UDK 811.163.42'367.4
811.163.42'242
Izvorni znanstveni članak
Prihvaćeno za tisak: 3. lipnja 2019.
https://doi.org/10.22210/suvlin.2019.087.07

Frane Malenica[1], *Marijana Kresić Vukosav*[1], *Dario Lečić*[2]
[1]*University of Zadar,* [2] *Croatian Science Foundation*
fmalenica29@gmail.com, mkresic@unizd.hr, dario.lecic@gmail.com

Disagreements over agreement – a comparison of naïve and expert intuitions*

This paper presents the results of a study comparing native speaker intuitions on sentences with Closest Conjunct Agreement (CCA) obtained from linguists and non–linguists and it functions as a continuation of a larger study of agreement patterns in the South Slavic languages. In this particular research, we used a sentence–picture matching experiment with a 0–100 scale, which the participants used to indicate the acceptability of a particular sentence with a CCA pattern. Our participants were two groups of native speakers of Croatian (of the Shtokavian dialect) with different levels of linguistic education: the non–linguists (N=30) were comprised of BA–level students of various subjects (excluding Croatian, Linguistics and Psychology), while the group of linguists (N = 30) was comprised of Croatian teachers and holders of PhD in Linguistics employed at elementary schools, high schools and universities. The difference between the results obtained from linguists and non–linguists as research participants has been a matter of scholarly debate for the past several decades (cf. Dąbrowska 2008, 2010) and our aim is to contribute to this discussion by providing data related to CCA. The results of the study indicate that there is a statistically significant difference in linguistic intuitions between the two groups and there are valid reasons to attribute this difference to the level of linguistic training received. This difference is most clearly visible in the fact that linguists provided lower ratings on a general scale, regardless of the type of sentence and CCA pattern. A small–sized effect of age was also established.

1. Introduction

In syntactic theory, conjunct agreement (CA) is usually referred to as the phenomenon that a conjoined, preverbal subject triggers agreement of the verb with one of its conjuncts in terms of gender and number. Closest conjunct agreement (CCA), i.e. verbal agreement with the conjunct closest to the verb, is the prevailing agreement pattern in the South Slavic languages Slovenian, Bosnian/Croatian/

* The authors would like to thank Jana Willer–Gold and two anonymous reviewers for their helpful suggestions on earlier drafts of this paper.

Serbian, and can occur either pre–verbally (1) or post–verbally (2), as the following examples show:

(1) Ravnala i olovke su pronađena/pronađene/pronađeni.
ruler.N.PL and pencil.F.PL AUX.PL found.N.PL/F.PL/M.PL

(2) Pronađena/pronađene/pronađeni su ravnala i olovke.
found.N.PL/F.PL/M.PL AUX.PL ruler.N.PL and pencil.F.PL
'Rulers and pencils have been found.'

(Arsenijević et al. 2019: 2)

The research presented here draws on the results of an experimental study (Arsenijević et al. 2019) which investigated CCA in South Slavic on the basis of a forced–choice picture matching study, testing the hypothesis whether CCA can be analysed as a result of "reduced clausal conjunction, and to simply display the agreement of the verb with a non–conjoined subject in the clause whose content survives ellipsis [...], i.e. whether closest conjunct agreement in these varieties may indeed be analyzed as entirely derived from conjunction reduction" (ibid.: 18). The main result of the study conducted by Arsenijević et al. (2019) is that CCA does not emerge exclusively from an elided biclausal structure, i.e. conjunction reduction, which is also the conclusion reached by previous research on CCA in South Slavic languages (Marušić, Nevins and Badecker 2015, Willer–Gold et al. 2016).

The research in this paper presents the results of a follow–up experiment conducted using the same experimental stimuli as Arsenijević et al. (2019) but with a different set of participants. The experiments conducted by Arsenijević at el. (2019) used undergraduate level university students who could be regarded as linguistically naïve (in the sense of Dąbrowska 2008 & 20101), while the experiments conducted in this paper involved a group of participants with size and dialectal background comparable to one of the 7 sites in Arsenijević et al. (2019) but with significantly higher levels of education and linguistic training – Croatian language teachers and linguists, i.e. persons involved in the primary, secondary and tertiary level of education. The aim of this paper is to provide additional insights on collecting empirical linguistic data (cf. Spencer 1973, Ferreira 2005, and others). The results from Dąbrowska (2008 with non–linguists as participants and 2010 with linguists as participants) indicate that linguists and non–linguists provide systematically different judgements of the same sentences and one of the claims in Dąbrowska (2010) is that the difference in ratings between linguists and non–linguists could

1 For the purpose of this paper, we regard as naïve participants all participants who had little or no explicit university–level linguistic training for the language in question. It should be noted that the term 'naïve participant' is used here in relation to the term 'expert participant' and does not form a strictly defined class of possible participants. As Dąbrowska (2010: 11) notes, an average university student might be less 'naïve' than an average native speaker, but is certainly more 'naïve' than a linguist with an MA or a PhD, which is why we believe university students still represent a valid baseline for comparison with the 'expert' group of participants.

be explained by the amount of exposure to sentences with long distance dependencies (the object of study in both Dąbrowska 2008 and 2010). The study presented in this paper aims to provide a comparison similar to that in (2010) and investigate whether two sets of participants differing only in the level of "expertise" could provide statistically different ratings of stimuli involving CCA.

The paper is structured as follows. In the following section 2, we discuss the issue of difference between data collected from naïve and expert participants, with a particular focus on differences between intuition studies and experimental studies. Section 3 briefly presents the aims and the motivation for this paper and its research hypotheses. In section 4, we provide the background for the experiment with naïve participants (Arsenijević et al. 2019), provide the methodology for the experiment with both naïve and expert participants, compare the results between the two groups, and discuss the possible factors behind our results. We finish with a conclusion in section 5.

2. Naïve vs. expert intuitions – is there a difference?

As Trask (1999: 88) pointed out, every speaker has intuitions about their own language, "about what is normal, acceptable, unusual, strange or impossible, or about what a given form means and when we might use it, if at all. The issue is how (if indeed any at all) much trust we should place in speakers' intuitions in compiling our descriptions of language.

If linguists rely exclusively on intuition data and disregard other types of data (usage data, experimental data, etc.), they may end up with dubious results. Experience tells us that in actual speech speakers rarely utter sentences with prototypical argument structure (i.e. having a subject, a verb and an object), even though such are the sentences that traditional grammars generate. Several previous works have demonstrated that items that linguists have labelled as 'ungrammatical' not only occur in actual language, but are also accepted by speakers, and vice versa. For instance, in her analysis of questions with long–distance dependencies (LDDs) Dąbrowska (2008) demonstrated that the sentences that are often used in generativist writing to support one rule about LDDs or another rarely actually appear in the language as such. In short, generative linguists claim that any number of clauses can intervene between the WHword and the main clause without having any effect on their comprehensibility or acceptability. Chomsky supported this claim by using the following example: "*Who did Mary hope that Tom would tell Bill that he should visit?"* Dąbrowska, on the other hand, shows that such constructed sentences differ to a great extent from **real–life** sentences (Dąbrowska's term), of which more than 70% conform to one of the following patterns: "*WH do you think S–GAP?"* or "*WH did NP say S–GAP?"* (2008: 392). In other words, unlike Chomsky's **constructed example,** the majority of LDD questions attested in real life only have one or two intervening clauses, contain the verb *think* or *say* and refer to the second person. This

finding made the author raise concerns that the informal data collection methods that were typical for syntactic theory of the past have led to unsound theories.

