Gender agreement and multiple referents

Chiara Finocchiaro, Bradford Z. Mahon & Alfonso Caramazza

We report a new pattern of usage in current, spoken Italian that has implications for both psycholinguistic models of language production and linguistic theories of language change. In Italian, gender agreement is mandatory for both singular and plural nouns. However, when two or more nouns of different grammatical gender appear in a conjoined noun phrase (NP), masculine plural agreement is required. In this study, we combined on-line and off-line methodologies in order to assess the mechanisms involved in gender marking in the context of multiple referents. The results of two pronoun production tasks showed that plural feminine agreement was significantly more difficult than plural masculine agreement. In a separate study using offline judgements of acceptability, we found that agreement violations in Italian are tolerated more readily in the case of feminine conjoined noun phrases (e.g., *la mela e la banana* 'the:FEM apple:FEM and the: FEM banana: FEM') than masculine conjoined noun phrases (e.g., *il fiore e il libro* 'the:MAS flower: MAS and the: MAS book:MAS'). Implications of these results are discussed both at the level of functional architecture within the language production system and at the level of changes in language use.*

1. Introduction

In gender-marked languages, grammatical gender is an intrinsic property of lexical items. When a given lexical item appears in context, its gender feature is inherited by all the elements with which it is in an agreement relationship and that require the gender feature. For instance, in the Italian sentence *Il cappello è rosso* 'The:MAS.SG hat:MAS.SG is red:MAS.SG', both the determiner and the adjective are marked with the masculine ending, as inherited from the word *cappello*, which is masculine. If the noun *cappello* has already been introduced in a common context, the speaker may use a pronoun, as in *lo prendo* 'I take it:MAS.SG', where, again, the pronoun *lo* is marked for masculine gender. The rules for gender-marking hold for plurals as well. The plural equivalent for both sentences would be *I cappelli sono rossi* 'The:MAS.PL hats:MAS.PL are red:MAS.PL'.

In the examples above, the gender feature is inherited from a lexical node by the elements in the sentence that require that feature: critically, in those examples, there is always one linguistic unit – a

Rivista di Linguistica 19.2 (2007), pp. 285-307 *(ricevuto nel marzo 2008)*

singular or a plural noun – triggering agreement. In this study we explored the situation in which agreement is triggered by the combination of more than one linguistic unit. This is the case for conjoined NPs (e.g., "the hat and the tie"). In Italian, the agreement resolution rule for a conjunction of nouns is as follows (see Dardano & Trifone 1997; Serianni & Castelvecchi 2004; Salvi & Vanelli 2004): when the nouns are all feminine (e.g., *la mela e la banana* 'the apple and the banana') they invoke feminine agreement (e.g., *sono buone* 'they are good:FEM.PL'); when the nouns are all masculine (e.g., *il fiore e il libro* 'the flower and the book'), they require masculine agreement (e.g., *sono belli* 'they are beautiful:MAS.PL"); when the nouns have different gender features (e.g., *il fiore e la pianta* 'the flower:MAS and the plant:FEM'), no matter how many feminine and masculine nouns there are, masculine plural agreement is required (e.g., *sono belli* 'they are beautiful:MAS.PL").

While the existence of the above described grammatical rule in Italian is not in dispute, informal observation indicates that native Italian speakers sometimes use masculine agreement when the rule would have required feminine agreement. Such 'errors' in the use of gender resolution rules may reveal the complex interplay of factors that are not considered in the rule as it is commonly stated (Corbett 1991). To our knowledge, there have been no systematic investigations of gender agreement for multiple referents in Italian, and specifically, of those situations that generate violations of the rule in the spontaneous speech of native Italian speakers. That is the subject of the present investigation.

One issue concerns the reasons why native Italian speakers, in some situations, 'violate' gender agreement rules in spontaneous speech. The available hypotheses can be roughly divided into claims about performance factors and claims about competence factors. For instance, one possibility is that over-generalization of the masculine form is due to contextual reasons in the concrete use of language. If one wants to refer to a collection of objects, and there are no particular stylistic demands, he or she may want to use the form that applies to the majority of cases – the masculine form – even though prescriptively incorrect. If such were the case, then observations that native Italian speakers over-generalize masculine agreement would not necessarily indicate that they represent grammatical rules that are different from those presumed to govern grammatical gender agreement in Italian. Another possibility, however, is that the masculine

generalization is indeed due to competence factors. In other words, it may be that native speakers of Italian do in fact represent grammatical rules that permit over-generalization of masculine agreement.

There is no direct measurement of speakers' competence. Traditionally, however, speakers' intuitions are assumed to reflect, more or less directly, language knowledge. The task that is typically used for such purposes requires participants to judge the grammaticality of a given sentence in a given language. The rationale is that participants rely on what they know of the language in order to accept or reject, as well formed, the sentence to which they are exposed.

In contrast, performance is usually assessed trough on-line measurements of language use – typically speed of response or error rates.

In this study, we use a combination of on-line and off-line tests. Specifically, we used both chronometric production of pronoun production as well as participants' grammaticality judgments gender agreement violations for conjoined noun phrases in Italian. In the first two experiments we used online tasks under time constraint. Participants in those experiments were asked to refer to pictures by using singular or plural pronouns. In the third experiment, we asked participants to rate on a seven-point scale the acceptability of sentences containing gender-agreement violations on verbs or pronouns for conjunctions of nouns. Nouns were conjoined trough the conjunction *e* 'and', since it is the only way to formally express positive coordination in Italian. This task was performed without time constraints.

Different predictions follow from the competence and the performance hypotheses.

According to the performance hypothesis, the feminine disadvantage should emerge in the on-line tasks, but not in the off-line tasks. The competence hypothesis predicts instead a consistent pattern in both on-line and off-line tasks. This is because, according to the hypothesis, participants' performance is driven directly by knowledge of the rules of the language.

We acknowledge that introducing multiple referents deictically (Experiments 1-2) and introducing them trough a coordination of NPs (Experiment 3) may differ in other respects than the on-line/off-line contrast. However, the focus here is not on noun conjunctions *per se*, but on the production/judgement of forms that need to agree in gender with a pair of referents. To that end, we assume that the two methodologies may be equivalent.

