FRANCIS JEFFRY PELLETIER

VACUOUS RELATIVES AND THE (NON-) CONTEXT-FREENESS OF ENGLISH

It is an elementary theorem of formal grammar theory that the intersection of a context-free stringset with a regular stringset is a context-free stringset. That is, where R is a regular stringset and $(A \cap R) = B$, if A is a context-free stringset then so is B. Or contrapositively, if B is not context-free, then A isn't.

Higginbotham (1984) uses this argument to attempt to prove that English is not a context-free stringset. In outline, he lets A be English, and

> R = the woman such that (the man such that)* she (gave (this ∪ him) to (this ∪ him))* left is here.

The result, B, is alleged to be

> B = {the woman such that (the man such that)^n she ((gave him to him) ∪ (gave him to this) ∪ (gave this to him) ∪ (gave this to this))^n left is here $|n \geq 0$ and – from left to right – the number of occurrences of 'this' never exceeds by more than 1 the number of occurrences of 'him'}.

Now B is clearly not a context-free stringset,[1] therefore, Higginbotham concludes that English is not either.

According to Higginbotham, the underlying justification for concluding that English intersected with R is B, is this (1984, pp. 227–228)

> In general, if Z is an ordinary English declarative sentence that contains an occurrence of a third-person pronoun which does not have to be taken as having its antecedent within Z, then, where N is any noun that agrees properly with the pronoun in number and gender, the expression
>
> the N such that Z
>
> is well-formed English NP.

As Pullum (1985, p. 292) notes in his response, this is not a sufficient reason to believe that (English ∩ R) will be B. Also needed is the claim that these are the *only* well-formed English NP's (which have occur-

Linguistics and Philosophy **11** (1988) 255–260.

rences of *him*, *her*, and *this* in their *such-that* complements). For, if English did have other ones, then, since all of the possible combinations are in R, (English $\cap$ R) would contain them too and would not be Higginbotham's B. This (English $\cap$ R) might very well be context-free.[2]

Pullum proceeds to give a long and varied list of acceptable English NPs that do not obey Higginbotham's constraint. This places considerable strain on the plausibility of the constraint when applied to the specific example Higginbotham is concerned with (*the man such that . . . . she gave this to him . . . gave him to this left*) but does not destroy it. What is needed is an example, using the specific construction under consideration, wherein one or the other of the two conditions given in B is violated. Pullum (1985, p. 295) gives his

(7b) The woman such that the man such that she gave this to him gave this to this left is here

in which "the count of *him* instances relative to *this* instances [is] not . . . maintained". As he remarks (1985, pp. 295fn), (7b) - as well as sentences which do obey Higginbotham's constraint –

> are prohibitively hard to process or to contextualize, . . . but we do not operate in such matters by attempting to render judgments of acceptablility on extreme cases. Rather, given a clear picture of what generalizations are operative in more natural cases [the ones Pullum had already given – FJP], we . . . extend those generalizations to the case where unaided intuition would fail.

In his reply to Pullum, Higginbotham (1985) calls sentences like

every triangle such that two sides are equal
every book such that it rains

vacuous relatives. These – especially the former – are the type of cases Pullum had been discussing when talking about the "generalizations [which] are operative in more natural cases". Pullum apparently believes (1985, p. 293) that they are both fully grammatical by "a rule expanding relative clause as *such that S*". In his (1985) reply, Higginbotham says that there are three possibilities: (a) *no* vacuous relatives are grammatical (clearly his position in 1984, and – perhaps not so clearly – his position in 1985 [personal communication]), or (b) some vacuous relatives are grammatical and some are not. He rejects Pullum's (c) all vacuous relatives are grammatical. In remarking on this (p. 300) Higginbotham says

> The construction I gave in (1984) requires only that (a) or (b) be true; for, if some vacuous relatives are ungrammatical, then the construction may be reproduced mutatis mutandis, using them instead of the ones that I choose.

This is really the crux of the matter; Higginbotham's further discussion of (what he calls) "split binding" is perhaps of interest but does not address itself to the question of whether "the construction can be reproduced mutatis mutandis" using the ungrammatical vacuous relatives of (b). (Higginbotham tells me [personal communication] that his discussion of "split binding" is designed to explain why he believes some vacuous relatives to be acceptable in spite of his belief that they are ungrammatical. His answer is that "acceptable vacuous relatives [are] in fact reinterpreted so as to be non-vacuous". To this position I have not much to say.)