Sampson (2007) believes there is no reason to assume that realities of a speaker's language will be reflected in their intuitive grammaticality judgments. Divjak (2016: 21) argues along the same lines, that "judgments may reflect properties of the rater rather than properties of the grammar." The author came to this conclusion after *rater generosity* came out of her analysis as one of the variables with the strongest effect. This variable means that participants who gave filler sentences high scores were more likely to give trigger sentences high scores as well.

A number of authors have shown that repetition of items may have an effect on acceptability of those items. Surprisingly, however, opposing effects have arisen from analyses by different authors. For instance, Luka & Barsalou (2005) find that being exposed to structures in an initial reading task increases the acceptability of those same structures in a subsequent rating task (the authors call this effect *habituation*). On the other hand, Nagata's (1988) findings show that repeated exposure to sentences can make the judgments more stringent (i.e. the more the subjects look at a sentence, the more problematic issues they find with it). Since both sets of authors found that grammaticality judgments were easily influenced by repetition and other variables, such as embedded context, this led them to conclude that linguistic intuitions underlying grammaticality judgments are not absolute but rather relative. The comparison of results between the two groups in our study seem to confirm this view. With regards to *satiation effects* (i.e. all sentences starting to look alike), Snyder (2000) shows that they only appear in certain types of sentences and dismisses this effect as a property of the judgment process which should not prevent the linguist from using acceptability as an explanatory tool. Various methods can be used to minimise such unwanted effects of order (such as counterbalancing the test material by adding unrelated filler sentences, etc.). All in all, as argued by Cowart (1997: 5), "the utility of introspective judgments in furthering research far outweighs any limitations ascribed to them." (Schütze 2011: 216) argues that acceptability judgments are "themselves data about human behaviour and cognition that need to be accounted for; they are not intrinsically less informative than, say, reaction time measures – many linguists would argue that they are more informative."

Another extremely problematic and widely debated issue with regards to intuition studies, which is also the main focus of the present article, is whether trained linguists should participate in any kind of intuition studies as respondents. In the works of linguists of the past, they traditionally relied only on their own judgments, or judgments from their colleagues, also linguists, as it was believed that those who know more about a topic provide more reliable judgments.[2] However, as

2 Valian (quoted in Schütze (2011: 212)) uses the analogy of wine tasting to argue in favour of using 'expert' judgments in psycholinguistic experiments. Namely, this skill relies on the acquired ability to detect subtle distinctions that inexperienced wine drinkers simply cannot make. He believes the same applies to judgments of linguists on linguistic matters.

Labov (1978: 199) argues, "linguists cannot continue to produce theory and data at the same time." A theory of language derived only from the data provided by the linguist necessarily describes only their own idiolect rather than the whole language. Both Gries (2002) and Dąbrowska (2010) have shown that the judgments of linguists and the judgments of non–linguists diverge to a great extent and therefore linguists' judgments should not be considered as being representative of the whole population. In Gries' (2002) analysis of the English genitive alternation, the linguists failed to predict the influence of several variables on the choice of 's–genitive versus of–genitive that the data from corpora and naïve speakers highlighted as significant. One finding that has arisen from Dąbrowska's (2010) experiments is that linguists tend to give more categorical judgments, whereas nonlinguists tend to use the full rating scale. Furthermore, linguists seem to be more lenient when evaluating less prototypical linguistic units (especially ungrammatical ones) as they encounter them more often in their work. Finally, Gibson & Fedorenko (2013) argue that the linguists' theoretical biases (e.g. generative versus cognitive) could influence their judgments.

However, as Culbertson & Gross (2009: 725) point out, linguists' judgments are also more reliable, i.e. "among linguists there is a greater tendency for sentences they judge *acceptable* to be *grammatical*." However, the authors also argue that this reliability should not be identified with consistency, which refers to the responses staying constant across different elicitations, regardless of accuracy. Snow & Meijer (1977) also demonstrated through experiments that linguists showed greater agreement with one another than non–linguists. They explain this in the following manner: "either linguists have learned to ignore minor irrelevant differences among sentences (such as their semantic plausibility) or they have learned to apply their theory to unclear cases." In opposition to all the evidence above, Sprouse & Almeida's (2012) survey discovered that the two sets of judgments differed by only 2%.

Even though a number of works have discovered a difference between the judgments of linguists and those of non–linguists, none of them have still managed to establish the exact cause of the differences. For instance, the main source of respondents' reliability, according to Culbertson and Gross's (2009) experiment, is not expertise in syntax but rather having task–specific knowledge. In other words, subjects who have previously come across similar types of experiments (e.g. in psychology or other disciplines of cognitive science) were much more successful than subjects who have not. Knowledge of linguistics was irrelevant in this case.

3. Motivation for this paper

Our motivation for this paper was to investigate whether a difference between judgments of linguistic experts and naïve participants could be found, which was pursued by replicating a previously conducted experiment on morphosyntactic

agreement with conjunct and single NP subjects. Thus, our study is in line with the experiments such as those conducted by Dąbrowska (2008, 2010). Our aim is to investigate whether there is a correlation between the participants' level of expertise and their linguistic intuitions. If acquiring a first language was something all human beings achieve effortlessly regardless of factors such as age, IQ level, education, social background and the like, we would expect the level of expertise to have no effect on acceptability ratings and ultimately, it would follow that no statistically significant difference should be observable between the grammaticality/acceptability judgements of individuals belonging to groups differing in the mentioned factors. However, if the two groups which differ only in terms of length of exposure to language and linguistic training provide statistically different results, this raises the question whether the results obtained from either group are generalisable to a wider population.[3] More importantly, such findings ultimately cast doubt on the validity of conducting studies using only participants from a single group (expert or naïve participants respectively) and potentially erode the applicability of acceptability judgements for linguistic studies. In order to achieve our goal, we intend to answer the following research questions:

1. Does the level of linguistic exposure have an effect on linguistic intuitions, i.e. is there a difference in ratings between expert and naïve participants?
2. If such a difference exists, is the difference noticeable for particular conditions and/or items or is the difference in ratings stable across all items/conditions?
3. Is there a correlation between age and acceptability rating for the critical conditions in the group of expert participants?

4. Comparing naïve and expert data

Background. The experiment with naïve participants was conducted as a part of the project Coordinated Research in the Experimental Morphosyntax of South Slavic Languages (EMSS) (cf. Arsenijević et al. 2019) and it focuses on postverbal mixed gender conjunction that is a subject of a collective predicate with which it agrees in CCA. The conditions in this experiment were designed according to the factors *predicate type* [Collective, Non–Collective] and *category of subject* [&P, NP_{PL}]. It is designed as a sentence–picture matching experiment in which participants' task is to evaluate to what degree a sentence in one of the four conditions *&P–Col-*

3 An anonymous reviewer disagrees with this claim and argues that relative ratings should be analysed instead (i.e. whether a difference between the conditions in the study is present among both groups). While we somewhat agree with this view, the point we are trying to make in our paper is that there is a significant difference between the two populations, not whether the two groups display different effects. While the latter finding would be fascinating from the perspective of our paper, it would also cast serious doubt on the validity of both the original study and the replication study conducted here. This, however, is not the case in our study, as we shall see in §4.

lective (= &P subject + collective predicate), *NP–Collective* (= NP_{PL} subject + collective predicate), *&P–non–Collective* (= &P subject + non–collective predicate), and *NP–non–Collective* (= NP_{PL} subject + non–collective predicate)[4] matches a given picture. The experiment was designed and administered using an online software, Ibex Farm (Drummond 2011), which automatically recorded the responses and exported them afterwards for statistical analysis according to the degree of value appointed to the four conditions. Apart from the experiment itself, a written consent form and comprehensive biographic questionnaire were administered off–line at the beginning of the experiment by the experimenter.

