Argument or no argument?*

Geoffrey K. Pullum and Kyle Rawlins
   University of California, Santa Cruz

Do natural languages ever exhibit unbounded syntactic reduplication? This paper addresses the issues raised by one of the very few constructions in English that would appear to motivate a positive answer, namely the type of adjunct seen in the underlined part of (1).

(1)   The North Koreans were developing nuclear weapons anyway,
        Iraq war or no Iraq war.

We will call such adjuncts instances of the *X or no X* construction It was suggested in the middle 1980s by Alexis Manaster-Ramer that well-formed instances of *X or no X* involve a string *X* of unbounded length preceding *or no*, with an exact copy following (for a brief allusion to it in print, see Manaster-Ramer 1986). This claim has received hardly any discussion at all in the subsequent literature. But if it were true, it would settle, in the negative, the question of whether English was context-free.

It has been widely assumed over four decades that this questions has already been settled, but in fact all published arguments for the negative answer have been flawed. The early arguments from sentences containing *respectively*, for instance, are criticized in Daly (1974) and Pullum and Gazdar (1982) on empirical grounds, and were dealt a final decisive blow by Dalrymple and Kehler (1995), who showed that counterexamples of the sort Pullum and Gazdar adduced are well attested in available corpora. In other cases the controversy has remained open, with both data questions and mathematical disputes proving hard to settle. For example, the literature on the *such that* argument (Higginbotham 1984) remains inconclusive.[1] Similar remarks could be made about the argument from comparative clause string inequality:[2] linguistic data and mathematical basis alike turned out to be problematic, and ultimately the controversy stalled.

Manaster-Ramer's suggested argument, however, is particularly clean, and initially looks very convincing. We first present Manaster-Ramer's argument as convincingly as we can, and then show that it is not convincing on empirical grounds. We also sketch an analysis of what is really going on in the *X or no X* construction. That analysis leads us to the conclusion that it does not involve syntactic reduplication.

Two crucial empirical assumptions are needed to set up the version of Manaster-Ramer's argument that is developed here (for which we take responsibility): that there are indefinitely many instances of the *X or no X* construction, and that they really involve string identity.

The phrases flanking the words *or no* are of the sort we will call *nominals*, as in Huddleston and Pullum (2002). (They are called $\overline{\text{N}}$ in early X-bar theory, and under the DP hypothesis they are the maximal projection of N.) The crucial claim is that English contains unboundedly many nominals that can occur in the *X or no X* construction (a countable infinity of them under the usual view). Actual uses of the construction tend to be very short, limited to single nouns or very short fixed nominals. Daniel Radzinski pointed out to us that the ACL/DCI *Wall Street Journal* corpus appears to contain 16 occurrences of the construction in about 44 million words. The longest of them are phrases like *Cold War or no Cold War*. But there seems intuitively to be no reason why a phrase like *string-identity-based non-context-freeness argument or no string-identity-based non-context-freeness argument* might not be

used, and suggesting a finite length cutoff would obviously trivialize any application of formal language theory argumentation to natural languages. So we will assume that arbitrarily long instances of the construction are grammatical.

The second crucial assumption is that the construction involves word-for-word syntactic identity between strings. This is of the essence, because it is mathematically demonstrable that a CF grammar cannot in general enforce such a requirement over an unbounded domain.

Speakers certainly have fairly strong reactions to arbitrary violations of the condition that the string before *or no* should be the same as the string after it. Everyone accepts the attested (1), but nobody seems inclined to accept anything like (2).

(2)   ??The North Koreans were developing nuclear weapons anyway, [Iraq war] or no [IAEA inspections].

The putative argument that English is not CF depends crucially on a empirical premise about such strings as (2), which for reference we state informally in (3):

(3)   The String-Identity Premise:
      Strings such as (2) are ill-formed for a syntactic reason — the lack of string identity between the bracketed substrings.