2. Experiment 1

In this experiment we investigated the production of verb phrases [V-accusative pronoun], where the verb form was kept constant (imperative, II sing, e.g., *porta-* 'bring [you:SG]'). In contrast, the accusative form of the pronoun varied depending on the gender and the number of the pictures presented. The pronouns that were used belonged to the group of clitics. Italian clitics are phonologically unstressed pronouns that can be placed only adjacent to the verb on which they depend.[1]

On every trial, either a single or two pictures would appear. In one case, participants were required to produce the singular pronoun form (*lo* 'it:MAS' or *la* 'it:FEM' depending on the gender of the pictured noun). In the other case, they were required to produce a plural pronoun corresponding to the integrated value of the genders of the two pictured nouns. The correct plural forms were *le* 'them:FEM' for two feminine gender nouns, and *li* 'them:MAS' for two masculine gender nouns, as well as for situations in which one feminine and one masculine picture were presented.

Of critical importance for addressing the causes of over-generalization of masculine agreement are plural trials: are participants slower to respond when they are presented with two feminine pictures compared to the other experimental conditions (congruent-masculine and mixed-gender trials)?

2.1. Method

2.1.1. Participants. Twelve native Italian speakers, students at Pisa University, took part in the experiment for payment. They had normal or corrected-to-normal vision and reported never having had any type of language disorder.

2.1.2. Materials and design. Forty-eight pictures were selected. They represented familiar and concrete objects, half with masculine gender names and half with feminine gender names. Masculine and feminine pictures did not differ with respect to length (mean length, mas: 6.8; fem: 6.3) or frequency of use as reported in Bortolini et al. (1971. Mean freq., mas: 19.2; fem: 16.7). Pictures could appear alone (i.e., with a blank frame: e.g., *gonna* 'skirt:FEM' – [blank frame]; *guanto* 'glove:MAS' – [blank frame]) or in pairs. Picture pairs could be congruent in gender (i.e. both masculine or both feminine: e.g., *gonna* 'skirt:FEM' – *cravatta* 'tie:FEM'; *guanto* 'glove:MAS' – *cappello* 'hat:MAS')

or incongruent in gender (i.e., one masculine and one feminine: e.g. *gonna* 'skirt:FEM' – *cappello* 'hat:MAS'). The complete set of picture pairs are reported in Appendix A. Pictures were divided in three blocks. A given picture appeared once per block. There were an equal number of congruent-gender conditions, mixed-gender conditions and singular (i.e., one picture presented) conditions in each block. In the singular condition, half of the pictures appeared on the right side of the computer screen, the other half on the left. Stimuli were randomized within blocks with the following constraints: (i) there was a maximum of two consecutive trials with the same number values; (ii) the two trials at the beginning and at the end of each block were singular condition trials. Six different inter-block randomizations and two different intra-block randomizations were used. An additional set of 24 picture pairs were selected. They served as practice trials before the experiment proper.

2.1.3. Procedure. Participants were tested individually in a testing room and seated at a distance of about 60 cm. from the computer screen. The experiment started with a naming task aimed at familiarizing participants with the pictures and their names. When participants produced a name other than that expected by the experimenter, they were corrected. Such instances were very rare. Participants were then instructed to orally produce the default verb *portare* 'to bring' in the second singular, imperative form. In addition, they had to 'attach' to the verb the clitic pronoun corresponding to the picture(s) that they would see (e.g., *portalo* 'bring it:MAS'). Thus, the response set was constituted by the following combinations: *portala* ('bring it:FEM'; singular condition, feminine), *portalo* ('bring it:MAS'; singular condition, masculine), *portale* ('bring them:FEM'; plural condition, two feminine pictures), *portali* ('bring them:MAS'; plural condition, two masculine pictures or two mixed-gender pictures). Instructions emphasized response speed and accuracy. Participants then performed a practice block, after which the experiment proper began.

2.1.4. Trial Structure. At the beginning of each trial, a question mark appeared in the center of the computer screen. As soon as participants pressed the space bar the question mark disappeared and it was immediately replaced by a fixation point (plus sign) for 500 ms. A blank screen of 200 ms followed, after which the picture(s) appeared. Pictures remained on the screen until participants responded or until 2000 ms had elapsed – whichever came first. Stimulus presentation was controlled by the program Psyscope (Cohen et al., 1993). Response latencies were measured by means of a voice key, from the onset of the picture stimuli(us). The experimenter was present

throughout the testing session and recorded participants' responses manually.

2.1.5. Analysis. Verbal dysfluencies, responses differing from those intended by the experimenter, no-responses, failures to record and outliers (RTs less than 300 ms. or greater than 3 Standard Deviations (s.d.) from each participant's mean) were scored as errors. Errors were removed from the analyses of response times and were submitted to separate analyses. Separate ANOVAs were conducted for the plural conditions and for the singular condition, treating subjects as a random variable (Raaijmakers et al., 1999). This choice is also motivated by the fact that the number of items, especially in Experiment 2, is not large enough to motivate generalization over items. In the singular condition, there was one within-participant factor, TARGET GENDER, with two levels (feminine *vs.* masculine). In the plural condition, there was one within-participant factor, GENDER CONGRUENCY, with three levels (congruent-feminine *vs.* congruent-masculine *vs.* mixed-gender). Since the observations for the mixed-gender condition were twice as many as the observations for each of the other conditions, only half of the trials for that condition were included in the statistical analysis. This was done in order to keep the error term constant across conditions and to fulfil the requirements of homogeneity of variance. However, in order to not bias the results by arbitrarily excluding half of the trials for the mixed-gender condition, the analysis was carried out using a software that randomly selected half of the trials from the relevant conditions for each iteration of the analysis. In this way, we automatically generated one-thousand random samples of 50% of the trials for the mixed-gender condition and computed the statistical differences among the three experimental conditions one-thousand times, once for each random sample generated for the mixed-gender condition. We will report the range, the mean, and the standard deviation of F values, and the percentages of cases in which p values were significant ($p < .05$) across the iterations of each analysis. A contrast will be considered significant when the p value is significant on 95% of the analyses. In case of significance, two-tailed t-test comparisons were conducted. On the other hand, mean RTs and percentage of error rates, when reported, refer to the whole data set.

2.2. Results

2.2.1. Singular Condition. A total of 3.6% of the trials in the experiment across all participants were errors and were discarded.

Errors were equally distributed between feminine (3.5%) and masculine pictures (3.8%). Response latencies did not significantly differ according the gender of the picture names (fem: 604, mas: 617, $p > .8$).