The experimental design is a variation of a standard sentence–picture matching experiment, using a sentence–picture matching task (0–100% scale) with on–line recording of responses. Participants were tested individually or in parallel. The goal of the experiment was to disproof the theoretical claim that postverbal CCA is a result of clausal ellipsis (Aoun, Benmamoun & Sportiche 1994), and to confirm the claim of most current accounts that assume that CCA arises in the context of phrasal coordination (e.g. Benmamoun, Bhatia & Polinsky 2010, Marušić, Nevins and Badecker 2015, Bošković 2009, Franks & Willer–Gold 2014, Puškar & Murphy 2015, among others). Collective predicates have been chosen as they show partial agreement with one conjunct, but semantically they must be predicated of the entire conjunct (Marušić, Nevins and Badecker 2015). Thus, the interpretation for (3b) is 'vozila su se smjenjivala s kočijama' (*'Vehicles crossed with carriages'*), and not 'vozila su se smjenjivala s vozilima' (*'Vehicles crossed with vehicles'*).

The hypothesis was that if CCA is derived by clausal ellipsis (3a), then CCA sentences cannot be matched with pictures that are incompatible with clausal ellipsis, depicting (3b).

(3a) Na mostu [su se smjenjivala vozila] i [su se smjenjivale kočije]. – clausal ellipsis
(*'On the bridge, crossed.N vehicles.N and crossed carriages.F'*)

(3b) Na mostu su se smjenjivala [vozila i kočije]. – nominal ellipsis
(*'On the bridge, crossed.N vehicles.N and carriages.F'*)

The results of this experiment showed that CCA was not a result of ellipsis (Arsenijević et al. 2019), according to the answers of the participants (N = 30), who were all linguistically naïve (undergraduate students). We decided to replicate the experiment with expert participants, Croatian language teachers (MA or higher) or speakers with a PhD in Linguistics.

Participants. The data gathered for the purpose of this study were collected during two experimental sessions – the data from the group of naïve participants were gathered in May 2016 as part of the EMSS project (Arsenijević et al. 2019),

4 The complete list of stimuli in these four conditions is provided in the Appendix of this paper.

while the data from expert participants were collected from February to July 2017 by the authors of this paper at the University of Zadar. Both experiments used groups of the same size (N = 30), and were structured equally with respect to the participants' sex (M = 12, F = 18). The naïve participants were all undergraduate level university students at the University of Zadar, with an average age of 20 years5. In order to ensure that they were truly naïve participants, the experiment included only those students who had not taken Croatian, Linguistics or Psychology as their major field of study, i.e. assumedly have not been exposed to formal instruction and reflexion on matters of morphosyntactic agreement in South Slavic and had not participated in any of the previous EMSS experiments. The group of expert participants included people who had obtained either an MA or PhD in Croatian language and literature (or an equivalent degree) or a PhD in linguistics, and their age ranged from 27 to 62 (mean = 39.67, SD = 9.16). None of the expert participants were involved with previous stages of the EMSS project in any capacity nor, to our knowledge, did they have any scientific interest in the experiment. All naïve participants were born in the Zadar County area and spent most or all of their youth in this region; 23 expert participants were born in the Zadar County area, while the remaining 7 participants were born in the neighbouring counties. All participants, both naïve and expert, were native speakers of the Shtokavian dialect spoken in the area.

Materials and Method. The sentence–picture matching experiment was created and conducted via the IBEX Farm platform for experiments (Drummond 2011). All experimental sessions were conducted on personal computers, in the presence of one of three researchers, in quiet computer rooms or offices with minimal external interfering factors, such as noise or visual distractions. The group size for naïve participants ranged from 2 to 7 participants, while sessions for expert participants included only 1 participant per session, as their schedules made it difficult to arrange larger groups.

For every stimulus, participants were shown an image and a sentence below it. They were instructed to use a scale to rate the extent to which the sentence provided a good description of the image. The scale was divided into coloured areas, encompassing different shades of green (more acceptable) to red (less acceptable) and thereby expressing different degrees of acceptability, which were subsequently turned into numerical values ranging from 0–100.

The experiment had a 2x2 design with two levels with two factors each – type of predicate (collective or non–collective) and type of subject (conjunct, or &P, and single NP), which yielded a total of 4 conditions and 32 stimuli (8 stimuli per condition). The factorial design of the experiment can be seen in Table 1.

5 Since the participants were university students 3 years apart at most, their age was not taken into account as a potential variable and was not recorded and factored in.

	NP	&P
collective	U bitci su se sudarala koplja. *'In the battle collided.N spears.N'*	U bitci su se sudarala koplja i sablje. *'In the battle collided.N spears.N and swords.F'*
non–collective	U dućanu su izložena ogledala. *'In shop were displayed.N mirrors.N'*	U dućanu su izložena ogledala i lampe. *'In shop were displayed.N mirrors.N and lamps.F'*

Table 1. A 2x2 design of the experiment with all 4 conditions

There were two types of fillers – one half of them (i.e. 16) involved a sentence which matched the picture, while the remaining half involved sentences which were ungrammatical due to mismatch in terms of number of objects, depiction of the NP in the subject or object position, or in terms of the adverbial used. Before starting the experiment, participants had to go through 6 practice items in order to familiarise themselves with the task and the interface for the experiment (no data was gathered from practice items). The items were randomized so that every item could appear only once and every participant was exposed to every stimulus only once. Overall, every participant had to go through 64 stimuli (32 fillers and 32 experimental items), which gave a total of 1920 data points (3840 with fillers included) from 60 participants (30 expert and 30 naïve).

Results. The data collected by the procedure described in the previous section were analysed using the R statistical software package (R Core Team 2015). The results obtained from IBEX Farm were converted into tabular format using standard spreadsheet software. Mean and SD values were calculated for both sets of participants (Table 2), and these values were used to remove all the outliers from further analysis (all observations over 2 SDs away from the mean for that particular condition), which resulted in removal of 6.5% of the data points (125/1920). The data were then aggregated for each participant across every condition so that the participant's value for each condition represents its mean score for that particular condition, and these results were plotted onto boxplots in Figure 1. A two–way ANOVA was conducted on these aggregated values for both groups of participants with Predicate type (Collective/Non–Collective) and Subject type (NP/&P) as the predictor variables. The results showed a main effect of Predicate type both for naïve ($F(1,116) = 12.14$, $p < 0.001$) and expert participants ($F(1,116)=20.92$, $p < 0.001$) and no effect of Subject type for both groups. As expected, these results are in line with the results in Arsenijević et al. (2019), who also obtained the same effects.

