Inflectional Competition and Interpretation: A Case Study on the Slovenian Dual*

[removed for blind review]

August 26, 2020

Abstract

This paper reports on acceptability judgment experiments that aim at uncovering the underlying semantics of the Slovenian dual. The results indicate that the interpretation of a dual noun in Slovenian is similar to that of a noun phrase with a numeral 'two'. We argue that these results are incompatible with a previous analysis proposed by Sauerland (2008), according to which the Slovenian dual is semantically compatible with singular and dual reference ('one or two'). In light of the literature on the semantics and pragmatics of numerals, we will discuss the following three alternative theoretical possibilities. According to the Lexical Ambiguity Theory, the underlying semantics of the dual is ambiguous between lower-bounded ('at least 2') and bilateral ('exactly 2') meaning. The Scalar Strengthening Theory holds that the underlying semantics of the dual is lower-bounded but can be strengthened to a bilateral reading. The Pragmatic Weakening Theory assumes that the underlying semantics of the dual is bilateral but can be optionally weakened to a lower-bounded reading. We argue that our experimental results are most straightforwardly explained by the Pragmatic Weakening Theory, and discuss the consequences for competition between inflectional categories of number.

1 Introduction

Plural noun phrases in languages like English typically receive a plural reading, but sometimes give rise to number-neutral readings in certain grammatical contexts. For example, the plural noun phrase in (1a) receives a plural reading, while the plural noun phrase in (1b) has a number-neutral reading, making the overall sentence not equivalent to (and in fact, stronger than) the negation of the meaning of (1a).

(1) a. The customer bought magazines.
b. The customer didn't buy magazines.

While the current theoretical literature contains competing analyses of this observation (Sauerland 2003, Sauerland, Anderssen & Yatsushiro 2005, Sauerland 2008, Spector 2007, Farkas & de Swart 2010, Grimm 2013, Zweig 2009, Ivlieva 2013, Martí 2020b, Mayr 2015, Križ 2017, Sudo 2019), it is generally agreed that the underlying semantics of plural noun phrases cannot simply be plural ('more than one'): If that were the case, the observed meaning of (1a) would be straightforwardly accounted for, but (1b) would mean that the customer didn't buy multiple magazines, a weaker meaning than what is actually observed.

*[removed for blind review]

Against this theoretical backdrop, this paper investigates the underlying meaning of the Slovenian dual, which is an inflectional number category along with the singular and the plural. It reports on acceptability judgment experiments, whose results indicate that the interpretation of dual noun phrases in Slovenian is quite similar to the interpretation of noun phrases with the numeral *two*. We will claim that these results are incompatible with a previous analysis put forward by Sauerland (2008), according to which the Slovenian dual is semantically compatible with singular and dual reference ('one or two'). In light of the literature on the semantics and pragmatics of numerals (see Spector 2013 for an overview), we will discuss the following three alternative theoretical possibilities. According to the Lexical Ambiguity Theory, the underlying semantics of the dual is ambiguous between lower-bounded ('at least 2') and bilateral ('exactly 2'). The Scalar Strengthening Theory holds that the underlying semantics of the dual is lower-bounded, but can get strengthened to a bilateral reading. The Pragmatic Weakening Theory assumes that the underlying semantics is bilateral, but can be optionally weakened to a lower-bounded reading. We argue that our experimental results are most straightforwardly explained by the Pragmatic Weakening Theory.

The present paper proceeds as follows. Section 2 provides a brief review of the literature on the semantics and pragmatics of nominal number, including a previous analysis of the Slovenian dual due to Sauerland (2008). In Section 3 we will review the literature on numerals and discuss the three alternative analyses of the dual mentioned above. The experiments and their results are presented and discussed in Sections 4–6. We conclude in Section 7.

2 The Semantics and Pragmatics of Number and Inflectional Competition

2.1 Unmarked Plural and Inflectional Competition

Plural noun phrases are so called because they are typically used to talk about multiple entities. For instance, the bare plural in (2a) gives rise to an entailment that the customer in question bought *multiple* books about climate change, and the definite plural noun phrase in (2b) gives rise to a presupposition that the building has *multiple* exits.

(2) a. The customer bought **books** about climate change.
b. **The emergency exits** of the building are clearly indicated.

At the same time, plural noun phrases in languages like English are known to receive number-neutral readings in certain grammatical contexts (Sauerland 2003, Spector 2007, Farkas & de Swart 2010, among others; see however Kiparsky & Tonhauser 2012 for some potential complications). For example, the following examples have readings where the bare plural noun phrases in bold have number neutral readings.[1]

(3) a. This applicant does not have **journal papers**.
b. If you have **electronic devices** in your bag, take them out and put them on a tray.
c. This plant can survive without **leaves** for several years.
d. We should clean up the mess before **customers** arrive.

These examples suggest that the interpretation of a plural noun phrase is not simply always plural.

[1] Bare plurals in polar questions are often mentioned in this connection, but their judgments seem to be more involved than usually assumed, and for this reason we will not include polar questions here. See Pearson, Khan & Snedeker (2010), Bale, Gagnon & Khanjian (2011) for discussion.

Similarly, plural definites containing bound variables give rise to inferences that would be unexpected if their semantics were inherently plural. To see this, consider (4).

(4) Each applicant submitted **their experimental papers** as part of their application.

If the semantics of the plural definite here were plural, (4) should presuppose that every applicant had multiple experimental papers. In reality, the perceived presupposition is weaker than this, and the sentence could be felicitously uttered against a situation where only some of the applicants have multiple experimental papers. Here, the interpretation is not completely number-neutral, as the sentence would be infelicitous in a situation where every applicant has exactly one paper, but the observed presupposition is weaker than what would be expected if the plural definite were inherently semantically plural.

How to capture these observations is a point of active debate in the current literature. For example, Farkas & de Swart (2010) and Grimm (2013) postulate a number-neutral reading and a semantically plural reading for plural noun phrases, but these analyses cannot straightforwardly account for the reading of (4) (see also Martí 2020b, who follows Harbour 2014).

On the other hand, Sauerland (2003, 2008) and Sauerland et al. (2005) propose a theory that is based on *inflectional competition*. The rough idea is that a plural noun phrase is always semantically number neutral, and that when the plural meaning arises it is due to competition with the singular version of the sentence, which is assumed to have a more specific meaning that is only compatible with singular reference.[2] We will not discuss here the details of the way in which Sauerland and his colleagues cash out this theory, because serious empirical problems have been pointed out for it (see Spector 2007 for details). Other ways of implementing the idea of inflectional competition between the singular and the plural have been proposed that circumvent these problems (Spector 2007, Zweig 2009, Ivlieva 2013, Mayr 2015, Sudo 2019), but they all involve some additional theoretical machinery, the details of which will not be crucial in the following discussion.[3] That being said, it is important to understand how a theory based on inflectional competition accounts for sentences like (4), in order to understand the logic behind our experimental design. The idea is that the presupposition of the sentence in (4) is actually number neutral (i.e. every applicant has at least one experimental paper), but it competes with the singular version of the sentence in (5), which by assumption presupposes that every applicant has exactly one paper.

(5) Each applicant submitted **their experimental paper** as part of their application.

Subsequently, a pragmatic principle requires use of the sentence with a stronger presupposition whenever its presupposition is satisfied.[4] Therefore, this principle prevents (4) from being felicitously used whenever (5) is felicitous, and as a consequence, it seems as if the presupposition of (4) is stronger than the number-neutral presupposition.

It is not our purpose here to contribute directly to the literature on the plural in English and other similar languages, but from this literature, a question naturally arises as to what happens in languages with more number categories. In this paper, we will focus on Slovenian, which has

[2]Farkas & de Swart's (2010) theory also makes use of the idea of inflectional competition, but they use it to account for the meaning of the singular, which they assume is underlyingly number neutral.

[3]Križ (2017) puts forward a view that does not make use of inflectional competition, but instead crucially refers to 'homogeneity'. However, it is an open question how this view can explain the reading of (4), and more importantly, how it can be extended to the Slovenian dual in order to account for our experimental results.

[4]The principle is usually called Maximize Presupposition (Heim 1991, Percus 2006) but its nature and formulation are actively debated in the current theoretical literature and the principle is sometimes given different names (see Spector & Sudo 2017, Marty 2017, Anvari 2019, for example).

dual number, in addition to singular and plural. Dual nouns in this language are typically used to describe pairs of entities, but as we will explain below, the current literature lacks convincing evidence as to what part of this meaning comes from the underlying semantics, and what part arises through inflectional competition or some other additional mechanism.

2.2 The Slovenian Dual

In this subsection we will review key facts about the Slovenian dual. Slovenian marks nominal number by inflection. Both nouns and pronouns are marked for one of three number categories: singular, dual, or plural. For example, the neuter noun *mesto* 'town' declines as in Table 1 (see Marušič & Žaucer to appear and references therein for more comprehensive descriptions of the Slovenian number system).