Any argument that English as a whole is not CF has to ensure that the crucial string-reduplication property of the *X or no X* construction carries over to English as a whole; showing that English has an infinite proper subset that is non-CF would not, of course, suffice. However, since intersection with a regular set of strings preserves CF status, if we can isolate an infinite proper subset of English that is of *X or no X* form by intersecting English with a regular stringset, then we can construct a non-CFness argument that applies to English as a whole.

There are many irrelevant ways in which sentences of English may contain *or no* flanked by non-identical strings; for example, $X_1 = $ *We have to keep our campaign promises* and $X_2 = $ *one will believe us next time* gives us X_1 *or no* $X_2 = $ *We have to keep our campaign promises or no one will believe us next time*, which is grammatical. We need a way to distinguish instances of the construction in question from non-instances.

It seems to be possible to pick out a set of strings each of which does represent an occurrence of the construction by taking an unbounded set Z of nominal expressions over a tightly restricted vocabulary, and intersecting English with the regular set of all strings of (say) the form *The show will go on,* X_1 *or no* X_2, where X_1 and X_2 are drawn from Z. Let Z be the set of all non-null strings composed of the words *box* and *bag*. (Any two nouns that combine promiscuously in N-N compounds would do just as well.) We will assume that every member of Z is a well-formed English N-N compound: the phrase *box bag* denotes a bag for holding boxes; *box box* denotes a box for holding boxes; *bag box* denotes a box for holding bags; *box box bag* denotes a bag for box boxes; *box bag bag box* denotes a box for bags that hold box bags; and so on without limit (we will assume), up to any arbitrary finite string length. And according to the string-identity claim, cases like (4a) in which $X_1 \neq X_2$ are syntactically deviant, so the intersection of R with English will not contain those; the grammatical word sequences of English will be a set of which (4b) is a typical member.

(4)   a.   *The show will go on, box bag bag or no box bag box.

      b.   The show will go on, box bag bag or no box bag bag.

The argument is now straightforward, and relies solely on elementary results in formal language theory. Let E be the word-stringset of English under the agreed empirical claims. Let R be the regular strings defined thus:

(5) $R = \{\text{We'll do it, } X_1 \text{ or no } X_2 \mid X_1, X_2 \in Z\}$

Ex hypothesi, $E \cap R$ contains just sentences where X_1 and X_2 are identical; thus $R \cap E = \{$ We'll do it, X or no $X \mid X \in Z\}$, which is clearly not CF.[3] Since $R \cap E$ is not CF, and being CF is preserved under intersection with regular stringsets, E is not CF.   **Q.E.D.**

The argument in this form is certainly valid. But we now argue that nonetheless it is not sound. The weak point is that there is reason to think the string-identity claim (3) is not true.

There are attested counterexamples to the claim of string identity, of several kinds. That is, we have found examples that are obvious instances of the construction in question but which have the form X_1 *or no* X_2, where X_1 and X_2 are distinct.[4] All examples come from the World Wide Web. It is true that all sorts of outright linguistic errors occur in prose published on the web, as in most corpora; but it should be clear that our examples could hardly all be dismissed as typos.

First, there are cases of complement ellipsis within one of the nominals, like this one: The phrase *war with Iraq or no war* occurs (apparently independently) as the headline of an article by Marvin Olasky in the *Austin American-Statesman* (October 16, 2002) and *World* (October 26, 2002), and on a film review web site:

(6)   a.   War with Iraq or No War, Innocent People Are Likely to Die. (`journalism.utexas.edu/faculty/vita/Olasky_vita.pdf`)

        b.   War with Iraq or no war, the Oscars show will go on, organizers vowed on Monday as they brought together the class of 2002 for the annual nominees lunch and delivered the traditional and often ignored warning about keeping acceptance speeches short and sweet. (`www.rottentomatoes.com/news-20399/`)

In both cases the phrase was written by a journalist, editorially scrutinized, and picked up by other writers and editors who did not 'correct' it. This suffices to show that string identity between the two nominal phrases in the construction is not an absolute requirement.