The first step in the analysis was to check whether there was any difference between the ratings of expert and naïve participants using the pertinent data in Table 2 and Figure 1. Several points are noticeable: i) both the mean and the median values are higher among the naïve participants across all four conditions, which might be taken as an indication of generally more conservative ratings by expert participants; ii) the data collected from the expert participants have a higher degree of dispersion in comparison to naïve participant data across all four critical conditions, as visible from SD values in Table 2 and the boxplots in Fig. 1; and iii) the values for the stimuli with non–collective predicates are rated more highly across both groups.[6]

	Expert		Naïve	
	mean	SD	mean	SD
Collective &P	67.6375	34.42896	81.475	25.92844
Collective NP	69.2875	34.5036	84.8125	24.31682
Non–collective &P	78.15	31.0496	88.50833	22.33934
Non–collective NP	73.57083	32.8494	89.3	20.45798

Table 2. Acceptability ratings for all items

A Shapiro–Wilk normality test was conducted on both the main data set and the two subsets of participants, showing that the data were not distributed normally in the main data set ($W = 0.8753$, $p < 0.001$), nor in the expert ($W = 0.94069$, $p < 0.001$) or the naïve ($W = 0.8997$, $p < 0.001$) subset alone, which is why non–parametric tests were used further in the analysis. After running the Wilcoxon rank sum (Wilcoxon–Mann Whitney) test[7] for all four conditions, observation i) was shown to be statistically quite significant, across all four conditions – *Collective &P* ($W = 160$, $p < 0.001$), *Collective NP* ($W = 179$, $p < 0.001$), *Non–collective &P* ($W = 265$, $p < 0.01$), and *Non–collective NP* ($W = 200.5$, $p < 0.001$), which clearly shows that there is a highly significant difference between the two groups of participants.

6 Since nothing hinges on observation iii), we report no additional information about it.

7 We decided to treat the acceptability ratings as an interval variable here, which is why the Wilcoxon–Mann Whitney test was used.

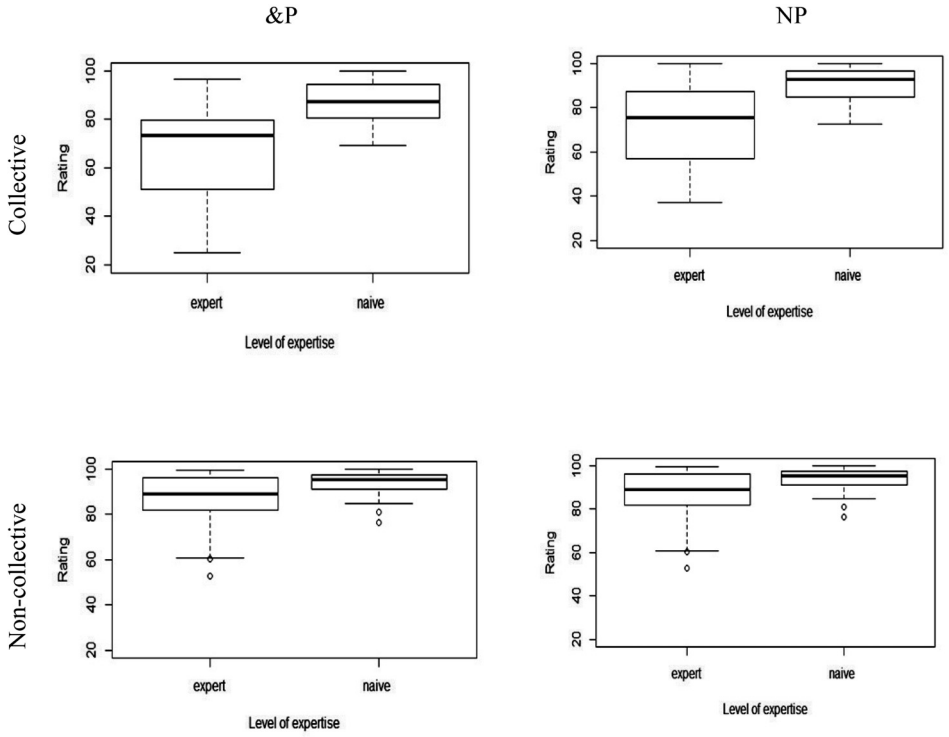


Figure 1. Results of expert and naïve participants across all four conditions

It is worth noting that, while there is a statistically significant difference between expert and naïve participants across all four conditions, the basic assumption of the experiment conducted by Arsenijević et al. (2019) holds true for both sets of participants. That is, there is no statistically significant difference between sentences with coordinated subjects and single NP subjects with collective predicates among naïve participants ($W = 346$, $p > 0.05$) and expert participants ($W = 360$, $p > 0.05$), and the same effect (or lack thereof) is observed with non–collective predicates among naïve ($W = 477$, $p > 0.05$) and expert participants ($W = 541$, $p > 0.05$) alike.

To check whether the data collected from the expert participants really have a higher degree of dispersion, we looked at the differences in the interquartile ranges between the groups (Table 3) and conducted the Levene's test. It confirms what can be seen from data in Table 2 and Table 3, i.e. that observation ii) is statistically significant across all four conditions – *Collective &P* (Df = 1, F value = 7.8931, Pr(>F) < 0.01), *Collective NP* (Df = 1, F value = 19.177, Pr(>F) < 0.001), Non–collective &P (Df = 1, F value = 8.3049, Pr(>F) < 0.01), and *Non–collective NP* (Df = 1, F value = 14.815, Pr(>F) < 0.001).

	Expert	Naïve
	IQR	IQR
Collective &P	26.3125	13.3125
Collective NP	29.71875	11.3676
Non–collective &P	13.94375	5.91518
Non–collective NP	19.34375	9.22768

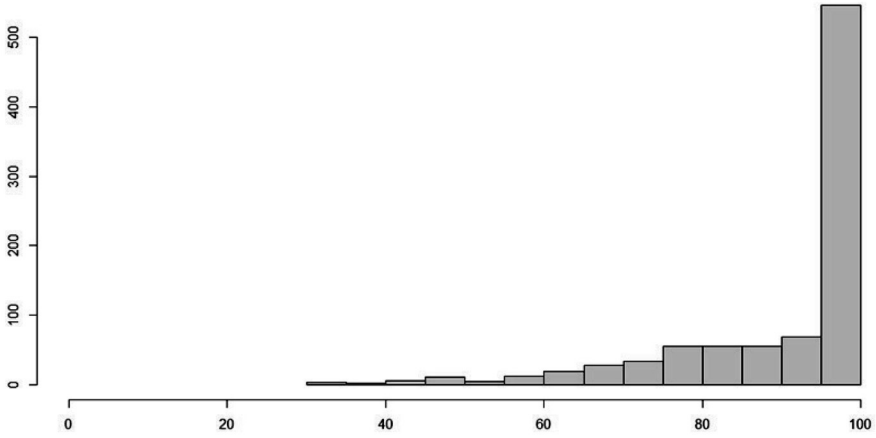
Table 3. IQRs for all conditions across both groups

An anonymous reviewer has raised the question of whether the lower variability among the ratings of naïve participants is a result of ceiling effects or whether experts used a wider array of ratings. The histograms in Figure 3. seem to indicate that these two possibilities are not mutually exclusive. The naïve participant data clearly show ceiling effects as a vast majority of their ratings are clustered around the 100% rating and very few data points are located at the lower end of the scale.[8] The expert data, while also clustered around the higher end of the scale, are more dispersed across the whole spectrum and the highest ratings (i.e. 100%) are not as predominant as for the naïve participants. We consider this an additional indication of wider dispersion of answers among the expert participants.

The difference in variance is perhaps even more visible in Fig. 3, which clearly shows the difference between the two populations. While some degree of dispersion could be explained by certain stimulus–related problems, as is probably the case with stimulus no. 19, it is fairly clear from the overall picture that the higher degree of variance in acceptability ratings among expert participants is systematic and spread across all four conditions. This belief is corroborated by Fisher's F–test which shows that the variances between the two groups are statistically significant across all four conditions – *Collective &P* ($F = 4.1651$, num df = 29, denom df = 29, $p < 0.001$), *Collective NP* ($F = 3.598$, num df = 29, denom df = 29, $p < 0.001$), *Non–collective &P* ($F = 4.7825$, num df = 29, denom df = 29, $p < 0.001$), and *Non–collective NP* ($F = 4.089$, num df = 29, denom df = 29, $p < 0.001$). While this could hypothetically be attributed to a number of factors, it is our belief that the difference in acceptability ratings would be best explained by the difference in the level of "expertise" between the two participant pools.