	Singular	Dual	Plural
Nominative/Accusative	mest**o**	mest**i**	mest**a**
Dative	mest**u**	mest**oma**	mest**om**
Instrumental	mest**om**	mest**oma**	mest**i**
Locative	mest**u**	mest**ih**	mest**ih**
Genitive	mest**a**	mest	mest

Table 1: Inflection of *mesto* 'town'

Determiners, adjectives and verbs obligatorily agree in number, as in (6). Any mismatch in agreement here would result in ungrammaticality.

(6) T**a** dv**a** stol**a** st**a** polomljen**a**
these.DU.M.NOM two.DU.M.NOM chair.DU.M.NOM be.3.DU.PRES broken.DU.M.NOM
'These two chairs are broken.' (Derganc 2003:168)

It is mentioned in the literature that bare dual noun phrases in Slovenian tend to receive specific/definite interpretations, unlike bare singular and plural noun phrases, which are simply underspecified in this regard (Jakopin 1966, Dvořák & Sauerland 2006, Marušič & Žaucer to appear). For instance, Jakopin (1966:99) observes that (7a) and (7b) can receive generic interpretations, while (7c) only has a specific or definite interpretation.

(7) a. Otrok se rad igra.
child.SG REFL glad.SG.M play.SG
'A child likes to play.'
b. Otroci se radi igrajo.
child.PL REFL glad.PL.M play.PL
'Children like to play.'
c. Otroka se rada igrata.
child.DU REFL glad.DU.M play.DU
'The/our two children like to play.'

To make matters more complex, Marušič & Žaucer (to appear) further point out that a non-specific indefinite interpretation of a dual noun is actually available in certain cases. For example, they observe that (8) allows for a non-specific indefinite reading of the dual noun (see

also the experimental results reported in Marušič, Žaucer, Saksida, Sullivan, Skordos, Wang & Barner 2019).

(8) A si že videl enojajčna dvojčka?
Q AUX already seen one-egg.ADJ.ACC.DU twin.ACC.DU
'Have you ever seen (two) identical twins?' (Marušič & Žaucer to appear:(30))

It is left open here what the nature of this interpretive constraint is.

One peculiar fact about the Slovenian dual is that plural nouns are used for entities that naturally come in pairs, for example *noge* 'feet', as in (9) (Derganc 2003, Dvořák & Sauerland 2006, Sauerland 2008, Marušič & Žaucer to appear).[5]

(9) **Noge** me bolijo.
foot.PL me hurt.3.PRES
'My feet hurt.' (Derganc 2003:172)

While this is an interesting property of the Slovenian dual, we will not attempt to explain it in this paper.

Modifiers like *dva* 'two' and *oba* 'both' require the noun they modify to be in dual, even if it is a noun for naturally paired objects (Derganc 2003, Dvořák & Sauerland 2006, Martí 2020a). Thus, the noun is in dual, rather than plural, in (10), unlike in (9).

(10) **Obe nogi** me bolita.
both foot.DU me hurt.3.DU.PRES
'Both my feet hurt' (Derganc 2003:172)

The last thing to note is that there is a considerable degree of dialectal variation with respect to the extent of the presence of dual morphology, and some southern dialects possibly even lack dual as a nominal number altogether (Marušič, Žaucer, Plesničar, Razboršek, Sullivan & Barner 2016, Jakop 2008). However, most dialects exhibit a fair amount of dual morphology; the most widespread and influential central Slovenian dialects, to which most contemporary speakers have some exposure via the media, also exhibit a fair amount of dual morphology; and the standard variety, to which virtually all contemporary speakers have substantial exposure via the media and education, also exhibits robust dual marking (Marušič & Žaucer to appear). The experiments we report on below used the standard variety, and all crucial nouns were masculine, which is the gender that most commonly has distinct dual morphology across dialects.

2.3 Sauerland on the Slovenian Dual

Extending the competition-based analysis of singular vs. plural developed in Sauerland (2003) and Sauerland et al. (2005), Sauerland (2008) proposes that the Slovenian dual is semantically compatible with singular and dual reference, i.e. it semantically means 'one or two', while the singular is only compatible with singular reference, and the plural is number neutral, as in English (see also Dvořák & Sauerland 2006). In support of this claim, Sauerland (2008:75) raises the following example, which is structurally similar to the English example in (4).

[5] As Derganc (2003) and Marušič & Žaucer (to appear) observe, there are exceptions to this. For instance, the noun for 'twins' can appear in dual (*dvojčka*), as well as in plural (*dvojčki*), e.g. (8). Similarly for the noun for 'parents', *starša* (dual) vs. *starši* (plural).

(11) Vsak študent je prinesel s seboj svoj-i knjig-i.
every student be.SG brought.MASC with self his-DL book-DL
'Every student brought his book(s).'

Sauerland reports that this sentence is accepted in a context where some students brought exactly one book, while all the others brought exactly two. If the judgments were as Sauerland reports, (11) would indeed support his analysis. The reason is similar to how he accounts for (4), which we explained above. Specifically, according to Sauerland's analysis, (11) presupposes that every student has at least one but no more than two books, but the singular version of the sentence has a stronger presupposition that every student has exactly one book, which makes (11) infelicitous whenever every student has exactly one book. Thus (11) should be accepted when at least some, and possibly all, of the students have exactly two books.[6]

However, an informal survey with several native speakers we consulted suggests that (11) is actually unacceptable in the context described above. As we will see, the results of Experiment 1 verified this. If this is the case, what is the semantics of the Slovenian dual? To gain further empirical insight into this, we designed an experiment so that it also tested the interpretation of numerals, allowing them to be directly compared against the inflectional dual. In the next section we will explain the theoretical motivation behind this design.

3 The Dual vis-à-vis Numerals

While dual is a nominal inflectional category in the Slovenian number system on a par with singular and plural, its semantic function seems similar to the numeral 'two'. Moreover, the pronominal paradigm contains a morphological indication of the connection between dual and numerals. Specifically, the dual nominative pronouns contain what looks like the numeral 'two', i.e. *dva* for masculine and *dve* for feminine, as shown in Table 2.[7] Furthermore, the acquisition

	Singular	Dual		Plural	
		masc.	fem.	masc.	fem.
1st	jaz	mi**dva**	me**dve**/mi**dve**	mi	me
2nd	ti	vi**dva**	ve**dve**/vi**dve**	vi	ve
3rd	ona	ona**dva**	one**dve**/oni**dve**	oni	one

Table 2: Nominative pronouns in Slovenian

study conducted by Marušič et al. (2016) suggests that acquiring a language with dual seems to accelerate the acquisition of the numeral 'two'. Given these connections between dual and numerals, a question arises as to how similar their meanings are.

[6]One complication here is that Sauerland (2008) actually observes that the singular version of the sentence is acceptable in a context where some of the students brought exactly one book but others brought exactly two. To explain this, Sauerland suggests that the singular noun phrase may have a disjunctive representation that looks like *dve ali eno knjig-o* ('two or one book-SG'), and argues that for this reason the judgments of the singular version of the sentence are not informative with respect to its underlying semantics. As we will see, our experimental results replicated this observation, but they also indicated that the singular is acceptable to a similar extent when some of the students brought exactly two books and the others brought exactly three books, which is unexpected under Sauerland's concealed-disjunction analysis. We will discuss an alternative analysis later.

[7]Dual pronouns in non-nominative cases do not contain *dva/dve*, e.g. *naju* is the dual 1st pronoun in accusative, genitive and locative. See Marušič & Žaucer (to appear) for the entire paradigm. Also note that dual feminine pronouns each have two forms that are considered to be prescriptively acceptable, and both are listed here.

Numerals have been very intensively studied in both the theoretical and experimental literature, and there are three groups of theories about their meaning (see Spector 2013 for an overview). These theories agree that numerals sometimes give rise to a lower-bounded ('at least') reading, and sometimes to a bilateral ('exact') reading, but differ as to which reading reflects the underlying semantics.

1. Lexical Ambiguity Theory (Geurts 2006): Numerals are lexically ambiguous between the lower-bounded and bilateral readings.

2. Scalar Strengthening Theory (Horn 1972): The underlying semantics is lower-bounded, and the bilateral reading is derived via scalar strengthening with respect to the next numeral, which also has a lower-bounded semantics. For instance, *two* has a lower-bounded semantics ('at least 2'), but competes with *three* and yields a reading that *three* is false, which amounts to a bilateral reading ('at least 2, but not 3').

3. Pragmatic Weakening Theory (Breheny 2008): The underlying semantics is bilateral, and the lower-bounded reading is derived via pragmatic weakening.

Each of these theoretical views could be adopted for the Slovenian dual. According to the Lexical Ambiguity Theory, a dual noun is ambiguous between the lower-bounded and bilateral reading. Under the Scalar Strengthening Theory, the underlying semantics of the dual would be lower-bounded, but can be strengthened. Notice, however, that a question arises as to what could drive the strengthening mechanism. For the numeral *two*, *three* is a natural competitor, but could a numeral also be a competitor for a dual? Based on the fact that the dual is an inflectional category, one might expect its competitors to be other inflectional categories, in which case a numeral wouldn't be a competitor. But even under this assumption, the crucial competitor that gives rise to strengthening could be the plural, if it happens to mean 'more than two' in Slovenian. Thirdly, the Pragmatic Weakening Theory would assign a bilateral semantics to the dual and derive the lower-bounded reading by a mechanism of pragmatic weakening.

