# You are *"the only one"*…how far do we go in search for referents?

Francesca Foppolo[*], Luisa Meroni[#] and Andrea Gualmini[#] (Arbiter)
*University of Milano-Bicocca, #Utrecht Institute of Linguistics-OTS

**Purpose of the visit**
Francesca Foppolo (1st participant) to OTS, Utrecht, to meet with Luisa Meroni (2nd participant) and Andrea Gualmini (arbiter): preliminary discussion about experimental design; preparation of experimental stimuli to be used with Tobii eye-tracker for the baseline experiment with adults.

**Work carried out during first visit**
discussion of experimental design and preparation of sample stimuli, as in the detailed report that follows.

## REPORT

**Our starting point.**
Our project builds on some findings by Crain, Ni and Conway (1994) on the different parsing strategies adopted by children and adults in solving ambiguities related to the phenomenon of "one-substitution", exemplified by (1):
(1) *The big elephant is the only one that is playing guitar*
The ambiguity relies on the fact that *one* might substitute "the only (one)" with either (a) or (b):
(a) the only thing/animal;
(b) the only elephant (among elephants).
Crain et al.'s results seem to suggest that the adults' parser is guided by a strategy of "minimal commitment", which prefers the interpretation that has more chances to be true (i.e. (b)) to avoid costly and unnecessary commitments. On the contrary, children seem to adopt a strategy of "maximal commitment", judging (1) true only in situation (a). This difference was interpreted in terms of "learnability", in line with the Semantic-Subset principle (Crain and Philip, 1993).

**Our project: summary**
In this project we will focus mainly on the results with adults. As for this, we cast doubts in generalizing the observed adults' behaviour in the experiment by Crain et al. We think that another possibility should be taken into account -namely- that the observed adults' interpretation could be due to the presence of a scalar adjective (such as 'big') in the target sentence. The use of a scalar adjective in fact, may induce and favour a within-set comparison in adults ("big" among elephants). The argument proposed is thus that the strategy of "minimal commitment" observed in adults by Crain and colleagues may be (at least for those trials in which a scalar adjective was used) induced by the presence of a scalar adjective. As shown by eye-tracking studies by e.g. Sedivy et al. (1996) and Sedivy (2001) adults promptly consider the presence of a contrastive object and converge on the target item more quickly if the adjective used in the instruction relates to material or scalar properties (eg. "Pick up the *tall* glass" in a visual scene in which a "tall" and a "small" glass are present) than when it refers to properties that are frequently redundantly encoded, such as color adjectives. To support this claim, a questionnaire study with 60 Italian adults on the Italian equivalents of sentences like (1) -in which no scalar adjective was used- showed no evidence of a tendency towards a strategy of "minimal commitment" (Foppolo, in prep).
Our proposal aims at working out in detail different predictions and parsing strategies with respect to the phenomenon of "one-substitution" discussed above and to test these predictions by means of an experiment designed to test our opposing claims. In particular, we will address the following experimental questions: how do adults (and children) resolve the ambiguity related to "one-substitution"?; may (linguistic, visual, etc…) context override/modulate adults (and children) parser's default strategy? Two opposing predictions will be formulated – namely the ones which

relate with the strategy adults (and children) will adopt, either a "minimal commitment" (Hp1) or a "maximal commitment" strategy (Hp2). A visual world paradigm with Tobii eye-tracker will be used with adults and may be extended to children. The results obtained with the proposed paradigm will either corroborate or refute Crain et al.'s findings about the ambiguity resolution in case of "one-substitution". Furthermore, it will shed some lights on the role of linguistic and extra-linguistic context in modelling the parser's preferences.