2.2.2. Plural Condition. A total of 21.4% of the trials in the experiment across all participants were errors and were discarded. In the analysis of errors, there was a main effect of Gender Congruency ($F(2,22) = 14.07$ to 32.3, mean $F = 21.40$ ($\pm$ 3.2), $p < .05$ in 100% of cases). Post-hoc paired-sample t-tests revealed that participants made more errors on feminine- than on masculine-gender pairs ($t(1,11) = 6$, $p < .0001$), on congruent-feminine gender pairs than on mixed-gender pairs ($t(1,11) = 1.3$ to 5.3, mean $t = 2.9$ ($\pm$ 0.6), $p < .05$ in 88.3% of cases), and on mixed-gender pairs than on congruent-masculine gender pairs ($t(1,11) = -2.3$ to -11, mean $t = -5$ ($\pm$ 1.3), $p < .05$ in 100% of cases).

In the analysis of latencies, there was a significant main effect of Gender Congruency ($F(2,22) = 3.2$ to 6.8, mean $F = 5.1$ ($\pm$ 0.5), $p < .05$ in 99.7% of cases). Paired-sample t-tests showed that participants were slower on feminine- than on masculine-gender pairs ($t(1,11) = 2.5$, $p = .03$). On the other hand, the mixed-gender condition did not significantly differ from either of the gender congruent conditions. Indeed, p values turned out to be significant on a minority of the analyses performed according to the procedure explained above (contrast mixed-gender *vs.* congruent masculine-gender: ($t(1,11) = 0.7$ to -3.1, mean $t = -1.7$ ($\pm$0.5), $p < .05$ in 14.7% of cases; contrast mixed-gender *vs.* congruent feminine-gender: ($t(1,11) = 0.1$ to 3.2, mean $t = 2$ ($\pm$ 0.5), $p < .05$ in 40.3% of cases). Results are summarized in Table 1.

Table 1. Mean RTs (with SE in parentheses) and error percentage observed in Experiments 1-2 for different experimental condition.

	Singular conditions		Plural conditions		
	F_SING (LA)	M_SING (LO)	F-F (LE)	M-M (LI)	MIXED (LI)
EXP 1 (RTs)	604 (32.7)	617 (38.1)	846 (93)	754 (61)	788 (72)
EXP 1 (ERR)	3.5%	3.8%	38.1%	5.6%	21%
EXP 2 (RTs)	644 (18.3)	677 (24.7)	854 (41)	827 (47)	819 (42)
EXP 2 (ERR)	4.1%	4.8%	30.6%	4.2%	17.3%

2.3. Discussion

The results of Experiment 1 demonstrate that participants are slower and more prone to error when producing feminine grammatical gender, but only for multiple referents. Indeed, the congruent-feminine condition was the most 'difficult' condition for participants, and the mixed-gender condition was more difficult than the congruent-masculine condition. Importantly, this pattern was present both in the error and RT analyses. One objection that may be raised is that, in Experiment 1, participants produced masculine-marked clitics more often than feminine marked clitics (i.e., *li*-response > *le*-response). This is because there were the same number of trials per condition (congruent-gender feminine, congruent-gender masculine, mixed-gender), and so the response *li* was required on 75% of the plural-picture trials. It may be argued that this difference in the proportion of responses that were marked for masculine versus feminine is what underlies the differences observed in RTs and errors. In order to address this concern, we ran Experiment 2, in which we balanced the proportion of trials in which participants produced masculine and feminine marked clitics.

3. Experiment 2

This experiment sought to replicate the findings of Experiment 1 while maintaining the same proportion of *li* and *le* responses.

3.1. Method

3.1.1. Participants. Twelve Italian speakers, recruited from the same population as in Experiment 1, took part in the experiment.

3.1.2. Materials, Design, Procedure. Twenty-four pictures (half masculine, half feminine) were selected and divided in pairs. Some pairs were the same as in the previous experiment, while some formed new pairs (see Appendix B). The name of the pictures did not significantly differ with respect to length (mean length, m: 6.5; f: 5.8) or frequency as reported in Bortolini *et al.* (1971. Mean freq., m: 23.3; f: 20.8). Pictures appeared alone (SINGULAR CONDITION), with a same-gender pictured noun (CONGRUENT-GENDER CONDITION), or with a different-gender pictured noun (MIXED-GENDER CONDITION) in separate blocks. In addition, we included 12 filler trials, in which two feminine pictures appeared. This was done in order to yield the same proportion of *li* and *le* responses. There were thus a total of 60 trials, 36 of which elicited plural responses. *Li* responses were elicited by

the masculine-gender condition (N = 6) and by the mixed condition (N = 12). *Le* responses were elicited by the feminine-gender condition (N = 6) and by the feminine-feminine filler trials (N = 12). The remaining trials (N = 24) elicited singular responses, equally divided between masculine and feminine (i.e., *lo* and *la*). All other aspects of the experimental design, procedure, and statistical analysis were the same as in Experiment 1. Filler trials were excluded from the statistical analysis.

3.2. Results

3.2.1. Singular condition. Following the same criteria for determining errors as in Experiment 1, 4.5% of the trials were removed. Errors were equally distributed between feminine (4.1%) and masculine pictures (4.8%; $p \geq .7$).

In the analysis of response latencies there was a main effect of Picture Gender ($F(1,11) = 13$, $p < .005$), showing that feminine pictures (644 ms) were responded to faster than masculine pictures (677 ms).

3.2.2. Plural condition. Discarded data accounted for 17.4% of the data. The main effect of Gender Congruency was significant ($F(2,22)$ = between 5.6 to 15, mean $F = 9.4$ (± 1.6), $p < .05$ in 100% of cases). As revealed by post-hoc paired-sample comparisons, participants made more errors on congruent-feminine than on congruent-masculine pairs ($t(1,11) = 4.4$, $p = .001$). There also was a trend for mixed-gender pairs to elicit more errors than masculine-gender pairs $t(1,11)$ = -1 to -4, mean $t = -2.5$ (± 0.5), $p < .05$ in 69.4% of cases). The difference between mixed-gender pairs and feminine-gender pairs reached the specified alpha level on a minority of cases ($t(1,11)$ = 0.7 to 3.9, mean $t = 1.9$ (± 0.6), $p < .05$ in 27% of cases). In the analysis of naming latencies, the effect of Gender Congruency was significant only on 5.3% of the iterations ($F(2,22)$ = 0.4 to 6.2, mean $F = 1.8$ (± 0.9), $p < .05$ in 5.3% of cases). Thus, we did not perform post-hoc analyses on the contrasts between the mixed-gender condition and each congruent condition.

