# The puzzle of donkey anaphora resolution*

Francesca Foppolo

University of Milano-Bicocca

## 1. Introduction: the puzzle of donkey anaphora

In this paper I will present an experimental study conducted with Italian adults concerning the interpretation of pronouns in "donkey sentences". Consider the standard example in (1):

(1) "Every" farmer who owns *a donkey* beats *it*

As is well known from the literature, the pronoun *it* in (1) admits of two interpretations, the *universal* (∀) one and the *existential* (∃) one, whose truth conditional import is reported in (2) and (3) respectively:

(2) $\forall x[[farmer(x) \wedge \exists y\ donkey(y) \wedge has(x,y)] \rightarrow \forall z[donkey(z) \wedge has(x,z) \rightarrow beats\ (x,z)]]$ [∀-reading]
= *Every farmer who owns a donkey beats* <u>*all the donkeys he owns*</u>

(3) $\forall x[[farmer(x) \wedge \exists y\ donkey(y) \wedge has(x,y)] \rightarrow \exists z[donkey(z) \wedge has(x,z) \wedge beats(x,z)]]$ [∃-reading]
= *Every farmer who owns a donkey beats* <u>*one of the donkeys he owns*</u>

There are many proposals as to how these readings come about. However, my concern here is not so much to choose among such proposals (though eventually, I believe that these results will be relevant to such an issue). My immediate concerns here are rather to experimentally test an interesting generalization regarding the distribution of ∀- and ∃-interpretations, put forth in Kanazawa (1994). According to Kanazawa, to whose work we must refer for details, the preferred interpretation of donkey pronouns is the one that preserves the monotonicity properties of the determiner. Consider the sample set given in (4): his point is that the interpretations in the last column are the only ones that preserve

---

*I would like to acknowledge Gennaro Chierchia, Maria Teresa Guasti and Maria Nella Carminati for helpful discussion and encouragement in the preparation of this work. A special thank goes to Fabio Del Prete, for his punctual comments on a previous version of this work.

(in a donkey anaphora context) the monotonicity properties of each lexical determiner, spelled out in the second column and for this reasons these are the interpretations that are selected by default:

(4)

| Det | monotonicity | | preferred interpretation |
|---|---|---|---|
| *Every* | ↓ | ↑ | ∀ |
| *No* | ↓ | ↓ | ∃ |
| *Some* | ↑ | ↑ | ∃ |

While there has been some experimental work on how donkey pronouns are interpreted (a.o. Yoon, 1996 and Geurts, 2002), no work has tried to experimentally probe Kanazawa's claim. Yet, if empirically supported, such a claim would be important, as it would show that the semantic processor must have access to an abstract formal property of an unprecedented kind (namely, monotonicity preservation in non C-command anaphora).

## 2. The experimental study

In the aim of investigating Kanazawa's generalization experimentally, three questions constituted my guiding lines: (i) is there a default preference in donkey anaphora resolution?; (ii) does this preference conform to Kanazawa's prediction, i.e. depends on the monotonicity properties of the head determiner?; (iii) can extra-linguistic context affect this (default) interpretation? These questions were specifically addressed in the two experiments I am going to present in the next sections.

### 2.1 Experiment 1: between *universal* and *existential* interpretation of the anaphora: is there a *default*?

#### 2.1.1 Participants

Thirty subjects participated in this experiment. They were mainly students at the Psychological Faculty of the University of Milano-Bicocca, and received credits for their participation.

#### 2.1.2 Material and procedure[1]

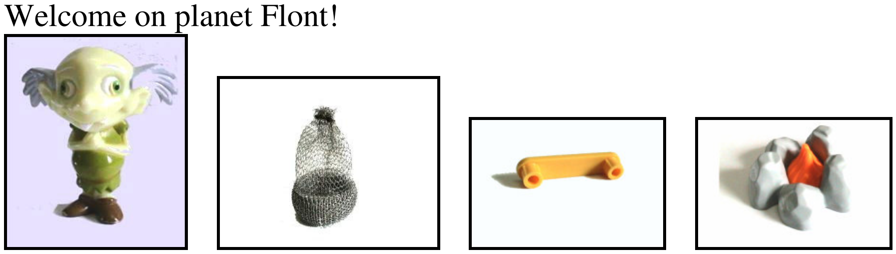
Participants were tested individually in a quiet room using a laptop. Their task was to evaluate sentences in certain situations, judging them "true" or "false" with respect to a scenario consisting of a block of four pictures to be considered as a whole. They were also told to be "charitable": whenever they encountered a sentence that could bear more than one interpretation, they should choose the one that rendered the sentence true, even if that interpretation was not their favoured one. To familiarize them with the procedure,