**Planned method.**
Visual world paradigm with Tobii eye-tracker to be used with adults.

**Experimental design: details**

4 factors varying on 2/3 levels:

1. Language
> {English, Italian, Dutch}                    Between subjects
> relevant factor: position of the adjective w.r.t. the noun
> English and Dutch: adjective preceded the noun
> Italian: adjective follows the noun

2. Adjective
> {(scalar (i.e., size), non-scalar (i.e., color)}        Within subjects

3. Context
> {C: <u>true</u> if "minimal commitment" and <u>false</u> if "maximal commitment" strategy; T: always <u>true</u>; F: always <u>false</u>}
>> Within subjects: C-T; C-F
>> Between subjects: T-F

4. List
> For each language group, two pseudo-randomized lists of subjects, tested on 6 different items per each critical condition (+ fillers as indicated):

| # | List 1  (C-T) | # | List 2 (C-F) |
|---|---|---|---|
| 6 | Scalar C | 6 | Scalar C |
| 6 | Scalar T | 6 | Scalar F |
| 6 | Non-scalar C | 6 | Non-scalar C |
| 6 | Non scalar T | 6 | Non scalar F |
| | | | |
| 16 | true fillers | 24 | true fillers |
| 24 | false fillers | 16 | false |

> Subjects to be tested in each list: at least 12

**Participants' task**
attend to sentences (presented acoustically) and judge them as TRUE or FALSE (by pressing button) with respect to a visual scene presented on the screen

**Experimental samples (examples of stimuli per condition)**
Factor: Adjective -- <u>Scalar</u> (S)
      Target sentence:     *The big circle is the only one that is reading*
      C (critical) → Fig.1
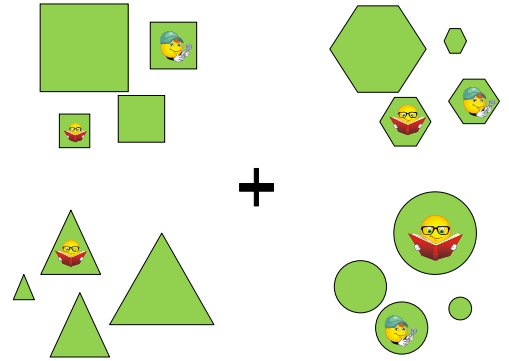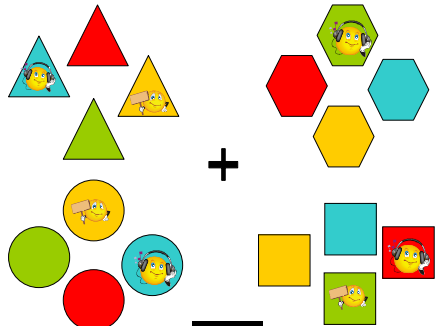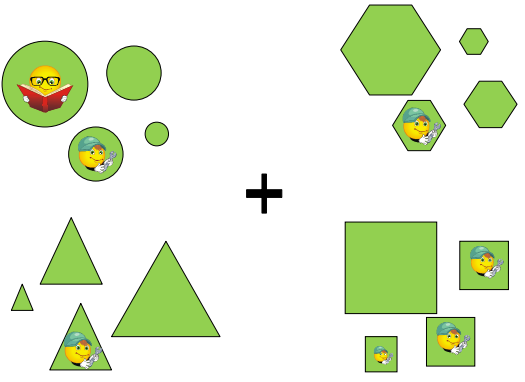      T (true) → Fig.2
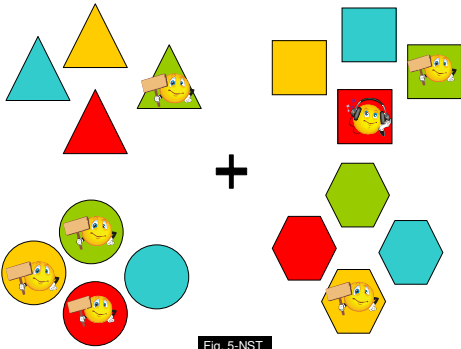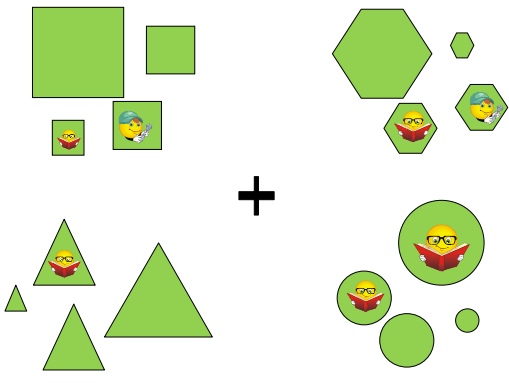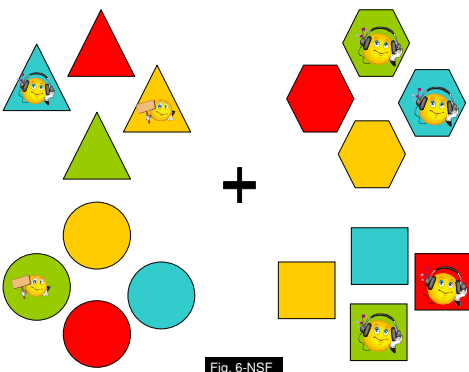      F (false) → Fig.3

