Experimental evidence for the interpretation of definite plural articles as markers of genericity – How Italian can help

Michela Redolfi
University of Verona, Italy
University of Konstanz, Germany
michela.redolfi@univr.it

Sergio Miguel Pereira Soares
University of Konstanz, Germany
sergio-miguel.pereira-soares@uni-konstanz.de

Anna Czypionka
University of Konstanz, Germany
anna.czypionka@uni-konstanz.de

Tanja Kupisch
University of Konstanz, Germany
University of Tromsø, Norway
tanja.kupisch@uni-konstanz.de

Appendix Detailed statistical analysis of the follow-up experiment

1 Predictions

Acceptance rates: We are interested in the five planned comparisons listed below. The element of contrast is marked in bold.

I. *Noncanonical-**Demonstrative**-Noncanonical vs. Noncanonical-**Definite**-Noncanonical*: If participants routinely consider both possible readings of definites, then acceptance rates should be lower for definites than for demonstratives in this comparison. If participants only consider the reading of definites that makes most sense with the given visual context, i.e., the specific reading, then acceptance rates for definites and demonstratives should be similar.

II. *Noncanonical-**Demonstrative**-Canonical vs. Noncanonical-**Definite**-Canonical*: If participants routinely consider both possible readings of definites, then acceptance rates should be higher for definites than for demonstratives. If participants only consider the specific reading of definites, then acceptance rates for definites and demonstratives should be similar.

III. ***Canonical**-Definite-Canonical vs. **Noncanonical**-Definite-Canonical*: If participants always consider the specific reading of definites, then the acceptance rates for the former condition should be higher than those for the latter condition. If they always consider the generic reading, then the acceptance rates for the former conditions should be lower than those for the second condition. If, on the other hand, they consider both possible readings, then no difference between the two conditions is expected.

IV. *Noncanonical-Definite-**Noncanonical*** vs. *Noncanonical-Definite-**Canonical**:* If participants always consider the specific reading of the definite article, then the acceptance rates for the first condition should be higher than those for the second condition. If participants always consider the generic reading of the definite article, then the acceptance rates for the first condition should be lower than for the second. If, on the other hand, they always consider both possible readings, then acceptance rates should be similar for the two conditions.

V. ***Canonical**-Definite-Canonical vs. **Noncanonical**-Definite-Noncanonical*: If participants always consider the specific reading or if they consider both possible readings, then the acceptance rates for the two conditions should be similar. If they always interpret the article as generic, acceptance rates for the first condition should be higher compared to the second one.

Reaction times: Participants can judge the felicity of the sentences either against the visual context or against their world knowledge. We expect participants to do the former with unambiguously specific conditions, and the latter with unambiguously generic conditions. In ambiguous conditions, participants need to choose between these two options. We therefore expect longer reaction times in ambiguous than unambiguous conditions, reflecting the additional workload associated with the former.

2 Results

Reaction times and acceptance rates from all participants were analyzed for the conditions relevant for the experiment (see Table 2 of the main publication). An overview of the results is given in Table 1 for both acceptance rates and reaction times.

Data were prepared for statistical analysis in R (R Development Core Team, 2019), using core functions and the packages *reshape* (Wickham 2007), *plyr* (Wickham 2011), and *car* (Fox & Weissberg 2011). Data were analyzed using the packages *lme4* (Bates et al. 2014, *glmer* function for acceptance rates and *lmer* function for reaction times) and *LMERConvenienceFunctions* (Tremblay & Ransijn 2015, *summary* function).

Table 1: Mean acceptance rates, reaction times and standard deviation per condition over participant for the follow-up experiment.

Cond.	COLOR-VISUAL	DETERMINATION	COLOR-AUDITORY	% acceptance	Mean RTs/ms (SD in parentheses)
1	*Canonical*	*Demonstrative*	*Canonical*	96.16	542 (339)
2		*Demonstrative*	*Noncanonical*	0.00	583 (352)
3		*Definite*	*Canonical*	98.96	548 (324)
4		*Definite*	*Noncanonical*	0.00	611 (393)
5	*Noncanonical*	*Demonstrative*	*Canonical*	5.32	641 (351)
6		*Demonstrative*	*Noncanonical*	93.55	533 (266)
7		***Definite***	***Canonical***	**42.20**	**715 (425)**

8		***Definite***	***Noncanonical***	**66.67**	**732 (418)**

2.1 Acceptance rates

In Table 1 we provided an overview of the mean acceptance rates per condition.

In general, when the visual stimulus presented a canonical property, the participants accepted sentences with the canonical color/pattern independently of the type of determiner. In the two *mixed* conditions that combined both types of visual stimuli (canonical and noncanonical), the participants accepted the sentences with the *partitive* independently of the property described in the auditory stimulus. However, only the *noncanonical* visual stimuli provide us with information on the interpretation of the determiners. More specifically:

- *Noncanonical-Demonstrative* conditions: Participants accepted the *noncanonical* auditory stimuli (93.55% acceptance) and rejected the *canonical* auditory stimuli (5.32% acceptance). In other words, when presented with yellow frogs, they accepted *These frogs are yellow* and rejected *These frogs are green*. The acceptance rates are compatible with a specific interpretation of the *demonstrative* article, as we expected.
- *Noncanonical-Definite* conditions: The acceptance rates in the two relevant conditions clearly indicate that the two conditions are ambiguous, although there is a slight preference for interpreting the *definite* article as specific. For example, when presented with a picture of yellow frogs, participants were more inclined to accept sentences like (*The) frogs are yellow* (66.67% acceptance), than sentences like (*The) frogs are green* (42.20% acceptance). After a closer look at individual differences in both *Noncanonical-Definite* conditions (i.e. when they heard *(The) frogs are green/yellow* while looking at yellow frogs), we observed that revealed that 37% of the participants were consistent in their responses. Of these participants, 11% consistently gave generic responses and 89% consistently gave specific responses.