[1]The procedure used in this experiment is analogous to the one reported in Foppolo, 2008 to test Scalar Implicatures.

they were shown a training session in which they were assisted by the experimenter. During this training, they encountered sentences that were clearly true in the scenario, some that were clearly false and some that were somehow ambiguous. For example, they were presented with the sentence “Two girls are sitting on a chair” in a scenario depicting four different girls, each of them sitting on a chair. Typically, many subjects would tend to judge the sentence false in such a scenario, interpreting “two” as “exactly two” instead of being “charitable” and accessing the logical “at least two” interpretation. Whenever this happened, the experimenter prompted the participant to be charitable and ask her to revise her interpretation of the sentence accordingly as to make it true, if she found a way to do it. To keep track of this operation, participants were also asked the following question whenever they answered “true”: “How much do you think the sentence is a good description of the situation represented in the pictures?” They were given a scale of response varying from 1 (bad) to 5 (good). In cases like the example above, we expected subjects that were prompted to be charitable to select a low score on the scale despite the fact that they accepted the sentence in the end.

The experiment proceeded as follows: each single sentence appeared in white at the top of a black screen. By pressing the space bar, a scenario consisting of four pictures appeared below the sentence. It’s important to remember that the four pictures appeared altogether on the screen and were to be considered as a whole, unique scenario, representing the whole world to be taken into account in order to judge the sentence. Participants had to evaluate the sentence in such a scenario, pressing a green key if they judged it “true” and a red key if they judged it “false”. Time taken to make a decision was recorded, starting from the moment they pressed the key to make the pictures appearing on the screen, till they pressed the answer key. Each subject was shown the complete battery of the material but saw only one occurrence per each critical item type, for a total of 17 test items, 9 of which were critical test sentences containing a “donkey” pronoun, and the others were controls and fillers. To avoid interferences from extra-linguistic factors, we only used fantasy names in the sentences during the experimental session. After the training session, subjects were told that they would explore different situations in planets different from Earth, meeting alien characters that used objects that are unfamiliar to inhabitants of Earth. They were also reassured that they were not required to memorize the names of these characters and objects, given that they would be provided with a description of each unfamiliar object immediately before each trial. Below, I provide an example of an introductory screen used before one test trial:

Welcome on planet Flont!



This is a Flont — This is a vilp — This is a murl — This is a tarp

**Fig. 1**. Introductory screen: an example

The main purpose of having unfamiliar objects was that of ruling out world knowledge as much as possible: as it is well known from the literature, expectations about how things go in usual circumstances may affect the interpretation of “donkey sentences”[2]. However, given that our world knowledge do not extend to planet Flont, we should bear no expectations whatsoever about situations that involve vilps, murls or other alien objects.

The experiment presented a 3×3 critical condition within subject design. First of all, three different types of sentences containing a “donkey” pronoun were presented, differing in the type of quantifier (and, crucially, in its monotonicity) that introduced them, as exemplified by the following examples (remember that only fantasy names were used in the experimental session):

(5) Condition I: quantifiers (left monotonicity)
Q1: <u>Every</u> Flont that has a vilp keeps it in a bin ↓
Q2: <u>No</u> Flont that has a murl preserves it in a jar ↓
Q3: <u>Some</u> Flont that has a tarp protects it with a leaf[3] ↓

Each sentence was presented in three different types of situation:

(6) Condition II: situations
S1: a situation which is compatible with both candidate interpretations of the anaphora, thus rendering the sentence true on both readings
(I will refer to this as the “<u>non-differentiating true</u>” situation = NDT)
S2: a situation which is incompatible with both candidate interpretations of the anaphora, thus rendering the sentence false on both readings
(I will refer to this as the “<u>non-differentiating false</u>” situation = NDF)
S3: a situation which is compatible only with one of the two possible resolutions of the anaphora, thus rendering the sentence true on one interpretation but crucially false on the other
(I will refer to this as the “<u>differentiating-critical</u>” situation = DC)

