Gender mismatches with NP ellipsis in Bosnian/Croatian/Serbian

Andrew Murphy & Zorica Puškar
andrew.murphy@uni-leipzig.de zorica.puskar@uni-leipzig.de

Universität Leipzig

21.06.16

1 Overview

In this presentation, we focus on hybrid nouns in Bosnian/Croatian/Serbian (BCS) which have grammatical feminine gender, but their natural gender can vary based on the gender of their discourse referent. There are various proposals for the variability in gender agreement with such nouns, some of which involve lexically pre-specified gender (Corbett 1991), semantic agreement (cf. Sudo and Spathas to appear and references therein) or interaction of syntactic operations (Puškar 2015). These approaches differ with regard to the presence of conflicting gender features on the noun, where unrealized features are either always present or only when reflected by agreement. We adopt the assumption of a syntactic identity condition on ellipsis (e.g. Merchant 2013) to try to distinguish between competing approaches to hybrid nouns in BCS.

2 Data: Hybrid nouns in BCS

We focus on a particular type of hybrid nouns, which can show gender agreement alternations based on their grammatical and/or natural gender.

- Grammatical gender of these nouns is always feminine, but the natural one can alternate depending on the discourse referent.
- Examples: *budala* 'fool', *varalica* 'cheater', all nouns ending in *-vođa* (*kolovođa* 'leader in traditional dances'), *mušterija* 'customer', *propalica* 'loser, failure', *pijanica* 'drunkard', *skitnica* 'wanderer, drifter', *sudija* 'judge', etc.
- With male referents, these nouns mostly trigger feminine agreement on the adjectives and predicates (grammatical agreement, (1a)).
- Some speakers also allow masculine agreement (1b).

(1) a. Milan nam je nova mušterija.
Milan us is new.F customer
'Milan is our new customer.'

b. %Milan nam je novi mušterija.
Milan us is new.M customer
'Milan is our new customer.'

- With female referents, these nouns only trigger feminine agreement.

(2) a. Marija nam je nova mušterija.
Marija us is new.F customer
'Marija is our new customer.'
b. *Marija nam je novi mušterija.
Marija us is new.M customer
'Marija is our new customer.'

- Such nouns can allow gender mismatches under ellipsis in two directions.
- If both referents are male, feminine agreement on both adjectives is preferred, while masculine agreement is not accepted by all speakers (3b). However, the status of (3c)-(3d) is not completely clear. Preliminary judgements we got from some speakers showed variation in acceptability.

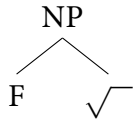
(3) a. Milan mu je stara mušterija, a Jovan mu je nova ⟨mušterija⟩.
Milan him is old.F customer but Jovan him is new.F ⟨customer⟩
'Milan is his old customer, and Jovan is his new one.'
b. %Milan mu je stari mušterija, a Jovan (mu je) novi ⟨mušterija⟩.
Milan him is old.M customer but Jovan him is new.M ⟨customer⟩
c. ??Milan mu je stari mušterija, a Jovan mu je nova ⟨mušterija⟩.
Milan us is old.M customer but Jovan us is new.F ⟨customer⟩
d. ??Milan mu je stara mušterija, a Jovan mu je novi ⟨mušterija⟩.
Milan him is old.F customer but Jovan him is new.M ⟨customer⟩

3 Theories of hybrid nouns

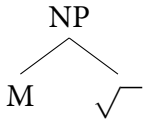
Hypothesis A: A lexicalist approach

- This approach would assume that hybrid nouns are simply listed in the lexicon twice, with two different gender features on the two entries (Corbett 1991; Merchant 2014).
- e.g. Merchant (2014:19) proposes such an analysis for Greek epicene nouns.

(4) Feminine noun:



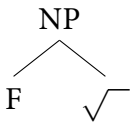
(5) Masculine noun:



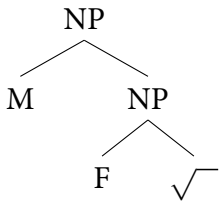
Hypothesis B: Natural gender is present only when reflected by agreement

- Grammatical gender is always present on the noun and natural gender is only there when it is reflected by agreement/concord (cf. Steriopolo and Wiltschko 2010; Matushansky 2013; Pesetsky 2014; Landau to appear).
- In such cases, we are dealing with two different structures, but one of them is a subset of the other, i.e. grammatical gender is always there (6), but in some cases natural is present above it and overwrites it (7).
- Alternatively, two different features are just represented differently, where feminine is always more marked than the masculine, i.e. masculine is a subset of the feminine (cf. Despić 2015; Kramer 2015; Willer-Gold et al. to appear)

(6) Feminine noun



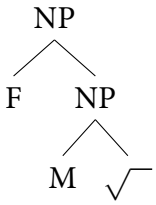
(7) Masculine noun



Hypothesis C: Both features always present on a noun

- Both the natural and grammatical gender are present on a hybrid noun. The structure itself is not responsible for different agreement patterns.
- What causes the mismatches is the Agree mechanism (Puškar 2015).
- Variability not caused by the structure of the DP, but by the structure of the probe and syntactic operations.