8 While the histogram in Figure 3 shows no ratings below 30% from naïve participants, this does not mean the naïve participants did not provide such low ratings at all. However, as most of the naïve participants' values were clustered around 100%, the values below 30% were marked as outliers and removed from further analysis, as described earlier in this section.

Naïve



Expert

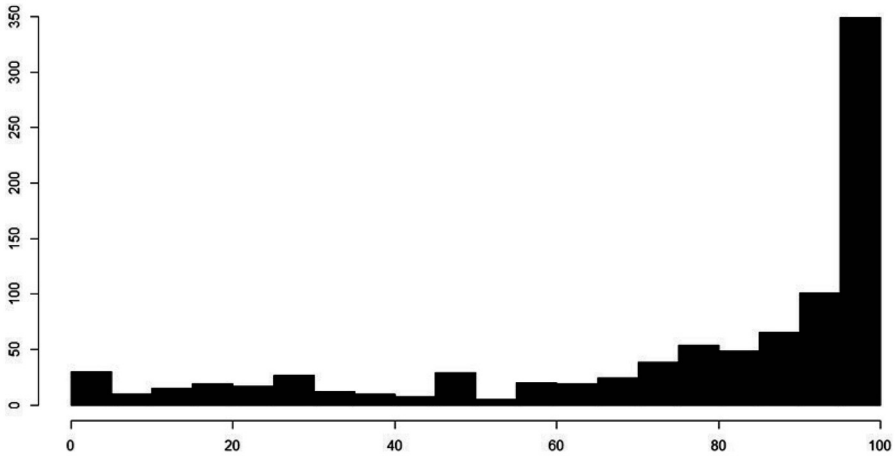


Figure 2. Distribution of ratings among naïve and expert participants

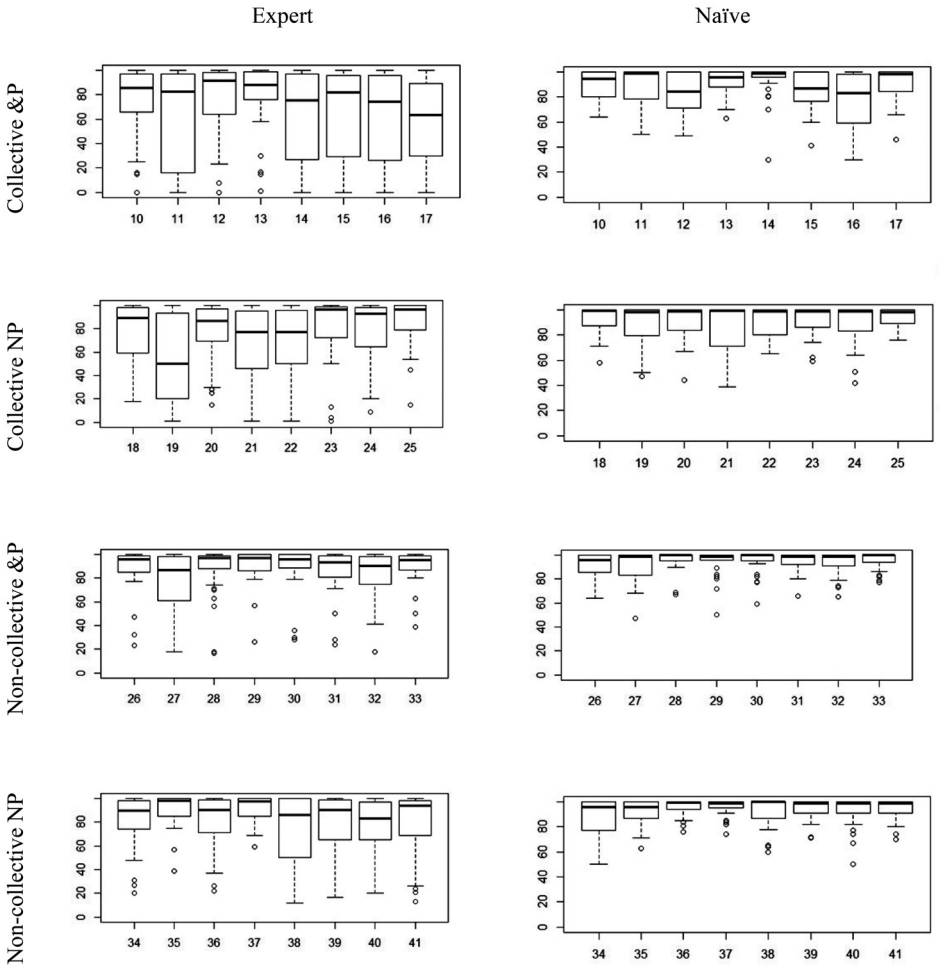


Figure 3. Results of expert and naïve participants across all four conditions

With regards to RQ3, we created a linear regression model to test the effects that age might have on acceptability ratings in the group of expert participants. As can be observed from Fig. 4, there appears to be an inverse correlation between age and acceptability rating across all four conditions in that the older participants have a tendency of giving lower overall ratings, with the slope having a steady downward trajectory in all conditions except for *Non–collective NP*. This idea is supported by the data in Table 3, which shows that age is a statistically significant predictor of acceptability ratings for all conditions except *Non–collective NP*, which fails to reach statistical significance.[9]

9 For the purpose of better visualization, the plots in Fig. 2 are based on mean scores for every participant across the four conditions instead of raw data, with a dot on the plot representing the mean scores for every participant. The calculations in Table 2 are based on raw data.

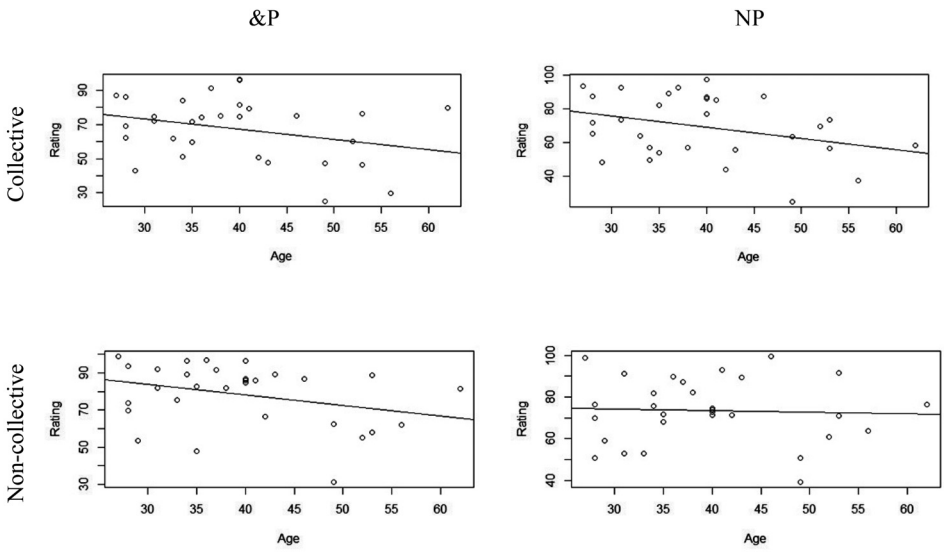


Figure 4. Linear models for all four experimental conditions based on mean scores

