

TECHNICAL CORRESPONDENCE

SUBJECT-VERB AGREEMENT IN RESPECTIVE COORDINATIONS AND CONTEXT-FREENESS

Langendoen (1977) advanced an argument against English being a context-free language involving cross-serial subject-verb agreement in *respectively* constructions such as (1).

- (1) The man and the women dances and sing, respectively.

As noted by Pullum and Gazdar (1982), however, and acknowledged subsequently by Langendoen (personal communication), such examples are unacceptable, and the argument collapses on empirical grounds.

However, at least some speakers reject examples like (2) as well.

- (2) The man and the women dance and sing, respectively.

This fact leads directly to a demonstration that there is, after all, a cross-serial dependency involving the grammatical number of subject NPs and verbs in *respectively* constructions. However, it is not clear at present how representative such speakers are, and so instead of making claims about English in general, we will confine them to just those varieties of the language that stigmatize examples like (2), which will be denoted as *English₁*, leaving to one side any varieties of which this may not be true (*English₂*).¹

In *English₁*, a verb that formally distinguishes singular from plural, i.e., a non-auxiliary present tense verb, cannot occur in a *respectively* construction if the corresponding subject NP is singular. This cannot be accounted for merely by barring marked singular verbs from occurring in coordinate predicates of *respectively* constructions. Such a move would correctly exclude examples like (1), but it would allow sentences like (2), with plural verbs corresponding to singular subjects.

Nor is it possible to simply bar singular subjects from occurring in *respectively* constructions, since they are perfectly possible provided the corresponding verb is either a past tense, as in (3a), or an auxiliary, as in (3b).

- (3) a. The man and the women danced and sing, respectively.
 b. The women and the man sing and can dance, respectively.

This means that a singular subject can only co-occur with a past tense or an auxiliary verb, whereas a plural subject can take a non-auxiliary present tense verb as well. The difference in the co-occurrence possibilities of singular as opposed to plural subject NPs amounts to a peculiar kind of number agreement.² This fact leads

quite directly to a demonstration that *English₁* is not context free.³

Consider the regular set (4).

- (4) {the man x and the women danced y and sing respectively | $x \in \{\text{the man, the women}\}^+$; $y \in \{\text{danced, sing}\}^+$ }

This is the set of all strings⁴ (only some of them grammatical in English) consisting of any number of occurrences in any order of the phrases *the man* and *the women*, with an *and* between the last two, followed by any number of occurrences in any order of the words *danced* and *sing*, with an *and* between the last two, followed by the word *respectively*.

According to what has been said, the intersection of (4) with *English₁* must be (5).

- (5) {the man x and the women danced y and sing respectively | $x \in \{\text{the man, the women}\}^+$; $y = \sigma_1(x) \in \{\text{danced, sing}\}^+$; $\sigma_1(\text{the man}) = \{\text{danced}\}$; $\sigma_1(\text{the women}) = \{\text{danced, sing}\}$ }

This is the set of all those strings of (4) that meet the additional condition that every occurrence of *the man* must be matched by an occurrence of *danced* and every occurrence of *the women* by an occurrence of either *danced* or *sing*. This matching is achieved by defining the substitution⁵ σ_1 of *the man* to be the set {*danced*} and that of *the women* to be the set {*danced, sing*} and requiring y to be equal to σ_1 of x .

We now define a substitution σ_2 such that

$$\begin{aligned} \sigma_2(\text{the man}) &= \{a\} \\ \sigma_2(\text{the woman}) &= \{b\} \\ \sigma_2(\text{danced}) &= \{c\} \\ \sigma_2(\text{sing}) &= \{d\} \\ \sigma_2(\text{and}) &= \{\lambda\}^6 \\ \sigma_2(\text{respectively}) &= \{\lambda\} \end{aligned}$$

This substitution maps (5) to (6).

- (6) {a x b c y d | $x \in \{a, b\}^+$; $y = \sigma_3 \in \{c, d\}^+$; $\sigma_3(a) = c$; $\sigma_3(b) = \{c, d\}$ & rbrc.

We now intersect (6) with the regular set

$$\{a^+ b^+ a^+ c^+ d^+ c^+\}$$

to obtain (7).

- (7) { $a^n b^m a^l c^k d^j c^i$ | $n \leq k$; $m \leq j$; $l \leq i$; $n + m + l = k + j + i$ }

The set in (7) can be shown trans-context-free by pumping. Take the string $z = a^k b^k a^k c^k d^k c^k$ (where k is the constant of the pumping lemma). On the one hand, if we pump only in the first or only in the second half of the string, the resulting string will violate the condition that $n + m + l = k + j + i$. On the other hand, since k

is the constant of the pumping lemma, the pumped substring cannot be longer than k , and therefore the only other place we might be able to pump would be in the middle as and the middle cs . But this would result in violating the condition that l may not be greater than i . Thus, z cannot be pumped without violating the pumping lemma, and hence (7) is not context free. Since context-free languages are closed under intersection with regular sets, it follows that (6) is not context-free either. Since context-free languages are also closed under substitution, this means that (5) is also not context-free. Finally, since (5) is the intersection of English, with the regular set (4), it follows that English₁ is not context-free. Q.E.D.