(8) Masculine noun and feminine noun



4 Syntactic identity under ellipsis

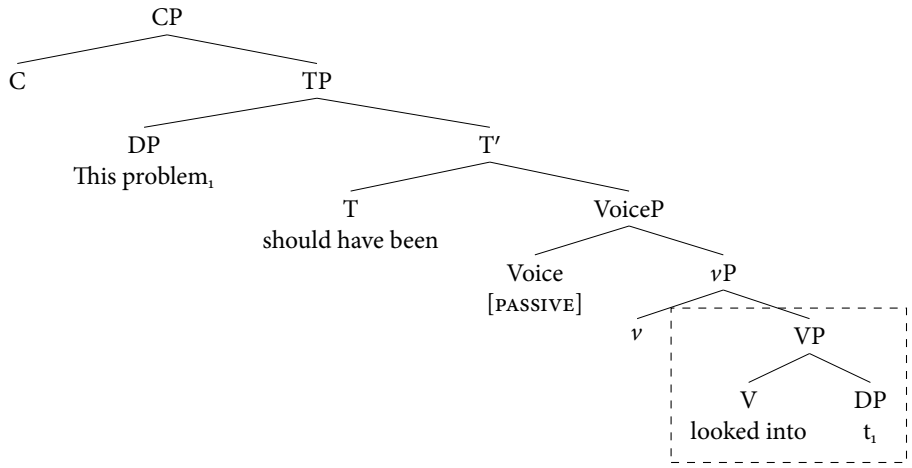
Idea: Test the three hypotheses above with a syntactic approach to ellipsis identity (cf. Merchant 2013; Murphy to appear).

- Syntactic approaches to ellipsis identity assume that there is no syntactic material in the ellipsis site not also present in the antecedent.
- E.g. Merchant (2013) illustrates the syntactic identity requirement through voice mismatches. VP ellipsis allows for voice mismatches, while TP ellipsis does not:

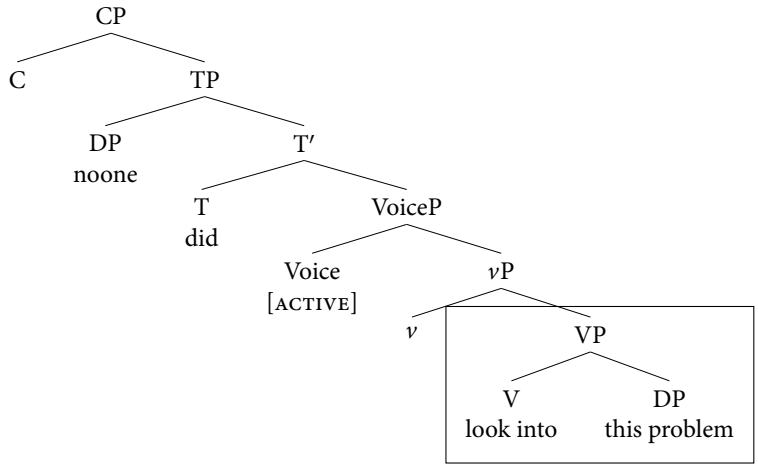
(9) a. This problem should have been looked$_{pass}$ into, but obviously no-one did ⟨ [$_{\text{VP}}$ look$_{act}$ into the problem] ⟩
b. *John was killed$_{pass}$, but we don't know who$_1$ ⟨ [$_{\text{TP}}$ t$_1$ killed$_{act}$ John] ⟩

- Merchant accounts for the difference by assuming that the mismatched feature (active vs. passive) is located in a VoiceP above vP.
- The crucial difference between VP and TP ellipsis is that TP will always include the VoiceP, so the mismatched feature will always be included in the ellipsis site.
- So for voice mismatches under VP ellipsis (9a), the voice feature is not included in the ellipsis site.