Since we were particularly interested in the contrast between the two congruent conditions (masculine *vs.* feminine), we conducted a planned comparison between congruent-gender masculine pairs and congruent-gender feminine pairs. The difference was marginally significant ($t(1,11) = 2$, $p = .07$). Results are summarized in Table 1.

3.3. Discussion

The results observed in Experiment 2 converge with the findings in Experiment 1. This means that the pattern initially observed in Experiment 1 is not due to the fact that participants in that experi-

ment produced masculine-marked responses more often than feminine-marked responses. In both experiments, the congruent-gender feminine and masculine conditions were the hardest and the easiest conditions respectively, with the mixed-gender condition in between the two. Moreover, the same pattern of results was observed in the analyses of errors and latencies – though marginally, for latencies, in Experiment 2.

It could be argued that the reported pattern of response time effects is due to the adoption of a response strategy on the part of subjects. Specifically, it could be that as soon as participants recognized an object with a masculine name, they responded without further processing of the second object. Thus, as soon as they perceived that two (instead of one) objects were presented, and that one of them was masculine, they interrupted further processing of the second object and responded with the masculine plural pronoun. On this account, the difference between the masculine-congruent *vs.* mixed-gender conditions could be explained by assuming that participants randomly chose which object to look at first. However, this argument has difficulty explaining the pattern of findings in the error rates. That argument would predict that as long as two feminine objects are identified, the accuracy level would not differ with respect to the other conditions. Contrary to this prediction, error rates were consistently higher in the congruent-gender feminine condition compared to the other conditions in both experiments.

In the light of these considerations, we assume that the observed pattern of results does indeed reflect how agreement is realized in natural language situations. Do these findings reveal aspects of competence or aspects of performance? Given the massive number of errors produced in the simple task used here, it would seem more reasonable to assume that they reflect competence and not merely performance factors. Indeed, the number of errors was surprisingly high in both experiments, and the error distribution closely mirrored the pattern of RTs. Thus, the feminine disadvantage most likely indicates that speakers' competence on this aspect of agreement does not exactly reflect the prescriptive rules of agreement.

Another finding that deserves attention is the result in the incongruent-gender condition, roughly intermediate between the congruent-masculine and the congruent-feminine conditions. This finding may be due to the experimental procedure: in order to produce the correct agreement form, people have to integrate the gender of two objects, one of which is masculine and one is feminine. One way

in which this integration may be thought of, is as a two-step process: people first suppress the feminine gender value in order to then correctly produce the masculine form. The suppression of the gender value would take time, thus slowing down the response and giving rise to more errors in comparison to the congruent-masculine condition.

Alternatively, it may reflect the fact that speakers' competence is such that it is relatively prone to accept the wrong agreement marking in a condition – such as the mixed-gender condition - in which the wrong form represents one of the two objects in a combination.

In order to better assess whether speakers' behaviour with congruent-feminine and mixed-gender combinations is due to performance or competence factors we carried out an off-line acceptability judgements study.

4. Experiment 3

In this experiment we set out to explore the mechanisms involved in gender-marking by asking participants to judge the grammatical acceptability of sentences violating gender agreement in different conditions. The rationale is that agreement violations should receive a higher acceptability rating the more the wrong agreement is considered acceptable in each condition. Conditions varied as a function of the gender (feminine *vs.* masculine) and number (one *vs.* two) of the referents. We also included a simple plural condition in which there was only one plural referent (feminine or masculine). This condition allows us to directly compare agreement violations with conjoined NPs (*e.g.*, *il bicchiere e il cappello* 'the glass and the hat') and agreement violations with plural nouns (*e.g.*, *i bicchieri* 'the glasses'). Both conditions require plural agreement, but only the first involves the integration of different gender features across nouns. On the other hand, agreement with a plural noun – as is the case for agreement with a singular noun – involves only one lexical node. In this way, it is possible to verify whether the single plural lexical entries are treated as the combination of two singular lexical entries, as far as the agreement process is concerned.

If the difference between congruent-masculine, congruent-feminine and mixed-gender conditions observed in the previous experiments results from the application of an experimental strategy due to time pressure, no differences among the conditions are predicted here. In other words, in the measure to which such differences are

observed, we may conclude that they index speakers' (competent) knowledge about gender agreement.

Two possibilities are considered concerning feminine agreement. According to the first one, the agreement marking of feminine plural is intrinsically more difficult compared to the plural masculine agreement marking. Therefore, agreement violations with conjunctions of feminine nouns (e.g., *la bottiglia e la cravatta* 'the bottle:FEM and the tie:FEM') and plural feminine nouns (e.g., *le bottiglie* 'the bottles') should pattern the same; namely, they should be tolerated more readily than agreement violations for the corresponding masculine conditions. According to the second possibility, the problem with the feminine gender lies in the integration process. That is, when two different feminine referents have to be integrated, the integration rule for feminine agreement is less strong than is dictated by prescriptive Italian grammar. If this were the case, agreement violations with conjunctions of feminine nouns (e.g., *la bottiglia e la cravatta* 'the bottle:FEM and the tie:FEM') should be tolerated more readily than agreement violations with conjunctions of masculine nouns (e.g., *il bicchiere e il cappello* 'the glass:MAS and the hat:MAS'). However, no difference in agreement violation ratings is predicted between plural feminine (e.g., *le bottiglie* 'the bottles') and masculine nouns (e.g., *i bicchieri* 'the glasses').

4.1. Method

4.1.1. Participants. Forty-three Italian speakers volunteered for the experiment. Thirty-eight of them were high-school students (age: 18-20) from the Pisa area.