So, there are three positions here: (i) no vacuous relative is grammatical [Higginbotham], (ii) all vacuous relatives are grammatical [Pullum], (iii) the acceptable vacuous relatives are grammatical and the unacceptable vacuous relatives are ungrammatical.[3] My purpose here is to cast doubt upon the view that a holder of position (iii) can "duplicate mutatis mutandis" Higginbotham's construction, as Higginbotham had suggested would be possible. I think, in fact, that this cannot be done; but I shall not attempt to prove this, rather I shall outline my doubts and throw the ball into the court of such a person so as to challenge him or her actually to "reproduce the construction".

The crucial items are these. (1) We assume that English has grammatical vacuous relatives along the lines outlined by Pullum and agreed to by the holder of position (iii).[4] (2) To "reproduce the construction", the set (English $\cap$ R) must simultaneously satisfy the conditions that (a) two distinct phrases occur the same number of times (for all n) and (b) the two different pronouns must bear some specific relationship to one another – as for example that the number of *this*'s never exceed by more than 1 the number of *him*'s. (3) The set R, which is to be intersected with the English that has some vacuous relatives and which intersection is to be such that it describes the language in (2), is to be a regular set.

We see from this description of the new problem that selection of the set R will be quite difficult. If it is chosen the way Higginbotham did originally, then the set B will not obey condition (2b). So R must be described in such a way that those vacuous relatives which lead to a violation of (2b) are not included in it. Then when it is intersected with English, such strings would not appear in the set B. But there seems to be a dilemma here: either (1) this specification (of which vacuous relatives are to be omitted) is not done in a regular manner, or (2) if something is done in a regular manner, there will be a violation of condition (2a).

What can be done in a regular manner? Well, if there were a finite list of ungrammatical vacuous relatives they could just be listed. This seems a forlorn hope, and in any case is not what upholders of position (b) have in mind. Otherwise we have $\cup$, $\cap$, $*$ to work with. But any time $*$, for example, is used to include (or exclude) certain constructions, then *all* reduplications of that construction are included (or excluded). Now intersect such a language with English (which ex hypothesi has some but not all of these vacuous relatives). If the $*$ was used to *include* a whole range of constructions, then the resulting language B will, as before, not obey the restriction on counting left-to-right the relative frequencies of *this* and *him*. If the $*$ was used to *exclude* a whole range of constructions, then the language B might not have the same number of occurrences of two distinct phrases. (Whether it does or not depends on both how R is chosen and whether in fact English is context-free. But it cannot be merely *assumed*, as our imagined theorist seems to do, that such a regular R can be given which will obey (2a) and (2b). That is tantamount to assuming what was to be proved: that English is not context-free).

In fact I think no such regular R can be given that will yield a language B obeying (2a) and (2b) when intersected with English. This is not so much because I am convinced that English *is* context-free as it is due to disenchantment with the particular example under consideration. If we follow Higginbotham's and Pullum's intuitions about which vacuous relatives are acceptable and which are not (and are thus believed by our theorist to be grammatical or ungrammatical, respectively), the whole issue seems not to be susceptible to any resolution. Higginbotham says (1985, p. 300),

> In his article, Pullum lists a number of examples of vacuous relatives .... These data were not in doubt; and it seems to me that (2) alone raises whatever questions there may be .... That there are numerous acceptable vacuous relatives is beside the question.

On the contrary, as I have said: if there were a finite number of ungrammatical vacuous relatives, or if there were a regularly-describable set of ungrammatical vacuous relatives (as Higginbotham had said in his 1984, namely all of them were ungrammatical), then one could describe the required regular language R. The merit of Pullum's article is not just to point to one acceptable vacuous relative, nor to just one class of them, but rather to show a variety of acceptable vacuous relatives. If all these acceptable vacuous relatives were grammatical, then, given the wide variety, it seems unreasonable to suppose that the required regular language R can be given. (I hasten to add that the complexity of the set of presumed-grammatical vacuous relatives in no way indicates that

English is not context-free; for any interesting context-free language has non-context-free sublanguages. To guarantee that the non-context-freeness of a sublanguage is inherited by the entire language, one needs to show that the sublanguage plus other things not in the whole language can be described in a regular manner.)