(10) *Antecedent*:

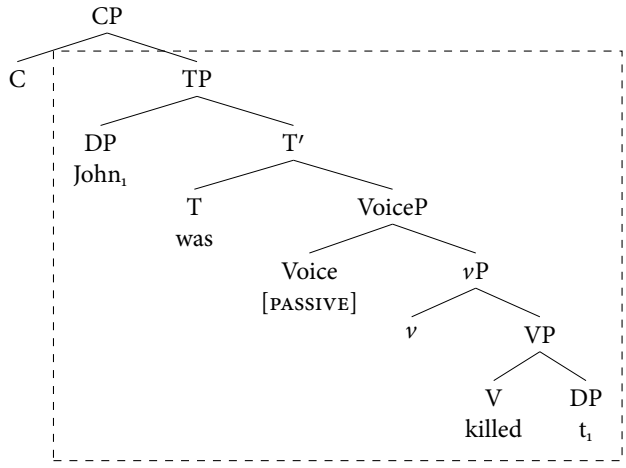


(11) *Ellipsis site*:

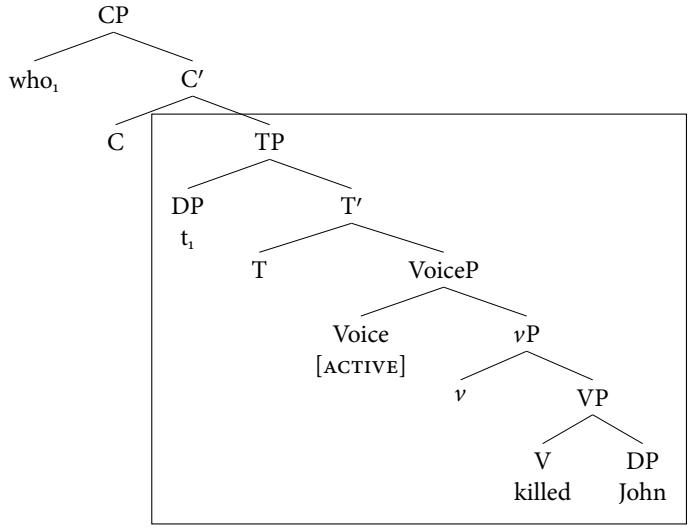


- However, for voice mismatches under TP ellipsis, the ellipsis site contains the mismatched voice feature:

(12) *Antecedent*:



(13) *Ellipsis site*:
I don't know…



Syntactic identity in gender mismatches:

- We utilize the assumption of syntactic identity to try and distinguish between the three types of structure for hybrid nouns.
- Under the assumption that the syntactic structures need to be identical in order to enable ellipsis, our three hypotheses for the structure of hybrid nouns make the following predictions:

1. **Prediction for Hypothesis A:** Mismatches impossible: ***nat** ↔ **gram** (because M≠F).
2. **Prediction for Hypothesis B:** One way mismatches possible: ***nat** → **gram**, ✓**gram** → **nat** (because M⊂F).
3. **Prediction for Hypothesis C:** Two-way mismatches possible: ✓**nat** ↔ **gram** (because M=F).

5 The Experiment

Aim: To test the grammaticality of sentences in which natural gender licenses the ellipsis of grammatical gender and vice versa, as well as to test whether there is a significant difference between the acceptability of the two combinations.

Task: A grammaticality judgement task in which participants were asked to rate sentences involving NP ellipsis on a 7-point Likert scale (1=completely bad, 7=sounds excellent). Each sentence contained two clauses, in both of which the predicate noun was a *customer*-type hybrid noun, present in the first clause as the antecedent, and elided in the second clause.

5.1 Method

5.1.1 Experimental conditions

- Factors: gender of the subjects and type of agreement on adjectives in the first and second clause
- Conditions: combinations of gender on the subjects (NP_1 & NP_2) and adjectives (Adj_1 and Adj_2) in each clause
- The following table shows the abstract design:

(14)

Condition						Agreement Type	
		NP_1	Adj_1	NP_2	Adj_2	S_1	S_2
1	MFMF	M	F	M	F	Gram	Gram
2	MMMM	M	M	M	M	Nat	Nat
3	MFMM	M	F	M	M	Gram	Nat
4	MMMF	M	M	M	F	Nat	Gram
5	FFFF	F	F	F	F	Na./Gr.	Na./Gr.
6	FMFM	F	M	F	M	—	—

Table 1: Test items

- NP_1 and NP_2 refer to the gender on the subject in the first and second clause, respectively.
- Adj_1 and Adj_2 refer to the gender agreement on the adjective in the first and in the second clause, respectively.
- S_1 and S_2 indicate the type of gender that is reflected by the agreement on the adjective (natural vs. grammatical)

5.1.2 Stimuli

- Sentences containing two clauses coordinated by the conjunction *a* 'but'.
  - Clause 1: a subject and a hybrid noun in the predicate position, modified by an adjective.
  - Clause 2: a subject and an elided hybrid noun, with an adjective as a remnant.
- 96 test items total.
- 48 test items. 4 conditions (MFMF, MMMM, MFMM, MMMF) x 12 items per condition. The subject in each condition was masculine and the agreement on the adjectives was varied.