It is also possible that dual nouns never receive a lower-bounded reading, in which case, none of the above three theoretical options are warranted. In that case, its semantics should be simply bilateral. But if numerals can receive a lower-bounded reading, while dual nouns cannot, then the difference between them should somehow be explained.

Thus, we have several alternative theoretical options besides Sauerland (2008), and it will be theoretically informative to compare the interpretive behavior of the dual and a numeral in sentences like (11), which is precisely what we aimed to do in our experiments.

4 Experiment 1

4.1 Design and Procedure

Our objective was to gather acceptability judgments of sentences like (12), which have the same structure as Sauerland's example given in (11).

(12) Vsak moški je opral svoj-a avtomobil-a.
Every man aux washed self's-DL car-DL
'Every man washed his (two) cars.' DL

We compared sentences like this containing dual nouns (DL) with the versions of the sentences where the relevant noun is in the singular (SG), in the plural (PL), and in the dual – but occurring alongside the numeral *dva* 'two' (NUM).

(13) a. Vsak moški je opral svoj avtomobil.
Every man aux washed self's-SG car-SG
'Every man washed his car.' SG

b. Vsak moški je opral svoj-e avtomobil-e.
Every man aux washed self's-PL car-PL
'Every man washed his cars.' PL

c. Vsak moški je opral svoj-a dv-a avtomobil-a.
Every man aux washed self's-DL two-DL car-DL
'Every man washed his (two) cars.' NUM

As explained in Section 2, *dva* selects for a dual noun, so the items in the NUM condition also contained a dual noun. We opted for using *dva*, rather than some other numeral, in order to minimize the semantic difference between DL and NUM. Crucially, for all intents and purposes, a phrase like *dva avtomobila* 'two cars' is interpreted like its English counterpart *two cars*.

These target sentences were judged against three different types of contexts that differ as to how many relevant objects the individuals in the domain of quantification possess. For the above sentences, the three contexts are:

- [1 or 2]: Some men have exactly one car, the others have exactly two.
- [2 or 3]: Some men have exactly two cars, the others have exactly three.
- [exactly 2]: Every man has exactly two cars.

In each trial, the context was introduced in a yes/no-question following the target sentence, e.g. a Slovenian translation of 'Can one use this sentence in a situation where some men have one car and some men have two cars?', and answers were given by 'yes' or 'no'.[8]

Each participant was randomly assigned one of the three target contexts and saw all four types of number marking, 6 items each. 24 lexicalizations were created. Each participant saw each lexicalization exactly once. There were also 24 filler items interspersed with the target items, and the order of presentation was pseudo-randomized for each participant. The experiment was conducted online using Drummond's (2013) Ibex Farm (http://spellout.net/ibexfarm/). All the items are provided in the supplementary materials for this article.

4.2 Predictions

All theories predict that DL should be accepted in [exactly 2]. Thus, this condition acts as one baseline. According to Sauerland (2008), it should also be accepted in [1 or 2], but should be rejected in [2 or 3]. By contrast, under the view that the dual is similar to the numeral *dva* 'two' in allowing for a lower-bounded reading (which includes the three sub-theories mentioned in the previous section), DL is expected to be accepted in [2 or 3], but to be rejected in [1 or 2]. However, if the dual never receives a lower-bounded reading – say, unlike a numeral – then DL and NUM should differ in [2 or 3], but look similar in [exactly 2].

Note that for cases where competition with another item is involved, the results might turn out to be not so clear cut, given what is observed in many experimental studies on scalar items (e.g. Bott & Noveck 2004, van Tiel, van Miltenburg, Zevakhina & Geurts 2016; see Chemla & Singh 2014a,b, Noveck 2018 for overviews). However, it should be kept in mind that mild acceptability does not necessarily indicate the presence of competition.

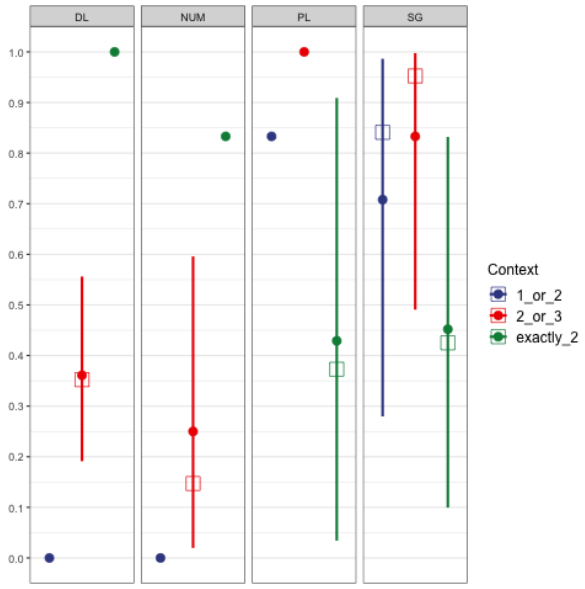
[8] Some items did not include the word *natančno* 'exactly' (cf. the supplementary file), but this had no effects in the results.

We are furthermore interested in the results with SG and PL. Sauerland (2008) predicts SG to be unacceptable in all three contexts, but remarks that it is actually accepted in [1 or 2] (see fn. 6). We would like to see whether or not that is replicated, and also whether or not SG behaves differently in [1 or 2] and [2 or 3]. The other theoretical options where the dual has a lower-bounded and/or bilateral meaning do not expect an effect of inflectional competition between SG and DL in any of the three contexts, on the assumption that the singular has bilateral ('exactly one') semantics and does not semantically overlap with the dual. As for PL, all theories predict some effects of competition in [2 or 3] and [exactly 2], as far as the semantics of the plural is as in English (see Marušič et al. 2016 for relevant discussion). However, it is possible that the presence of the dual in this language has some effect on the interpretation of the plural. In particular, if the plural means 'more than two', it should be rejected in all three contexts (modulo effects of inflectional competition), and in that case, the Scalar Strengthening Theory for the dual could rely on the plural as the crucial competitor for scalar strengthening.

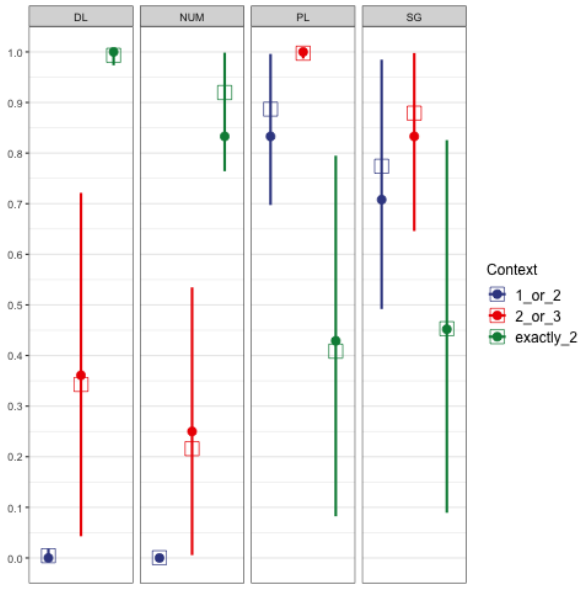
4.3 Participants

30 self-reported native speakers of Slovenian participated in the experiment. We excluded 9 of them for coming from a region where duals are not often used. The following statistical analyses are based on the data from the remaining 21 participants. Among them, 8 were assigned to [1 or 2], 7 to [exactly 2], and 6 to [2 or 3].

4.4 Results



(a) The filled dots indicate the observed proportions of *yes* answers, and the squares and the bars are the estimated proportions of *yes* answers and 95% confidence intervals predicted by the mixed effects logistic regression models, as described in the text.



(b) The filled dots indicate the observed proportions of *yes* answers, and the squares and the bars are the estimated proportions of *yes* answers and 95% credible intervals estimated by the Bayesian hierarchical logistic regression model with a weakly informed prior.

Figure 1: The observed and estimated proportions of *yes* answers in Experiment 1 with 95% confidence/credible intervals.

			MELR		Bayesian	
NUMBER	CONTEXT	*%Yes*	*%Yes*	95% CI	*%Yes*	95% CI
DL	[1 or 2]	0	—	—	0.4	[< 0.0001, 1.9]
	[2 or 3]	36.1	35.2	[19.1, 55.6]	34.3	[4.3, 72.1]
	[exactly 2]	100	—	—	99.3	[97.4, > 99.9]
NUM	[1 or 2]	0	—	—	< 0.001	[< 0.0001, 0.2]
	[2 or 3]	25	14.7	[2.0, 59.6]	21.6	[0.6, 53.5]
	[exactly 2]	100	—	—	92.0	[76.4, 99.8]
PL	[1 or 2]	83.3	—	—	88.7	[69.7, 99.6]
	[2 or 3]	100	—	—	99.7	[98.7, > 99.9]
	[exactly 2]	42.9	36.3	[1.6, 95.2]	40.9	[8.2, 79.5]
SG	[1 or 2]	70.8	84.1	[28.0, 98.6]	77.4	[49.2, 98.4]
	[2 or 3]	83.3	95.2	[49.1, 99.8]	87.9	[64.6, 99.8]
	[exactly 2]	45.2	42.5	[1.0, 83.2]	45.3	[8.9, 82.6]

Table 3: The observed proportions of *yes* answers in Experiment 1, and the estimations made by the mixed effects logistic regression models (MELR) and the hierarchical Bayesian model (Bayesian). 95% CI stands for 95% confidence intervals for the former and for 95% credible intervals.