To well understand the experimental design, it’s important to keep in mind that the two alternative interpretations of the “donkey sentences” are not logically *in*dependent of one another, given that one always entails the other and that the direction of entailment crucially depends on the monotonicity properties of the quantifier that introduces them: if

---

[2] Think, for instance, at classical examples like (i) and (ii):
(i) *No* man who has an umbrella leaves it at home in a rainy day [∀-reading, not ∃]
(ii) *Every* person who had a credit card paid his bill with it [∃-reading, not ∀]
The interpretation we normally assign to these examples seem to depend on our knowledge on how things go in usual circumstances, independently of the initial determiner: if it rains we don’t leave <u>all</u> our umbrella at home; to pay our bills we use <u>only one</u> of our credit cards.

[3] In order to remain as neutral as possible with regards of the *total/partial* or *stative/eventive* distinctions (cf. Yoon, 1996), we used only verbs of the same form (*conservare in un barattolo* (lit. to preserve in a jar), *tenere dentro un secchio* (lit. to keep in a bin), *proteggere con una foglia* (lit. to protect with a leaf)), all conjugated in the Simple Present tense and displaying the same number of characters. The choice of using “real” verbs instead of fantasy predicates was made in the end because it seemed that the use of imaginary names for both actions and objects could be too confounding in understanding the pictures.

the donkey sentence is introduced by the universal quantifier "Every" (Q1), then the *universal* reading is the strongest and entails the *existential*; if the donkey sentence is introduced by the negative quantifier "No" (Q2), then the *existential* reading is the strongest and entails the *universal*; if the donkey sentence is introduced by the existential quantifier "Some" (Q3), then the *universal* reading is the strongest and entails the *existential* (crucially, this is the only case in which the predicted preferred reading in the weakest one and for this reason the material used for "Some" is different from the other quantifiers, as shown below). Considering these entailment patterns, the NDT (S1) and the DC situations (S3) for each quantifier were, for instance, the ones reported below (presented to different subjects):

(7) *Every Flont that has a vilp keeps it in a bin* (Q1)

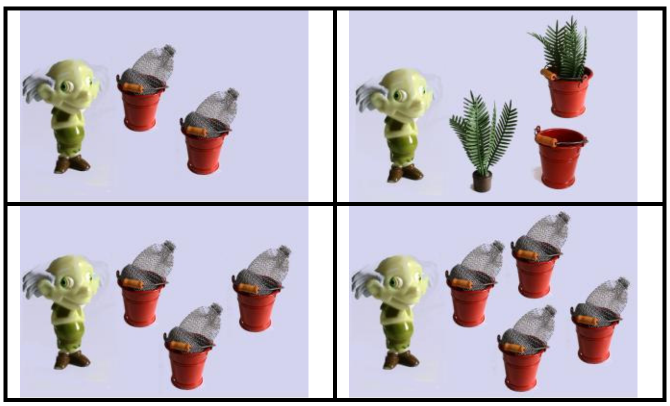


**Fig. 2**: S1 (NDT) for Q1: compatible with both readings (∀-reading entails ∃)

**Fig. 3**: S3 (DC) for Q1: only compatible with ∃-reading (∀-reading predicted)

(8) *No Flont that has a murl preserves it in a jar* (Q2)

**Fig. 4**: S1 (NDT) for Q2: compatible with both readings (∃-reading entails ∀)

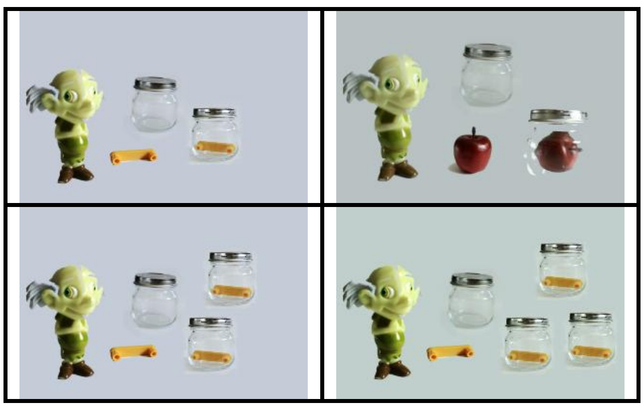


**Fig. 5**: S3 (DC) for Q2: only compatible with ∀-reading (∃-reading predicted)

(9) *Some Flont that has a tarp protects it with a leaf* (Q3)

