Croatian Mixed-Gender Conjunct Agreement: An ERP Study

Marijan Palmović and Jana Willer-Gold

Abstract: In a recent elicited-production study with native speakers of Slovenian, Marušič, Nevins, and Saksida (2007) and Marušič, Nevins, and Badecker (2015) show that there are three distinct variously attested gender-agreement grammars. In this study, the high temporal-resolution of the ERP (event-related potential) technique was used to detect neurological components and measure the processing cost of the three gender-computing mechanisms. The study is comprised of two acceptability-judgment experiments, using a factorial design with nonmasculine mixed-gender conjuncts. Experiment 1 contrasts two strategies, Distant- (DCA) and Closest-Conjunct Agreement (CCA), to question whether the linear distance between a participle and the two conjuncts is language- or memory-related. The Experiment 1 results show behaviorally an overall significant effect of gender; and neurologically a memory-related component, the P300. Experiment 2 sets out to detect alternations to the processing cost when default (Def) agreement is added to the experimental paradigm. The Experiment 2 results indicate no gender effects; instead, two language-related components, N250 and N450, were observed, statistically picking out DCA once again. We argue that in an ecologically valid environment where all three grammatical options are made available, processing of DCA is no longer supported by a general cognitive mechanism, such as memory, but is rather computed by language-related processes.

1. Single- Versus Multiple-Agreement Dependencies

The rapid increase in studies on gender, number, and person features in the last decade has affirmed that agreement provides valuable insight into language structure and language processing.

Feature agreement, in particular, has dominated experimental investigations into the mechanisms underlying agreement dependencies in psycho- and neurolinguistic studies. In these studies, a single (type of) agreement dependency is taken to hold in subject-verb, pronoun-verb, or (determiner)-noun-(adjective) relations; hence, feature mismatches, agreement violations, production errors, and ERP effects provide a useful tool for fleshing out the underlying cognitive processing of agreement (Bock and Miller 1991; Franck et al. 2002, 2006; Franck 2011; Gillespie 2011; Santesteban 2013; Nevins et al. 2007; Mancini et al. 2011; Mancini et al. 2014; Quiñones et al. 2014; among others).

However, recent linguistic investigations of agreement dependencies where the subject/object is a coordinated phrase have drawn attention to the possibility of a single argument taking part in multiple grammatical agreement dependencies (Aoun, Benmamoun, and Sportiche 1994, 1999; Benmamoun 1992; Benmamoun, Bhatia, and Polinsky 2010; Munn 1999; van Koppen 2005, 2007; Bhatt and Walkow 2014; Walkow to appear; Bošković 2009; Marušič, Nevins, and Saksida 2007; Marušič, Nevins, and Badecker 2015; among others). These findings suggest a less clear-cut grammatical–ungrammatical agreement divide and consequently demand an explicit experimental design that will allow disentangling of each one of the multiple agreement dependencies and show their independent contributions in language processing.

2. Gender Agreement with Conjuncts in South Slavic Languages

A recent production study by Marušič, Nevins, and Badecker (2015) on coordinated subject-verb agreement in Slovenian accentuates the possibility of experimentally investigating a structure that can undergo three agreement dependencies with a varying degree of frequency in production. This observation is most clearly exemplified by combinations of mixed-gender conjuncts, namely neuter (N) plus feminine (F) (henceforth NF), and feminine (F) plus neuter (N) (henceforth FN). Apart from the most widely attested default agreement option (Def), as in (1), in which the verb/participle agrees with the coordinated phrase, an agreement dependency can hold between the participle and the first or hierarchically highest conjunct (HCA), as in (2), or between the participle and the second or linearly closest conjunct (CCA), as in (3).

(1) Polja   i   livade   su   obasjani   proljetnim
     field$_{N.PL}$ and meadows$_{F.PL}$ AUX$_{3PL}$ illuminated$_{M.PL}$ spring
     suncem.
     sun

(2) Polja   i   livade   su   obasjana   proljetnim
     field$_{N.PL}$ and meadows$_{F.PL}$ AUX$_{3PL}$ illuminated$_{N.PL}$ spring
     suncem.
     sun

(3) Polja   i   livade   su   obasjane   proljetnim
     field$_{N.PL}$ and meadows$_{F.PL}$ AUX$_{3PL}$ illuminated$_{F.PL}$ spring
     suncem.
     sun

     'The fields and meadows are illuminated by the spring sun.'

2.1. Theoretical Predictions

In their account of the reported data, Marušič, Nevins, and Badecker argue for a two-step Agree, with Agree-Link, an operation which establishes Probe-Goal relations and Agree-Copy, an operation which retrieves feature values from the Goal to be copied onto the Probe.

Default agreement in (1) and agreement with the hierarchically closest conjunct in (2) follow from the assumption that Agree takes place in the syntax, where it applies to a hierarchically structured coordinated subject, &P. Once Agree-Link has applied, if the value for gender is missing on &P when Agree-Copy applies, either the default value is inserted, resulting in (1), or Agree-Copy retrieves a value from the closest available conjunct, the hierarchically closest conjunct, resulting in (2).

The third agreement option is explained once Agree is considered in the context of a "multiple components" account (Benmamoun, Bhatia, and Polinsky 2010; Marušič, Nevins, and Saksida 2007; Marušič, Nevins, and Badecker 2015; van Koppen 2005, 2007; Bhatt and Walkow 2014; among others), whereby Agree-Link applies in Syntax and Agree-Copy at PF, after Linearization of the two coordinated conjuncts. Once Agree-Link and Linearization have applied, Agree-Copy retrieves a gender-feature value from the closest available conjunct, the linearly closest conjunct, resulting in (3).

Basing our predictions solely on Marušič, Nevins, and Badecker's account, we would predict that if we observe comparable effects for Def and HCA, this should reflect a syntactic basis for agreement, whereas effects observed for CCA, depending on whether they are comparable to Def and HCA, can be taken to reflect either a syntactic or a PF basis for agreement. In other words, if we observe that Def and HCA behave similarly yet different to CCA, this is predicted by the timing of the application of Agree-Copy or, in other words, by the locus of feature retrieval in syntax. However, if we observe further variation in the grouping of Def and HCA, this is predicted by Agree-Copy retrieving feature values in a consecutive top-to-bottom fashion in the syntax, first probing &P for a gender value and only consequently the hierarchically closest conjunct.

2.2. Experimental Predictions

Marušič, Nevins, and Badecker make two additional observations—the grammaticality of three agreement patterns and the varying degrees of frequency with which each one of them is produced—that will allow us to make more explicit predictions about the cost of processing Def, HCA, and CCA.

Agreement with mixed conjuncts that are masculine (M) and neuter (N) or feminine (F) would result in a violation, as in (4) and (5), if agreement on

the predicate is valued by a gender that is not found on either of the two conjuncts, i.e., F and N, respectively.