Condition		β	SE	z	p
DL	Intercept	−0.609	0.425	−1.431	0.1530
NUM	Intercept	−1.758	1.095	−1.605	< 0.0001
PL	Intercept	−0.5199	1.4394	−0.361	0.7180
SG	Intercept	−0.3021	0.9698	−0.312	0.7554
	[1 or 2]	1.6659	1.3326	1.250	0.2113
	[2 or 3]	−2.9953	1.547	−1.936	0.0529

Table 4: The β-values, standard errors, z-values, and p-values of the logistic mixed effects models.

The proportion of *yes* answers in each experimental condition is as in Table 3 and visually represented by the filled dots in Figure 1. The squares and bars in Figure 1(a) are estimates of mixed effects logistic regression models fitted as described below and in Table 4, and those in Figure 1(b) are based on a Bayesian hierarchical logistic regression model with a weakly informative prior. The logistic mixed effects models were fitted using the `glmer` function of the `lme4` package (v. 1.1-23) (Douglas Bates and Martin Mächler and Ben Bolker and Steve Walker 2015) for the R statistics program (R Core Team 2020). The Bayesian model uses a weakly informative prior. The predictor variables are all dummy coded such that DL is the baseline for NUMBER, and NUM is in Experiment 3 and [exactly 2] is the baseline for CONTEXT. The

specification of the model for Experiment 1 looks like the following.

$$
\begin{aligned}
&\text{Answer}_i \sim \text{Binomial}(n_i, p_i) \\
&\text{logit}(p_i) = \alpha + \alpha_{\text{Subject}[i]} + \beta_{12} * C_{12} + \beta_{23} * C_{23} \\
&\qquad + (\beta_{NUM} + \beta_{NUM*C12} * C_{12}) * N_{\text{NUM}} + (\beta_{NUM} + \beta_{NUM*C_{23}} * C_{23}) * N_{\text{NUM}} \\
&\qquad + (\beta_{PL} + \beta_{PL*C_{12}} * C_{12}) * N_{\text{PL}} + (\beta_{PL} + \beta_{PL*C_{23}} * C_{23}) * N_{\text{PL}} \\
&\qquad + (\beta_{SG} + \beta_{PL*C_{12}} * C_{12}) * N_{\text{SG}} + (\beta_{SG} + \beta_{SG*C_{23}} * C_{23}) * N_{\text{SG}} \\
&\alpha_{\text{Subject}} = \text{Normal}(0, \sigma_{\text{Subject}})
\end{aligned}
$$

The prior distributions for the coefficients are all identical, being the normal distribution with $\mu = 0$ and $\sigma = 10$, while the prior distribution for σ_{Subject} is a half Cauchy distribution with $x_0 = 0$ and $\gamma = 1$. These are weakly informative priors (see McElreath 2020). The posterior distributions were estimated using three Hamilton Monte Carlo Markov Chains implemented in `Stan` constructed with 4000 samples each (of which 1500 were used for warm up).

The reason why two sets of statistical analyses are reported here is because the former cannot be fitted to the entire results of Experiment 1 due to (quasi-)separation caused by (near-)zero variance in some of the conditions (cf. Kimball, Shamtz, Eager & Roy 2019). We could have only reported the Bayesian model, but we decided to also report analyses based on mixed effects regression models, because they are currently more widely used, and also because what prior to use in Bayesian analyses might be a point of controversy. In the latter respect, we follow McElreath (2020) and use weakly informative prior, as specified above. Generally, we follow the advice of Cumming (2012, 2014) and Kruschke & Liddell (2018), and report point estimates and 95% confidence and 95% credible intervals (= 95% highest posterior density intervals), and do not exclusively rely on *p*-values (although we do not follow Cumming's advice to report no *p*-values). It should be noted that one added advantage of the Bayesian model for our purposes is that multiple comparisons can be performed without worrying about Type I error (Gelman, Hill & Yajima 2012, Kruschke 2014, Gelman, Carlin, Stern, Dunson, Vehtari & Rudin 2020). For this reason, we do not perform corrections for multiple comparisons for the mixed effects logistic regression models. As we will see, the two types of analyses point to the same conclusions here, as well as in the next two sections.

Let us start with DL. Unsurprisingly, it is accepted 100% of the time in [exactly 2], but crucially, it was never accepted in [1 or 2], and moderately accepted 36.1% of the time in [2 or 3]. Due to separation, a logistic regression cannot be fitted to the entire data of DL, but a mixed effects logistic regression with random intercepts for subjects was fitted to the data from [2 or 3].[9] According to this model the 95% confidence interval is [19.1%, 55.6%] with an estimated proportion of *yes* answers being 35.2%. Moreover, the 95% credible intervals estimated by the Bayesian model do not overlap with each other in this condition. We therefore conclude that the judgments for the three contexts are sufficiently distinct from each other.

The results of NUM look quite similar to the results of DL. Here too, a mixed effects logistic regression model cannot be fitted to all three contexts due to separation caused by the uniform data in [1 or 2] and the small variance in [exactly 2]. Instead, a mixed effects logistic regression model with random intercepts for subjects was fitted to the data from [2 or 3].[10] The 95% confidence interval for [2 or 3] is [2.0%, 59.6%] with an estimate proportion of *yes* answers being 14.7%, and the lower bound is close to 0%, reflecting the fact that three out of the six

[9] Including random intercepts for items seems to be superfluous, as indicated by an increase of 2.00 in AIC, and an increase of 3.58 in BIC.

[10] Including random intercepts for items results in an increased AIC (+1.84) and BIC (+3.43).

subjects answered *ne* 'no' to all six items in this condition. Similarly the 95% credible interval estimated by the Bayesian model is quite wide with the lower bound being at 0.6%. Thus, there is only weak evidence that [1 or 2] and [2 or 3] are different from each other in this case. On the other hand, the credible interval for [exactly 2] is very narrow and does not overlap with the other two conditions.

Let us now turn to PL. Again, due to separation, we cannot fit a logistic regression model to the entire data in this condition. A mixed effects logistic regression to the data of [exactly 2] with random intercepts for subjects resulted in a very wide confidence interval for this condition.[11] The 95% credible interval estimated by the Bayesian model for this condition is also wide, but does not overlap with that for [2 or 3], and only shows small overlap with [1 or 2] (about five percentage points). What is therefore notable here is that PL is perfectly accepted in [2 or 3] and its acceptability is quite high in [1 or 2] with variation. On the other hand, the acceptability in [exactly 2] is generally quite low, but large variation is observed and is not clearly rejected.

Lastly, let us look at SG. A mixed effects logistic regression model with random intercepts for subjects was fitted with [exactly 2] as the reference level.[12] The 95% confidence intervals estimated by this model for the three contexts are quite large, and overlap with each other by a large extent. So do the 95% credible intervals estimated by the Bayesian models. Therefore, there is no evidence for a difference among these three conditions. However, since these intervals do not include 0, it can be concluded that SG is not completely rejected in these conditions.

4.5 Discussion

To summarize the main findings (see also Figure 1):

- The results for DL and NUM are quite similar: They were rejected in [1 or 2], accepted in [exactly 2], and received intermediate acceptability in [2 or 3].
- PL is completely accepted in [2 or 3]. There is more variation in judgements in [1 or 2], but its acceptability is high. On the other hand, its acceptability is intermediate in [exactly 2].
- SG is mildly accepted in all three contexts.

The results for DL clearly speak against Sauerland's view, according to which the dual should be accepted in [1 or 2] but not in [2 or 3], the opposite of the observed pattern. Rather, they indicate that DL is *not* compatible with singular reference. Furthermore, the mild acceptability of DL in [2 or 3] suggests that the dual can have a lower-bounded interpretation, again contrary to Sauerland's proposal.

Of course, the fact that the acceptability of DL in [2 or 3] is not perfect calls for an explanation. Here, it is instructive to compare DL with NUM, which behaved similarly in our results. Recall the three theories of numerals discussed in Section 3: the Lexical Ambiguity Theory, the Scalar Strengthening Theory, and the Pragmatic Weakening Theory. All these theories are compatible with our results for NUM, at least with certain auxiliary assumptions. In order to account for the mild acceptability of NUM in [2 or 3], the Lexical Ambiguity Theory would need to assume that the lower-bounded reading is less readily accessible, at least in our experimental setting. The Scalar Strengthening Theory would explain this mild acceptability by assuming that the scalar strengthening is computed by default, but can be cancelled with some additional

[11] Including random intercepts for items results in overfit: +0.06 in AIC and +1.80 in BIC.