		Estimate	Std. Error	t value	Pr(>|t|)
Collective &P	(Intercept)	91.4939	9.7745	9.36	<2e–16
	AGE	–0.6014	0.2401	–2.505	0.0129
		Multiple R–squared: 0.02569, Adjusted R–squared: 0.02159 F–statistic: 6.274 on 1 and 238 DF, p–value: 0.01292			
Collective NP	(Intercept)	95.7900	9.7661	9.8080	< 2e–16
	AGE	–0.6681	0.2399	–2.7850	0.0058
		Multiple R–squared: 0.03156, Adjusted R–squared: 0.02749 F–statistic: 7.757 on 1 and 238 DF, p–value: 0.005782			
Non–collective &P	(Intercept)	100.567	8.8051	11.421	< 2e–16
	AGE	–0.5651	0.2163	–2.613	0.00955
		Multiple R–squared: 0.02569, Adjusted R–squared: 0.02159 F–statistic: 6.274 on 1 and 238 DF, p–value: 0.01292			
Non–collective NP	(Intercept)	76.79336	9.44576	8.13	2.36e–14
	AGE	–0.08124	0.23203	–0.35	0.727
		Multiple R–squared: 0.0005148, Adjusted R–squared: –0.003685 F–statistic: 0.1226 on 1 and 238 DF, p–value: 0.7266			

Table 4. Summaries of linear models across all for conditions

Discussion. The data presented in this paper indicate that the acceptability of sentences, as rated by the participants in the replicated experiment, can be attributed to the difference in level of linguistic 'naïveness' or 'expertness'. However, the results seem to go in the opposite direction of those reported by Dąbrowska (2008 and 2010) in the sense that her expert participants provided higher ratings for all grammatical sentences and lower ratings for all ungrammatical sentences in comparison to naïve participants, whereas our group of expert participants provided lower ratings across all four conditions. However, we believe this difference should be attributed to the types of constructions used in both studies. Dąbrowska's participants were asked to rate different constructions involving long–distance dependencies across several clauses, and these constructions ranged from more prototypical and relatively frequent to less prototypical, borderline ungrammatical and rare to nonexistent in everyday language (Dąbrowska 2010: 5–6). In this scenario, Dąbrowska hypothesizes that exposure to such constructions through linguistic training can (and indeed does) result in higher ratings for the unprototypical constructions, as naïve participants simply do not encounter them frequently enough to become entrenched.[10] In fact, Snyder (2000) found that a certain period of exposure to unprototypical constructions can increase their reported acceptability, which he regarded as the 'syntactic satiation effect'. However, the majority of the stimuli we used in our research were all monoclausal sentences with frequent occurrence in natural speech, which rules out the difference in type of construction as an explanation for the difference in ratings. We do, however, leave open the possibility, as pointed out by one anonymous reviewer, that some of the stimuli have been rated lower based on semantic grounds. We are aware that a sentence such as *Na zidu su se dodirivale grafike i ulja* is not something one would hear or produce spontaneously every day in natural speech.

One possible factor which we have taken into consideration and have shown to be statistically significant is *age* (Figure 4 and Table 3). In Table 4, we used the data provided in Table 3 to calculate the naïve participants' mean scores predicted by the linear model for the three conditions which reached statistical significance[11] and compared them to the to the mean scores from Table 2. As can be seen, the values which the linear model predicts come very close to the values we actually got from our naïve participants. The observation about expert participants rating both sets of fillers lower than naïve participants also provides strong support for this claim. We take these facts to represent solid indication that age is a good predictor of acceptability ratings – as age increases, average acceptability rating decreases.

However, this should not be taken to mean that age alone is responsible for the difference in acceptability and we should entertain the possibility that other

10 This issue obviously relates to the question of the difference between nativist and usage–based approaches to morphosyntactic analysis, an issue which is beyond the scope of this paper.

11 The equation used for the predicted values is *Intercept–(Slope(AGE)*20)*.

factors, such as education and social status, might also be responsible for this. It is interesting to note that correlation between acceptability ratings of different agreement patterns and age was also noticed for Russian and this effect was also attributed to education (Panov 1968 qtd. in Corbett 1991: 252).[12] For the future, we suggest to conduct the same experiment with two additional participant groups, older naïve participants and younger expert participants, in order to test for the significance of the age factor.

One reviewer has pointed out that the discrepancy between the two groups could be explained by the effects of prescriptivism on expert participant ratings. While it would hypothetically be possible to check this by looking at potential differences in ratings between expert participants who are teachers of Croatian and those who are general linguists (as the reviewer suggests), there are two main reasons why we think this comparison would not be valid here. First of all, the number of general linguists in the current sample (N = 5) is not high enough for a valid comparison to be made, so no comparison could go beyond the realms of speculation. Secondly, while one might expect a higher degree of prescriptivism from Croatian teachers than from general linguists, it would be more appropriate to quantify the degree of prescriptivist tendencies via a separate method instead of assuming them from type of education/training. For instance, a questionnaire containing 10–20 statements with Likert–scale ratings that would measure the extent of individual prescriptivist tendency could be administered to participants before or after the main experiment. Since we did not plan to include this factor into our analysis, such a method was not used in our study. However, we strongly encourage using this or a similar method for testing the potential effects of prescriptivism in subsequent research.

	Predicted	Obtained
Collective &P	79.4659	81.475
Collective NP	82.428	84.8125
Non–collective &P	89.265	88.50833

Table 5. Comparison of predicted and obtained values for naïve participants

Our opinion is that these results ultimately raise the question of what exactly is being measured with acceptability ratings of any kind, i.e. which criteria native speakers use when assessing whether a sentence is more or less acceptable – whether a sentence conforms to the rules of grammar, whether a sentence makes sense, whether a sentence is something someone would *actually* say, or by some other criteria. Some of the expert participants that the authors talked to after the completion of the experiment said that they gave lower ratings to some of the sentences

12 We would like to thank Greville Corbett for drawing our attention to this fact.

due to other criteria, which were not the subject of this experiment (position of the clitic, lexical choices in some of the materials, etc.). Another potential influencing factor which was identified on the basis of individual expert participants' comments after the experimental sessions is the impact of prescriptivist grammar on the Croatian–speaking experts. Some of them commented to have evaluated sentences on the basis of the criterion whether they conform to the grammatical rules of the Standard Croatian language.

Finally, some of the variance might be attributed to the unfamiliarity of the task at hand. Namely, rating how well a picture describes a sentence is not a task that many participants are likely to perform on a regular basis (cf. e.g. Stefanowitsch 2006: 73).

Conclusion

An issue discussed in the literature on data collection (cf. section 2 of this paper) relates to the question of differences in grammaticality ratings given by experts and non–experts in linguistic studies. Gries (2002) and Dąbrowska (2010) have both shown that judgments of linguists and the judgments of non–linguists diverge to a great extent, concluding that their judgments cannot be considered as representative of the population as a whole. Gibson & Fedorenko (2013), for example, argue that the theoretical biases of linguists could influence their judgments. In this paper, we report on a replicated experimental study which investigated whether CCA is based on clausal ellipsis or not, in which we compared the answers of a group of non–experts (cf. Arsenijević et al. 2019) and the answers of a group of experts (linguists with a PhD in the field and Croatian language teachers). The results of this study reveal that the acceptability of sentences, as rated by the expert participants in the replicated experiment, can be attributed to the difference in level of linguistic 'naïveness' or 'expertness'. However, the results seem to go in the opposite direction of those reported by Dąbrowska (2008 and 2010) in the sense that her expert participants provided higher ratings for all grammatical sentences and lower ratings for all ungrammatical sentences in comparison to naïve participants, whereas our group of expert participants provided lower ratings across all four experimental conditions (*Collective &P*, *Collective NP*, *Non–collective &P*, and *Non–collective NP*). These differences might be attributed to a higher degree of conservatism, i.e. compliance to the prescriptive rules of the Croatian Standard variety, to a longer exposure and also potential metalinguistic knowledge with respect to sentences and constructions as presented in the experimental stimuli, as well as to age, a factor requiring further exploration in future studies.