(4) *Polja   i   kanjoni   su   obasjane   suncem.
     fields$_{N.PL}$ and   canyons$_{M.PL}$ AUX$_{3PL}$ illuminated$_{F.PL}$ sunlight$_{INST}$

(5) *Kanjoni   i   livade   su   obasjana
     canyons$_{M.PL}$ and   meadows$_{F.PL}$ AUX$_{3PL}$ illuminated$_{N.PL}$

     suncem.
     sunlight$_{INST}$

As demonstrated in (1–3), this is never the case for instances of agreement dependencies with non-masculine mixed-gender conjuncts. This point is further supported by the additional evidence of agreement dependencies at the inter- and intraindividual level. Marušič, Nevins, and Badecker show that all three agreement grammars, exemplified in (1–3), can be available to a single speaker of a language. Assuming that all three agreement grammars are simultaneously available to a single speaker's subject-verb agreement-computing mechanism, no gender mismatches, agreement violations, or production errors are expected to arise when processing these grammatical instances of agreement dependencies (cf. Bošković 2009, Puškar, and Murphy 2015 for Serbo-Croatian; Franks and Willer-Gold 2015 for Croatian).

While grammatical, these three agreement strategies nevertheless show variation in their frequency of production. Reviewing the breakdown of the average percentages provided by Marušič, Nevins, and Badecker for Slovenian, we observe a preference for Def in the NF condition (M: 39%, F: 26%, N: 26%) and for CCA in the FN condition (M: 20%, F: 22%, N: 52%), showing that all three grammars, when available, are produced to a varying degree of frequency and that this variation is further conditioned by a mixed-gender combination of the two conjuncts, leaving open the question of why this is so.

Importantly, these two observations were further validated in a larger study replicating Marušič et al.'s experimental design with six South Slavic language varieties in Sarajevo, Zadar, Zagreb, Niš, Novi Sad, and Ljubljana (see Willer-Gold et al., this volume). The average percentages across all six varieties of the three agreement grammars is given for the NF condition (Mpl: 46%, Fpl: 36%, Npl: 18%), the FN condition (M: 36%, F: 11%, N: 53%) and for the relevant Croatian variety spoken in Zagreb, for the NF condition (Mpl: 51%, Fpl: 28%, Npl: 18%) and the FN condition (M: 39%, F: 6%, N: 54%).

Assuming as the null hypothesis a general processing parallelism between production and comprehension, none of these three strategies should show the profile of a violation per se. However, the production data does predict an increase in processing cost for HCA.

In summary, Marušič, Nevins, and Badecker's account argues for Def and HCA gender-feature values are computed in syntax (before Linearization), while in the case of CCA, gender-feature values are copied in the postsyntactic component after Linearization. Experimental data from two production studies, Marušič, Nevins, and Badecker and Willer-Gold et al., provide insight into the graded preferences for three agreement grammars relative to the position of the F and N conjuncts in mixed-gender coordinates. There is one final factor relevant for the processing cost of HCA and CCA: the linear or surface distance between the two conjuncts, whether Conj2 is closer to the participle and Conj1 is more distant (henceforth DCA). The assumption that linear distance plays a role in processing, reflected in the longer retention of gender-feature values in working memory, predicts that DCA processing would consume more power than CCA. The consequences that these two complementary areas of investigation have for the design and interpretation of a comprehension study are discussed in the next two sections.

3. An ERP Study of Mixed-Gender Conjunct Agreement

The aim of this comprehension study is to tap into the neurological processing of conjunct agreement, in particular gender agreement, with native speakers of Croatian. A special focus is placed on mixed-gender conjuncts in preverbal subject position and the processing underlying the three grammatical subject-verb agreement dependencies. To achieve this goal, we use event-related potentials (ERP), a widely used electrophysiological technique, to record the average brain activity related to the processing of a particular agreement strategy, allowing us to estimate the processing cost during comprehension of Def, HCA, and CCA and learn which processing difficulty has occurred and when it occurred.

3.1. Motivation

Our study is a partial replication of the production study for Slovenian reported in Marušič, Nevins, and Badecker 2015. In line with the Slovenian study, we use a factorial design to draw inferences about speakers' conjunct-agreement strategies, but also to verify whether the same factors can explain variance across two newly designed experiments (see below). The methodology applied called for adjustments to the experimental design.

First, the sentence comprehension and evaluation task called for agreement in the stimulus. The value of gender on the participle varies with respect to the experimental question. We restricted our design to the most informative or transparent combinations of mixed-gender conjuncts, NF and FN. These nonmasculine mixed-gender combinations are the only two predicted

to allow for all three genders on the participle; compare M-Def in (1), F-DCA in (2), and N-CCA in (3).

Second, instead of a single experiment with two conditions, NF and FN, we expanded the stimulus design and in view of the nature of the comprehension task created two new experiments. To test for the effect of distance and its origin in agreement dependency processing, we designed the first experiment with two-level closeness factors: Distant (hierarchically closest) or Close (linearly closest), where the participle on the verb takes either a feminine or neuter gender value. To investigate the interaction of Distant and Close with Def, we created a second experiment by adding Default as an independent factor, where all three gender values appear on the participle, masculine, feminine, and neuter. Note that testing the same agreement strategies in two parallel experiments with a factorial design tests the limits of the factorial design and the prediction that the same agreement strategy will produce equal results, irrespective of the levels of closeness or independent variables.

Croatian is an inflectional language, hence, comprehension of subject-verb agreement is fully recoverable once the gender value on the participle has been processed. Therefore, although a single agreement strategy, e.g., Def, might be responsible for predicting the value of gender agreement on the predicate and hence serve as a baseline against which to evaluate any deviation in processing cost, we assume that the gender value of both conjuncts has to be kept active in memory at least until the gender value of the participle has been integrated in the built-up context (i.e., irrespective of reanalysis).

Consequently, the null hypothesis predicts that equal amounts of processing cost are consumed in comprehension of Def, DCA, and CCA sentences. However, considering the production-study data showing a varying degree of preferences for each agreement grammar relative to the gender combination of the two conjuncts, the null hypothesis should be easily falsified. For Experiment 1, Marušič, Nevins, and Badecker predicted that DCA and CCA will show variation, as DCA is fully processed in syntax, while the processing of CCA, being spread across two language components, turns out to be more costly. On the other hand, the activation of working memory would predict the opposite result if the gender value of the first conjunct has to be retained for longer or reaccessed at greater distance than the second conjunct. In Experiment 2 since masculine is the most prominent gender (Def for all 9 uniform and mixed-gender combinations) and according to the theory that for Def Agree-Link and Agree-Copy probe the same Goal, &P, the least processing cost is expected to be consumed for Def. We summarize the discussion in terms of theory-related hypotheses (the locus or timing of the Agree-Copy operation; see (ia), (iia)) and memory-related hypotheses (longer or shorter retention of a conjunct's gender value in memory; see (ib), (iib)) for Experiment 1 in (i) and Experiment 2 in (ii):

(i) a. If DCA and CCA vary with respect to language-component distance (syntax vs. syntax + PF), processing effects should be higher for CCA.

   b. If DCA and CCA vary with respect to processing distance (long/first vs. short/second), memory effects should be higher for DCA.

(ii) a. If Def (&P) and DCA/CCA (single conjunct) vary with respect to operations (Agree vs. 2-step Agree), agreement-violation effects should be higher for DCA/CCA.

   b. If Def (&P) and DCA/CCA (single conjunct) vary with respect to distance (retention of both conjuncts vs. retention + reaccess of single conjunct), memory effects should be higher for DCA/CCA.

Finally, due to the ERP methodology and its temporal rather than spatial precision, we note that the results from this comprehension study provide only indirect evidence bearing on questions raised by the account presented: Where do Agree-Link and Agree-Copy operate—in the syntax or the postsyntactic component? And what are the neurophysiological correlates of Agree-Link and Agree-Copy?