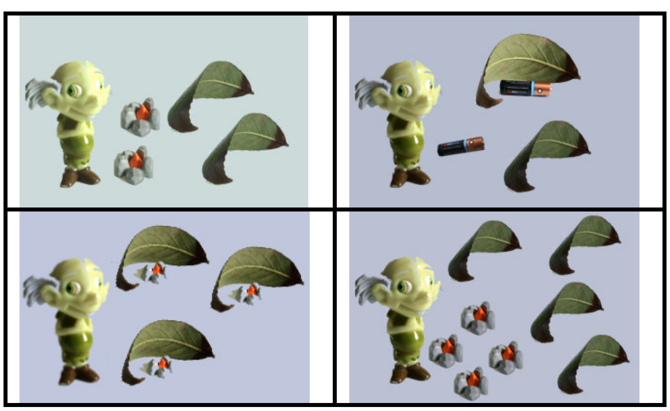


**Fig. 6**: S1 (NDT) for Q3: compatible with both readings (∀-reading entails ∃)

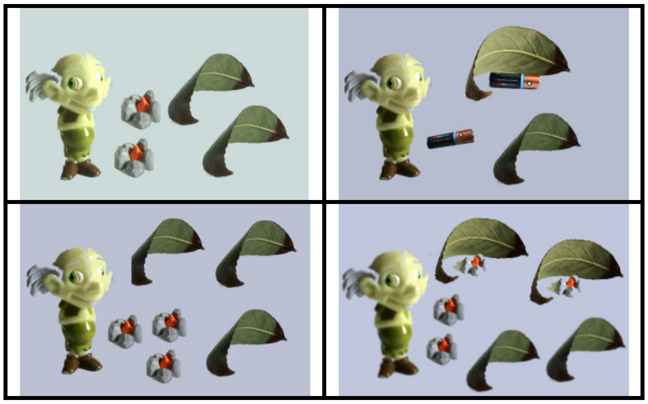


**Fig. 7**: S3 (DC) for Q3: only compatible with ∃-reading (<u>∃-reading predicted!</u>)

An additional control condition was added, that made the sentence false on any interpretation of the donkey pronoun (S2=NDF).

As we have discussed above, each sentence in Condition I is associated with two possible interpretations depending on which interpretation is assigned to the donkey anaphora they contain. One first question addressed in this experiment is whether one of the candidate interpretations constitutes the default one. To answer this question, data from the DC condition (S3) for each sentence type will prove crucial. With respect to this condition, two hypotheses can be made, each one deriving contrasting predictions that can be tested experimentally:

- $H_0$: there is no default, thus both interpretations are generated and are equally accessible
  → prediction: subjects should split
- $H_1$: there is a default interpretation, which can be compatible or not with the scenario presented in the DC condition
  → prediction: subjects should go for it (no split)

As for $H_1$, if Kanazawa's predictions are confirmed, then we should expect the following distribution:

- $H_K$: the default interpretation depends on the left monotonicity properties of the initial quantifier: *universal* reading is predicted as preferred for "donkey sentences" introduced by "Every" (Q1) and *existential* reading for sentences introduced by "No" (Q2) and "Some" Q3 (cf. (4)
  → prediction: in the corresponding DC condition (S3): subjects should reject Q1 but accept Q3 (only compatible with ∃-reading) and reject Q2 (only compatible with ∀-reading)

No difference instead is expected between $H_{1/K}$ and $H_0$ with respect to the other conditions (NDT and NDF) for which, respectively, true and false answers at ceiling are predicted.

### 2.1.3 Results and Discussion

Results are summarized in Table 1 below: column 3 reports the percentage of "true" answers; column 4 the rate assigned on the scale; columns 5-7 report respectively: the response times (RTs, in ms.) to answer "true" and "false" and the mean total time per condition (in parentheses the number of cases included in the analysis is shown).[4]

