory—perhaps we *do* need our

[15] For details and justification of this claim, see Kobele (2006).

pragmatic theory to be able to compute something like identity of syntactic structures. This theory, however, has yet to be developed, and remains therefore mysterious. The syntactic analysis I have sketched here, on the other hand, is perfectly clear.

A Reconstructing naturality

The English *X or no X* construction can be thought of as involving copying of potentially unboundedly large subexpressions (NPs). In this section I will examine TAGs and MGs, and conclude that they are incapable of describing this generalization in an elegant way. We'll see that it is easy to modify MGs (by adding 'copy movement') so as to allow them to describe this kind of copying insightfully. Again, I am not claiming that TAGs or MGs are incapable of describing the data. The basic patterns of copying, even when there are other (simple) constraints on the copied strings (such as selectional restrictions, etc), are well known to be describable by these so-called 'mildly context-sensitive' grammar formalisms. Nor am I claiming that the structures assigned to expressions with copies cannot support a compositional semantics. (Indeed, if compositionality is taken just to mean that the meaning of the whole is determined by the meanings of its immediate parts as a function of how these parts were combined to form the whole, then *any* grammar can be given a compositional semantics (Zadrozny, 1994).[16]) Instead, I will be trying to show that neither TAGs nor MGs are able to provide a *simple* and *revealing* description of the *X or no X* construction in English (or of similar constructions elsewhere, for that matter). It should be obvious what I intend upon comparison of the descriptions of copying in these different frameworks with the description obtained using CMGs. As a working definition, I will distinguish between simple, revealing analyses and analyses which are not simple and revealing in terms of how novel, open class, words are introduced. In particular, we'll see that whereas CMGs allow the addition of novel words with just a single lexical entry (say, a new proper name, *Bill*), TAGs and MGs require the addition of multiple new lexical items in order to have this new name behave as the others. This is due to the fact that neither TAGs nor MGs are expressive enough to describe the distribution of an expression if this expression can be copied. And this fact, in fact, is the source of the lack of insightfulness of TAG and MG analyses of copying.

[16]Clearly, what Zadrozny's investigation reveals is that this is not all that we have in mind when we talk about compositionality. Kracht (2007) proposes that we not only place limits on what we consider to be appropriate modes of semantic composition, but also speak of compositionality *relative to a particular meaning space*.

A.1 The Gold Standard

I said I was claiming that neither TAGs nor MGs are able to provide an elegant description of 'the *X or no X* construction in English.' But, what exactly is that? P&R devoted a significant portion of their paper to uncovering new and subtle details of this construction; I certainly don't want to be in the position of being forced to give a complete analysis of every facet of this construction before being able to compare the respective grammars descriptions in terms of their elegance. Natural languages are far too complex to make this a feasable task. I will instead base my claim on a simplified, artificial language which is like English, in that it allows for copying in an *X or no X*-like construction, but which places no other syntactic restrictions on its sentences. Clearly, if a grammar formalism can't adequately describe even *this* language, it will have no chance of describing the actual *X or no X* construction in all of its gory glory.

This simplified language is as follows. Given that I just want to model the copying in this construction, no selectional restrictions, or word order constraints or anything of that sort will be enforced. Thus, *any* string of words w will be treated as a legitimate NP. The aspect of the *X or no X* construction that interests us here is the copying of unboundedly large NPs. Accordingly, I will also admit, for any sentence (i.e. NP) w in our language thus far, the sentence (i.e. instance of the *X or no X* construction) $w\mathsf{c}w$, where c is a grammatical marker of the copy construction (modeled after the *or no*), which occurs nowhere else in the language. To make things as simple as possible, I will assume that there are only two words (other than the grammatical marker c) in the language: a and b. Our target language is now the set of all strings w and $w\mathsf{c}w$, where $w \in \{\mathsf{a}, \mathsf{b}\}^*$, which I shall henceforth refer to as $w + w\mathsf{c}w$. This language can be thought of as modeling the fact that, in English, you don't have to copy (w), but you can ($w\mathsf{c}w$). It broadly isolates a major difficulty MGs and TAGs have with copying constructions. Clearly, if we impose tougher conditions on 'w' (i.e. syntactic constraints), the problem only gets harder.