3.2. ERP Methodology

To our knowledge, no ERP study investigating conjunct agreement has yet been attempted on the group of South Slavic languages under investigation, and hence there is no baseline for stating predictions or evaluate data in this study. We expect to observe EEG components in the areas of interest that are standardly reported in the ERP literature in studies on grammatical gender agreement violations (see Molinaro, Barber, and Carreiras 2011 for an overview; for individual studies, see Barber and Carreiras 2005; Deutsch and Bentin 2001; Nevins et al. 2007; Caffarra 2014, 2015). However, considering that disagreement in gender, similar to person mismatches or number attraction, has been shown to induce weaker ERP effects or to be manifested in a different way from proper ungrammatical violations, e.g., case violations, if we were to observe ERP correlations of agreement violations, we do not expect them to reach the highest degree of processing cost.

   In agreement studies, determiner/adjective-noun (subject/object) and subject-verb dependency violations are reported to induce late positivity effects (P600, 500–800ms) with larger amplitudes over centroposterior regions, or (early) anterior negativity effects ((E)LAN, 300–500ms) followed by late positivity effects (P600, 500–800ms) with larger amplitudes over centroposterior regions. LAN effects have been thought to reflect mismatches between the grammatical features of two elements in an agreement-dependency relation (Friederici 1995; Hagoort and Brown 1999) or difficulties induced by the in-

tegration of disagreeing elements in the existing sentential context (Gunter, Friederici, and Schriefers 2000). P600 effects have been associated with the costs of structure building, checking, and reprocessing (Friederici 1995; Van de Meeredonk et al. 2009), costs related to attempts at reanalysis and repair (Osterhout and Mobley 1995), or to late non-syntactic integration processes (Brouwer, Hartmund, and Hoeks 2012). Consequently, we take LAN and P600 to reflect the detection of gender-agreement inconsistencies in terms of processing an unexpected gender-agreement strategy and subsequent structural reanalysis for the purposes of potential integration of this agreement dependency in the existing sentential context (iia–b).

When investigating the processing of sentences containing an agreement dependency, the formal cues (morphemes) provided by an inflexionally rich system such as Croatian should not be overlooked (Leinonen et al. 2008; Caffarra 2014, 2015). Moreover, understanding Agree as a feature-value retrieval mechanism acting on a nonlocal dependency relation implies that it has to retain the morphologically marked gender value of a conjunct to be matched against the morphologically marked gender value of the predicate in order for the latter to be integrated into the built-up context. Therefore, we might expect to observe cost related to the use of short-term or working memory during sentence processing reflected in a sustained LAN (Coulson, King, and Kutas 1998; Molinaro, Barber, and Carreiras 2011; Gouvea et al. 2010), in slow wave amplitude increase (Johnson 1995), or an increase in working-memory load due to distance of information retrieval reflected in the P300 (Friedmann 1990; Polich 2007; cf. (ia–b)).

Finally, the N400 component, taken to reflect semantic violations or semantic gender mismatches (Barber and Carreiras 2005), has been reported for studies that vary the animacy of the noun (Deutsch and Bentin 2001). Therefore, to avoid these confounding variables and in line with the original Marušič, Nevins, and Badecker (2015) study, all the stimuli in this study were inanimate nouns. Similarly, in order to a priori exclude the possibility of gender-number interaction or the possibility of the observed (E)LAN-P600 being the result of a number-agreement violation (see Molinaro, Barber, and Carreiras 2011 for an overview), the number value plural is kept constant across conjuncts and participles.

3.3. Methods and Design

3.3.1. Participants

In total, 21 participants were included in the study (14 females and 7 males). All of them were university students aged 20–24. Most of them were psychology students or students of speech and language pathology, as well as some volunteers who signed up for the experiment on an electronic scheduling

system. Two participants were eventually excluded due to extensive artifacts. All of them had normal or corrected-to-normal vision. Prior to the ERP measurement each participant was given the Edinburgh Handedness Questionnaire (Oldfield 1971). Only right-handed participants were included in the ERP measurement. Once a participant was chosen, (s)he took part in several experiments that were carried out in the lab, always including both experiments described in this study. The whole procedure lasted two to three hours per participant. The experiments were scheduled so that Experiments 1 and 2 were never adjacent to each other, i.e., there was always a third experiment (not a part of this study) in between, together with a short break.

3.3.2. Materials

In both experiments 40 sentences per condition were presented visually to the participants. The sentences had the same structure: an initial conjunction of neuter and feminine nouns followed by a participle, an auxiliary (together constituting the perfect tense), and the final phrase (an adverb or an object NP). The initial conjunction consisted of nouns in the plural. All nouns were inanimate and had a transparent gender suffix to avoid animacy and transparency effects. Examples are given in (1–3). In both experiments the order of the conjuncts was manipulated as well as the gender of the participle. In Experiment 1 possible combinations of the order of the conjuncts and the gender of the participle yielded four experimental conditions allowing for a simple 2×2 design, as shown in Table 1.

Table 1. Experiment 1: Conditions and Factors

Conditions		Factors	
Order of the conjuncts	Gender of the participle	Distance	Gender
F + N	F	DISTANT	FEM.
N + F	F	CLOSE	FEM.
F + N	N	CLOSE	NEUT.
N + F	N	DISTANT	NEUT.

In Experiment 2 an extra option was added, namely agreement with the participle in masculine gender. This created two more experimental conditions and required a different design, as shown in Table 2. In this design the Distance factor was defined as distance in the underlying linguistic structure, i.e., distance from the Conj1, Conj2, or &P nodes. If the participle is masculine, the Gender factor was defined as the gender of the first conjunct, not the par-

ticiple. This was necessary because a default masculine participle is in fact at no distance from anything.

Table 2. Experiment 2: Conditions and Factors

Conditions		Factors	
Order of the conjuncts	Gender of the participle	Distance	Gender
F + N	F	DISTANT	FEM.
N + F	F	CLOSE	NEUT.
F + N	N	CLOSE	FEM.
N + F	N	DISTANT	NEUT.
F + N	M	DEFAULT	FEM.
N + F	M	DEFAULT	NEUT.

This means that the statistical analysis will not be strictly parallel in both experiments. Finally, in both experiments, 40 sentences of similar but slightly different structures, half of them ungrammatical, were added as fillers. The initial conjunct contained animate nouns or nouns of the same gender in the singular, and sometimes there was no participle but a finite verb form, e.g., 3rd person plural present. These filler sentences served two purposes. First, because all experimental sentences were actually grammatical and the participants had to make a grammaticality judgement, some easily recognizable ungrammatical sentences had to be added to avoid participants pressing just one button without much thought. Second, these different structures prevented participants from forming strategies and hypotheses about what was actually under investigation in the experiment.