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| **Quant.** | **Sit.** | **True** | **Scale*** | **RTs for T** | **RTs for F** | **Mean RTs** |
| **Q1 Every** | S1 (NDT) | 100% | 4.57 | 7486 (30) | - | 7486 (30) |
| | S2 (NDF) | 93% | - | - | 8605 (28) | 8697 (29) |
| | S3 (DC) | 57% | 3.41 | 12323 (16) | 10615 (12) | 11591 (28) |
| **Q2 No** | S1 (NDT) | 100% | 4.83 | 7238 (29) | - | 7238 (29) |
| | S2 (NDF) | 97% | - | 7569 (1) | 9554 (26) | 9835 (27) |
| | S3 (DC) | 93% | - | 6323 (2) | 10903 (26) | 10576 (28) |
| **Q3 Some** | S1 (NDT) | 80% | 3.38 | 8761 (24) | 12421 (4) | 9284 (28) |
| | S2 (NDF) | 97% | - | 13455 (1) | 8516 (26) | 8699 (27) |
| | S3 (DC) | 87% | 4.38 | 7942 (26) | 17467 (2) | 8622 (28) |

**Table 1** (*only in case of acceptance)

These data were submitted to statistical analysis using the program for data analysis STATISTICA 6. We first analysed the acceptance rate of the critical sentences in the different conditions, starting from control conditions (S1 and S2). Data obtained for the control-false condition (Condition S2-NDF) showed that the overall percentage of correct responses was 96%, indicating that subjects understood the task and performed correctly. A one way analysis of variance ANOVA showed no significant difference among quantifiers in this condition, as expected ($F(2)=.244$, $p=.78$, n.s.). A similar result was obtained for the control-true condition (Condition S1-NDT), where the percentage of correct responses overall was also attested around 96%. In this case, however, the ANOVA revealed a significant difference among quantifiers ($F(2)=7.25$, $p<.01$). Post-hoc analysis revealed that this effect was due to a lower rate of acceptance of sentences introduced by the existential quantifier "Some" (Q3) with respect to sentences introduced by the other quantifiers (80% for Q3 vs. 100% for Q1 and Q2, $t(29)=2.69$, $p<.01$). This difference, however, can be explained by the fact that, in case of "Some", the interpretation represented in the NDT scenario (cf. Fig. 6) is the *universal* one, which is not only the strongest but it is also crucially the one predicted as the disfavoured one in Kanazawa's generalization. Despite the fact that the sentence was "true" under the *existential* interpretation (because this was actually entailed by the *universal* interpretation represented in the scenario), it's conceivable that some of the subjects noted the "lack" of the preferred interpretation, thus judging the sentence "false". This explanation is also supported by the observation that the mean score on the scale

[4]Anomalous effects on RTs were curtailed in two steps: first, we excluded RTs exceeding 2.5 times the mean item time; then, values above individual cut-off (mean + 2 SD) were smoothed (over the total, 3,33% items were excluded and 3,75% smoothed).

attributed to the subjects who accepted the sentence as "true" (i.e. 3.38) is significantly lower if compared to the mean score given by subjects in the same NDT condition to sentences introduced by Q1 and Q2, which was 4.57 and 4.83 respectively ($t(52)=4.02$ and $t(52)=5.34$, $p<.0001$).

I will now turn to the analysis of the results obtained in the critical DC (S3) for each quantifier. In case of "donkey sentences" introduced by the negative and the existential quantifier (Q2 and Q3 in situation S3), subjects showed a preference for the *existential* interpretation: 93% of the subjects rejected the critical sentence introduced by "No" in the scenario representing the *universal* interpretation of the anaphora (Fig. 5), while 87% of subjects accepted the critical sentence introduced by the existential quantifier "Some" in the scenario compatible with the *existential* interpretation of the anaphora (Fig. 7). Subjects' distributions were compared with chance performance and in both cases the proportions revealed significantly different from chance ($p<.0001$). These results conform to Kanazawa's generalization, contra $H_0$. On the contrary, in case of sentences introduced by the universal quantifier "Every", subjects split: 43% of them actually rejected the critical statement in the scenario only compatible with the *existential* interpretation of the anaphora (Fig. 3), while 57% accepted it, and this distribution equals chance ($p<.05$). This last result, as it is, seems incompatible with Kanazawa's generalization. However, I believe that it deserves a closer look.