Pullum, it will be recalled, apparently believes that all vacuous relatives are grammatical, but that some of them are unacceptable because they are "bizarre". Higginbotham (1985, p. 303) marshalls arguments to the effect that there is no good explanation for why they are "bizarre". These same arguments can be used against the claim of our position (iii) theorist (who thinks that they are ungrammatical) that he or she can describe the ungrammatical vacuous relatives (=Pullum's bizarre ones) in a regular way. This, together with the unclarity of the notion of "split binding" (which might be used by our position (iii) theorist) to describe the grammatical vacuous relatives, should convince anyone that the ungrammatical vacuous relatives cannot be described regularly.

To put my point in a more pictorial mode: Let E be the set of well-formed English sentences, S be the set of all "*such that* sentences", and V be the vacuous ones of these. Then the picture which we get from our position (iii) theorist who thinks that he or she can "reproduce Higginbotham's proof" is this. Our theorist claims that B is not context free and can be generated by intersecting E with some regular set. But given the difficulty he has in distinguishing the grammatical vacuous

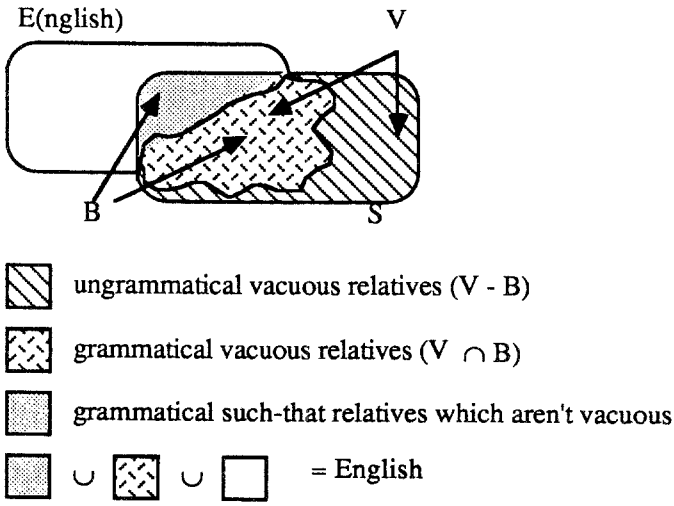

Fig. 1.

relatives ($B \cap V$) from the ungrammatical ones ($V - B$), it seems unlikely that a regular R could be found such that $(R \cap E) = B$. Intuitively, the content of my claim is: if you think that there are some grammatical and some ungrammatical vacuous relatives, but you cannot distinguish between them in a regular manner, then you cannot use the argument of Higginbotham (1984) or any easy modification of it.

Of course this is no proof that no argument using vacuous relatives can ever be given, but the difficulties Pullum has in explaining why certain ones are "bizarre" even though grammatical, and the difficulties our imaginary theorist has in attempting to account for the grammaticality of the acceptable ones by "split binding", seem to show that it is quite unlikely. Given the recalcitrant nature of vacuous relatives, my recommendation is that this is not an area where one should look to find a proof of the non-context-freeness of English.[5]

NOTES

[1] Because a context-free grammar cannot simultaneously "keep track" of the twin requirements (a) that the n's be the same, and (b) that the 'this' not exceed the 'him' by more than 1.

[2] It is clear enough that Higginbotham (1984) recognized the point; it follows from the style in which his construction proceeds.

[3] The *Linguistics and Philosophy* referee remarks that there are other possible positions: perhaps the acceptable vacuous relatives are ungrammatical and the unacceptable ones are grammatical. The referee continues: "[this is] a position so silly that even the participants to this debate have failed to attack or defend it".

[4] In particular, included in these are to be some selection of ones like Pullum's (7b).

[5] Thanks to Jim Higginbotham, Geoff Pullum, and an anonymous *Linguistics and Philosophy* referee for their comments on an earlier version of this paper. This research was supported in part by a Canadian Natural Science and Engineering Research Council grant #A5525.

REFERENCES

Higginbotham, J.: 1984, 'English is Not a Context-Free Language', *Linguistic Inquiry* **15**, 119–126.

Higginbotham, J.: 1985, 'Reply to Pullum', *Linguistic Inquiry* **16**, 298–304.

Pullum, G.: 1985, '*Such That* Clauses and the Context-Freeness of English', *Linguistic Inquiry* **16**, 291–298.