3.3.3. Methods

The participants were asked to sit in front of a 19″ LCD computer screen where the stimuli were presented with a vertical visual angle of around 2° depending on the distance to the screen (70–100 cm). The experiments were encoded with the E-Prime 2 program (Schneider, Eschman, and Zuccolotto 2002). Participants were instructed to read the sentences and make a grammaticality judgment at the end of the sentence by pressing the left or right button on the Serial Response Box (part of the E-Prime package). Each sentence was divided into four parts: an initial conjunction (duration 500ms), an auxiliary word (300ms), a participle (300ms) and a final phrase (500ms). After each segment there was a 300ms pause before the onset of the next segment ([CONJUNCTION$_{500ms}$] 300ms [AUXILIARY$_{300ms}$ PARTICIPLE$_{300ms}$] 300ms

[FINAL PHRASE$_{500ms}$]). This also served for the collection of the ERP response in the middle of the sentence. After the sentence ended there was a period of 1000ms for collecting the response. The next segment started with the cross in the middle of the screen. Its duration was 300ms, with another 300ms of blank screen before the onset of the conjunction to allow the participant to prepare for the next trial. The duration of each segment was established after several off-line trials in order to achieve a comfortable reading speed. The trigger for averaging the EEG signal was set at the onset of the participle.

The EEG signal was recorded with a Brain Products 32 channel ActiCap electrode cap with electrodes arranged according to the 10-10 standard and the BP QuickAmp amplifier using average reference. The signal was sampled with 1kHz. Vertical and horizontal eye movement artifacts were recorded with two bipolar electrodes (VEOG and HEOG). Off-line, the EEG signal was re-referenced to both mastoids and filtered with the bandpass 0.1–30 Hz. Ocular artifacts were corrected with an ICA-based artifact removal tool. The signal was then averaged, corrected for the baseline –200–0ms, and averaged for each condition. Statistical analysis (repeated measure ANOVA) was performed on the behavioral data, with reaction time and accuracy as dependent variables. Since there were no ungrammatical stimulus sentences (only fillers), the percentage of sentences judged ungrammatical by the participants is, in fact, the percentage of sentences they found unacceptable. Statistical analysis on the ERP data was performed on the mean amplitudes in time intervals relevant for the detected ERP components.

Due to these slight differences in design, the same analysis could not be performed across experiments. Behavioral results (reaction time and acceptance rates) were tested on two factors, Distance and Gender (with a slight difference in their definition, as explained). The ERP data in Experiment 1 was tested with two additional factors relevant for the ERP technique: Lobe (frontal, central, and parietal) and Hemisphere (left, central, and right). In Experiment 2, such an analysis was not possible; instead, the factor Condition with three levels (close, distant, and default agreement) was used for the analysis, together with the Lobe and Hemisphere factors.

3.4. Results

3.4.1. Behavioral Results

In Experiment 1, reaction times and acceptance rates for the grammaticality-judgment task were measured on the last word in the sentence. A repeated measure ANOVA was used for the analysis with Gender (neuter, feminine) and Distance (close, distant) as two-level factors. No main effects of Gender and Distance were found. However, the interaction Gender × Distance was confirmed as statistically significant (F[1, 20] = 4.4, p = 0.049). For the ac-

ceptance rate, a main effect of gender was found (F[1, 20] = 200.1, p < 0.001, η_p2 = 0.19). The results are shown graphically in Figure 1.

These results indicate that generally gender matters: neuter participles were significantly more accepted, and participants were faster in making grammaticality judgments. Additionally, these results indicate that the participants did not make their choices based on distance. However, they were quicker in accepting the neuter participle as grammatical.

In Experiment 2, the same analysis could have not been performed on the resultant data since the default distance level of the Distance factor would not have been independent from the Gender; it is, in fact, a participle in masculine gender. Therefore, the Gender factor is defined as the gender of the first conjunct in order to preserve some gender information in the analysis. The Distance factor is now defined as the distance between the participle and a particular node in the conjunction structure (Conj1, Conj2, or &P nodes). This means that the repeated measure ANOVA was performed on a 3×2 design (Distance × Gender) with close, distant, and default as the three levels of the Distance factor and feminine and neuter as the two levels of the Gender factor. Two dependent measures were included in the analysis: reaction times and acceptance rates, as in the first experiment. No main effects were found to be statistically significant (Distance: F[1, 18] = 1.33, p = 0.28; Gender: F[1, 18] = 0.64, p = 0.43). However, the Distance × Gender interaction was found to be statistically significant (F[2, 36] = 8.5, p < 0.001). A post-hoc test (Tukey) reveals that a difference was obtained between the close distance and distant feminine levels (p = 0.022), the default distance and close neuter levels (p = 0.013), and the default distance and distant feminine levels (p < 0.001). For the acceptance rate variable, neither Distance nor Gender were found to be statistically significant
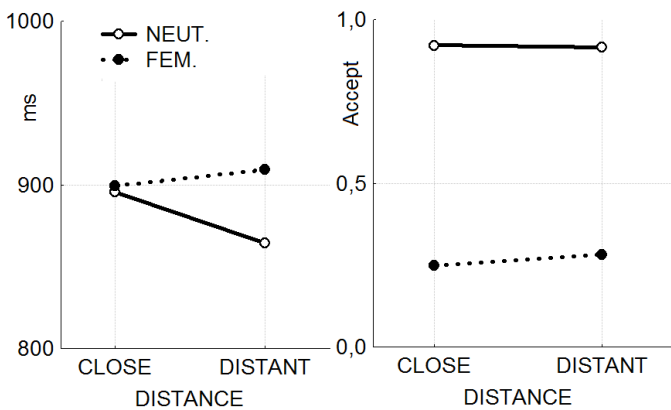

Figure 1. Behavioral results in Experiment 1
(left: reaction times; right: acceptance rate)

(Distance: F[1, 18] = 3.01, p = 0.06, Gender: F[1, 18] = 0.24, p = 0.62). However, a post-hoc test (Tukey) reveals a difference between the close and default levels of the Distance factor (p = 0.049). The Distance × Gender interaction was not found to be significant (F[2, 36] = 0.23, p = 0.79). The results are graphically represented in Figure 2 on page 150, and because they did not reach statistical significance, they can serve only to show trends in the results, e.g., that the acceptance rate is slightly higher for the close feminine and distant neuter conjuncts, while being irrelevant for default distance.

The participants were somewhat more confident in a default choice for the participle. However, as statistical significance was not found anywhere, these results must be taken with caution and only as trends.

3.4.2. ERP Results

In Experiment 1 the results show a weak P300 effect (marked with an arrow in Figure 3 on page 150) measured on the onset of the participle.

The ERP waveforms represent the averaged brain response to the participle in two experimental conditions. The point 0 is the onset of the participle. The largest effect obtained on the Pz electrode is marked with an arrow. Note that the experimental conditions differ both in the NF or FN combinations of the conjunct and in the gender of the participle. Because point 0 represents the onset of the participle regardless of its gender, the ERP waveforms obtained for two experimental conditions differ only regarding the position of the conjunct that agrees with the participle (close or distant) and not regarding some characteristics of the participle itself.