First of all, if we consider the scores assigned on the scale in case of affirmative answers, an interesting comparison emerges between S1 and S3 for sentences introduced by "Every": subjects that accepted the critical sentences introduced by "Every" in the NDT condition (S1, representing the $\forall$-reading) gave a high score on the scale when judging sentence's adequacy to the scenario (namely, 4.57). Conversely, when judging the adequacy of the same type of sentence with respect to the scenario in S3 (DC condition, representing the $\exists$-reading of the anaphora), those subjects that accepted the sentence (roughly half of the sample) judged it significantly less adequate (4.57 vs. 3.41, $t(45)=4.28$, $p<.001$). This result seems to indicate that, despite the fact that half of the subjects accepted the donkey sentence introduced by "Every" in a scenario only compatible with the *existential* reading, they didn't judge it fully appropriate, as they did when the scenario represented the *universal* reading of the anaphora instead. In the second place, to better understand the behaviour of subjects in evaluating the sentences introduced by the universal quantifier (Q1) in the critical situation (S3), it's interesting to turn our consideration to RT measures. In particular, these are, in my opinion, the relevant comparisons among RTs to be made that would be compatible with Kananzawa's generalization: in the first place, I would expect significantly higher RTs for subjects that accept the sentences introduced by "Every" (Q1) than for those that accepted the sentences introduced by "Some" (Q3) in condition S3 that, being only compatible with *existential* interpretation, should be the favoured one for "Some"-sentences but not for "Every"-sentences; at the same time, I won't expect a difference in an analogous comparison in condition S1 (NDT); secondly, I would expect significantly lower RTs for subjects that accepted the sentences introduced by "Every" (Q1) in S1 (NDT, representing the *universal*-favoured interpretation) than those that accepted the same sentences in S3 (only compatible with *existential*-disfavoured interpretation), while I won't expect a difference in an analogous comparison between RTs of subjects that rejected the sentences introduced by "Every" (Q1) in S2 (NDF) and those that rejected the same sentences in S3.[5] Statistical analysis confirmed these predictions (for convenience, statistically significant comparisons are marked with an *): Q1.S3(True) vs. Q3.S3(True): $t(40)=2.37$, $p<.05$* & Q1.S1(True) vs. Q3.S1(True): $t(52)=1.18$, $p=.25$, n.s.; Q1.S1(True) vs. Q1.S3(True): $t(44)=-2.85$, $p<.01$* & Q1.S2(False) vs. Q1.S3(False): $t(38)=1.04$, $p=.31$, n.s.

To sum up, we can conclude that a default preference for the *existential* interpretation emerged in case of "donkey sentences" introduced by negative (Q2) and existential (Q3) quantifiers and that this preference is in fact the one predicted by Kanazawa's generalization. Statistical analysis confirmed that the preference expressed in case of "donkey sentences" introduced by these quantifiers is different from chance, against $H_0$. On the contrary, the results obtained for "donkey sentences" introduced by the universal quantifier "Every" (Q1) seem not to conform to Kanazawa's predictions. As we have seen, in the critical situation (S3) subjects split: roughly half of them rejected the "donkey sentence" in the scenario only compatible with *existential* interpretation of the pronoun, as predicted, while the other half accepted it. As such, this finding seems to argue against a default interpretation of "donkey sentences" when they are introduced by the universal quantifier, thus supporting $H_0$ and rejecting $H_{1/K}$ in this case. However, I think that this conclusion, though plausible, needs to be handled with care. As we have seen, judgements on the scale and RT measures show that the decision to accept "donkey sentences" introduced by "Every" in a scenario only compatible with *existential* interpretation of the pronoun is made at a "cost": the grade of acceptability of these sentences is low and the processing time to accept them is high compared to other relevant conditions. These results may indicate that these subjects are adopting the strategy suggested by the experimental instructions and are trying to be charitable whenever they find an interpretation that renders the sentence true (in this case, they turn to the ∃-reading). To account for these facts, the following proposal can be put forth: there actually is a default/preferred interpretation even in case of "donkey sentences" introduced by "Every" (as stated by $H_1$ and contra $H_0$), this being the *universal* (as predicted by $H_k$). However, for some reasons that we won't investigate here, the alternative *existential* interpretation is more easily accessed in case of "donkey sentences" introduced by "Every". This hypothesis has been addressed in a follow-up experiment that I am going to present in the next session, in the attempt to make the results discussed here in a way clearer.