[12] As above, including random intercepts for items is superfluous, as indicated by increased AIC (+2.00) and BIC (+4.84).

processing cost. Due to this cost it was not cancelled uniformly in our experiment, resulting in mixed acceptability. Lastly, the Pragmatic Weakening Theory would have to assume that to arrive at the lower-bounded reading, one needs to go through some extra pragmatic reasoning, which can be assumed to be costly and/or not readily available in our experimental setting.

It is not our goal here to decide which of these theoretical possibilities best explains the results of NUM. Rather, we would like to discuss below whether or not each of the three theories extended to the dual can account for the results of DL.

Firstly, the Lexical Ambiguity Theory would assign two different underlying meanings to the dual, the lower-bounded meaning ('at least 2') and the bilateral meaning ('exactly 2'). This theory will be compatible with the data of DL under the assumption that the lower-bounded semantics is less readily accessible in DL, just like in NUM. However, there is a serious issue when the other conditions are taken into consideration. Recall, in particular, the results for PL, which show that the plural is accepted quite well in [1 or 2] and [2 or 3] (although the latter received more uniform judgments than the former). In particular, the high acceptability of PL in [1 or 2] suggests that the underlying semantics of the plural in Slovenian is number-neutral, just as in English (cf. Marušič et al. 2016). The results of Experiment 2, to be presented in the next section, confirm this. Furthermore, the current results also show that the plural is not perfectly accepted in [exactly 2], which is unlike in English (see the appendix for the version of the experiment run in English).

It seems reasonable to us to attribute the degraded acceptability in [exactly 2] to the presence of the dual in this language, which was uniformly accepted in [exactly 2] in our results. Thus, the results of [exactly 2] can be taken to suggest that the dual and the plural compete and the former is preferred in [exactly 2], as it has a more specific meaning. However, if that is the case and if the underlying semantics of the dual is ambiguous between the lower-bounded and bilateral interpretation, a similar competition effect should be expected in [2 or 3] as well, at least with respect to the lower-bounded interpretation. But no such competition effect was observed, and the plural was in fact uniformly accepted in [2 or 3]. For this reason, we think the Lexical Ambiguity Theory is not well supported by our data.

Secondly, let us consider the Scalar Strengthening Theory applied to the dual. By assumption, the underlying semantics of the dual will be lower-bounded ('at least 2'), and in order to account for the mild acceptability in [2 or 3], it will have to be additionally assumed that the lower-bounded semantics is enriched by default to the bilateral reading via scalar strengthening. However, if the scalar strengthening is due to a competition with another item with a stronger underlying meaning, it is unclear how that can be done with the dual, which, unlike a numeral, does not seem to have a natural competitor with a stronger meaning. In particular, as we remarked above, the results of PL suggest that the underlying semantics of the plural is number-neutral and weaker, rather than stronger.

To achieve scalar strengthening with the dual, we can think of two possible ways, both of which turn out to have further issues, however. One is that the relevant competitor looks like *tri* N_{pl} 'three N_{pl}', whose underlying semantics under the current hypothesis is lower-bounded at three ('at least three'). That is, just as it derives the bilateral reading of *dva* N_{dl} 'two N_{dl}', it derives the bilateral reading of the dual noun phrase. However, this possibility is not free from issues. In particular, it is known that the space of possible competitors for scalar strengthening needs to be sufficiently constrained, and generally, it appears that phrases that are structurally more complex do not give rise to scalar strengthening (see Katzir 2007, Breheny, Klinedinst, Romoli & Sudo 2018 and references therein). Obviously, *tri* N_{pl} is structurally more complex than a dual noun phrase without a numeral, and so is not expected to be a competitor for the latter. Furthermore, if such a noun phrase with a numeral competes with the dual, a similar

competition should happen to the plural. In particular, if *dva* N_{dl} 'two N_{dl}' competes with the plural, the latter would end up being upper-bounded at one ('exactly 1'), and should be at least only mildly accepted in [2 or 3]. However, in our results for PL, the plural was perfectly accepted in this condition, showing no such competition effects.[13] The other way to arrive at the necessary scalar strengthening is to assume that the mechanism for scalar strengthening does not require a linguistic alternative and simply turns the lower-bounded meaning of the dual ('at least 2') to an upper-bounded one ('exactly 2') (for such a view for scalar implicatures, see van Rooij & Schulz 2004, van Rooij 2017, for example). However, this would have to somehow explain why similar strengthening would not apply to the plural, turning it from lower-bounded ('at least 1') to upper-bounded ('exactly 1').

In sum, the Scalar Strengthening Theory would be compatible with the data under the assumption that scalar strengthening is computed by default, but there is a general question about how to achieve scalar strengthening to begin with in the absence of a natural competitor to the dual with a stronger meaning. In particular, it should somehow selectively apply to the dual, but not to the plural.

Finally, the Pragmatic Weakening Theory would assign the bilateral reading ('exactly 2') as the underlying semantics of the dual. Assuming that this is the default reading and that weakening it to the lower-bounded reading ('at least 2') incurs some cost, the mild acceptability of DL in [2 or 3] will be straightforwardly explained. Note that this theory, unlike the previous one, would not overgenerate for the plural. The plural has a number-neutral semantics, which is already very weak and cannot be weakened further. Thus, the overall conclusion is that our experimental data is best explained by the Pragmatic Weakening Theory.

It should also be mentioned that the results for SG are quite surprising. On the view that singular is only compatible with singular reference, i.e. it has a bilateral ('exactly 1') meaning, it should be unacceptable in all three contexts that we tested. However, it is actually accepted to a significant degree in all three contexts, especially in [1 or 2]. Importantly, we do not think these unexpected results indicate a flaw in the experimental design or procedure.[14] For one, NUM behaved as expected from what has been observed repeatedly for numerals across different experimental tasks. In addition, Sauerland (2008) reports similar intuitions of the singular in Slovenian (see fn. 6). Thus, it seems to us to be likely that the number inference of the singular in Slovenian is simply not as strong as one might expected under the standard view. In order to buttress this point, we ran two more experiments, which are reported in the next two sections. Their results also provide further support to the other conclusions we drew above.

5 Experiment 2

5.1 Design and Procedure

Experiment 2 is identical to Experiment 1, except that the contexts are changed to the following three.

- [exactly 2]: Every man has exactly two cars.

[13] Note that adding more alternatives would not help here. For instance, *exactly one N* is also an alternative, then that being a symmetric alternative with respect to *two Ns* with the lower-bounded reading, there would be no scalar implicature. But that means that the plural would not have a plurality inference anywhere, which is an unwelcome result. Also, if *exactly one N* is an alternative, *exactly two Ns*, *exactly three Ns*, should also be alternatives. In this case too, the plural would not have a plurality inference in any context.

[14] We thank Paul Marty (pers.comm.) for very helpful discussion on this point.

- [exactly 3]: Every man has exactly three cars.
- [3 or 4]: Some men have exactly three cars, the others have exactly four.

All the items are given in the supplementary document.

5.2 Predictions

The first context, [exactly 2], is meant to replicate the results of Experiment 1. In particular, we would like to replicate the effects of the competition with DL on PL. We also expect DL and NUM to exhibit perfect acceptability in this context, as in Experiment 1.

In the other two contexts, PL should be perfectly acceptable, and DL and NUM should show intermediate acceptability, as they should be able to optionally receive lower-bounded readings.

For SG, the theoretical prediction is that it should be unacceptable in all three contexts, but given the results of Experiment 1, it wouldn't be surprising to find intermediate acceptability in all three contexts.

5.3 Participants

35 self-reported native speakers of Slovenian were recruited on Prolific.ac and were paid £2.50 for their participation. The entire experiment took 14m08s on average (SD = 8m53s), which makes the average hourly rate 10.61£/hour. We excluded one participant for coming from an area where duals are not often used. Three other participants were also excluded for providing correct answers to less than 75% of the filler items. The following statistical analyses are based on the results from the remaining 31 participants. Among them, 10 were assigned to [exactly 2], 9 to [exactly 3], and 12 to [3 or 4].

5.4 Results

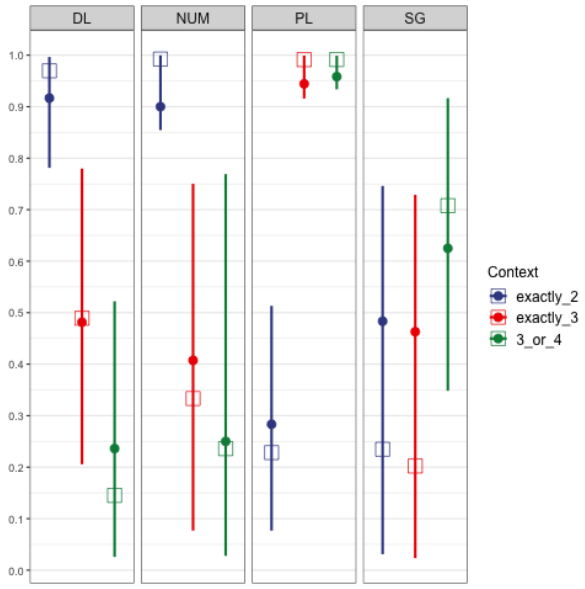
The results are summarized in Figure 2 and Table 5. Although there is no issue of separation, we also report an analysis based on a Bayesian hierarchical model for the sake of uniformity. Also, as mentioned in the previous section, Bayesian models allow for more straightforward interpretations of multiple comparisons. The model specification is essentially the same as in the model for Experiment 1, except that the context variables are renamed. The numerical details of the mixed effects logistic regression models are given in Table 6.