The P300 component can be interpreted as an electrophysiological trace of the distance between the controller (the conjunction) and the target. Statistical analysis (repeated measure ANOVA with two 3-level factors [Lobe: frontal, central, parietal; and Hemisphere: left, central, right] and two 2-level factors [Distance: close, distant; and Gender: neuter, feminine]) found a statistically significant main effect of Distance (F[1, 19] = 7.58, p = 0.013, η_p^2 = 0.29). There was no statistically significant main effect of Gender (F[1, 18] = 0.98, p=0.33, η_p^2 = 0.05). Nor was any statistically significant interaction found. The results are

---

[1] Reaction time has proven to be a weakly informative measure in detecting neurological processes underlying a particular agreement grammar (Staub 2009, 2010). This might be due to a bias that reaction time in an acceptability task reflects the speed of acceptance or rejection of a presented stimulus, rather than a measure of processing a sentence that can be attributed to a specific agreement-dependency interpretation. In more detail, reaction time was not measured at the predicate, which might be responsible for not providing the exact processing value of the specific agreement dependency. Moreover, if all the options are equally available, the question is which one, if any, will be rejected and based on what criteria in a context-free task.
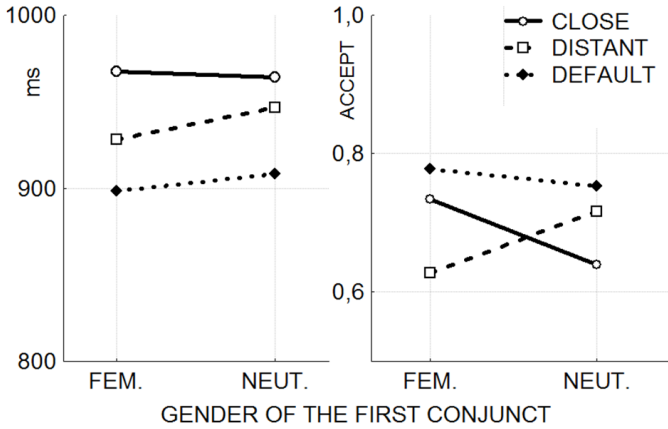
Figure 2. Behavioral results in Experiment 2
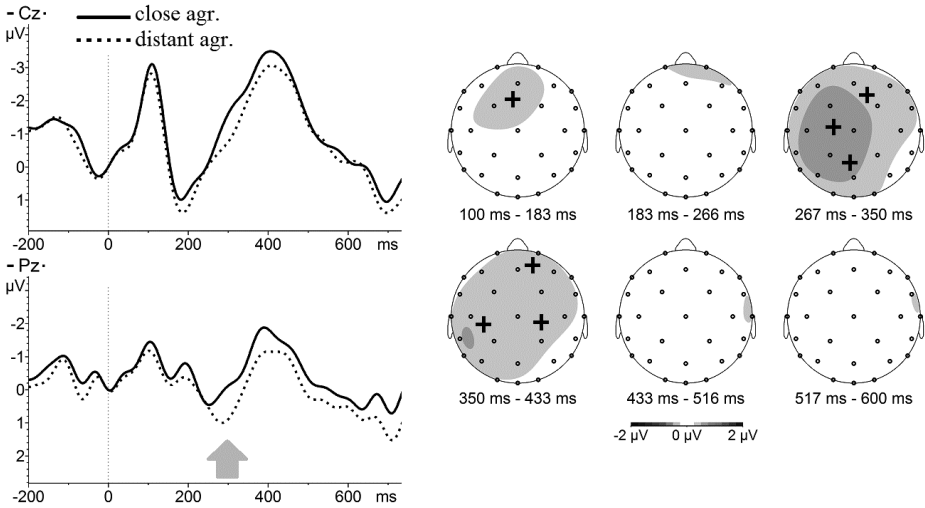(left: reaction times;[1] right: acceptance rate)

Figure 3. ERP waveforms (left) and difference maps (right) for Close and Distant agreement conditions in Experiment 1. The difference map is obtained by simple subtraction of the amplitudes of the ERP waveforms in the Distant condition from the amplitudes in the close condition on all electrodes in given time intervals showing the scalp distribution of the P300 effect.

shown graphically in Figure 4 and clearly show that the results can mainly be explained by the variance in Distance, not Gender, as partial eta squared (η_p^2) suggests.

Furthermore, the P300 component can be directly related to an enhanced memory load for Distant agreement; one has to carry the gender information over a longer time interval.

In the second experiment, with default masculine gender on the participle added, all options available in Croatian were offered to the participants. The third option, however, changed the results in a profound way. Two negative deflections were elicited in this experiment, one around 250ms and distributed frontocentrally, and the second peaking around 450ms with a similar distribution. Both effects are larger for the distant agreement condition, i.e., close agreement and default masculine agreement elicited similar ERP responses. The results are shown graphically in Figure 5 on page 152.

Statistical analysis was performed on the two time intervals, 250–300ms and 470–520ms. Masculine gender on the participle was taken as a third "distance," i.e., the default distance, so a 3×3×3 repeated measure ANOVA was used for the statistical analysis, the three factors being Lobe (frontal, central, parietal), Hemisphere (left, central, right), and Condition (close agreement, distant agreement, default masculine agreement).

In the first interval, a significant main effect of Condition was found ($F[1, 19] = 3.4$, $p = 0.04$, $\eta_p2 = 0.15$). The post-hoc test (Tukey) shows a statistically significant difference between the Close and Distant agreement conditions ($p = 0.03$) while the differences between Close and Distant agreement with a Default agreement condition were not found to be statistically significant (p
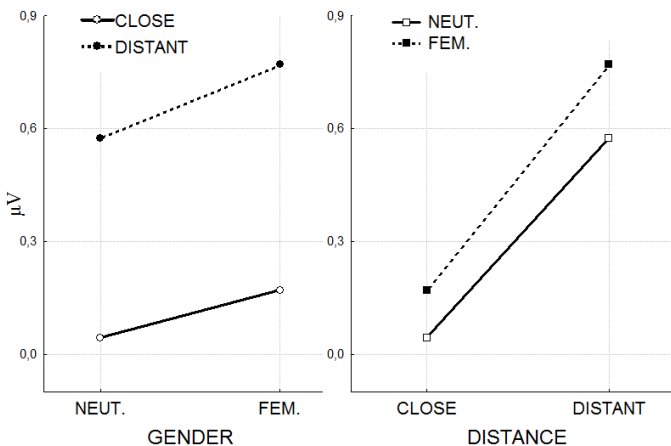