Second, there are attested cases of full nominal ellipsis, where the second of the two nominals has been elided under identity of sense with the first, that also represent instances of the construction:[5]

(7)   . . . its willingness to print this story, anonymous source or no, would seem to suggest there's some legitimacy to it . . .

Third, in written English the first nominal may be expanded with a clarification such as an abbreviation expansion:[6]

(8)   DSB (Deep Sand Beds) or no DSB?

And fourth, most spectacularly, there is a large class of examples where the two nominals are of different lengths because of interpolated expressive expletives. An abundance of examples (often offensive in content and tone, of course, and with typing errors characteristic of the thoroughly informal writing in which they are used) can be found on the web. The following phrases are typical; some of them occur several times independently.[7]

(9) a. day trip or no bloody day trip

b. rain or no damn rain

c. Canada Council or no damn Canada Council

d. money or no fucking money

e. pork or no fucking pork

f. beard or no goddam beard

g. Empty Campus Day or no goddamn Empty Campus Day

h. end of the world or no goddamn end of the world

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. And in some of them an expletive interrupts either the first or the second nominal internally; for example:

(10) *I* say let's bring her down and END this like we should have, right from the beginning. New powers or NO new fucking powers.

[shifting-sands.alara.net/vnwu10.htm]

Here we have an instance with X_1 = *new powers* and X_2 = *new fucking powers*.

Presumably multiple expletives would be grammmatical too, though we have not yet found such a case. For instance, (11a) is actually attested, but presumably (11b) would also be possible:

(11) a. end of the world or no goddamn end of the world   [attested]

b. end of the frigging world or no goddamn end of the fucking world [conjectured]

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. The mathematical basis for a non-context-freeness argument appears to be unchanged thereby. For let A be an alphabet of 2 or more symbols and let $\approx$ denote the relation that holds between two strings over A iff one of them the result of inserting extra arbitrary-length strings from A^+ into a copy of the other at arbitrary points, and consider L_1 and L_2 defined thus:

$$L_1 = \{xcw|(x, w \in A^+) \wedge (c \notin A) \wedge (\exists y[(y \in A^+) \wedge (w = xy)])\}$$
$$L_2 = \{xcw|(x, w \in A^+) \wedge (c \notin A) \wedge (x \approx w)\}$$

Both L_1 and L_2 are easily shown to be non-CF. Intuitively, the point is that answering 'Is w is an arbitrary-length prefix of x?' or 'Does w contains a possibly interrupted copy of x?' is just as hard as answering 'Does $w = x$?', and possibly harder. L_1 and L_2 are not going to fall in a simpler language-theoretic class than $\{xcx|x \in A^+ \wedge c \notin A\}$.[8]

The set of strings containing all and only the fully acceptable instances of the construction thus probably is not CF. The new data does not bear on that. Rather, it bears on whether English requires a syntactic description of supra-CF power. We argue that if deviations from string identity are permitted at all, the strong tendency for the string-identity claim to be respected stands in need of a different explanation. We claim that such an explanation is available, and it makes the syntactic string-identity claim redundant. We will outline a

semantic account of the restriction which makes no reference to syntactic reduplication, and provides an explanation of why there would be counterexamples to string-identity of the sort just seen.

The *X or no X* construction seems to be closely related to certain other adjunct constructions involving an obligatory *or*-coordination:

(12)   a.   whether you like it or whether you don't

     b.   whether you like it or not

     c.   like it or not

     d.   whether he goes or I do

     e.   sink or swim

     f.   friend or foe

     g.   for good or ill

In each case, the constituent on the right side of the disjunction denotes approximately the opposite of the one on the left side, but in these examples there does not need to be an explicit negation.