2.2 Statistical analysis of acceptance rates

In general, the results for acceptance rates are very clear already descriptively. We therefore pursued only the five planned comparisons outlined above. Differences were analyzed with binomial generalized mixed models. For comparisons (I) and (II), DETERMINATION was specified as fixed effect, participants and items as random intercepts, and DETERMINATION as random slope for participants and items. For comparisons (III) and (IV), we specified COLOR-VISUAL as fixed effect, participants and items as random intercepts, and COLOR-VISUAL as random slope for participants and items. For comparison (V), we specified COLOR-AUDITORY as fixed effect and with participants and items as random intercepts, and COLOR-AUDITORY as random slope for participants and items. Only statistically significant results will be reported in detail below (alpha = .05), unless specifically stated otherwise.

I. *Noncanonical-**Demonstrative**-Noncanonical vs. Noncanonical-**Definite**-Noncanonical* (93.55% acceptance vs. 66.67% acceptance): There is no statistically significant difference.

II. *Noncanonical-**Demonstrative**-Canonical vs. Noncanonical-**Definite**-Canonical* (5.32% acceptance vs. 42.20% acceptance): There is no statistically significant difference.

III. ***Canonical**-Definite-Canonical vs. **Noncanonical**-Definite-Canonical* (98.96% acceptance vs. 42.20% acceptance): There is a statistically significant difference (z = 4.44, p < .001).

IV. *Noncanonical-Definite-**Canonical** vs. Noncanonical-Definite-**Noncanonical*** (42.2% acceptance vs. 66.67% acceptance): There is a marginally significant difference (z = 1.81, p = .07).

V. ***Canonical**-Definite-**Canonical** vs. **Noncanonical**-Definite-**Noncanonical*** (98.96% acceptance vs. 66.67% acceptance): There is a statistically significant difference (z = 2.68, p < .01).

To summarize, for the conditions with *noncanonical* visual stimuli (e.g. yellow frogs):

Demonstrative: participants accepted the sentences (93.55% acceptance) with the noncanonical property (i.e. *These frogs are yellow*) and rejected the sentences (5.32% acceptance) with the canonical property (e.g. *These frogs are green*). This is in line with a specific interpretation of demonstratives, as expected.

Definite: acceptance rates were close to 50%, both for sentences with the canonical and the noncanonical property. Participants accepted 42.20% of the sentences with the *canonical* auditory stimulus (*The frogs are green*) and 66.67% of the sentences with the *noncanonical* auditory stimulus (*The frogs are yellow*). This is in line with an ambiguous interpretation of definites.

2.3 Reaction times

Before data analysis, we removed reaction times shorter than 200 ms and longer than 6000 ms, leading to the removal of 11.64% of the data. Reaction times per condition were analyzed for the responses matching the expected interpretation (for unambiguous demonstratives). This led to the removal of 3.68% of the data. Moreover, we calculated two standard deviations from the mean as a cut-off, leading to discard another 3.10% of the data.

2.4 Statistical analysis of reaction times

Log-transformed RTs were analyzed using linear mixed effects models. We specified DETERMINATION and COLOR-AUDITORY as fixed effects with full interactions and participant and item as random effects. In addition, DETERMINATION and COLOR-AUDITORY were defined as random slopes. We found a significant main effect of the COLOR-AUDITORY (t = -4.86, p < .001) and an interaction of DETERMINATION and COLOR-AUDITORY (t = 2.417, p = 0.02). The main effect of

DETERMINATION was statistically significant for COLOR-AUDITORY *noncanonical* (t = -3.017, p < .01) (*Noncanonical-**Demonstrative**-Noncanonical* – 533ms – was shorter than *Noncanonical-**Definite**-Noncanonical* – 732ms –); but not for COLOR-AUDITORY *canonical* (p > .2) (*Noncanonical-**Demonstrative**-Canonical* – 641ms – was not significantly shorter than *Noncanonical-**Definite**-Canonical* – 715ms –).

To summarize, the statistical analysis revealed that, when looking at noncanonical pictures, there is an effect of DETERMINATION on conditions with *noncanonical* sentences: Reaction times to unambiguously specific sentences are shorter than those to sentences that are ambiguous between specific and generic readings. On conditions with *canonical* sentences, there is no such effect, possibly caused by the fact that these sentences are usually rejected in the *demonstrative* conditions, but are sometimes accepted and sometimes rejected in the *definite* conditions. Taken together, these findings are in line with the results of the acceptability ratings, which suggest that definites are ambiguous.

References

Bates Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI: 10.18637/jss.v067.i01.

Fox, John & Sanford Weisberg. 2011. An R companion to applied regression. Thousand Oaks, California. *Online: http://socserv. socsci. mcmaster. ca/jfox/Books/Companion*.

R Development Core Team. 2005. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Tremblay, Antoine & Johannes Ransijn. 2015. *LmerConvenienceFunctions: Model selection and post-hoc analysis for (g)lmer models*.

Wickham, Hadley. 2007. Reshaping data with the reshape package. *Journal of Statistical Software* 21(12). 1–20. http://www.jstatsoft.org/v21/i12/paper.

Wickham, Hadley. 2011. The split-apply-combine strategy for data analysis. *Journal of Statistical Software* 40(1), 1–29. http://www.jstatsoft.org/v40/i01/.