Alexis Manaster-Ramer
 Computer Science Department
 Wayne State University
 Detroit, MI 48202

REFERENCES

Kac, M.B.; Manaster-Ramer, A; and Rounds, W.C. 1987 Simultaneous-Distributive Coordination and Context-Freeness. *Computational Linguistics* 13(1-2): 25-30.
 Langendoen, D.T. 1977 On the Inadequacy of Type-3 and Type-2 Grammars for Human Languages. In Hopper, P.J., Ed., *Studies in Descriptive and Historical Linguistics: Festschrift for Winfred P. Lehman*. John Benjamins, Amsterdam: 159-172.
 Manaster-Ramer, A. 1983 The Soft Formal Underbelly of Theoretical Syntax. *CLS* 19: 256-262.
 Manaster-Ramer, A. In press. Dutch as a Formal Language. *Linguistics and Philosophy* 10.
 Pullum, G. and Gazdar, G. 1982 Context-Free Languages and Natural Languages. *Linguistics and Philosophy* 4: 471-504.

NOTES

1. Pullum and Gazdar (1982) state that they "can tolerate" examples like (2), and Langendoen (personal communication) agrees.
2. In other terms, we must be able to tell which verb would agree with which subject given the chance, and disallow just those combinations where the result would be a marked singular verb.
3. Ignoring, for the sake of simplicity, the arguments advanced in Manaster-Ramer (1983; in press) about the need to state formal results about natural language in other than weak generative capacity terms.
4. Ignoring, for the sake of readability, the punctuation that would normally be required in written English and the suprasegmental features that would occur in the spoken language.
5. In the discussions of formal properties of natural languages, substitutions have not figured at all prominently, whereas homomorphisms, which are just a special case of substitutions, have. It may be helpful, therefore, to point out that a substitution is a mapping like a homomorphism except that it is usually denoted by σ rather than h and that it may associate each element in the vocabulary of a language with a whole set (possibly infinite) of strings rather than with just one string, as in the case of a homomorphism. In the present case, we needed to employ a (slightly more general kind of) substitution in order to be able to associate *the women* with *sing* as well as *danced*. It should also be noted that, while *the man* and *the women* are linguistically analyzable, we have for technical convenience treated them as single elements of the terminal vocabulary in defining the substitutions.
6. λ denotes the empty string.

A NOTE ON A STUDY OF CASES

This note describes and illustrates a study of deep cases using a large sample of sentences. The purpose of the note is to draw attention to the value of the source material used for those interested in case-based representations of sentence meaning, and to indicate the potential utility of the study results.

The purpose of this note is to draw attention to the utility of a specific source of data relevant to the use of case-based meaning representations in language processing, by illustrating the way we have exploited this source.

Like many others, we have used a language analyser that builds meaning representations expressing semantic case roles; specifically, Boguraev's (1979) analyser builds dependency trees with word senses defined by **semantic category primitive** formulae, and with case labels, i.e., **semantic relation primitives**, on the constituents of verb (and some other) structures.

Using the analyser for more varied and demanding purposes than Boguraev's original tests (see, e.g., Boguraev and Sparck Jones 1983) left us dissatisfied with the original set of case relations. We therefore carried out a detailed analysis of a large sample of English sentences to evaluate our proposals for a better-founded and more comprehensive set of case relations. This study exploited F.T. Wood's "English prepositional idioms" (Wood 1967), which provides a careful account, supported by extensive examples, of the uses of English prepositions and preposition-like terms. For instance,

WITHIN

(1) Inside

Within the house all was quiet.

The Kingdom of God is within you.

(2) Amongst the members of a group.

Opinion within the profession is divided.

(3) Inside specified bounds or limits.

They were ordered to remain within the precincts of the college.

The scholarship is open to anyone residing within fifty miles of the university.

He always strove to live within his income.

...

Our study was intended to establish both the justification for each case relation individually, by reference to a range of sentences, and the plausibility of the set of relations as a whole, by reference to the complete set of sentences. Looking at Wood's description of a preposition's sense, and its accompanying illustration(s), we tried to assign a case label to the link between the sentence elements made by the preposition which we felt captured the essential nature of that link, at the level of generality represented by a set of 20-30 cases. Thus "location" would be the label associated with a number of specific space-relation prepositions, e.g., *above*, *at*, *by*. The study was primarily concerned with prepositionally-