

# Complexity, ellipsis and subject-verb agreement puzzles in NP coordination

Rui P. Chaves Jeruen E. Dery

{rchaves,jedery}@buffalo.edu

Department of Linguistics, University at Buffalo – SUNY

## 1. Problem: an agreement puzzle

Hoeksema (1988) first noticed the determiner-sensitive subject-verb agreement puzzle in (1) and (2).

- (1) **Non-universally quantified conjuncts**
- The boy and the girl were given sweets.
  - \*The boy and the girl was given sweets.
  - An ethnic group and a tribe were documented.
  - \*An ethnic group and a tribe was documented.
- (2) **Universally quantified conjuncts**
- Every boy and every girl were given sweets.
  - Every boy and every girl was given sweets.
  - Each ethnic group and each tribe were documented.
  - Each ethnic group and each tribe was documented.

## 2. Why is there such a subject-verb agreement difference?

- Hypothesis 1: no explanation, it's just the way it is.**

The grammar of subject-verb agreement is idiosyncratic and contains a special provision which is sensitive to the determiner type of NP conjuncts.

**Problem:** such a peculiar stipulation is unexpected, and it is more likely that the contrast follows from independently motivated factors.

- Hypothesis 2: lexical ambiguity of 'and'.**

There are two 'and' coordinators; the one in (1) forms a plurality, but the one in (2) does not.

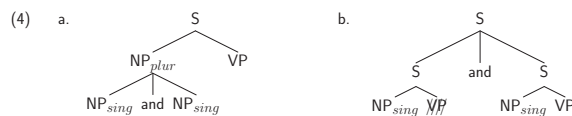
**Problem:** conjunction can form pluralities regardless of the determiner type, as shown in (3). It is unlikely that the contrast between (1) and (2) results from different coordinator meanings.

- (3)  $\left\{ \begin{array}{l} \text{Every boy and every girl} \\ \text{The boy and the girl} \end{array} \right\}$  talked to each other.

- Hypothesis 3: competence & processing conspire.**

Chaves (2007) proposes that the grammar of subject-verb agreement behaves in a systematic way, and that the agreement puzzle is due to the interaction of performance and the grammar:

- the plural verb agreement** in (1a,c) and (2a,c) is due to a **direct NP coordination parse** like (4a). **The singular verb** in (2b,d) obtains with an **elliptical S coordination parse** like (4b).



**Note:** there is independent motivation for this kind of ellipsis (known as 'right-node raising'):

- Pre- and post-war Germany were very different.
- Signals move from a pre- to a post-synaptic neuron.
- I know a man who loves and a woman who hates any operative system other than MS.

**Prediction:** the NP coordination parse and the S coordination parse are impossible in (6).

- \*[Every boy and every girl]<sub>plur</sub> is<sub>sing</sub> talking to each other.
- \*[Every boy<sub>i</sub> is talking to each other<sub>i</sub>] and [Every girl<sub>j</sub> is talking to each other<sub>j</sub>].

- the unacceptability of a singular verb** in (1b,d) is due to the **preemption of the S coordination parse (4b) by the direct NP coordination parse in (4a)**. Crucially, such preemption is not triggered by the universally quantified conjuncts in (2).

That is, once the parser runs into a sequence 'NP and NP', the presence of singular determiners like *a*, *the*, or *one* overwhelmingly biases the parser towards a NP coordination parse, preempting alternative (and more complex) elliptical S coordination parses. The presence of a singular verb causes so much surprisal that no backtracking is done. If so, then **why is there such a syntactic preemption?**

## 3. Pragmatic conditions, semantic ambiguity, and frequency

- Pragmatics: peculiar contexts are required**

Special situational contexts are needed for a universally quantified conjunction. Thus, these structures are more unexpected and pragmatically marked than non-universally quantified conjunctions.

- Semantics: complex and highly ambiguous**

Unlike non-universally quantified conjuncts, universally quantified conjuncts yield complex ambiguities. Sentence (7a) either means that boys shook hands with girls, or that kids shook each other's hands. In (7b) the indefinite NP can be interpreted specifically (a unique particular guard for each inmate) or non-specifically (possibly different guards for each inmate), and (7c) is even more complex, as it allows the entire group of individuals to meet, or just pairs of inmates and lawyers.

- Every boy and every girl shook hands.
- Every inmate and a guard met.
- Every inmate and each of his lawyers met.
- An inmate and a lawyer met.
- A boy and a girl shook hands.

Note that **none of these ambiguities exists** in existentially quantified conjunctions like (7d,e).

Thus universally quantified conjunctions are extremely rare when compared to non-universally quantified conjunctions, due to (i) their peculiar contextual requirements and (ii) the tendency for speakers to avoid complex semantic ambiguities. This frequency difference is observable in the BNC corpus:

NP coordination type	BNC absolute frequency
<i>the ...<sub>n</sub> and the ...</i>	28,952
<i>a ...<sub>n</sub> and a ...</i>	4,682
<i>one ...<sub>n</sub> and one ...</i>	472
<i>every ...<sub>n</sub> and every ...</i>	74
<i>each ...<sub>n</sub> and each ...</i>	36

Table 1. Occurrence of [*Det ...<sub>n</sub> and Det ...*] conjunctions, where  $1 < n < 4$ .