A.1.1 A TAG Account of Copying

The reader unfamiliar with TAGs should consult a standard reference (Joshi (1987) is one). Nothing I say here depends on knowledge about how TAGs work. The most obvious way of defining the language $w + w\mathsf{c}w$ is via the lexicon in figure 1, which consists of six lexical entries.

In order to add another word with the same distribution as a and b to our language, we require *two* new lexical items. Furthermore, a situation

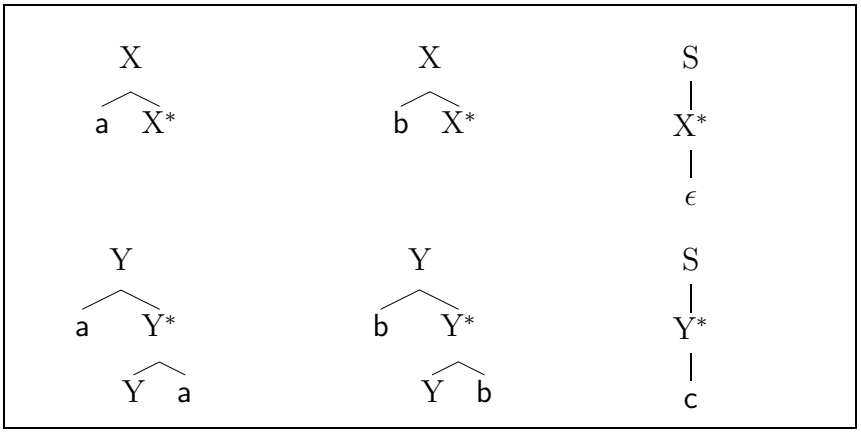

Figure 1: A TAG lexicon for $w + w\mathsf{c}w$

in which there exist expressions in every way like a and b except that they may not appear in copies, or that they may *only* appear in copies, would be straightforward to describe. Indeed, such an unnatural situation would be *easier* to describe than the apparently actual situation in which all such words appear in both copied and non-copied structures, as it would require the addition of just one lexical item.

A.1.2 A MG Account of Copying

Minimalist syntax, with an operation of phrasal movement, provides us with enough in the way of expressive power to be able to describe patterns of copying. Again, as with TAGs, the 'very same' expression is forced to be split into multiple lexical items, which describe its distribution within, and without, copied structures.[17]

To describe the language $w + w\mathsf{c}w$ using nothing but phrasal movement to establish long-distance dependencies between elements, we require the following *eleven* lexical items.

a::=t t	ϵ::t
b::=t t	ϵ::=t s

a::=x +z A -z	a::=A +y x -y	ϵ::=x +y +z s
b::=x +z B -z	b::=B +y x -y	ϵ::w -z
	c::=w x -y	

Figure 2: A MG lexicon for $w + w\mathsf{c}w$

[17] Again, nothing hinges upon knowledge of how minimalist grammars work. The interested reader should consult Stabler (1997).

Here, in order to add another expression with the same distribution as a or b to the language, we need *three* new lexical items: one governing its behaviour in non-copied structures, and two governing its behaviour in the right- and left-hand sides respectively of copied structures. Thus, to add the expression d to the language, we need the following lexical items.

d::=t t d::=x +z C -z
d::=C +y x -y

Furthermore, it is straightforward to define unnatural patterns like $w + w\mathsf{c}w^r$ (figure 3), and such patterns are no more complex than the more natural ones we actually find attested.

a::=t t	ϵ::t	a::=s A -z	a::=A +z s	c::s
b::=t t	ϵ::=t s	b::=s B -z	b::=B +z s	

Figure 3: A MG lexicon for $w + w\mathsf{c}w^r$

A.1.3 Extending the MG Formalism

The basic problem which manifested itself in different ways above is that copying is not a particularly natural derived operation in the TAG or MG formalisms. Thus, not only were expressions artificially forced to be distinguished based on their surrounding context (i.e. whether they found themselves within copies or not), but furthermore the grammars for copying were so complex as to be unable to be formally distinguished from those deriving linguistically unnatural patterns. Introducing a primitive operation of syntactic copying, as formalized in terms of the copy-theory of movement in (Kobele, 2006), allows us to overcome both of these problems.

We can derive the non-copying fragment of the language $w + w\mathsf{c}w$ using the following four lexical items.

a::x ϵ::=x =x x
b::x ϵ::=x s

To add copying to the language, we need just two more lexical items; one to do the copying, and one to 'put the copies into the tree'.