Factor: Adjective – <u>Non - Scalar</u> (N)
      Target sentences:     *The red square is the only one that is listening to music*
      C (critical) → Fig.4
      T (true) → Fig.5
      F (False) → Fig. 6

| Scalar | Non Scalar |
|---|---|
| The big circle is the only one that is reading | The red square is the only one that is listening to music |
| <br>Fig. 1-SC | <br>Fig. 4-NSC |
| <br>Fig. 2-ST | <br>Fig. 5-NST |
| <br>Fig. 3-SF | <br>Fig. 6-NSF |

3

**Rationale for the preparation of the stimuli**

- Each scene has four groups of objects, each consisting of four objects of the same type. The four groups are also related. For example, a scene with sixteen geometrical figures, divided in four groups of the same type. Other possibilities are: sixteen consonants (please use consonants that look and sound differently in the same trial, for instance, P, Q, T, S but not B with V or P with R), sixteen numbers (same criteria as above, e.g. 3 and 8 not on the same trial).

- Only "abstract" objects for critical trials (numbers, consonants, geometrical figures) but for fillers both abstract and real (related) objects can be used (e.g. 4 shovels, 4 buckets, 4 shells, 4 beach umbrellas).

- In order to vary the type of predicates in our experiment, we decided to use "emoticons". Emoticons are simple images that can easily represent all type of predicates, even those that refer to feelings and emotions, otherwise difficult to represent visually. We think emoticons are nowadays so commonly used and known and have the advantage of being immediately interpreted.

- In all conditions, at least one item of each set has to have an emoticon. This maneuver is to encourage the exploration of the whole scene, not relying only on peripheral vision. For the same reason, two different actions have to be mixed in the same scene (keeping the ratio of target/alternative action constant across similar trials.

- The number and type of emoticons has to be kept constant across trials of the same type, only varying their position with respect to their set and the whole scene. Crucially, though, no emoticon should appear in the objects that share the property of the target (e.g. in case the target is the red square, no emoticons should appear in any other "red" object in the scene)

- In the critical condition, one member of the target set has to do an alternative action from the target member (eg: in the critical trial "the red square is the only one that is smiling" there will be the red square that is smiling and some other (random) square that is doing something else). The same for condition 2 (true). In case of condition 3 (false), two members of the target set do the same action (so as to make it "not the only one").

- In the Scalar condition, the color of the items should be kept constant across sets and the same for the difference in size. For example, the big square in Fig 1 should be as big as the big triangle in the same scene.

- In the Non-scalar condition all the items should be of the same size but they will vary in color.

**Randomization**

- The position of each member within its own set has to be randomized across trials.(i.e the biggest square not always in the upper left corner in the set of squares).