(15) Jovan je star-a mušterija, a Marko potencijaln-a _____.
Jovan is old-F customer but Marko potential-F
'Jovan is an old customer and Marko a potential one.' MFMF

(16) Jovan je star-i mušterija, a Marko potencijaln-i _____.
Jovan is old-M customer but Marko potential-M
'Jovan is an old customer and Marko a potential one.' MMMF

(17) Jovan je star-a mušterija, a Marko potencijaln-i _____.
Jovan is old-F customer but Marko potential-M
'Jovan is an old customer and Marko a potential one.' MFMM

(18) Jovan je star-i mušterija, a Marko potencijaln-a _____.
Jovan is old-M customer but Marko potential-F
'Jovan is an old customer and Marko a potential one.' MMMF

- 48 control items. 2 conditions x 24 items per condition. These included the same type of sentences, only with feminine subjects.
  - FFFF (the good baseline, the combination that was expected to be judged as very good)

– FMFM (the bad baseline, the combination was expected to be judged as completely bad, due to the complete lack of agreement)

(19) Slavica je tešk-a pričalica, a Bojana umerenij-a _____ .
Slavica is heavy-F talker but Bojana moderate-F
'Slavica is a big talker and Bojana is less of one.' FFFF

(20) Slavica je tešk-i pričalica, a Bojana umerenij-i _____ .
Slavica is heavy-M talker but Bojana moderate-M
'Slavica is a big talker and Bojana is less of one.' FMFM

5.1.3 Procedure

- The experiment was coded using LimeSurvey (LimeSurvey Project Team 2012) and run online via the LimeService platform.
- Sentences were presented one by one in a random order.
- Each participant saw all 96 sentences.
- The participant was asked to give a grammaticality judgement on a 7-point Likert scale (1=completely bad, 7=sounds excellent) by dragging a slider from the middle of the scale towards the number signalling their response.

(21)

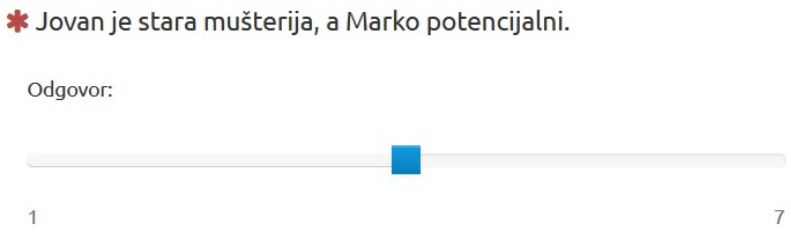


Figure 1: Example experimental item

5.1.4 Participants

- A total of 50 volunteers, 12 male, 38 female, aged 15-55.
- None of the participants were paid or otherwise compensated for their participation.

5.2 Results

- After removing the speakers who performed badly with the baselines (*bad* (FMFM): ≥3, *good* (FFFF): ≤5)), n = 25 participants.