These examples are, in turn, parallel to the construction that Huddleston and Pullum (2002: 760–765) call the open (i.e., *wh-*) ungoverned exhaustive conditional, as in *We're going ahead, whatever you say*. This is discussed in Izvorski 2000, Gawron 2001, and other work cited there.[9]

The morpheme *-ever* in the open exhaustive conditional, as in *whoever it turns out to be*, is obligatory, and contributes "a type of modality that is independent of the verbal system by enforcing universal quantification over epistemic alternatives to the world of evaluation" (Dayal 1997). That is, the *wh-ever* word in an exhaustive conditional quantifies exhaustively over the (relevant) set of possible worlds, and the *-ever* contributes the exhaustivity.

The exhaustivity requirement that correlates with the presence of *-ever* in open exhaustive conditionals applies as well to closed ungoverned exhaustive conditionals, which (as noted by Gawron 2001 and Huddleston and Pullum 2002: 762) must contain a disjunction (contrast *We're going ahead, whether you like it or not* with the ungrammatical **We're going ahead, whether you like it*).

More specifically, each disjunct introduces a new alternative, and it is a semantic condition on the construction that these alternatives must as a set jointly exhaust all the possibilities (that is, all the possibilities relative to the contextual domain at hand). If they do not, the disjunct is semantically anomalous.[10]

Adjuncts like *whether Q or not* have a particular pragmatic point. They convey a point in a rhetorically effective way: by inviting the hearer to see that there is a proof of it. Some point P is claimed to hold whether Q is true or not. This amounts to asserting

$$(Q \Rightarrow P) \wedge ((\neg Q) \Rightarrow P)$$

which is truth-conditionally equivalent to

$$(Q \vee \neg Q) \Rightarrow P$$

And since $Q \vee \neg Q$ can be taken as given (it is a classical tautology), the hearer can immediately infer P via modus ponens.

Even on a non-classical semantics for disjunction (such as that of Zimmermann 2000) the construction can be described in approximately the same way — the hearer is invited to consider different assumptions about Q, and told that P follows from any of them.

The *X or no X* construction is a type of adjunct with exactly the same required presence of a disjunction and the same semantic property of partitioning of the alternatives to the world of evaluation. The requirement that the disjuncts determine a partition, i.e. establish an exhaustive classification of disjoint possibilities, gives a clear reason why in most cases, the two *X*s in the *X or no X* construction need to be string-identical. By assumption the construction involves two disjuncts, one negated. If this is so, the only way the whole disjunction will satisfy the exhaustivity requirement (along with mutual exclusivity) is if the positive disjunct and the negative one pick out sets that exactly complement each other. This can only happen when the first and the second *X* have exactly the same denotation. The unacceptability of an example like (2b) stems from the fact that its two disjuncts fail to exhaust the relevant set of worlds: the worlds in which there is war with Iraq and the worlds in which the Olympic Games are taking place do not exhaust the possibilities relative to any natural contextual domain. The disjuncts are also not mutually exclusive: both the Iraq war and the Olympic games can happen in the same world (for instance, in the actual world).

But in cases such as (6), where there is not exact string identity but the sentence is acceptable, the difference in the word string is not reflected in the denotation. In *war with Iraq or no war*, it is implicit, but necessary, that the war must be the Iraq war, and so the denotation of the right-hand disjunct is the complement of the denotation of the left.

Even the slightest difference in denotation is too much: we do not find cases like ??*pretty girls or no beautiful girls*, or ??*insects or no bugs*. 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.[11] 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.

In other words, we are claiming that when $X_1 \neq X_2$, a use of X_1 *or no* X_2 is exactly as unlikely to be judged acceptable as the phrase *not just* X_1 *but* X_2 is to be judged unacceptable. Examples like *war with Iraq or no war* are among the relatively rare cases in which, because of the understood ellipsis in this case, the string non-identity does not lead to an assumption of semantic distinctness; and notice, ??*not just war with Iraq but war* sounds inane.