4.1.2. Materials. Fourteen correct sentences were created. Half of them explored gender agreement on verbs, the other half explored gender agreement on clitics.[2] For four sentences there was singular agreement, while for ten sentences there was plural agreement. For each number agreement, agreement on both verbs and clitics was explored. Four different agreement conditions were used: 1. SINGULAR CONDITION (in four sentences); 2. SIMPLE PLURAL CONDITION (in four sentences); 3. COMPLEX PLURAL CONGRUENT CONDITION (in four sentences); 4. COMPLEX PLURAL MIXED CONDITION (in two sentences). The first condition explores agreement on verbs or clitics with a singular feminine (e.g., *una bottiglia* 'a bottle') or masculine noun (e.g., *un quaderno* 'a copybook'). The second condition explores agreement on verbs or clitics with a plural feminine (e.g., *due bottiglie* 'two bottles') or masculine

noun (e.g., *due quaderni* 'two copybooks'). The third condition explores agreement on verbs or clitics with a conjunction of two singular NPs of the same gender (e.g., *una bottiglia e una forchetta* 'a bottle and a fork', both feminine, or *un libro e un quaderno* 'a book and a copybook', both masculine). The fourth condition explores agreement on verbs or clitics with a conjunction of two singular NPs of different gender (e.g., *una bottiglia e un libro* 'a bottle:FEM and a book:MAS').

Using the same syntactic structures we generated fourteen 'incorrect' sentences for which there were errors in gender agreement. Number agreement was always preserved. Incorrect sentences used the same lexical items as correct sentences but those lexical items were arranged differently. This was done in order to use the same materials across correct and incorrect sentences; the other advantage of this approach is that participants cannot use strategies to judge the validity of a presented sentence based on the mere presence of a given word. For instance, the incorrect version of the sentence *Il pinguino è balzato in mare* 'The penguin:MAS jumped:MAS in the sea' was *Il pesce è balzata in mare* 'The fish:MAS jumped:FEM in the sea', in which the subject noun is masculine, but the past participle is incorrectly marked with the feminine ending. All of the incorrect sentences are reported in Appendix C. The order of the 28 sentences was randomized for each participant with the following constraints: 1. No more than three consecutive sentences corresponded to the same experimental condition; 2. There were no more than three consecutive correct or incorrect sentences.

4.1.3. Procedure. Participants were presented with the 28 sentences on a printed sheet.

They were asked to rate the grammatical acceptability of each sentence on a seven-point scale (from 1: completely unacceptable, to 7: completely acceptable). Participants were encouraged to use the entire scale. They were also explicitly instructed to focus on grammatical acceptability while neglecting any personal preference on semantic grounds.

4.1.4. Analysis. The precondition for a participant to be included in the analysis was that the mean rating for correct sentences be higher than the mean rating for incorrect sentences. All participants satisfied this criterion. Overall, the mean rating for correct and incorrect sentences was 6.1 [±1.2] and 2.2 [±1.2] respectively. Ratings for correct sentences were not considered any further. Scores for incorrect sentences were analysed by performing an ANOVA on the subjects' mean ratings for the sentences where the verb or the clitic had to agree with a singular noun or two different nouns sharing

their gender feature. Two variables were considered: CONDITION (with three levels: Singular; Simple Plural; Complex plural congruent) and NOUN(S) GENDER (with two levels: Feminine *vs.* Masculine). Two-tailed t-test comparisons were also performed between the mean ratings for the complex agreement conditions: COMPLEX PLURAL MIXED CONDITION *vs.* COMPLEX PLURAL FEMININE CONDITION *vs.* COMPLEX PLURAL MASCULINE CONDITION.

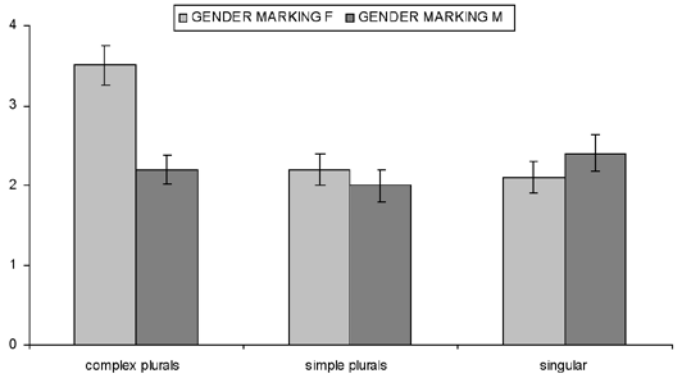
4.2. Results

A repeated-measures ANOVA showed a significant main effect of Condition ($F(2,41) = 15.8$, $p < .00001$). T-Test comparisons revealed that the Complex plural congruent condition (2.84) was rated, on average, as more acceptable than the Singular (2.23; $t(1,85) = 3$, $p < .003$) and the Simple plural condition (2.07; $t(1,85) = 5.2$, $p < .00001$). On the other hand, the Singular and the Simple plural conditions were not significantly different ($t = -1.18$, $p > .1$. See Table 2 and Figure 1). There also was a main effect of Noun(s) Gender ($F(1,42) = 10.3$, $p < .003$), showing that incorrect agreement on feminine nouns (2.59) was rated as more acceptable than incorrect agreement on masculine nouns (2.18). These findings are qualified by a significant two-way interaction Condition * Noun(s) Gender ($F(2,41) = 13$, $F < .00001$). This interaction shows that the difference between ratings on feminine and masculine nouns is modulated by condition. T-test comparisons revealed that, in the Complex plural congruent condition, incorrect agreement on feminine nouns was rated as more acceptable than incorrect agreement on masculine nouns (3.53 *vs.* 2.15; $t(1, 42) = 5.9$, $p < .00001$). On the other hand, no difference between incorrect agreement on feminine and masculine nouns was observed in the Singular (2.06 *vs.* 2.4; $t(1, 42) = -1.5$, $p > .1$) or in the Simple plural Conditions (2.16 *vs.* 1.98; $t(1, 42) = 1.1$, $p > .1$). The Complex plural mixed condition was then compared with the other two Complex plural conditions (feminine and masculine) by means of separate paired t-Test. Results showed that incorrect agreement on mixed-gender nouns was rated as more acceptable than incorrect agreement on gender congruent masculine nouns (2.71 *vs.* 2.15; $t(1, 42) = -2.8$, $p = .007$), but as less acceptable than incorrect agreement on gender-congruent feminine nouns (2.71 *vs.* 3.53; $t(1, 42) = 3.7$, $p = .0007$. See Figure 2).

Table 2. Participants' mean ratings (with S.E. in parentheses) for sentences violating agreement across different experimental conditions (Experiment 3).