References

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

Arsenijević, Boban, Jana Willer-Gold, Nadira Aljović, Nermina Čordalija, Marijana Kresić Vukosav, Nedžad Leko, Frane Malenica, Franc Lanko Marušič, Tanja Milićev, Nataša Milićević, Petra Mišmaš, Ivana Mitić, Anita Peti-Stantić, Branimir Stanković, Jelena Tušek, and Andrew Nevins (2019). Elided Clausal Conjunction Is Not the Only Source of Closest-Conjunct Agreement: A Picture-Matching Study. *Syntax*, synt.12171. https://doi.org/10.1111/synt.12171

Benmamoun, Elabbas, Archna Bhatia, and Maria Polinsky (2009). Closest conjunct agreement in head final languages. *Linguistic Variation Yearbook* 9: 67–88, https://doi.org/10.1075/livy.9.02ben

Bošković, Željko (2009). Unifying first and last conjunct agreement. *Natural Language & Linguistic Theory* 27(3): 455–496, https://doi.org/10.1007/s11049–009–9072–6

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

Cowart, Wayne (1997). *Experimental syntax: applying objective methods to sentence judgements*. Thousand Oaks, Calif: Sage Publications

Culbertson, Jennifer, and Steven Gross (2009). Are Linguists Better Subjects? *The British Journal for the Philosophy of Science* 60(4): 721–736, https://doi.org/10.1093/bjps/axp032

Dąbrowska, Ewa (2008). Questions with long–distance dependencies: A usage–based perspective. *Cognitive Linguistics* 19(3), https://doi.org/10.1515/COGL.2008.015

Dąbrowska, Ewa (2010). Naive v. expert intuitions: An empirical study of acceptability judgments. *The Linguistic Review* 27(1): 1–23, https://doi.org/10.1515/tlir.2010.001

Divjak, Dagmar (2017). The Role of Lexical Frequency in the Acceptability of Syntactic Variants: Evidence From *that*– Clauses in Polish. *Cognitive Science* 41(2): 354–382, https://doi.org/10.1111/cogs.12335

Drummond, Alex (2011). IBEX Farm (Version 0.3.9). Retrieved from http://spellout.net/ibexfarm/

Ferreira, Fernanda (2005). Psycholinguistics, formal grammars, and cognitive science. *The Linguistic Review* 22(2–4): 365–380

Franks, Steven, and Jana Willer–Gold (2014). Agreement Strategies with Conjoined Subjects in Croatian. Witkoś, Jacek, and Sylwester Jaworski, eds. *New Insights into Slavic Linguistics*. In *New Insights into Slavic Linguistics Band 3*, Frankfurt am Main: Peter Lang, 91–113

Gibson, Edward and Evelina Fedorenko (2013). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes* 28(1–2): 88–124, https://doi.org/10.1080/01690965.2010.515080

Gries, Stefan Th. (2002). Evidence in linguistics: Three approaches to genitives in English. Brend, Ruth M., William J. Sullivan, and Arle R. Lommel, eds. *LACUS Forum XXVIII: What Constitutes Evidence in Linguistics?*. Fullerton: LACUS, 17–31

Labov, William. (1978). *Sociolinguistic Patterns*. Oxford: Blackwell

Luka, Barbara J., and Lawrence W. Barsalou (2005). Structural facilitation: Mere exposure effects for grammatical acceptability as evidence for syntactic priming in comprehension. *Journal of Memory and Language* 52(3): 436–459, https://doi.org/10.1016/j.jml.2005.01.013

Marušič, Franc, Andrew Ira Nevins, and William Badecker (2015). The Grammars of Conjunction Agreement in Slovenian. *Syntax* 18(1): 39–77, https://doi.org/10.1111/synt.12025

Nagata, Hiroshi (1988). The relativity of linguistic intuition: The effect of repetition on grammaticality judgments. *Journal of Psycholinguistic Research* 17(1): 1–17, https://doi.org/10.1007/BF01067178

Panov, Mixail V. (ed.) (1968). *Morfologia i sintaksis sovremenogo russkogo literaturnogo jazyka (Russkij jazyk i sovetskoe obščestvo: Sociologo–lingvističeskoe issledovanije: III*. Moscow: Nauka

Puškar, Zorica, and Andrew Murphy (2015). Closest Conjunct Agreement in Serbo–Croatian: A Rule–Ordering Account. Assmann, Anke, Sebastian Bank, Doreen Georgi, Timo Klein, Philipp Weisser, and Eva Zimmermann, eds. *Topics at InfL*. *Vol 92*. *of Linguistische Arbeitsberichte (LAB),* Universität Leipzig, 441–482

R Core Team. (2015). *R: A language and environment for statistical computing*. Retrieved from https://www.R–project.org/

Sampson, Geoffrey R (2007). Grammar without grammaticality. *Corpus Linguistics and Linguistic Theory* 3(1): 1–32, https://doi.org/10.1515/CLLT.2007.001

Schütze, Carson T. (2011). Linguistic evidence and grammatical theory: Linguistic evidence and grammatical theory. *Wiley Interdisciplinary Reviews: Cognitive Science* 2(2): 206–221, https://doi.org/10.1002/wcs.102

Snow, Catherine, and Guus Meijer (1977). On the Secondary Nature of Syntactic Intuitions. Greenbaum, Sidney, ed. *Acceptability in Language*. https://doi.org/10.1515/9783110806656–012

Snyder, William (2000). An Experimental Investigation of Syntactic Satiation Effects. *Linguistic Inquiry* 31(3): 575–582, https://doi.org/10.1162/002438900554479

Spencer, Nancy J. (1973). Differences between linguists and nonlinguists in intuitions of grammaticality–acceptability. *Journal of Psycholinguistic Research* 2(2): 83–98

Sprouse, Jon, and Diogo Almeida (2012). Assessing the reliability of textbook data in syntax: Adger's Core Syntax. *Journal of Linguistics* 48(03): 609–652, https://doi.org/10.1017/S0022226712000011

Stefanowitsch, Anatol (2006). Negative evidence and the raw frequency fallacy. *Corpus Linguistics and Linguistic Theory* 2(1), https://doi.org/10.1515/CLLT.2006.003

Trask, Robert Lawrence (1999). *Key concepts in language and linguistics*. London; New York: Routledge

Willer–Gold, Jana, Boban Arsenijević, Mia Batinić, Nermina Čordalija, Marijana Kresić, Nedžad Leko, Franc Lanko Marušič, Tanja Milićev, Nataša Milićević, Ivana Mitić, Andrew Nevins, Anita Peti–Stantić, Branimir Stanković, Tina Šuligoj, and Jelena Tušek (2016). Conjunct agreement and gender in South Slavic: From theory to experiments to theory. *A* 24(1): 187–224

Neslaganje oko slaganja – usporedba jezičnih procjena stručnih i nestručnih ispitanika