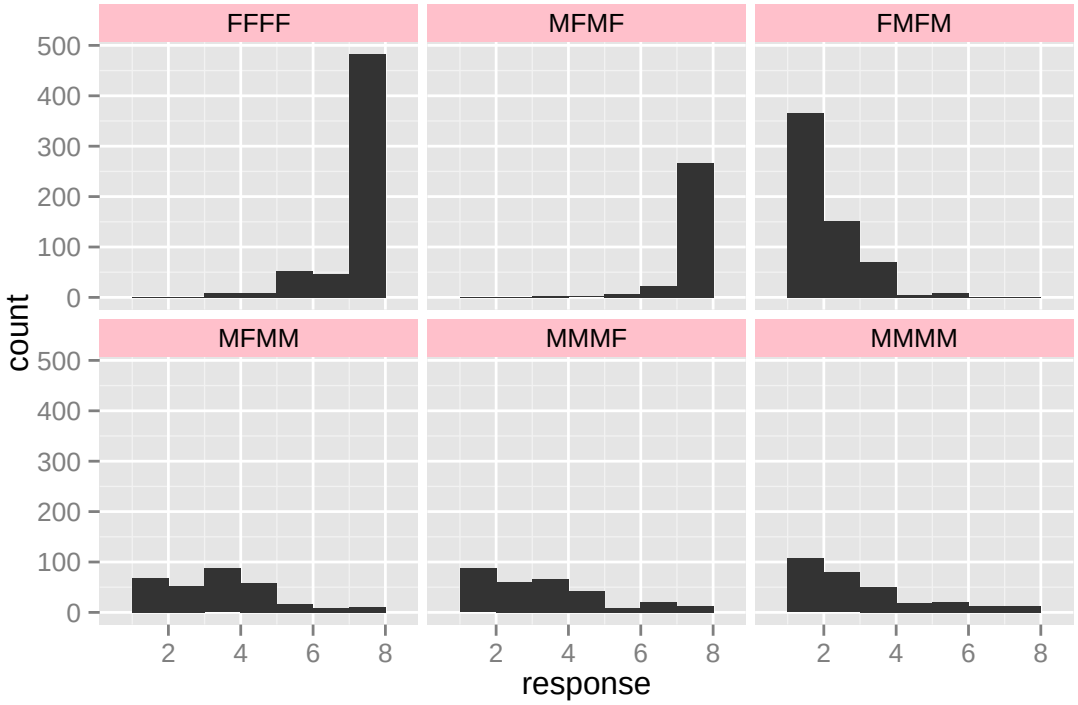


Figure 2: Total responses per condition

- As Figure 1 shows, the distributions of MFMM and MMMF seem to be almost identical.
- To test whether there is a difference between the two, we fitted a model with the formula `response ~ combination + (1|participant)` (Bayesian mixed effects model).

	post.mean	l-95% CI	u-95% CI	pMCMC	
(Intercept - MFMM)	2.89339	2.66968	3.10867	<0.001	**
combination FFFF	3.75542	3.61688	3.91681	<0.001	**
combination FMFM	-1.32946	-1.47508	-1.17964	<0.001	**
combination MFMF	3.93493	3.76130	4.11543	<0.001	**
combination MMMF	-0.09145	-0.29315	0.07554	0.347	
combination MMMM	-0.37642	-0.54405	-0.19304	<0.001	**

Table 2: Formula `response ~ combination + (1|participant)` using MCMCglmm package (Hadfield 2010)

- The results show the posterior mean (roughly equivalent to the coefficient in a traditional model), together with the confidence intervals and the estimated MCMC p value (which is calculated based on the posterior distribution of the confidence intervals).
- The reference level in the intercept (the starting point of the model) is MFMM, against which the other combinations are compared.
- We see that most levels are different from the intercept **except** for MMMF which is not statistically different.

6 Discussion

- Recall that our three hypotheses from the beginning made different predictions about possible mismatches under NP ellipsis depending on the type of gender present on the hybrid noun.
- **Hypohesis A:**
  - Prediction: Mismatches impossible.
  - We expect all mismatches to be simply ungrammatical and therefore we expect no difference between FMFM (ungrammatical control) and test items.
  - Result: ☹ There was a significant difference between MFMM and FMFM ($p < 0.001$).
  - Hypothesis A thus arguably is not the right one to account for the structure of hybrid nouns.
- **Hypothesis B:**
  - Prediction: One-way mismatches possible.
  - We expect the hybrid noun to contain more structure/features with one of the genders, and consequently, we expect a one-way alternation and the difference in grammaticality judgements between MFMM and MMMF (i.e. one of them is less probable to license the other).
  - Result: ☹ There was no significant difference between MFMM and MMMF ($p = 0.347$).
- **Hypothesis C:**
  - Prediction: Two-way mismatches possible.
  - Since this hypothesis assumes identity of structure, we predict no difference between MFMM and MMMF.
  - Result: ☺ There was no significant difference between MFMM and MMMF ($p = 0.347$).
- If syntactic identity is correct, then our results are incompatible with the hypothesis that gender-variable nouns have two distinct syntactic structures (Hypotheses A & B).
- The assumption that hybrid nouns always contain both gender features can be maintained (Hypothesis C). Under this approach, variation in agreement can be attributed to other factors, e.g. relativized probing (cf. Puškar 2015).
- Possibly also compatible with semantic approaches to ellipsis identity (e.g. Merchant 2014), depending on whether gender is encoded semantically or not.