**The proposal:** in the case of high-frequency strings (like 'the N and the N'), speakers are so highly biased towards the direct coordination parse in (4a) that the alternative (and more complex) elliptical S-coordination parse (4b) is preempted, and thus not recoverable in the presence of a singular verb. On the other hand, because the coordination of universally-quantified NPs is rather infrequent, no significant bias for the direct parse (4a) exists, and the elliptical parse (4b) is also available.

**Prediction:** the **more frequent** the NP coordination type, the **less acceptable** the presence of a singular verb is. Our experimental findings support this hypothesis.

## 4. Magnitude Estimation experiments

We conducted five Magnitude Estimation experiments (Bard et al. 1996; Cowart 1997), manipulating determiner type (*a*, *one*, *the*, *each*, and *every*), and verb number (*singular* vs. *plural*) as experimental factors. All of our items had the following form:

[DET NOUN<sub>1</sub> and DET NOUN<sub>2</sub>  $\left\{ \begin{array}{l} \text{were} \\ \text{was} \end{array} \right\}$  PRED]

The two determiners (DET) were identical for each item, the two nouns were different singular nouns for each item, and PRED was a scenario-appropriate adjective, participial verb, or prepositional phrase. This constitutes a 2-factor  $5 \times 2$  design. We constructed 100 test items, each having ten versions, and counterbalanced across ten presentational lists. The nouns and adjectives were different for each item, and equally frequent (Kučera and Francis 1967). All ten lists had the same reference sentence (the modulus), *the patient examined by the doctor was discharged yesterday*. 40 participants were recruited, all undergraduate students native speakers of American English. The responses of the participants were grouped according to category, and log-transformed prior to analysis.

## 5. Coordination frequency $\approx$ difference in verb acceptability

As usual in Magnitude Estimation, the raw ratio scores of the participants were grouped according to category, and log-transformed prior to analysis to normalize distribution. The five t-tests are seen below.

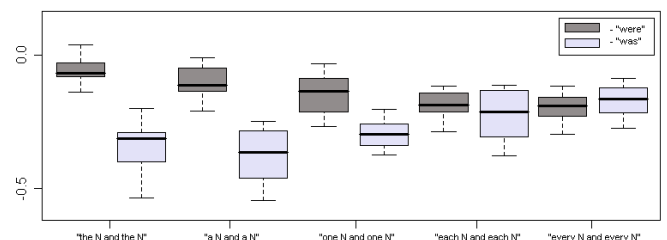


Figure 1: Acceptability magnitude judgements for coordination types and verb agreement

A **significant** difference was found for the conjunctions with universal quantifiers, but **no significant** difference was found for non-universal ones. Because five different experiments were piggy-backed to each other, we use a Bonferroni-corrected **p-value of .01**.

Coordination type vs. Verb form	t[798]	d	r	p
'the N and the N'	17.038	1.206	.516	.0001
'a N and a N'	14.088	.997	.446	.0001
'one N and one N'	9.220	.653	.31	.0001
'each N and each N'	2.390	.169	.084	.017
'every N and every N'	-1.256	-.089	.044	.209

Table 2. t-tests: coordination type vs. verb number agreement

The **Pearson's correlation** between the effect-size *r* for universally quantified conjuncts ( $M=0.28$ ,  $SD=0.1889$ ,  $N=5$ ) and the proportions in the BNC frequency values in Table 1 ( $M=6843.2$ ,  $SD=11191.2649$ ,  $N=5$ ) approached significance ( $r(3)=-.719$ ,  $p=0.085$ ).

This suggests that the less frequent the coordination type, the less likely for comprehenders to distinguish between singular and plural agreement, and is consistent with the hypothesis that the more frequent determiner-NP-coordination type, the greater the bias towards an NP coordination parse. Thus, unacceptable sentences like *\*the boy and the girl was thrilled about it* are worse than *\*one boy and one girl was thrilled about it* (although in our analysis both are grammatical).

## Acknowledgements

Many thanks to Doug Roland for comments and discussion.

## References

- Bard, E. G., D. Robertson, and A. Sorace (1996). Magnitude estimation of linguistic acceptability. *Language* 72, 32–68.
- Chaves, R. P. (2007). *Coordinate Structures – Constraint-Based Syntax-Semantics Processing*. Phd dissertation, University of Lisbon, Portugal.
- Cowart, W. (1997). *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. Thousand Oaks, CA: Sage Publications.
- Hoeksema, J. (1988). The semantics of non-boolean *and*. *Journal of Semantics* 6, 16–40.
- Kučera, H. and W. N. Francis (1967). *Computational Analysis of Present-Day American English*. Providence: Brown University Press.