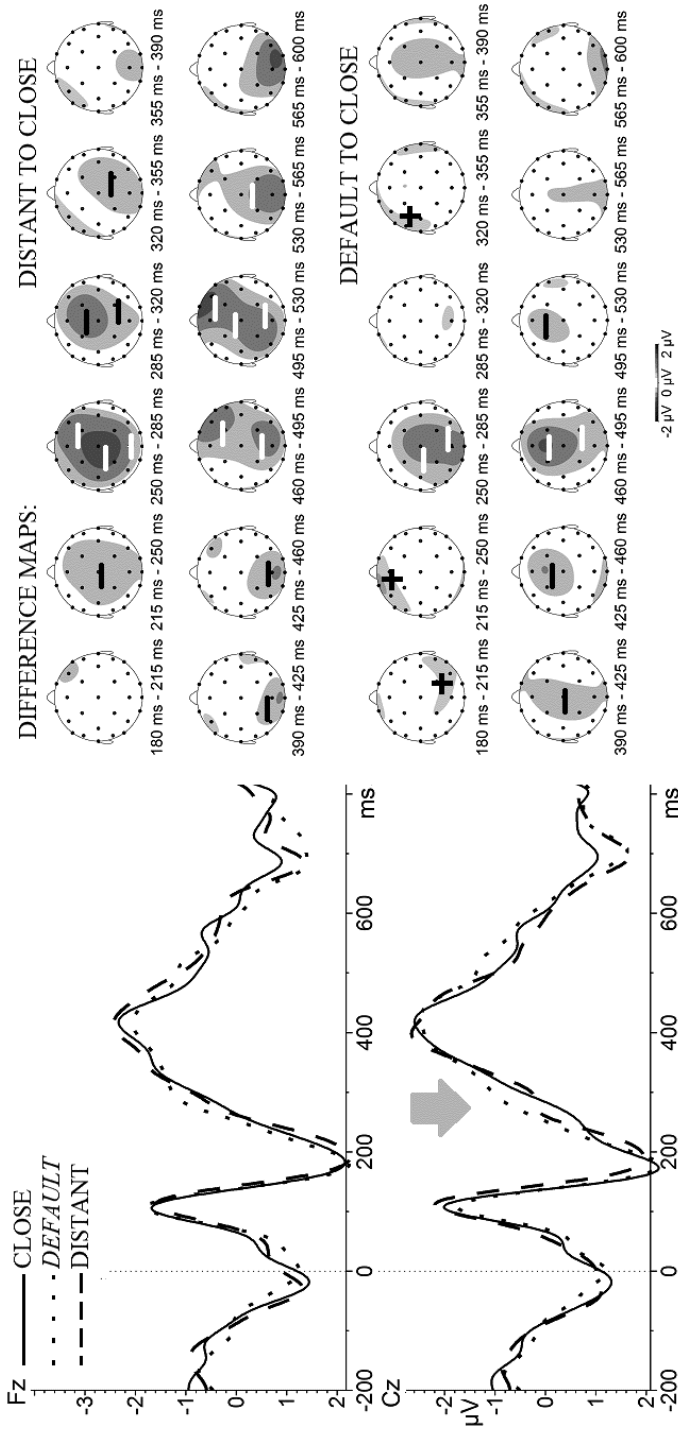**Figure 4.** Weighted means of the averaged amplitudes for Distance and Gender

Figure 5. ERP waveforms (left) and difference maps (right) in the relevant time intervals. The grey arrow indicates the largest effect. Some intervals in the distribution maps are missing; these intervals were omitted in order to conserve the full interval from 170–550ms and to keep the presented intervals as narrow as possible to avoid smearing the effect.

= 0.32 and p = 0.49, respectively). No statistically significant interaction was found. The results of the statistical analysis are shown graphically in Figure 6.

In the later interval (470–520ms), the main effect of Condition was not statistically significant (F[1, 19] = 2.19, p = 0.12). Nor were significant interactions found. The results show overall tendencies similar to the first interval (see Figure 7).
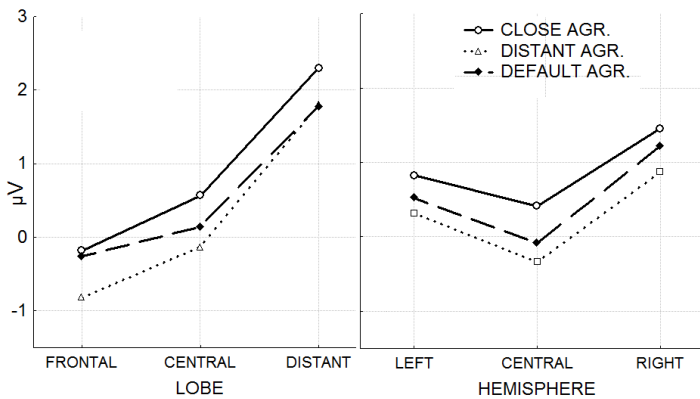

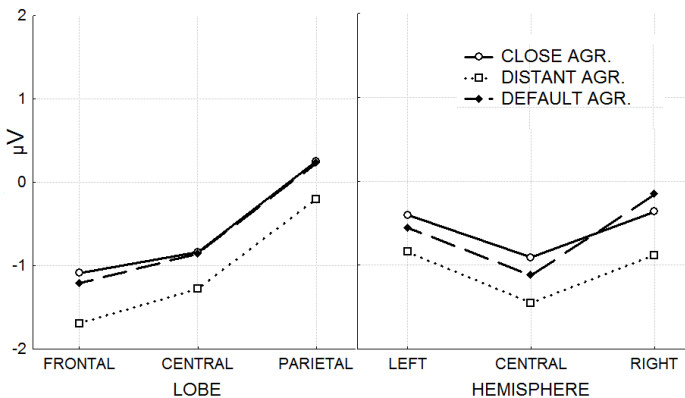**Figure 6.** Weighted means of the averaged amplitudes for the experimental conditions in the 250–300ms interval

Figure 7. Weighted means of the averaged amplitudes for the experimental conditions in the 470–520ms interval

4. Discussion

This ERP study addresses the issue of the neurological processes underlying coordination agreement given mixed-gender conjuncts with native Croatian speakers. It investigates variation in the degree of processing cost for three agreement grammars attested in two behavioral production studies. A factorial design was used to disentangle the three agreement grammars. In Experiment 1, a Closeness factor was manipulated to differentiate hierarchical and linear agreement with the gender on the participle. In Experiment 2, a Default factor was introduced to verify the effects of Def, a language-specific agreement grammar, against the other two grammatical conjunct-agreement dependencies in Croatian.

Under the assumption, predicted by the factorial design, that the same factors should account for any observed variation, the significance of the results reported in this study lies in the variability of the data collected in Experiments 1 and 2, as well as in the variability of results with respect to behavioral, reaction-time, and acceptability judgments, and the neurophysiological EEG components. The variation in data collected with the same conditions in the two experiments (in which one has only one additional condition) indicates that there is more to conjunct agreement, or gender agreement, in terms of top-down processes, speakers' expectations, or task-related processing than predicted by the basic null hypothesis (see section 3.1). As this variability is in part methodological, we contend that future experimental work investigating conjunct agreement should consider a more elaborate psycholinguistic model and accordingly apply better experimental control, e.g., a between-subject design with participants divided according to their prevailing strategies or a simpler 2×2 match-mismatch design.

The Experiment 1 findings show a memory-related effect—a weak P300 is interpreted as the activation of memory during the processing of DCA but not CCA. Considering the linear distance from the participle to each of the two conjuncts, the results suggest that either the gender value of the first conjunct is retained in memory for longer, until it is evaluated against the gender value of the participle (cf., Wagers and McElree 2011 and references therein) or integration of the gender value of the participle has proven to be inconsistent with the built-up context, and hence the distance to reaccess the gender value of the first conjunct is greater than the second. The latter possibility is less convincing, as we would expect additional effects of inconsistency with contextual or language-driven predictions, which is not the case. Therefore, the difference between DCA and CCA is reflected in a higher memory load for DCA, and argue that processing of DCA, but not CCA, is supported by memory. This interpretation confirms (ib) and falsifies (ia), or at least does not provide more subtle evidence for its confirmation.

The Experiment 2 findings show a combination of two language-related effects—negative deflections at 250ms and at 450ms with frontocentral distribution, where again DCA differs from CCA and Def. Interestingly, a comparison of results from the two experiments suggests that the introduction of Default to the experimental paradigm ensures a linguistic context, activating language-specific processes which take over the role of memory. In a more ecologically valid experiment, where all three grammatical options are made available and are equally acceptable statistically and judged with equal speed (see the behavioral results for Experiment 2), processing of DCA is no longer supported by a general cognitive mechanism, such as memory, but is rather being processed by language-related processes (compare reaction time tendencies in Experiment 1, where CCA and DCA were judged with equal speed, and Experiment 2, where gradience in judgment speed was slowest for CCA and fastest for Def). This interpretation falsifies (iia–b), but is in line with gradable preferences observed in production studies and processing models arguing for frequency-based parallelism in production and comprehension (e.g., Levy 2008; MacDonald 2013).