	SING	2.06 (0.20)
FEMININE	SIMPLE PL	2.16 (0.20)
	COMPLEX PL	3.53 (0.24)
	SING	2.41 (0.23)
MASCULINE	SIMPLE PL	1.98 (0.19)
	COMPLEX PL	2.15 (0.19)
MIXED GENDER	COMPLEX PLURAL	2.71 (0.22)

Figure 1. Participants' mean ratings for sentences violating gender agreement (Experiment 3). Violations of feminine agreement are tolerated significantly more readily than violations of masculine agreement only in the case of conjoined NPs (complex plurals). The mixed-condition is not considered. Error bars depict standard errors of the means.



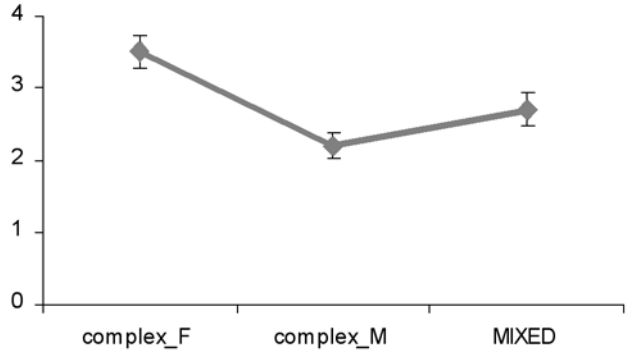


Figure 2. Participants' mean ratings for incorrect sentences requiring agreement with complex plurals. (Experiment 3). Mean acceptability ratings for incorrect agreement in the mixed condition are significantly higher than in the congruent-gender-masculine condition, and significantly lower than in the congruent-gender-feminine condition. Error bars depict standard errors of the means.

4.3. Discussion

The main findings of the present experiment can be summarized as follows:

a. Agreement violations with conjunction of feminine NPs were tolerated significantly more readily than agreement violations with conjunction of masculine NPs or conjunction of NPs of different gender. On the other hand, no difference between simple feminine and simple masculine plurals was observed.

b. Agreement violations with conjunction of NPs of different gender were tolerated significantly more readily than agreement violations with conjunction of masculine NPs.

The mere fact that acceptability ratings varied across conditions is at odds with the view that the results from the previous experiments are exclusively due to an experimental strategy related to time pressure.

Both findings (a) and (b) may be assumed to reflect speakers' competence on aspects of gender agreement. As to finding (a), the contrast between complex and simple feminine plurals should be particularly emphasized. This finding rules out the hypothesis of an intrinsic greater difficulty for the plural feminine marking. On the other hand, it supports the hypothesis that the greater difficulty associated with processing feminine gender lies in the integration process (i.e., in the process that integrates information from more than one lexical node). Specifically, these findings suggest that when two different feminine referents have to be integrated, the resolution rule for feminine agreement is not so strong as to completely exclude the possibility for the masculine agreement marking. With respect to finding (b), it is worth mentioning that the relative order of feminine and masculine nouns in mixed conjunctions may affect participants' acceptability and production of noun-verb agreement. There is in fact independent evidence suggesting that the closer (within a sentence) that a noun appears next to a verb, the higher is the probability that the noun evokes incorrect agreement on the verb (Haskell & MacDonald (2005; but see Vigliocco & Nicol (1998) and Franck et al. (2002) for evidence against a role of noun-verb proximity in influencing agreement). Our results do not directly bear on the issue of noun-verb proximity, as the feminine noun was always the first noun in mixed conjunctions – that is, the noun more distant from the verb. However, if anything, we may have underestimated the extent to which Italian speakers tolerate feminine agreement with a conjunction of nouns having different genders. In other words, in the meas-

ure to which proximity may influence judgments of acceptability, one would expect that incorrect feminine agreement would be rated even more acceptable if the feminine noun were the noun nearer to the verb in a conjunction of mixed-gender nouns.

5. General Discussion

The objective of this study was to explore the mechanism of gender agreement with pairs of pictured objects or conjunctions of written nouns; in particular, we asked whether the occasional reports of the over-generalization of the masculine form even when the feminine form would have been required by rule, is due to competence- or performance-related reasons. The results from the on-line experiments and the off-line task reported in this study converge to indicate a competence-based process.

The main findings can be summarized as follows: 1. People are less accurate and slower to produce, as well as more inclined to tolerate violations in comprehension, for conjunctions of feminine referents (e.g., *la bottiglia e la tazza* 'the bottle:FEM and the cup:FEM') compared to conjunctions of masculine or mixed-gender referents. 2. People are less accurate and slower to produce (in Exp. 1) as well as more inclined to tolerate violations in comprehension, for conjunctions of mixed-gender referents (e.g., *la bottiglia e il bicchiere* 'the bottle:FEM and the glass:MAS') than for conjunctions of masculine referents (e.g., *il libro e il bicchiere* 'the book:MAS and the glass:MAS'). Additionally, as is evidenced by the acceptability ratings, the difference between feminine and masculine gender is restricted to the case of conjoined NPs: There is no difference in the acceptability of incorrect gender agreement between masculine and feminine plurals (e.g., *due bottiglie* 'two bottles:FEM' vs. *due bicchieri* 'two glasses:MAS').

We have argued that the convergence of on-line and off-line data constitutes support for the competence hypothesis outlined in the Introduction. Alternatively, one may argue that the masculine gender is over-extended just because it is the default value in the Italian gender system. This may be easily appreciated if we consider those nouns that allow for both gender values. Since those nouns typically refer to naturally sexed entities, the choice of the specific value usually depends on the semantic gender of the referent. For instance, the Italian word for 'friend' is *amico* or *amica* depending on the sex – male or female, of the referent. Importantly, whenever the sex of the referent is left unspecified, the masculine form is used. Note,

however, that the notion of default has to be applied on a case-by-case basis (Corbett, p.c.). That is, the fact that a particular morphological realization is the default in a given case, does not entail that it is a default in all cases. This thus reduces to saying that the mere fact that the masculine gender may work as the default in the specific case of *amico*/*amica* 'friend:MAS/FEM', does not entail that it must work as the default gender in all the other cases.