There is a corpus-testable prediction here. When a case of X_1 *or no* X_2 is found with X_1 and X_2 synonymous but non-identical, our claim is that locutions like *not just* X_1 *but* X_2 and/or *not just* X_2 *but* X_1 are (ceteris paribus) likely to be acceptable and thus perhaps attested. One case in which the prediction is borne out concerns the apparently synonynous adjective *likely* and *probable*. We would agree that ??*likelihood or no probability* is thoroughly unacceptable. Why, if the senses are the same? Our proposal is that the issue of whether speakers of English are truly taking the two senses to be the same should be reexamined. It then becomes interesting and germane that both *not just likely but probable* and *not just probable but likely* can be found attested on the web. Speakers are apparently taking

the two words to have slightly different senses, differing in degree, even though they do not all agree on the direction in which the difference holds.

The expressive epithets are a particularly interesting test case for our claim about sense identity because, as argued by Potts (2005), they make their contribution toward expressing speaker attitude via a conventional implicature, and are utterly without semantic effect on the sense of the phrase they are attached to. *The baggage handlers stole my damn iPod* is true if and only if the baggage handlers stole my iPod. Use of the interpolated epithet conveys personal irritation, but not directed at or motivated by the iPod. Semantically, *my damn iPod* has the sense that *my iPod* has. The epithet *damn* contributes to an entirely different dimension of meaning (Potts gives a detailed and precise account of the logic involved).

Yet such epithets clearly count as present for purposes relating to the syntactic string. For example, they matter for selecting the right form of the indefinite article, so we get *I don't want a goddam argument about it* because *goddamn* begins with a consonant, not **I don't want an goddam argument about it* on the grounds that *argument* begins with a vowel.

Hence the strict string-identity claim (3) lead us to expect that interpolated expressive epithets will disrupt grammaticality, while the strict sense identity claim leads us to expect otherwise. It is the latter expectation that is borne out.

Given the attested cases where the string-identity condition is not respected, and the prima facie plausibility of the semantics we have informally sketched, no reason remains for regarding a non-CFness argument based on *X or no X* as persuasive. Hence those computational linguists who essentially always take the problem of parsing English to be a problem in CF parsing (which is to say, virtually all of them) do so with considerable justification.

Of course, there may be unbounded-length reduplication constructions in other languages that truly depend on string identity — perhaps, for example, the *X-o-X* construction in the lexical morphology of Bambara, as discussed in Culy 1985, is one of them. And of course at least one dialect of one Germanic language appears to offer a sound argument for non-CFness in the phrasal syntax (Zurich Swiss German, as described by Shieber 1985). It seems likely that some natural languages have a CF word-stringset and some do not. 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. There may be other reasons for adopting a framework with greater expressive power than CF-PSGs, such as the derivational minimalism of Stabler (1997; see Harkema 2001, Michaelis 2001, Kobele and Michaelis 2005), but such a course does not appear to be mandated by mere weak generative capacity considerations in English.

This conclusion has all the more force when we consider strong generative capacity considerations — the structures assigned — rather than weak. Hitherto we have talked about two nominals separated by the sequence *or no*, as if expressions of natural languages were merely strings. But when we consider what structures the relevant expressions have, it seems fully clear that the structure we should assign to an adjunct like *box bag or no box bag* is that of a disjunctive coordination of noun phrases (DPs, if the DP hypothesis is accepted), with *or* as the coordinator in an ordinary coordinate structure and *no* as the determiner in the second noun phrase. That is, the bracketing should be [[*box bag*] [*or* [*no box bag*]]]. 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. A detailed discussion of this topic would call for a separate paper, but preliminary investigation suggests to us that tree adjoining grammars, though capable of generating stringsets like

$\{X \text{ or no } X | X \in (bag \cup box)^+\}$, cannot assign an appropriate structure, and the same seems to be true for the formalization of minimalism presented in Stabler (1997). The most obvious case of a framework that probably does have the requisite power is lexical-functional grammar: the techniques employed in Bresnan et al. (1982) would apparently be equal to the task.

But *X or no X* adjuncts do not provide any support for the adoption of such a framework, if our conclusions in this paper are correct.