U ovom su radu opisani rezultati istraživanja putem kojeg su uspoređene jezične procjene izvornih govornika o rečenicama koje sadrže slaganje, odnosno sročnost s najbližim konjunktom (engl. *Closest Conjunct Agreement*). Ovo istraživanje predstavlja nastavak većeg projekta koje proučava obrasce sročnosti u južnoslavenskim jezicima (Arsenijević i sur. 2019, Willer–Gold i sur. 2016), a primarna svrha eksperimenta bila je ispitati tvrdnju kako je elipsa uzrok sročnosti glagola s najbližim konjunktom u postverbalnom položaju (Aoun, Benmamoun & Sportiche 1994). U istraživanju prikazanom u ovom radu uspoređivane su procjene lingvista i nelingvista putem eksperimentalne platforme IBEX Farm (Drummond 2011). Korištena je metoda mjerenja prihvatljivosti rečenice za opis prizora na slici upotrebom ljestvice od 0 do 100 kojom su ispitanici izražavali prihvatljivost rečenica. Ispitane su dvije skupine izvornih govornika hrvatskog jezika (štokavskog dijalekta) različitih razina jezičnog obrazovanja – skupinu nelingvista činili su studenti (N = 30) različitih studija preddiplomske razine (osim kroatistike, lingvistike i psihologije), dok su skupinu lingvista činili nastavnici hrvatskog jezika i lingvisti zaposleni na osnovnoškolskoj, srednjoškolskoj i sveučilišnoj razini (N = 30). Sve su skupine ispitanika bile izložene istim eksperimentalnim podražajima u istim ili usporedivim uvjetima. Razlike u rezultatima prikupljenima od lingvista i nelingvista kao sudionika u istraživanju tema su brojnih znanstvenih istraživanja u posljednjih nekoliko desetljeća (v. Culbertson & Gross 2009, Dąbrowska 2010, Gibson & Fedorenko 2013, ali i Sprouse & Almeida 2012 za suprotno mišljenje). Naš je cilj pridonijeti toj raspravi upotrebom podataka o sročnosti s najbližim konjunktom. Rezultati prikupljeni u ovom istraživanju pokazuju kako postoji statistički značajna razlika između jezičnih procjena tih dviju skupina ispitanika te je opravdano tu razliku objasniti upravo razlikom u razini jezičnog obrazovanja. Ta se razlika najviše ogleda u činjenici da su lingvisti davali niže procjene prihvatljivosti na razini svih podražaja, neovisno o vrsti rečenice i sročnosti. Također je zabilježen i utjecaj dobi na razinu prihvatljivosti.

Keywords: Closest Conjunct Agreement, naïve vs. expert participants, native speaker intuitions, sentence–picture matching experiment, Croatian

Ključne riječi: sročnost s najbližim konjunktom, stručni i nestručni ispitanici, procjene izvornih govornika, eksperiment uparivanja rečenice i slike, hrvatski jezik

Appendix

Number	Type	Sentence
10	Collective+ConjP	Na zidu su se dodirivale grafike i ulja.
11	Collective+ConjP	Na mostu su se smjenjivala vozila i kočije.
12	Collective+ConjP	Po kanalu su se susretale jedrilice i motorna plovila.
13	Collective+ConjP	U bitci su se sudarala koplja i sablje.
14	Collective+ConjP	U poštanskom sandučiću su se izmiješale razglednice i pisma.
15	Collective+ConjP	Na suprotnim stranama rijeke su se pružala polja i gore.
16	Collective+ConjP	Na gomilu su bile izdvojene prepone i kladiva.

17	Collective+ConjP	U moru su se zapetljala debla i mreže.
18	Collective+NP	Na zidu su se dodirivale grafike.
19	Collective+NP	Na mostu su se smjenjivala vozila.
20	Collective+NP	Po kanalu su se susretale jedrilice.
21	Collective+NP	U bitci su se sudarala koplja.
22	Collective+NP	U poštanskom sandučiću su se izmiješale razglednice.
23	Collective+NP	Na suprotnim stranama rijeke su se pružala polja.
24	Collective+NP	Na gomilu su bila izdvojena kladiva.
25	Collective+NP	U moru su se zapetljale mreže.
26	Non–Collective+ConjP	Na kostim su zašivene ruže i pera.
27	Non–Collective+ConjP	Nakon parade su počišćena parkirališta i kuće.
28	Non–Collective+ConjP	Na stol su odložena srca i mašne.
29	Non–Collective+ConjP	Na zidu su visjele medalje i priznanja.
30	Non–Collective+ConjP	U ladicu su stavljene vizitke i nalivpera.
31	Non–Collective+ConjP	Za koncert su pripremljena čela i note.
32	Non–Collective+ConjP	Na brod su ukrcane gajbe i sidra.
33	Non–Collective+ConjP	U dućanu su izložena ogledala i lampe.
34	Non–Collective+NP	Na kostim su zašivene ruže.
35	Non–Collective+NP	Nakon parade su počišćena parkirališta.
36	Non–Collective+NP	Na stol su odložena srca.
37	Non–Collective+NP	Na zidu su visjele medalje.
38	Non–Collective+NP	U ladicu su stavljene vizitke.
39	Non–Collective+NP	Za koncert su pripremljena čela.
40	Non–Collective+NP	Na brod su ukrcane gajbe.
41	Non–Collective+NP	U dućanu su izložena ogledala.
42	Filler_Mismatch	Na nebu su se sjajile zvijezde.
43	Filler_Mismatch	U frižideru su se sušile kobasice.
44	Filler_Match	Na zid su se naslanjale police.
45	Filler_Match	Kroz brda su se provlačile rijeke.
46	Filler_Mismatch	U vitrini su izložene krune.
47	Filler_Mismatch	Na polici su stajale knjige.
48	Filler_Match	Na stolu su otvorene karte.
49	Filler_Match	Na stolicama su čekale violine.
50	Filler_Mismatch	Kroz ključanicu su se vidjela sazviježđa.

51	Filler_Mismatch	Na lutkama su se zračila krzna.
52	Filler_Match	Na jarbolu su se sušila jedra.
53	Filler_Match	U pećnici su se pekla peciva.
54	Filler_Mismatch	Vodu iz špine su lovila korita.
55	Filler_Mismatch	Na zid su obješena ogledala.
56	Filler_Match	U čamcu su ostavljena vesla.
57	Filler_Match	Potocima su povezana jezera.
58	Filler_Mismatch	Jedan nasuprot drugoga su se parkirali kamioni.
59	Filler_Mismatch	Ispred dućana su se prodavali bicikli.
60	Filler_Match	Na parkiralištu su se prevrnuli motori.
61	Filler_Match	Preko rijeke su se uzdizali mostovi.
62	Filler_Mismatch	Po nebu su vijugali avioni.
63	Filler_Mismatch	Jedan pored drugoga su složeni kompjutori.
64	Filler_Match	Na stolu su stajali telefoni.
65	Filler_Match	Na pozornici su postavljeni klaviri.
66	Filler_Mismatch	Na križanju su se sudarili vlakovi i autobusi.
67	Filler_Mismatch	U selidbi su se polomili klarineti i kontrabasi.
68	Filler_Match	Na zgradi su se pojavili grafiti i grbovi.
69	Filler_Match	Od hladnoće su se smrznuli prsti i nokti.
70	Filler_Mismatch	Za proslavu su ukrašeni muzeji i paviljoni.
71	Filler_Mismatch	U kolicima su dovezeni kaktusi i kokosi.
72	Filler_Match	Na model su zašiveni rukavi i gumbi.
73	Filler_Match	U kutiju su odloženi ključevi i kablovi.