## 2.2 Experiment 2: degrees of accessibility of the alternative readings

If we want to talk about default interpretations, then we need to attest experimentally the access to alternative readings of donkey pronouns in situations that specifically induce such readings. In the literature we can find several examples that testify the effective ambiguity of "donkey sentences" (see also footnote 2). It seems that the accessibility of

[5]Given that differences in RT may depend on the response-type (true vs. false responses), I will only consider direct comparisons of RTs across subjects who gave the same (true or false) response.

these alternative interpretations depends much by the specific example, by the type of predicate (cf. Yoon, 1996) and by the quantifier that precedes the anaphora (cf. Geurts, 2002). Intuitively, it seems easier to recreate situations that render the alternative (*existential*) reading more easily accessed in case of sentences introduced by the universal quantifier. On the contrary, in case of sentences introduced by negative and existential quantifiers, it seems that very marked examples are to be construed to access the alternative (*universal*) interpretation of the anaphora.[6]

Aim of the experiment I am going to present in this section is twofold: in the first place, it aims at supporting the hypotheses that alternative readings are actually available, a necessary step to uphold the default-hypotheses advocated in Experiment 1; secondly, it aims at testing the hypotheses that the degree of accessibility of these alternative readings may vary accordingly to the type of quantifier that introduces the "donkey sentence". Please note that this experiment is to be intended as a completion to Experiment 1 described in the preceding section and is to be analyzed in comparison with the results discussed there.

### 2.2.1 Participants

A different group of thirty-six subjects participated in this experiment. They were mainly students at the Psychological Faculty of the University of Milano-Bicocca, and received credits for their participation.

### 2.2.2 Material and procedure

The material and the procedure employed here are the same used in Experiment 1, except for two manipulations. In the first place, the "donkey sentences" introduced by the three different quantifiers (Q1, Q2, Q3 as described above) had to be evaluated in 2 (instead of 3) critical situations: the NDF (S2) situation that acted as a control and the critical DC situation (S3), using the same material of Experiment 1. The second manipulation is the critical one: a biasing context was added before the presentation of each critical sentence. Crucially, aim of this context was to bias subjects towards the interpretation represented in the scenarios reported above in Fig. 3 for "Every", Fig. 5 for "No" and Fig. 7 for "Some". For example, below is the context added before the "donkey sentence" introduced by the universal quantifier "Every" (Q1) to induce subject to interpret the pronoun *existentially* and thus to accept the sentence in the critical scenario S3. Note that this context is modelled after Casalegno's suggestion reported in Chierchia, 1992:

(10) *On planet Flont there's the plague of the termites. To face it, Flonts use special traps, the vilps. Fighting termites causes a lot of stress to the Flonts. To improve the situation, they consulted a famous magician who suggests them a remedy: they have to sacrifice a vilp putting it into a bin full of water; a vilp in the water*

---

[6]Think, for instance, at examples like the following, to induce the ∀-reading of the "donkey sentence" introduced by "Some":

(i) Some boy that has an apple doesn't share it with his schoolmates ≅ he doesn't share any *doesn't function as a trap anymore, but releases an anti-stress substance. Let's see if Flonts are less stressed given that…*

As we said, context (10) preceded sentence (7), that had to be evaluated in the scenario reported in Fig. 3 (DC condition for Q1), representing the ∃-reading of the pronoun, with the aim of biasing subjects towards that interpretation. Analogous biasing contexts were created for each situation and each quantifier. Subjects were trained to take into account (whenever they could) the suggestions given in the context when evaluating the sentence in the corresponding scenario.

### 2.2.3 Results

As expected, the biasing context had no significant effect on the percentage of "false" answers provided by subjects overall in the control S2 condition (NDF), that was around 90% in both experiments (p=.29, n.s.). Interesting comparisons between the two experiments emerged from the analysis of the proportions of acceptance in the critical DC condition (S3) for the sentences introduced by the different quantifiers, as summarized in the table below:

| Quantifier | | Situation | Exp. 1 | Exp. 2 | difference |
|---|---|---|---|---|---|
| Q1 | Every | S3-true | 57% | 87% | p<.01* |
| Q2 | No | S3-true | 7% | 17% | p=.22, n.s. |
| Q3 | Some | S3-false[7] | 13% | 17% | p=.65, n.s. |

**Table 2**

Considering "donkey sentences" introduced by the universal quantifier "Every" (Q1), the acceptance rate in a scenario only compatible with *existential* interpretation of the anaphora rose from 57% in Experiment 1 to 87% in Experiment 2, and this difference is statistically significant (p<.01). Conversely, the acceptance rate of "donkey sentences" introduced by the negative quantifier "No" (Q2) in a scenario only compatible with *universal* interpretation of the anaphora didn't benefit that much by the presence of a biasing ∀-context, rising from 7% to 17%, but not reaching statistical significance (p=.22, n.s.). Analogously, the presence of a biasing context towards the *universal* (disfavoured) interpretation of the anaphora (remember that, in case of "Some", the *universal* interpretation entails the *existential*) didn't have a significant effect on the percentage of subjects that rejected the "donkey sentence" introduced by the existential quantifier "Some" (Q3) in the critical condition S3, only compatible with the *existential* reading of the anaphora (p=.65, n.s.).