References

Bresnan, Joan W.; Ronald M. Kaplan; P. Stanley Peters; and Annie Zaenen (1982). Cross-serial Dependencies in Dutch. *Linguistic Inquiry* **13**, 613–635.

Chomsky, Noam (1963). Formal properties of grammars. In *Handbook of Mathematical Psychology*, Vol. II, ed. by R. Duncan Luce and Robert R. Bush and Eugene Galanter. New York: John Wiley.

Culy, Christopher (1985). The complexity of the vocabulary of Bambara. *Linguistics and Philosophy*, **8**, 345–351.

Dalrymple, Mary and Andrew Kehler (1995). On the constraints imposed by *respectively*. *Linguistic Inquiry*, **26**, 531–536.

Daly, Robert Timon (1974). *Applications of the Mathematical Theory of Linguistics*. The Hague: Mouton.

Dayal, Veneeta (1997). Free Relatives and Ever: Identity and Free Choice Readings. In *Proceedings of SALT VII*, 99–116. Ithaca, NY: CLC Publications.

Gawron, Jean Mark (2001). Universal Concessive Conditionals and Alternative NPs in English. In *Logical Perspectives on Language and Information*, ed. by Cleo Condoravdi and Gerard Renardel de Lavalette, Stanford: CSLI Publications.

Geurts, Bart (2005). Entertaining alternatives: disjunctions as modals. *Natural Language Semantics*, **13**, 383–410.

Harkema, Henk (2001). A characterization of minimalist languages. In *Logical Aspects of Computational Linguistics: 4th International Conference*, ed. by Philippe de Groote, Glyn Morrill and Christian Retoré, 193–211. Berlin: Springer Verlag.

Higginbotham, James (1984). English is not a context-free language. *Linguistic Inquiry*, **15** 119–126.

Higginbotham (1985). Reply to Pullum. *Linguistic Inquiry*, **16**, 298–304.

Hopcroft, John E. and Jeffrey D. Ullman (1979). *Introduction Automata Theory, Languages, and Computation*. Reading, MA: Addison-Wesley.

Huddleston, Rodney and Geoffrey K. Pullum et al. (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Izvorski, Roumyana (2000). Free Adjunct Free Relatives. In *Proceedings of the 19th West Coast Conference on Formal Linguistics*, ed. by Roger Billerey and Brook Danielle Lillehaugen, 232–245. Somerville, MA: Cascadilla Press.

Kobele, Gregory M. and Jens Michaelis (2005). Two Type 0-variants of minimalist grammars. Presented at the 2005 FG-MoL (Formal Grammars/Mathematics Of Language) *FG-MoL '05: The 10th Conference on Formal Grammar and the 9th Meeting on Mathematics of Language*, University of Edinburgh. Proceedings to be published by CSLI Publications, Stanford, California.

Manaster-Ramer (1986). Copying in natural languages, context-freeness, and queue grammars. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, 85–89. New York: Columbia University.

Manaster-Ramer, Alexis (1991). Vacuity. *Linguistics and Philosophy*, **14**, 339–348.

Michaelis, Jens (2001). Transforming linear context-free rewriting systems into minimalist grammars. In *Logical Aspects of Computational Linguistics: 4th International Conference, LACL '01*, ed. by Philippe de Groote, Glyn Morrill and Christian Retoré, 228–244. Berlin: Springer Verlag.

Pelletier, Francis Jeffrey (1988). Vacuous relatives and the (non-)context-freeness of English. *Linguistics and Philosophy*, **11**, 255–260.

Potts, Christopher (2005). *The Logic of Conventional Implicatures*. Oxford: Oxford University Press.

Pullum, Geoffrey K. and Gerald Gazdar (1982). Natural languages and context-free languages. *Linguistics and Philosophy*, **4**, 471–504.

Pullum, Geoffrey K. (1985). *Such that* clauses and the context-freeness of English. *Linguistic Inquiry*, **16**, 291–298.

Shieber, Stuart (1985). Evidence against the context-freeness of human language. Linguistics and Philosophy 8, 333-343.