Starting from DL, it is accepted almost perfectly in [exactly 2] with the observed proportion of *yes* answers being 91.7%, and shows mild acceptance in [exactly 3] and [3 or 4] with the observed proportions of *yes* answers being 48.1% and 23.6%, respectively. A mixed effects logistic regression model with random intercepts for subjects was fitted to the data with [exactly 3] as the reference level.[15] The confidence/credible intervals in Figure 2 and Table 5, as well as the p-values in Table 6, indicate that [exactly 2] behaves differently from [exactly 3], while there is notevidence that [exactly 3] and [3 or 4] differ from each other.

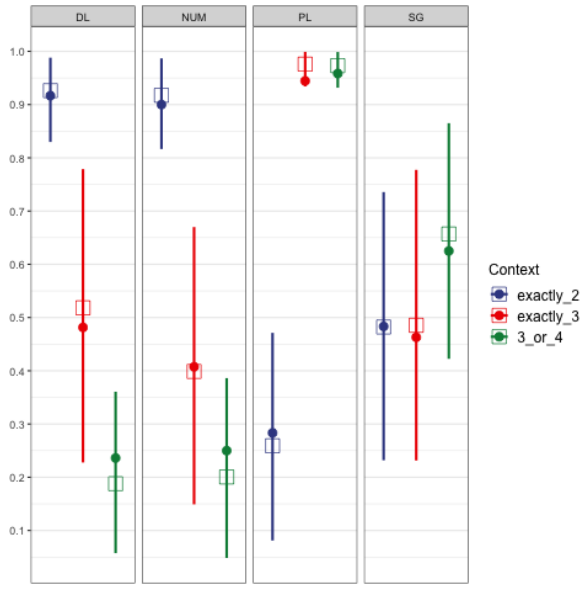
The results of NUM are quite similar to those of DL. A mixed effects logistic regression model similar to the one above was fitted to the data.[16] The confidence/credible intervals in Figure 2 and Table 5, as well as the p-values in Table 6, suggest that NUM is accepted far better

[15] A model with random intercepts for subjects resulted in a higher AIC (+2.00) and a higher BIC (+5.23), but points to the same conclusion.

[16] A model with random intercepts for subjects resulted in a higher AIC (+2.01) and a higher BIC (+5.23), but points to the same conclusion.



(a) The filled dots indicate the observed proportions of *yes* answers, and the squares and the bars are the estimated proportions of *yes* answers and 95% confidence intervals predicted by the mixed effects logistic regression models, as described in the text.



(b) The filled dots indicate the observed proportions of *yes* answers, and the squares and the bars are the estimated proportions of *yes* answers and 95% credible intervals estimated by the Bayesian hierarchical logistic regression model with a weakly informed prior.

Figure 2: The observed and estimated proportions of *yes* answers in Experiment 2 with 95% confidence/credible intervals.

			MELR		Bayesian	
NUMBER	CONTEXT	*%Yes*	*%Yes*	95% CI	*%Yes*	95% CI
DL	[exactly 2]	91.7	96.7	[78.1, 99.7]	92.6	[83.0, 98.8]
	[exactly 3]	48.1	48.9	[20.6, 78.0]	51.8	[22.8, 77.9]
	[3 or 4]	23.6	14.5	[2.6, 52.2]	18.8	[5.7, 36.1]
NUM	[exactly 2]	90.0	99.2	[85.5, 99.9]	91.7	[81.7, 98.7]
	[exactly 3]	40.7	33.3	[7.7, 75.0]	39.9	[14.9, 67.0]
	[3 or 4]	25.0	23.6	[2.8, 76.9]	20.1	[4.8, 38.6]
PL	[exactly 2]	28.3	22.8	[7.7, 51.3]	25.9	[8.1, 47.1]
	[exactly 3]	94.4	99.1	[91.6, 99.9]	97.6	[93.4, 99.9]
	[3 or 4]	95.8	99.2	[93.4, 99.9]	97.3	[93.2, 99.9]
SG	[exactly 2]	48.3	23.5	[3.1, 74.6]	48.2	[23.2, 73.6]
	[exactly 3]	46.3	20.2	[2.3, 72.9]	48.5	[23.2, 77.7]
	[3 or 4]	62.5	70.8	[34.8, 91.7]	65.7	[42.2, 86.5]

Table 5: The observed proportions of *yes* answers in Experiment 2, and the estimations made by the mixed effects logistic regression models (MELR) and the hierarchical Bayesian model (Bayesian). 95% CI stands for 95% confidence intervals for the former and for 95% credible intervals.

Condition		β	SE	z	p
DL	Intercept	−0.04306	0.66720	0.065	0.94854
	[exactly 2]	−3.46446	1.11799	−3.099	0.0019**
	[3 or 4]	−1.77364	0.95001	−1.867	0.0619
NUM	Intercept	−0.6933	0.9147	−0.758	0.4485***
	[exactly 2]	−48814	1.5870	−3.076	0.0021**
	[3 or 4]	−1.1745	1.2132	−0.968	0.3330
PL	Intercept	−1.2180	0.6484	−1.878	0.0603
	[exactly 3]	4.7409	1.2014	3.946	< 0.0001***
	[3 or 4]	4.7681	1.0834	4.401	< 0.0001***
SG	Intercept	0.885	0.771	1.148	0.2510
	[exactly 2]	−1.182	1.153	−1.025	0.3050
	[exactly 3]	−1.372	1.204	−1.139	0.2550

Table 6: The β-values, standard errors, z-values, and p-values of the logistic mixed effects models.

in [exactly 2] than in the other two contexts, and there is no evidence that its acceptance differs in [exactly 3] and [3 or 4] .

Turning to PL, it exhibits mild acceptability in [exactly 2] with *yes* answers being provided 28.3% of the time, while it is more or less perfectly accepted in the other two contexts. A mixed effects logistic regression model with random intercepts for subjects was fitted to the data, with [exactly 2] being the reference level.[17] As the confidence/credible intervals in Figure 2 and Table 5 and the p-values in in Table 6 indicate, [exactly 2] is noticeably different from the other two contexts. The 95% confidence and credible intervals for this condition are relatively wide but do not overlap with 0%, which suggests that PL is mildly acceptable in this context.

Finally, SG seemed to be accepted mildly in all three contexts. We fitted a mixed effects logistic regression model with random intercepts for subjects to the data.[18] Since it exhibited somewhat higher acceptance in [3 or 4] than in the other two contexts, we used [3 or 4] as the reference level in the statistical model. The 95% confidence/credible intervals are quite wide and overlap with each other quite a bit. Therefore, there is no evidence for any difference among these three conditions. But that the 95% credible intervals do not contain 0% suggests that SG is not completely unacceptable in any of them.

5.5 Discussion

To summarize the main findings of Experiment 2:

- The results for [exactly 2] essentially replicated the same condition in Experiment 1.
- As in Experiment 1, DL and NUM behaved similarly. Both of them are accepted in [exactly 2], but showed low acceptability in [exactly 3] and [3 or 4].

[17] To avoid non-convergence, the `nAGQ` parameter of the `glmer` function was set to 0. Also, a model with random intercepts for subjects resulted in a higher AIC (+2.00) and a higher BIC (+5.23), but points to the same conclusion.

[18] A model with random intercepts for subjects resulted in a higher AIC (+2.00) and a higher BIC (+5.22), but points to the same conclusion. To avoid non-convergence, the `nAGQ` parameter of the `glmer` function was set to 0.

- The results of PL in [exactly 2] can be seen as showing mild acceptability, as in Experiment 1. In [exactly 3] and [3 or 4], PL is accepted.

- In all three contexts SG showed mild acceptability.

Let us start with PL. That PL is accepted almost perfectly in [exactly 3] and [3 or 4] is expected under any view. We take its mild acceptance in [exactly 2], which was also observed in Experiment 2, to be due to competition with the dual. Overall, PL behaved as expected, which is evidence that there is no significant flaw in the experiment.

We also take the results of DL and NUM to be supporting the conclusions from the previous section. As in Experiment 1, they were almost perfectly accepted in [exactly 2]. The other two conditions are somewhat difficult to interpret, but the observed proportions of *yes* answers are comparable to the results of [2 or 3] in Experiment 1. Recall that according to the Pragmatic Weakening Theory, DL and NUM can be optionally weakened to have lower-bounded readings, hence these mild to low acceptance rates in [exactly 3] and [3 or 4] are as expected.