- Also, the position of different sets should be randomized across trials (e.g. set of squares not always in the same position).

- The position of the target set should vary across trials.

**Fillers**

- Half fillers (24) require that subjects explore all four sets of objects on the scene (e.g. All the yellow numbers are yawning); the others require that subjects explore only the relevant set (e.g. All the 3s are sleeping).

- Type of items used for fillers: Quantifiers{all, no, some, three}

- 12 sentences per quantifier; 6 for each of the following sentence type; each presented in two different contexts: one making it true and one making it false.

Examples of filler sentences

1. All the yellow numbers are yawning     (exploration of the whole scene → Fig.7)
2. All the 3s are sleeping     (exploration of the relevant set → Fig.8)

3. No red letter has fever
4. No bucket wears eye-glass
5. Some red letters are reading
6. Some buckets play the trumpet

7. Three green numbers are in love
8. Three shovels are singing

**Contrasting Hps and Predictions**

Dependent Variables considered for statistical analysis:
- reflecting automatic processes: response time (RTs) to give yes/no answer
- Eye-gaze and fixations: across-set comparison (all items on the screen) or within-set comparison (only members of the relevant set))
- behavioural: type of answer provided: True vs. False

$H_0$: adults rely on a strategy of <u>minimal commitment</u> (Crain and Conway, 1994), thus preferring a within-set comparison (only looks at the critical set, i.e. the square that is big among squares, not among the whole set of geometric figures)

$H_1$: adults DO NOT always rely on a strategy of <u>minimal commitment</u> (Crain and Conway, 1994); they can be induced to consider only a comparison within-set (i.e. the square that is big among squares, not among the whole set of geometric figures) by the presence of a <u>scalar adjective</u> such as "big" – in case of a <u>non-scalar adjective</u> (e.g. "red") they might consider the whole set of objects (the square that is red among the whole set of geometric figures), thus relying on a strategy of <u>maximal commitment</u> (contra Crain and Conway, 1994)

**corollary to $H_1$**: the position of the adjective with respect to the noun might prompt subjects that speak different languages to (initially) rely on different strategies: when the adjective precedes the noun (as in English), we expect subjects to be more easily induced towards a strategy of maximal commitment (resulting in a comparison across sets: the square that is big/red among the whole set of figures); conversely, when the adjective follows the noun (as in Italian), we expect subjects to be more easily induced towards a strategy of minimal commitments (resulting in a comparison within the relevant set: the square that is big/red among the squares).

Factor: Language

$H_0$:
– no effect of the position of the adjective wrt noun is in principle expected
  → no difference in the 3 languages is predicted with respect to the strategy adopted by participants

$H_1$:
– due to the position of the adjective wrt noun in the 3 languages: we expect a difference between English/Dutch and Italian, and no difference between English and Dutch.
– we expect Italian subjects to rely (significantly) more on the strategy of Minimal Commitment, only considering the members of the relevant set (the name of the set comes first); conversely,

we expect more Dutch/English subject to explore the whole scene, relying more on the strategy of Maximal Commitment (the adjective – that may apply to members of different sets - comes first and the noun specifying the relevant set comes second)

→ overall, less looks to other sets in case of Italian vs. English/Dutch in the critical condition C

→ overall, more TRUE answers in case of Italian vs. English/Dutch in the cirtical condition C


Factor: Adjective:
Scalar vs. Non-Scalar

$H_0$:
– no effect of type of adjective is in principle expected

→ no difference across Scalar-Non Scalar condition is expected in any condition (C, T, F)


$H_1$:
– we expect more subjects to rely on the strategy of Minimal Commitment (only within-set comparison) in case of Scalar than Non-Scalar Adj.

→ overall, less looks to other sets in case of Scalar vs. Non-Scalar (and for this reason, faster RTs in case of Scalar vs. Non-Scalar)

→ more TRUE answers in the critical condition C in case of Scalar vs. Non-Scalar