[7]Remember that, in case of "Some", the DC scenario represents the ∃-reading, which is the weakest (and favoured), while context biases towards the ∀-reading.

## 3. Conclusive Remarks

As we have discussed above, the results obtained in our first experiment seem to support Kanazawa's generalization: there seems to be a default interpretation of "donkey sentences" and this seems to be the one that preserves natural inferential patterns based on the left monotonicity properties of the head determiner. At least, this conclusion was sharp in case of "donkey sentences" introduced by the negative and the existential quantifier, for which the *universal* interpretation of the pronoun was clearly rejected. Results seem instead less clear-cut for what concerns "donkey sentences" introduced by the universal quantifier "Every": at least half of the subjects accepted these sentences in a scenario representing the ∃-reading of the anaphora, contra the predictions made in Kanazawa. As we already observed, however, this choice may be the result of the instructions given to the subjects (namely, "be charitable" whenever possible) and, interestingly, subjects that made this choice did it at a "cost", as revealed by the low rate assigned on the scale and by the high RTs taken to accept these sentences in the ∃-reading condition. Considering these facts altogether, the picture that emerges from the analysis of "donkey sentences" introduced by "Every" seems not sufficient, per se, to reject Kanazawa's claim or other interesting proposals that has been attempted to account for possible counterexamples to Kanazawa's generalization (cf. Del Prete, 2004 in which he accounts for unexpected readings as the effect of a pragmatic restriction on structural domains of quantifiers) or, ultimately, to revise, from inside, the logical apparatus of "donkey sentences" as a whole. As a first step, to account for the split observed in case of "donkey sentences" introduced by "Every", we suggested that the availability of the alternative interpretation and the effect of context over it may depend on the type of quantifier that embeds the anaphora (without speculating here on which factors can be responsible for the peculiarity of the universal quantifier in this respect). By means of Experiment 2, we aimed at testing this hypotheses, adding a biasing context to induce subjects to access the alternative, otherwise dispreferred, reading of the donkey pronoun. All in all, the results obtained in this second experiment seem compatible with this suggestion, showing that subjects are much less prone to shift their interpretation in case of sentences introduced by the existential or negative quantifiers than they are in case of sentences introduced by the universal quantifier. To what extent this difference depends on intrinsic properties of the quantifiers involved or is linked to other extrinsic factors is still an open question that, not only for reasons of space, we cannot deal with here.

## References

Chierchia, G. 1992. Anaphora and Dynamic Binding. Linguistics and Philosophy. 15:111-183

Chierchia, G. 1995. Dynamics of meaning: anaphora, presupposition, and the theory of grammar. Chicago, University of Chicago Press.

Del Prete, F. 2003. On a Unified Semantic Treatment of Donkey Sentences in a Dynamic Framework. ESSLLI.

Del Prete, F. 2004. Universal and Existential Readings of Donkey-Sentences and the Role of a Structural Form of Domain Restriction in the Explanation of Some Distributional Anomalies. Proceedings of the 14th Amsterdam Colloquium.

Foppolo, F. 2007. The logic of pragmatics. An experimental investigation with children and adults. Unpublished Ph.D dissertation. University of Milano-Bicocca, Milan.

Foppolo, F. 2008. Between "cost" and "default" of Scalar Implicature: the cost of embedding. In Grønn, A. (ed.). Proceedings of SuB12, Oslo

Geurts, B. 2002. Donkey Business. Linguistics and Philosophy. 25: 129-156.

Kanazawa, M. 1994. Weak vs. Strong Readings of Donkey Sentences and Monotonicity Inference in a Dynamic Setting. Linguistics and Philosophy. 17: 109-158.

Yoon, Y. 1996. Total and Partial Predicates and the Weak and Strong Interpretations. Natural Language Semantics. 4: 217-236.

Department of Psychology
University of Milano-Bicocca
Viale dell'Innovazione, 10 - ed. U9
20125 Milan, Italy

francesca.foppolo@unimib.it