Simons, Mandy (2001). Disjunction and alternativeness. *Linguistics and Philosophy*, **24**, 597-619.

Stabler, Edward P. (1997). Derivational minimalism. In *Logical Aspects of Computational Linguistics: LACL '96*, ed. by Christian Retoré, 68–95. Berlin: Springer Verlag.

Ullian, Joseph (1966). Failure of a conjecture about context-free languages. *Information and Control*, **9**, 61-65.

Zimmermann, Thomas Ede (2000). Free choice disjunction and epistemic possibility. *Natural Language Semantics*, **8**, 255-290.

Zwicky, Arnold M. and Jerrold M. Sadock (1985). A note on *xy* languages. *Linguistics and Philosophy*, **8**, 229-236.

Notes

[1] Pullum (1985) claimed Higginbotham's argument failed on straightforward empirical grounds; Higginbotham (1985) disputed that; Pelletier (1988) disagreed with Higginbotham's response; Manaster-Ramer (1991) introduced further complications; the dispute was never satisfactorily resolved.

[2] The story may be tracked through Pullum and Gazdar 1982 and Zwicky and Sadock 1985, referring back to Chomsky 1963 and Ullian 1966.

[3] A fully rigorous demonstration can be completed by following the strategy of Hopcroft and Ullmann (1979:136, example 6.5). Let h be the homomorphism defined by $h(\textit{We'll do it}) = e$, $h(\textit{bag}) = a$, and $h(\textit{box}) = b$. Note that $h(R) = (a \cup b)^* c (a \cup b)^*$, and thus $h(R \cap E) = \{xcx | x \in (a \cup b)^*\}$. Intersecting the latter with $a^* b^* c a^* b^*$ we get $\{a^i b^j c a^i b^j | i, j \geq 0\}$. The pumping lemma can be used directly to show that this is not CF (Hopcroft and Ullmann 1979:128, example 6.2). But the class of CFLs is closed under intersection with regular stringsets, so this means that $h(R \cap E)$ is not CF; and the class is also closed under homomorphisms, which means that $R \cap E$ is not CF.

[4] For convenience, we will continue to refer to the construction below For convenience, we will continue to refer to the construction as *X or no X* rather than X_1 *or no* X_2, but from now on the reader should be aware that it is just a name, not a characterization.

[5] We omit the irrelevant preceding and following context for brevity, but the reader can check the source at `http://metatalk.metafilter.com/mefi/10225#247599`.

[6] The source is `saltaquarium.about.com/b/a/142817.htm`.

[7] We omit source URLs; the reader who wishes to check the contexts will find it easier to locate the phrases with Google than to type out the lengthy URLs of the pages where the examples were found.

[8] Thanks to Stuart Shieber for pointing this out. The intuitive observation we have made does not amount to a proof, of course. But constructing the proof seems straightforward, so we simply concede that L_1 and L_2 are non-CF at this point. (If they are not, of course, then the X *or no* X construction does not support a non-CFness argument at all, independently of the considerations presented in the remainder of the paper.)

[9] The construction has often been taken to involve an adjoined free relative. This has become controversial, though not in ways that affect our concerns here: Huddleston and Pullum (2002: 985–9) present detailed argumentation to support the claim that it is syntactically a species of embedded interrogative clause functioning as adjunct; note the parallel with *We're going ahead, whether you like it or not*, where *whether* signals an interrogative. Izvorski's analysis, despite the paper's title, actually treats instances of this construction much like questions semantically, and following literature does as well.

[10] Like any set of alternatives introduced by disjunction, the alternatives also need to be mutually exclusive/distinct, related, and non-trivial; see Zimmermann 2000, Simons 2001, and Geurts 2005 for discussion of such constraints. The interesting property of the *X or no X* construction, however, is that exhaustivity is clearly a condition on felicitous use, and this property is not shared by disjunction at large (though disjunction often asserts exhaustivity, with the right intonation). For this reason we focus on the exhaustivity condition.