Finally, we elaborate on two observations based on the behavioral findings for Experiment 1. Considered together with ERP findings, behavioral data can provide a valuable insight into the peculiarity of DCA and the importance of an ecologically valid environment. Although reaction-time results did not reach significance, we are pleased to report on them in the context of issues raised in this paper and this volume. Comparing Distant and Close, the latter indicates no preference for either FN or NF, while the former seems to be sensitive to the order of the two genders within the coordination. This tendency, a weak one at best, points to the possibility that Distant, unlike Close, refers to the structure and by extension is sensitive to the internal order of the two conjuncts with respect to gender. Acceptability data show a clear preference for sentences where neuter rather than feminine is found on the participle. To explain this clear gender differentiation, we suggest that in an ecologically free environment, i.e., no Def, neuter assumes the role of default masculine gender. This is not unexpected, as it has been extensively argued that South Slavic neuter is not only the least marked or underspecified gender, but a gender with no gender values (Franks and Willer-Gold 2014). Furthermore, although we could argue that inanimate nouns have no biological gender, they do have a referent, regardless of whether they are morphologically neuter or feminine. However, in a non-restrictive and referent-free dependency environment, such as subject-verb, neuter is the gender found on the verb, never feminine (see Willer-Gold et al., this volume).

An analysis contra Marušič, Nevins, and Badecker (2015), arguing that Def agreement is the result of the same process underlying CCA, could account for the observed clustering of Def and CCA, apart from DCA. Assuming agreement is licensed by probing and finding a local Goal which is a max-

imal projection that can singlehandedly value number and gender features
on the Probe (Franks and Willer-Gold 2014; van Koppen 2005, 2007), the Goal
is Conjunction Phrase or ConjP in the case of Def, the second conjunct or NP2
in the case of CCA, and the Goal is the first, but also hierarchically highest,
conjunct or NP1 in the case of HCA. Therefore, CCA shows the lowest pro-
cessing cost or least working-memory cost out of the three agreement gram-
mars, which is to be expected considering it is structurally small and linearly
closest, and in memory the most recent Goal. Hence for CCA and Def, unlike
HCA, a more economical single-conjunct agreement strategy argues against
markedness-retrieval accounts that take the locus of markedness to be a fea-
ture value (Franck 2011, Wagers and McElree 2011, among others), rather than
opting for a specific agreement strategy.

5. Conclusion

In sum, memory-related processing effects observed for CCA in Experiment 1
have been substituted with language-related processing effects in Experiment
2 in an ecologically valid environment where all three grammatical possibil-
ities were made available. CCA and Def showed the lowest processing costs
and DCA the highest, which is in line with gradable preferences for three
agreement grammars observed in Marušič, Nevins, and Badecker (2015) and
Willer-Gold et al.'s (this volume) production studies. Interparticipant variabil-
ity most probably contributed simply to weak effects being observed.

　Acceptability-judgment results for Experiment 1 show a significant pref-
erence for agreement with the participle in neuter rather than in feminine,
as suggested in an ecologically-free environment where neuter takes the role
of default masculine gender. Reaction-time results weakly indicated that all
local agreement is highly acceptable, while hierarchical agreement is more
acceptable if the first conjunct is neuter, tentatively arguing for the possibility
that DCA, unlike CCA, refers to structural relations, and by extension is sen-
sitive to the internal order of the two conjuncts with respect to Gender.

　We stress that the observed variation in the results collected with off-line
and on-line measures, ERP electrophysiological, and temporal data add an ex-
tra level of analytic complexity for a more complete approach to understand-
ing gender agreement strategies with coordinated subjects, aiming to map
out three agreement grammars which may underlie coordination-participle
agreement dependencies in South Slavic languages. However, importantly,
the observed difference in the results of the same conditions in two experi-
ments (in which only one has an additional condition) indicate that there is
more to it than envisaged in terms of top-down processes, speakers' expecta-
tions, or task-related processing. Consequently, these methodological impli-
cations argue for better experimental control in future conjunct-agreement
experiments.

References

Alsina, Alex and Boban Arsenijević. (2012a) "Hierarchies and competing generalizations in Serbo-Croatian hybrid agreement". Miriam Butt and Tracy Holloway King, eds. *Proceedings of the 12th LFG Conference.* Stanford, CA: CSLI, 301–17.

———. (2012b) "The two faces of agreement". *Language* 88: 369–79.

Aoun, Joseph, Elabbas Benmamoun, and Dominique Sportiche. (1994) "Agreement, word order, and conjunction in some varieties of Arabic". *Linguistic inquiry* 25(2): 195–220.

———. (1999) "Further remarks on first-conjunct agreement". *Linguistic inquiry* 30(4): 669–81.

Barber, Horacio and Manuel Carreiras. (2005) "Grammatical gender and number agreement in Spanish: An ERP Comparison". *Journal of cognitive neuroscience* 17(1): 137–53.

Benmamoun, Elabbas. (1992) *Functional and inflectional morphology: Problems of projection, representation and derivation.* PhD dissertation, University of Southern California, Los Angeles.

Benmamoun, Elabbas, Archna Bhatia, and Maria Polinsky. (2010) "Closest-conjunct agreement in head-final languages". *Linguistic variaition yearbook* 9: 67–88.

Bhatt, Rajesh and Martin Walkow. (2013) "Locating agreement in grammar: An argument from agreement in conjunctions". *Natural language and linguistic theory* 31(4): 951–1013.

Bock, Kathryn and Carol A. Miller. (1991) "Broken agreement". *Cognitive psychology* 23: 45–93.

Bošković, Željko. (2009) "Unifying first- and last-conjunct agreement". *Natural language and linguistic theory* 27: 455–96.

Brouwer, Harm, Hartmund Fitz, and John Hoeks. (2012) "Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension". *Brain research* 1446: 127–43.

Caffarra, Sendy, Niels Janssen, and Horacio Barber. (2014) "Two sides of gender: ERP evidence for the presence of two routes during gender agreement processing". *Neuropsychologia* 63: 124–34.

Caffarra, Sendy, Anna Siyanova-Chanturia, Francesca Pesciarelli, Francesco Vespignani, and Cristina Cacciari. (2015) "Is the noun ending a cue to grammatical gender processing? An ERP study on sentences in Italian". *Psychophysiology* 52: 1019–30.

Corbett, Greville. (1991) *Gender.* Cambridge: Cambridge University Press.

Coulson, Seana, Jonathan W. King, and Marta Kutas. (1998) "Expect the unexpected: Event-related brain response to morphosyntactic violations". *Language and cognitive processes* 13: 21–58.

Deutsch, Avital and Sholomo Bentin. (2001) "Syntactic and semantic factors in processing gender agreement in Hebrew: Evidence from ERPs and eye movements". *Journal of memory and language* 45: 200–24.

Franck, Julie. (2011) "Reaching agreement as a core syntactic process. Commentary of Bock and Middleton 'Reaching Agreement'". *Natural language and linguistic theory* 29(4): 1071–86.

Franck, Julie, Gabriella Vigliocco, and Janet Nicol. (2002) "Subject-verb agreement errors in French and English: The role of syntactic hierarchy". *Language and cognitive processes* 17(4): 371–404.