For the agreement phenomena considered here, the use of the masculine agreement with conjunctions of feminine nouns was never permitted, at least according to the grammar of Standard Italian (cf. Dardano & Trifone (1997: 201), in relation to Adjective-Noun agreement (free translation): "When it refers to a conjunction of nouns having the same gender, all singular, all plural, or some singulars and some plurals, the adjective takes the gender of the nouns and is usually plural"). Since gender agreement is always mandatory, the over-extension of the masculine gender cannot be attributed to the use of a default value. This, however, cannot be the whole story, as the data also document an inclination towards the representation of the gender of each referent in the response (relative acceptability of the wrong agreement with mixed referents). This is also at odds with the *dictata* of the grammar: cf. again Dardano & Trifone (1997: 201), in relation to Adjective-Noun agreement (free translation): "When it refers to a conjunction of nouns having different gender and numbers, the adjective is usually masculine plural". See also Serianni & Castelvecchi (1988: 169;355) on this point). This means that the rule for gender agreement in its complex cannot be reduced to the new creation of a default.

Our results may reflect the beginning of the erosion of the system for gender agreement in Italian. This hypothesis would fit nicely with the commonly accepted view that language changes start from peripheral cases. The reasoning would go as follows. Our results do not align with respect to the application of a rule only in the case of multiple referents (*e.g.,* the misuse of the masculine agreement with feminine nouns is restricted to the case of conjunctions of feminine nouns). On the reasonable assumption that gender is a property of lexical nodes, in the case of agreement with a single noun the gender feature is inherited from a single lexical node. On the other hand, in the case of agreement with a conjnction of nouns, the gender feature is inherited from the value that arises due to the application of a rule regarding the integration of all the nouns in the conjunction. Whereas the first case is assumed to be the *central* or *prototypical* case of gender agreement – as it is lexically-dependent, the second case is assumed to be

the *peripherical* case of gender agreement – as it cannot be resolved within a single lexical node, but requires the considerations of two or more lexical nodes. Thus, if a process of erosion of the gender agreement system is in fact occurring in Italian, one reasonable hypothesis is that it started from outside of the lexical boundaries, that is, from the 'peripheral' case where the features of different lexical nodes have to be integrated into the value-triggering agreement. One way to tackle the hypothesis of gender erosion could be to test people of different age ranges: if older people tolerated the masculine agreement with multiple feminine referents, and the feminine agreement with mixed-gender referents less readily than young people, this would speak in favor of the hypothesis of a language change in the Italian gender system.

An alternative hypothesis would be that the language change specifically concerns the function or usage of Italian gender-marked pronouns, and not the general underlying system of gender agreement. However, were this the case, we would have expected a difference in the acceptability of violations of agreement depending on the word (verb or pronoun) on which gender is marked. The results of Experiment 3 do not support this prediction.

Leaving aside the hypothesis of gender erosion, agreement with multiple referents appear to be sensitive to two distinct kinds of pressure: on the one hand, a bias towards the masculine response, and on the other hand, an inclination towards the representation of the gender of each referent in the response. This means, in other words, that there is confusion regarding the agreement rules that operate over conjunctions of nouns, as is revealed by the fact that Italian speakers do not apply the rule consistently except when the nouns in the conjunctions are both masculine.

It might be a challenging issue for future research to better characterize the nature of the inconsistency in the application of a rule as a function of context. For instance, it would be interesting to explore the limits of the generalization of the masculine gender to multiple feminine referents. Can the on-line activation of incorrect masculine agreement forms (or the possibility for incorrect masculine agreement in an off-line task) be modulated by the absolute number of feminine referents, or by the combination of natural and grammatical gender? For instance, one could make the hypothesis that when the referents have both grammatical and natural feminine gender (e.g., *la nonna e la zia* 'the grandmother and the aunt'), people are less prone to accept the masculine agreement than when the referents are only grammatically marked (e.g., *la bottiglia e la tazza* 'the bottle:FEM and the cup:

FEM'). This prediction is corroborated by data from Serbian/Croatian/Bosnian (e.g., Wechsler & Zlatić, 2000; 2003). In this language, the gender resolution rule would be similar to the Italian one, except that Serbian/Croatian/Bosnian has a three-way gender system (masculine/feminine/neuter). The feminine is restricted to the case in which all of the nouns in a conjunction are feminine, whereas the masculine has to be used in all other cases. A more careful analysis, however, has revealed that the masculine form may be used even when all the nouns in a conjunction are feminine, unless they are also marked for natural gender (Corbett, 1991: 299-303). When nouns are both grammatically and semantically marked for feminine, the masculine cannot be used as a default.

Similar questions could also be addressed with mixed-gender referents. Does the possibility for feminine agreement vary as a function of the absolute number of feminine referents in a conjunction, or the relative proportion of masculine and feminine referents in a conjunction, or, even, the relative order of masculine and feminine referents (e.g., *la bottiglia e il bicchiere* 'the bottle:FEM and the glass:MAS' vs. *il bicchiere e la bottiglia* 'the glass:MAS and the bottle:FEM')? One reasonable hypothesis could be that the probability for the misselection of the feminine gender increases with the (absolute or relative) number of feminine referents in a mixed-gender conjunction.

To conclude, our findings may set the stage for future research in multiple directions: the first, more diacronically-oriented, focuses on a better characterization of instability in the gender agreement system in Italian; the second, more synchronically-oriented, focuses on the interplay between gender agreement and other contextual factors.

Address of the first Author:

Chiara Finocchiaro, Center for Mind/Brain Sciences - CIMeC, University of Trento, Polo di Rovereto, Corso Bettini, 31, 38068 Rovereto (TN), Italy, chiara.finocchiaro@unitn.it

Appendix A. Materials used in Experiment 1. The pictures in the first two columns were paired with the pictures in the other two columns of the same row on different trials.

FEM PICs (1)	MAS PICs (1)	FEM PICs (2)	MAS PICs (2)
banana 'banana'	fungo 'mushroom'	zucca 'pumpkin'	carciofo 'artichoke'
bilancia 'balance'	vaso 'vase'	pistola 'gun'	piatto 'plate'
foca 'seal'	topo 'mouse'	giraffa 'giraffe'	pinguino 'penguin'
foglia 'leaf'	fiore 'flower'	carota 'carrot'	rastrello 'rake'
gonna 'skirt'	guanto 'glove'	cravatta 'tie'	cappello 'hat'
mucca 'cow'	serpente 'snake'	farfalla 'butterfly'	canguro 'kangaroo'
pipa 'pipe'	fiocco 'bow'	bottiglia 'bottle'	coltello 'knife'
ruota 'wheel'	ombrello 'umbrella'	campana 'bell'	ventaglio 'fan'
scarpa 'shoe'	pettine 'comb'	chitarra 'guitar'	lucchetto 'padlock'
sedia 'chair'	tavolo 'table'	chiave 'key'	casco 'helmet'
tenda 'tent'	quadro 'painting'	valigia 'luggage'	orologio 'watch'
tromba 'trumpet'	tamburo 'drum'	racchetta 'racket'	libro 'book'

Appendix B. Materials used in Experiment 2. The pictures in the first two columns were paired with the pictures in the other two columns of the same row on different trials.