7 Future directions

A follow-up study:

- A mismatch becomes possible only in the case where the referent in the clause with NP ellipsis is feminine.
  - In (22a), the agreement on the adjective can reflect either the grammatical gender of the noun in the ellipsis site, or the natural gender of the referent, but since the values are the same, and they match the value of the antecedent, the sentence is grammatical.

– In (22b) (for speakers who allow for masculine agreement), the mismatch is also tolerated in the direction natural masculine–(natural/grammatical)feminine.
– (22c) is only grammatical in the direction (natural/grammatical)feminine–grammatical feminine, but not in the direction (natural/grammatical)feminine–natural masculine.

(22) a. Milan mu je stara mušterija, a Marija mu je nova ⟨mušterija⟩.
Milan him is old.F customer but Marija him is new.F
b. ?Milan mu je stari mušterija, a Marija mu je nova ⟨mušterija⟩.
Milan him is old.M customer but Marija him is new.F
c. Marija mu je stara mušterija, a Milan mu je nova/*novi ⟨mušterija⟩.
Marija him is old.F customer but Milan him is new.F/new.M

- Idea: test the acceptability of sentences of the type in (22), with subjects of different genders and see whether there are differences in mismatches.

8 Acknowledgements

We would like to thank Peter Staroverov for the help in the experimental design, Joanna Zaleska for the help in coding the experiment and Matías Guzmán Naranjo for the help in analysing the results. We would also like to thank the anonymous participants in the experiment, as well as Boban Arsenijević and Jelena Stojković. This work was completed as part of the DFG-funded graduate school Interaktion Grammatischer Bausteine 'Interaction of Grammatical Building Blocks' (IGRA).

References

Bobaljik, Jonathan D. and Cynthia Levart Zocca (2011): 'Gender Markedness: The anatomy of a counterexample', *Morphology* **21**, 141–166.

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

Despić, Miloje (2015): Investigations on mixed agreement: Polite plurals, hybrid nouns and coordinate structures. Unpublished manusript. Cornell University.

Hadfield, Jarrod D (2010): 'MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package', *Journal of Statistical Software* **33**(2), 1–22.
URL: *http://www.jstatsoft.org/v33/i02/*

Kramer, Ruth (2015): *The morphosyntax of gender*. Oxford University Press, Oxford.

Landau, Idan (to appear): 'DP-internal semantic agreement: A configurational analysis', *Natural Language and Linguistic Theory* .

LimeSurvey Project Team, Carsten Schmitz (2012): *LimeSurvey: An Opeen Source survey tool*. LimeSurvey Project, Hamburg, Germany.
URL: *http://www.limesurvey.org*

Matushansky, Ora (2013): Gender confusion. *In:* L. Cheng and N. Corver, eds, *Diagnosing Syntax*. Oxford University Press, Oxford, pp. 271–294.

Merchant, Jason (2013): 'Voice and ellipsis', *Linguistic Inquiry* **44**(1), 77–108.

Merchant, Jason (2014): 'Gender Mismatches under Nominal Ellipsis', *Lingua* **151**, 9–32.

Murphy, Andrew (to appear): 'Subset relations in ellipsis licensing', *Glossa* .

Pesetsky, David (2014): *Russian case morphology and the syntactic categories*. MIT Press, Cambridge, MA.

Puškar, Zorica (2015): Interactions of gender and number agreement: Evidence from Bosnian/Croatian/Serbian. Unpublished manusript. University of Leipzig.

Steriopolo, Olga and Martina Wiltschko (2010): Distributed GENDER hypothesis. *In:* G. Zybatow, P. Dudchuk, S. Minor and E. Pschehotskaya, eds, *Formal Studies in Slavic Linguistics. Proceedings of FDSL 7.5*. Peter Lang, New York, pp. 155–172.

Sudo, Yasutada and Giorgos Spathas (to appear): Nominal ellipsis and the interpretation of gender in Greek. In: *Proceedings of Sinn und Bedeutung 20.* .

Willer-Gold, Jana, Boban Arsenijević, Mia Batinić, Nermina Čordalija, Marijana Kresić, Nedžad Leko, Lanko Marušič, Tanja Milićev, Nataša Milićević, Ivana Mitić, Andrew Nevins, Anita Peti-Stantić, Branimir Stanković, Tina Šuligoj and Jelena Tušek (to appear): 'Conjunct Agreement and Gender in South Slavic: From Theory to Experiments to Theory'.