Lastly, SG behaved unexpectedly again. Contrary to the theoretical prediction that it should be rejected in all three contexts, it was accepted to some extent in all of them. Recall that it was expected to be rejected in all three conditions of Experiment 1 as well, but was actually accepted to similar extents. We therefore think that the number inference of the singular in Slovenian is actually weak, at least in this experimental task. Before discussing possible reasons behind this, we would like to know whether or not the singular in Slovenian is special. To find out, we ran a version of Experiment 1 in English. The results of this experiment also lend further support to the claim that the dual and the plural compete with each other in [exactly 2].

6 Experiment 3

6.1 Design and Procedure

Experiment 3 is constructed from Experiment 1 by translating all the materials into English. Since English has no dual, there are three types of target sentences, SG, NUM, and PL.

(14) a. Every man washed his car. SG
b. Every man washed his two cars. NUM
c. Every man washed his cars. PL

Thus, each participant saw 18 target sentences, instead of 24. As in Experiment 1, there are 24 filler items. All the items are in the supplementary document.

6.2 Predictions

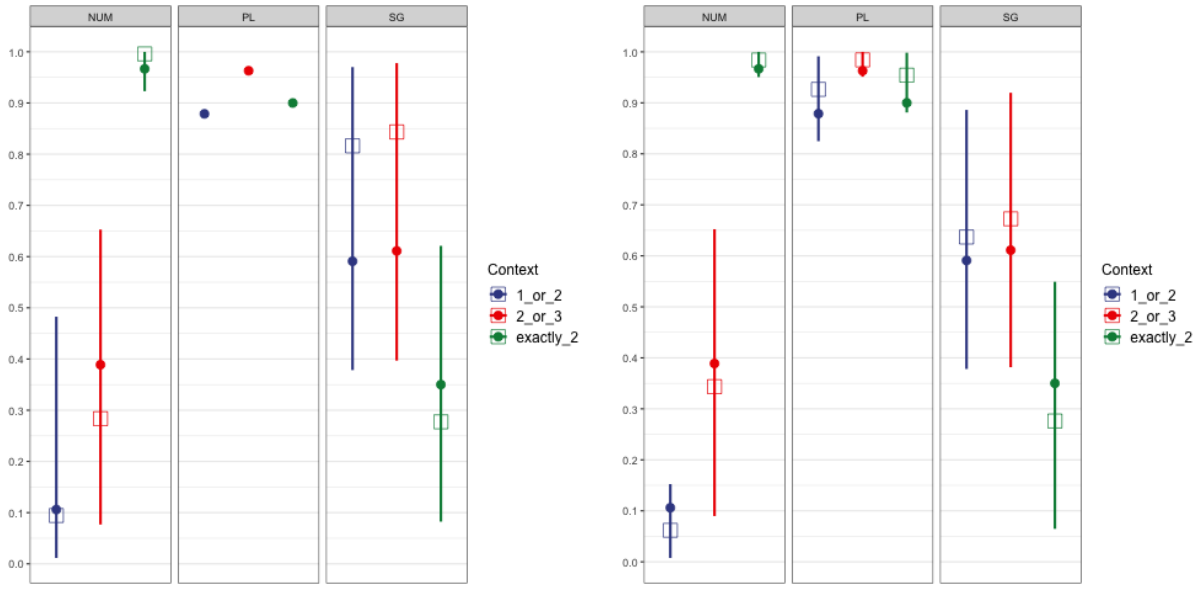
The lack of dual in English leads to two crucial predictions. Firstly, there should be no competition between the plural and the dual, so PL should be accepted in [exactly 2] as well as in [1 or 2] and [2 or 3], unlike in Experiment 1. Secondly, unlike in the Slovenian experiments, the NUM condition does not involve a dual noun, but a plural noun instead. This is important because strictly speaking, NUM in Experiments 1 and 2 involved dual nouns, so its similarity to DL could be attributed to the nominal number, rather than to the numeral. If NUM in the present experiment behaves similarly to DL in Experiment 1, that will give further support to the claim that the meaning of the dual is similar to the meaning of the numeral *two*.

Also, as we mentioned at the end of the previous section, we would like to see if the singular in English behaves similarly to the singular in Slovenian, given its unexpected behavior in Experiments 1 and 2. If that turns out to be the case, then the singular perhaps should be given a weaker semantics than usually assumed; if the two languages differ, on the other hand, its unexpected behavior in the Slovenian experiments should be given a language specific explanation, perhaps in relation to the presence of the dual in Slovenian.

6.3 Participants

32 self-reported native speakers of English were recruited on Prolific.ac, and were paid £2 for their participation. The entire experiment took 9m19s on average (SD=4m3s), which makes the average hourly rate 12.88£/hour. We excluded two participants for providing correct answers to less than 75% of the filler items. The following statistical analyses are based on the results from the remaining 30 participants. Among them, 11 were assigned to [1 or 2], 9 to [2 or 3], and 10 to [exactly 2].

6.4 Results



(a) The filled dots indicate the observed proportions of *yes* answers, and the squares and the bars are the estimated proportions of *yes* answers and 95% confidence intervals predicted by the mixed effects logistic regression models, as described in the text.

(b) The filled dots indicate the observed proportions of *yes* answers, and the squares and the bars are the estimated proportions of *yes* answers and 95% credible intervals estimated by the Bayesian hierarchical logistic regression model with a weakly informed prior.

Figure 3: The observed and estimated proportions of *yes* answers in Experiment 3 with 95% confidence/credible intervals.

The results are summarized in Figure 3 and Table 7. As in Experiment 1, the data were analyzed using mixed effects logistic regression models and a Bayesian hierarchical logistic regression model with a weakly informed prior. The numerical details of the mixed effects logistic regression models are given in Table 8. The model specification for the Bayesian model is just like in Experiment 1 except that the terms for DL are all removed. The main reason for

			MELR		Bayesian	
NUMBER	CONTEXT	*%Yes*	*%Yes*	95% CI	*%Yes*	95% CI
NUM	[1 or 2]	10.6	9.4	[1.2, 48.3]	6.2	[0.8, 15.2]
	[2 or 3]	38.9	28.3	[7.7, 65.3]	34.3	[9.0, 65.2]
	[exactly 2]	96.7	99.6	[92.3, > 99.9]	98.4	[95.0, 99.9]
PL	[1 or 2]	87.9	—	—	92.6	[82.5, 99.1]
	[2 or 3]	96.3	—	—	98.4	[95.1, 99.9]
	[exactly 2]	90.0	—	—	95.4	[88.1, 99.8]
SG	[1 or 2]	59.1	81.6	[37.8, 97.0]	63.7	[37.8, 88.6]
	[2 or 3]	61.1	84.3	[39.7, 97.8]	67.3	[38.2, 92.0]
	[exactly 2]	35.0	27.8	[8.2, 62.1]	24.6	[6.5, 54.9]

Table 7: The observed proportions of *yes* answers in Experiment 3, and the estimations made by the mixed effects logistic regression models (MELR) and the hierarchical Bayesian model (Bayesian). 95% CI stands for 95% confidence intervals for the former and for 95% credible intervals.

Condition		β	SE	z	p
NUM	Intercept	−0.9274	0.7953	−1.166	0.2436
	[1 or 2]	−2.2606	1.1179	−2.022	0.0432*
	[exactly 2]	5.4767	1.5267	3.587	0.0003***
SG	Intercept	−0.9577	0.7406	−1.293	0.1960
	[1 or 2]	1.4916	1.0143	1.471	0.1410
	[2 or 3]	1.6841	1.0727	1.570	0.1160

Table 8: The β-values, standard errors, z-values, and p-values of the logistic mixed effects models.

reporting both types of statistical analyses is because the PL condition shows ceiling effects, and the three levels of the fixed effect would be highly correlated, resulting in unreasonably high standard errors. We therefore did not fit a mixed effects logistic regression model to the data in this condition.

The results of NUM are quite similar to Experiment 1. A mixed effect logistic regression model with random slopes for subjects was fitted to this data with [exactly 2] being the reference level.[19] As the 95% confidence intervals in Table 7 and the p-values in Table 8 indicate, its acceptability is higher in [exactly 2] than in [2 or 3], which in turn is significantly higher than kin [1 or 2]. Although the 95% credible intervals predicted by the Bayesian model for [1 or 2] and [2 or 3] overlap by 6.2 percentage points, we take these results as suggesting that NUM is rejected in [1 or 2], mildly accepted in [2 or 3], and perfectly accepted in [exactly 2]. This is the same pattern as in Experiment 1.

Turning to PL, the observed proportions of *yes* answers are all high. As mentioned above, the high correlation among the different levels of the fixed effects prevents us from fitting a logistic

[19]Including random intercepts for items results in an increase in AIC (+2.00) and in BIC (+5.19). Also, that model results in similar p-values to what is reported in Table 8.

regression model to the data here, but the Bayesian model shows that there is no evidence for any difference among these three conditions. Crucially, this is a marked difference from Experiment 1, where PL was only mildly accepted in [exactly 2].