FEM PICs (1)	MAS PICs (1)	FEM PICs (2)	MAS PICs (2)
banana 'banana'	fungo 'mushroom'	zucca 'pumpkin'	carciofo 'artichoke'
bottiglia 'bottle'	coltello 'knife'	pipa 'pipe'	fiocco 'bow'
giraffa 'giraffe'	pinguino 'penguin'	foca 'seal'	topo 'mouse'
foglia 'leaf'	fiore 'flower'	carota 'carrot'	rastrello 'rake'
valigia 'luggage'	orologio 'watch'	tenda 'tent'	quadro 'painting'
sedia 'chair'	tavolo 'table'	chiave 'key'	casco 'helmet'

Appendix C. Sentences violating gender agreement used in Experiment 3.

C'è una bottiglia. Portalo.	'There is a bottle:FEM. Bring it:MAS.'
La pantera è fuggito dai cacciatori.	'The panther:FEM escaped:MAS from the hunters. '
C'è un quaderno. Prendila.	'There is a copybook:MAS. Take it:FEM.'
Il pesce è balzata in mare.	'The fish:MAS jumped:FEM in the sea. '
Ci sono due forchette. Portali.	'There are two forks:FEM. Bring them:MAS. '
Le giraffe sono fuggiti dai cacciatori.	'The giraffes:FEM escaped:MAS from the hunters. '
Ci sono due quaderni. Prendile.	'There are two copybooks:MAS. Take them:FEM. '
I pinguini sono balzate in mare.	'The penguins:MAS jumped:FEM in the sea. '
Ci sono una forchetta e una bottiglia. Portali.	'There are a fork:FEM and a bottle:FEM. Bring them:MAS. '
La giraffa e la pantera sono fuggiti dai cacciatori.	'The giraffe:FEM and the panther:FEM escaped:MAS from the hunters. '
Ci sono un libro e un quaderno. Prendile.	'There are a copybook:MAS and a book:MAS. Take them: FEM. '
Il pinguino e il pesce sono balzate in mare.	'The penguin:MAS and the fish:MAS jumped:FEM in the sea. '
Ci sono una gomma e un quaderno. Prendile.	'There are a rubber:FEM and a copybook:MAS. Take them: FEM.'
La lontra e il pinguino sono balzate in mare.	'The otter:FEM and the penguin:MAS jumped: FEM in the sea'.

Notes

* We would like to thank Luigi Lombardi and Francesco Vespignani for their assistance with the statistical analysis. B.Z. Mahon was supported in part by an NSF Graduate Research Fellowship. The research was supported in part by the Fondazione Cassa di Risparmio di Trento e Rovereto.

1 Italian clitics can precede (i.e., they are proclitics) or follow (i.e., they are enclitics) the verb (cf. *lo porto* 'I bring it:MAS' vs. *portalo* 'bring it:MAS', depending on the given verb form (finite tenses yield proclisis, whereas non-finite tenses and affirmative imperatives yield enclisis). In this study, only enclitic object forms were used (e.g., *portalo* 'bring it:MAS').

2 Verbs are not usually marked for gender in Italian. There are three exceptions: 1) verbs requiring the auxiliary verb essere 'to be' when used in the present perfect. In this case, the past participle of the main verb has to agree with the subject (e.g., *la mia amica è partita* 'my friend:FEM left:FEM'); 2) all verbs when used in passive analytic forms (e.g., *la porta è aperta dalla segretaria* 'the door is opened by the secretary'). As in the previous case, agreement is with the subject; 3) transitive verbs requiring the auxiliary verb *avere* 'to have' when used in the present perfect with direct object clitic pronouns. In this case, the main verb has to agree with the direct object (e.g., *le ho prese* 'I took [FEM] them [FEM]'). In this experiment, only the first condition of verb agreement was considered.

Bibliographical References

BORTOLINI Umberto, Carlo TAGLIAVINI & Antonio ZAMPOLLI 1971. *Lessico di Frequenza della Lingua Italiana Contemporanea.* Milano: Garzanti.

COHEN Jonathan D., Brian MACWHINNEY, Matthew FLATT & Jefferson PROVOST 1993. PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments and Computers* 25. 257-271.

CORBETT Greville G. 1991. *Gender.* Cambridge: Cambridge University Press.

CORBETT Greville G. 2006. *Agreement.* Cambridge: Cambridge University Press.

DARDANO Maurizio & Pietro TRIFONE 1997. *La nuova grammatica della lingua italiana*. Milano: Zanichelli.

FRANCK Julie, Gabriella VIGLIOCCO & Janet L. NICOL 2002. The role of syntactic tree structure and complexity in subject-verb agreement. *Language and Cognitive Processes* 17. 371-404.

HASKELL Todd R. & Maryellen MACDONALD 2005. Constituent structure and linear order in language production: Evidence from subject verb agreement. *Journal of Experimental Psychology: Learning, Memory and Cognition* 35. 891-904.

RAAIJMAKERS Jeroen G.W., Joseph M.C. SCHRIJNEMAKERS & Frans GREMMEN 1999. How to deal with 'the language-as-a-fixed-effect fallacy': common misconceptions and alternative solutions. *Journal of Memory and Language* 41. 416-426.

SALVI Giampaolo & Laura VANELLI 2004. *Nuova grammatica italiana.* Bologna: Il Mulino.

SERIANNI Luca & Alberto CASTELVECCHI 1988. *Grammatica italiana. Italiano comune e lingua letteraria*. Torino: Utet.

VIGLIOCCO Gabriella & Janet L. NICOL 1998. Separating hierarchical relations and word order in language production. Is proximity concord syntactic or linear? *Cognition* 68. 13-29.

WECHSLER Stephen & Larisa ZLATIĆ 2000. A theory of agreement and its application to Croatian. *Language* 76. 799-832.

WECHSLER Stephen & Larisa ZLATIĆ 2003. *The many faces of Agreement*, Stanford: CSI Publications.