ϵ::=x s -z c::=$\hat{\mathsf{s}}$ +z s

Adding new words with the distribution of a and b to our language is easy. We can add d to the language with just the following lexical item.

d::x

In fact, it would be quite difficult to add a word like a or b but which occured only in non-copies, or only in copies (doing so would require a drastic revision of the whole grammar). Moreover, copying is now distinguished from other, unnatural, patterns as well. The language $w + w\mathsf{c}w^r$, which is indistinguishable in terms of lexical complexity from $w + w\mathsf{c}w$ in the MG and TAG frameworks, is vastly more complex to describe in CMGs than its copying counterpart.

References

Baker, M. (1988). *Incorporation: a theory of grammatical function changing*. Cambridge, Massachusetts: MIT Press.

Brody, M. (2000). Mirror theory: Syntactic representation in perfect syntax. *Linguistic Inquiry 31*(1), 29–56.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, Massachusetts: MIT Press.

Chung, S. (2006). Sluicing and the lexicon: The point of no return. In R. T. Corver and Y. Kim (Eds.), *BLS 31: General Session on Parasession on Prosodic Variation and Change*, Berkeley, CA, pp. 73–91. Berkeley Linguistics Society.

Gärtner, H.-M. and J. Michaelis (2003). A note on countercyclicity and minimalist grammars. In G. Jäger, P. Monachesi, G. Penn, and S. Wintner (Eds.), *Proceedings of Formal Grammar 2003*, pp. 103–114.

Groenink, A. V. (1997). Mild context-sensitivity and tuple-based generalizations of context-free grammar. *Linguistics and Philosophy 20*(6), 607–636.

Hale, J. T. and E. P. Stabler (2005). Strict deterministic aspects of minimalist grammars. In P. Blache, E. P. Stabler, J. Busquets, and R. Moot (Eds.), *Logical Aspects of Computational Linguistics*, Volume 3492 of *Lecture Notes in Computer Science*, Chapter 11, pp. 162–176. Springer.

Joshi, A. K. (1987). An introduction to tree adjoining grammars. In A. Manaster-Ramer (Ed.), *Mathematics of Language*. Amsterdam: John Benjamins.

Kobele, G. M. (2002). Formalizing mirror theory. *Grammars 5*(3), 177–221.

Kobele, G. M. (2006). *Generating Copies: An investigation into structural identity in language and grammar*. Ph. D. thesis, University of California, Los Angeles.

Kracht, M. (2007). The emergence of syntactic structure. *Linguistics and Philosophy 30*(1), 47–95.

Lebeaux, D. (1988). *Language Acquisition and the Form of Grammar*. Ph. D. thesis, University of Massachusetts.

Mahajan, A. (2000). Eliminating head movement. In *The 23rd Generative Linguistics in the Old World Colloquium*.

Manaster-Ramer, A. (1986). Copying in natural language, context-freeness, and Queue Grammars. In *Proceedings of the 24th conference on Association for Computational Linguistics*, New York, pp. 85–89.

Manaster-Ramer, A. (1993). Capacity, complexity, construction. *Annals of Mathematics and Artificial Intelligence 8*, 1–16.

Marantz, A. (1997). No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In A. Dimitriadis, L. Siegel, C. Surek-Clark, and A. Williams (Eds.), *Proceedings of the 21st Annual Penn Linguistics Colloquium*, Volume 4 of *Penn Working Papers in Linguistics*, pp. 201–225.

Merchant, J. (2001). *The Syntax of Silence: Sluicing, Islands, and the Theory of Ellipsis*, Volume 1 of *Oxford Studies in Theoretical Linguistics*. New York: Oxford University Press.

Pullum, G. K. and K. Rawlins (2007). Argument or no argument? *Linguistics and Philosophy 30*(2), 277–287.

Seki, H., T. Matsumura, M. Fujii, and T. Kasami (1991). On multiple context-free grammars. *Theoretical Computer Science 88*, 191–229.

Stabler, E. P. (1997). Derivational minimalism. In C. Retoré (Ed.), *Logical Aspects of Computational Linguistics*, Volume 1328 of *Lecture Notes in Computer Science*, pp. 68–95. Berlin: Springer-Verlag.

Zadrozny, W. (1994). From compositional to systematic semantics. *Linguistics and Philosophy 17*, 329–342.