Franck, Julie, Glenda Lassi, Ulrich H. Frauenfelder, and Luigi Rizzi. (2006) "Agreement and movement: A syntactic analysis of attraction". *Cognition* 101: 173–216.

Franks, Steven and Jana Willer-Gold. (2014) "Agreement strategies with conjoined subjects in Croatian". Sylwester Jaworski and Jacek Witkoś, eds. *New insights into Slavic linguistics*. Frankfurt am Main: Peter Lang, 91–115.

Friederici, Angela D. (1995) "The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data". *Brain and language* 3: 259–81.

Friedmann, David. (1990) "ERF's during continuous recognition memory for words". *Biological psychology* 30: 61–87.

Gillespie, Maureen and Neal J. Pearlmutter. (2013) "Against structural constraints in subject-verb agreement production". *Journal of experimental psychology: Learning, memory, and cognition* 39: 515–28.

———. (2011) "Hierarchy and scope of planning in subject-verb agreement production". *Cognition* 118: 377–97.

Gouvea, Ana C., Colin Phillips, Nina Kazanina, and David Poeppel. (2010) "The linguistic processes underlying the P600". *Language and cognitive processes* 25(2): 149–88.

Gunter, Thomas C., Angelica D. Friederici, and Herbert Schriefers. (2000) "Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction". *Journal of cognitive neuroscience* 12: 556–68.

Hagoort, Peter and Colin Brown. (1999) "Gender electrified: ERP evidence on the syntactic nature of gender processing". *Journal of psycholinguistic research* 28: 715–28.

Johnson, Ray. (1995) "Event-related potential insights into the neurobiology of memory systems". François Boller and Jordan Grafman, eds. *Handbook of neuropsychology*. Vol. 10. Amsterdam: Elsevier Science, 135–63.

Leinonen, Alina, Pauli Brattico, Miika Järvenpää, and Christina M. Krause. (2008) "Event-related potential (ERP) responses to violations of inflectional and derivational rules of Finnish". *Brain research* 1218: 181–93.

Levy, Roger. (2008) "Expectation-based syntactic comprehension". *Cognition* 106: 1126–77.

MacDonald, Maryellen C. (2013) "How language production shapes language form and comprehension". *Frontiers in psychology* 4: 226.

Mancini, Simona, Nicola Molinaro, Luigi Rizzi, and Manuel Carreiras. (2011) "When persons disagree: An ERP study of unagreement in Spanish". *Psychophysiology* 48(10): 1361–71.

Mancini, Simona, Francesca Postiglione, Alessandro Laudanna, and Luigi Rizzi. (2014) "On the person-number distinction: Subject-verb agreement processing in Italian". *Lingua* 146: 28–38.

Marušič, Franc, Andrew Nevins, and Bill Badecker. (2015) "The grammars of conjunction agreement in Slovenian". *Syntax* 18(1): 39–77.

Marušič, Franc, Andrew Nevins, and Amanda Saksida. (2007) "Last-conjunct agreement in Slovenian." Robert Compton, Magda Goledzinowska, and Ulyana Savchenko, eds. *Formal approaches to Slavic linguistics: The Toronto meeting.* Ann Arbor: Michigan Slavic Publications, 210–27.

Molinaro, Nicola, Horacio A. Barber, and Manuel Carreiras. (2011) "Grammatical agreement processing in reading: ERP findings and future directions". *Cortex* 47: 908–30.

Munn, Alan. (1999) "First-conjunct agreement: Against a clausal analysis". *Linguistic inquiry* 30(4): 643–68.

Nevins, Andrew, Brian Dillon, Shiti Malhotra, and Colin Phillips. (2007) "The role of feature-number and feature-type in processing Hindi verb agreement violations". *Brain research* 1164: 81–94.

Oldfield, R. Carolus. (1971) "The assessment and analysis of handedness: The Edinburgh inventory." *Neuropsychologia* 9(1): 97–113.

Osterhout, Lee and Linda A. Mobley. (1995) "Event-related potentials elicited by failure to agree". *Journal of memory and language* 34: 739–73.

Phillips, Colin, Matt Wagers, and Ellen Lau. (2011) "Grammatical illusions and selective fallibility in real-time language comprehension". Jeffrey Runner, ed. *Experiments at the interfaces.* Bingley, UK: Emerald, 153–86. [*Syntax and semantics*, 37.]

Polich, John. (2007) "Updating P300: An integrative theory of P3a and P3b". *Clinical neurophysiology* 118(10): 2128–48.

Puškar, Zorica and Andrew Murphy. (2015) "Closest-conjunct agreement in Serbo-Croatian: A rule-ordering account". Anke Assmann, Sebastian Bank, Doreen Georgi, Timo Klein, Phillipp Weisser, and Eva Zimmermann, eds. *Topics at InfL.* Leipzig: Institut für Linguistik der Universität Leipzig, 441–82. [*Linguistische Arbeitsberichte (LAB)*, 92.] Available at http://www.uni-leipzig.de/~asw/lab/lab92/LAB92_15_puskar_murphy.pdf.

Quiñones, Ileana, Nicola Molinaro, Simona Mancinia, Juan Andrés Hernández-Cabrera, and Manuel Carreiras. (2014) "Where agreement merges with disagreement: fMRI evidence of subject verb integration". *NeuroImage* 88: 188–201.

Santesteban, Mikel, Martin J. Pickering, and Holly P. Branigan. (2013) "The effects of word order on subject-verb and object-verb agreement: Evidence from Basque". *Journal of memory and language* 68(2): 160–79.

Schneider, Walter, Amy Eschman, and Anthony Zuccolotto. (2002) *E-Prime user's guide*. Pittsburgh: Psychology Software Tools, Inc.

Staub, Adrian. (2009) "On the interpretation of the number attraction effect: Response time evidence". *Journal of memory and language* 60: 308–27.

———. (2010) "Response time distributional evidence for distinct varieties of number attraction". *Cognition* 114: 447–54.

van de Meeredonk, Nan, Herman H. J. Kolk, Dorothee J. Chwilla, and Constance Th. W. M. Vissers. (2009) "Monitoring in language perception". *Language and linguistics compass* 3(5): 1211–24.

van Koppen, Marjo. (2005) *One probe—two goals: Aspects of agreement in Dutch dialects*. PhD dissertation, Leiden University, Utrecht.

———. (2007) "Agreement with coordinated subjects: A comparative perspective". *Linguistic variation yearbook* 7: 121–61.

Walkow, Martin. (to appear) "When can you agree with a closest conjunct?" *Proceedings of the 31st West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Press.

Wagers, Matthew and Brian McElree (2011) "Memory for linguistic features: Evidence from the dynamics of agreement". Unpublished ms., University of California, Santa Cruz and New York University.

Willer-Gold, Jana et al. (this volume) "Conjunct agreement and gender in South Slavic: From theory to experiments to theory". *Journal of Slavic linguistics* 24(1).

Marijan Palmović
University of Zagreb
Laboratory for Psycholinguistic Research
Zvonimirova 8
10000 Zagreb
Croatia
marijan.palmovic@erf.hr

Jana Willer-Gold
University College London
Chandler House
2 Wakefield Street
London, WC1N 1PF
United Kingdom
j.willer-gold@ucl.ac.uk