Finally, SG shows intermediate acceptance in all three conditions. A mixed effect logistic regression model with random slopes for subjects was fitted to this data.[20] [exactly 2] was taken to be the reference level in the model, as it shows somewhat attenuated acceptance compared to the other two contexts. As the overlapping 95% confidence intervals and the high *p*-values in Table 8 indicate, there is no evidence that there is a difference among the three contexts. The 95% credible intervals predicted by the Bayesian model also point to the same conclusion. Therefore, the results of SG essentially replicated the results of the same condition in Experiment 1.

6.5 Discussion

The main findings of this experiment are:

- As in Slovenian, NUM is rejected in [1 or 2], accepted in [exactly 2], and shows intermediate acceptability in [2 or 3].
- Unlike in Slovenian, PL is accepted in all three contexts.
- We replicated the results of SG in English.

The similarity of the results of NUM between Experiment 1 and the present experiment provides evidence that the similarity between DL and NUM in Experiment 1 is not explained by the use of dual nouns in these conditions. Bare dual nouns in Slovenian are similar in meaning to *two Ns* in English as well.

Another important finding is that PL is accepted in all three contexts, unlike in Experiment 1. This provides further support to the idea that the lower acceptability of PL in [exactly 2] in Experiment 1 is due to competition with the dual.

Lastly, in English too, SG is neither completely acceptable nor unacceptable in these three contexts, just as in Slovenian. This strongly suggests that this unexpected behavior of the singular is independent from the presence of the dual in Slovenian. One could interpret the intermediate acceptability of SG in all the contexts of all three experiments to be evidence that the core semantics of the singular is actually number neutral across languages. Farkas & de Swart (2010) put forward a theory based on this idea. They postulate some additional mechanisms to explain why the singular is typically used for singular reference, and why the plural often receives a plural reading in languages like English. As the semantics of the singular is not of our central concern in this paper, we will leave this question open here.

7 Conclusion

This paper reported on acceptability-judgment experiments whose results shed light on the underlying semantics of the Slovenian dual. As far as we know, this is the first experimental study on this topic (see Marušič et al. 2016, 2019 for acquisition studies). The discussion in the previous section resulted in a conclusion that the default reading of the dual is bilateral ('exactly 2') but it can optionally be weakened to a lower-bounded reading ('at least 2'). Recall, furthermore, that the acquisition study conducted by Marušič et al. (2016) suggests that acquiring a language

[20] Including random intercepts for items resulted in a higher AIC (+2.00) and a higher BIC (+5.19).

with dual seems to accelerate the acquisition of the numeral 'two'; in the present paper, we have now experimentally confirmed strong parallels between the semantics of the Slovenian dual and those of the numeral 'two', thereby furnishing a theoretical foundation for these previous empirical findings.

It is important to note that the bilateral semantics for the dual is compatible with a claim that a whole dual noun phrase may receive, and even prefers, a lower-bounded reading (as claimed by Marušič et al. 2019, for example). This is because even if the dual noun itself has a bilateral reading, an existential quantifier can turn the meaning of the whole DP lower-bounded. Concretely, even if the dual noun *avtomobila* 'car.DL' has a bilateral reading and is only true of pairs of cars, an existential sentence like *John has car.DU*, will have lower-bounded truth-conditions: the sentence will be true if there is a plurality that consists of exactly two cars that John owns, which is to say, John has at least two cars. Note that in our experiments, we only tested possessive constructions, which are generally definite (but see the discussion of the results for the singular) and more revealing with respect to the underlying semantics of the dual.

Needless to say, our experimental results do not answer all questions about the semantics and pragmatics of the dual. In particular they have very little to say about the exact nature of the mechanism for pragmatic weakening, other than that it should not be available very freely, at least in experimental settings like ours. For instance, Breheny (2008), who focuses on the interpretation of numerals in English, suggests that the lower-bounded ('at least') interpretation comes about via background implicature or Stalnaker's (1978) diagonalization applied to specific readings, but our experimental results are not informative with respect to the feasibility of this idea.[21] In fact, the results are compatible with an alternative mechanism for pragmatic weakening, e.g. domain restriction. The idea is that the dual showed mild acceptance in [2 or 3] because implicit domain restriction was performed, with which the bilateral reading becomes true in [2 or 3]. For instance, for (12), repeated here, what needs to be accommodated is some natural way to map each (relevant) man to two of his cars.

(12) Vsak moški je opral svoj-a avtomobil-a.
Every man aux washed self's-DL car-DL
'Every man washed his (two) cars.'

We have to leave questions about the mechanism of pragmatic weakening open for future research.

In closing, perhaps one of the conclusions of broadest interest are what the pattern of results, and indeed, the presence of a dual number category in Slovenian reveals about the plural. Recall that PL showed intermediate acceptability in [exactly 2] contexts in Slovenian, but not in English. This suggests that the presence of a dual incurs inflectional competition, and that while the morphosemantics of the plural may be uniform across these two languages, their morphopragmatics are not.

References

Anvari, Amir. 2019. *Aspects of Contextual Enrichment*: Ecole Normale Supérieure dissertation.
Bale, Alan, Michaël Gagnon & Hrayr Khanjian. 2011. On the relationship between

[21] It should however be noted that the mechanisms that Breheny (2008) suggests might not be straightforwardly applied to the sentences we tested. Background implicatures are not meant to explain this type of sentences to begin with. Furthermore, it is not immediately clear how the idea of diagonalization can be applied to sentences with quantifiers.

morphological and semantic markedness. *Morphology* 21(2). 197–221. doi:10.1007/s11525-010-9158-1.

Bott, Lewis & Ira A. Noveck. 2004. Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language* 51(3). 437–457. doi:10.1016/j.jml.2004.05.006.

Breheny, Richard. 2008. A new look at the semantics and pragmatics of numerically quantified noun phrases. *Journal of Semantics* 25(2). 93–139. doi:10.1093/jos/ffm016.

Breheny, Richard, Nathan Klinedinst, Jacopo Romoli & Yasutada Sudo. 2018. The symmetry problem: current theories and prospects. *Natural Language Semantics* 26(2). 85–110. doi:https://doi.org/10.1007/s11050-017-9141-z.

Chemla, Emmanuel & Raj Singh. 2014a. Remarks on the experimental turn in the study of scalar implicature, Part I. *Language and Linguistics Compass* 8(9). 373–386. doi:10.1111/lnc3.12081.

Chemla, Emmanuel & Raj Singh. 2014b. Remarks on the experimental turn in the study of scalar implicature, Part II. *Language and Linguistics Compass* 8(9). 387–399. doi:10.1111/lnc3.12081.

Cumming, Geoff. 2012. *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. New York: Routledge.

Cumming, Geoff. 2014. The new statistics: why and how. *Psychological Science* 25(1). 7–29. doi:10.1177/0956797613504966.

Derganc, Aleksandra. 2003. The dual in Slovenian. In Janez Orešnik & Donald D. Reindl (eds.), *Slovenian from a Typological Perspective*, 165–181. Berlin: Akademie Verlag.

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

Drummond, Alex. 2013. Ibex farm. *Online server: http://spellout.net/ibexfarm* .

Dvořák, Boštjan & Uli Sauerland. 2006. The semantics of the Slovenian dual. In Hana Filip, James Lavine, Steven Frans & Mila Tasseva-Kurktchieva (eds.), *Proceedings of FASL 14*, 98–112.

Farkas, Donka & Henriëtte de Swart. 2010. The semantics and pragmatics of plurals. *Semantics & Pragmatics* 3. 1–54.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David Dunson, Aki Vehtari & Donald B. Rudin. 2020. *Bayesian Data Analysis*. Boca Raton, FL: CRC Press 3rd edn.

Gelman, Andrew, Jennifer Hill & Masanao Yajima. 2012. Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness* 5(2). 189–211. doi:10.1080/19345747.2011.618213.

Geurts, Bart. 2006. Take 'five': The meaning and use of a number word. In *Non-definiteness and plurality*, John Benjamins.

Grimm, Scott. 2013. Plurality is distinct from number-neutrality. In *Proceedings of NELS 41*, Amherst, MA: GLSA.

Harbour, Daniel. 2014. Paucity, abundance and the theory of number. *Language* 90(1). 185–229. doi:10.1353/lan.2014.0003.

Heim, Irene. 1991. Artikel und Definitheit. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung/Semantics: An International Handbook of Contemporary Research*, 487–535. Berlin: de Gruyter.

Horn, Laurence. 1972. *On the Semantic Properties of the Logical Operators*: University of California Los Angeles dissertation.

Ivlieva, Natalia. 2013. *Scalar Implicatures and the Grammar of Plurality and Disjunction*:

Massachusetts Institute of Technology dissertation.
Jakop, Tjaša. 2008. *The Dual in Slovene Dialects*. Bochum: Universitätsverlag Brockmeyer.
Jakopin, Franc. 1966. Slovenska dvojina in jezikovne plasti. *Jezik in Slovstvo* 11(4). 98–104.
Katzir, Roni. 2007. Structurally-defined alternatives. *Linguistics and Philosophy* 30(6). 669–690. doi:10.1007/s10988-008-9029-y.
Kimball, Amelia E., Kailen Shamtz, Christopher Eager & Joseph Roy. 2019. Confronting quasi-separation in logistic mixed effects for linguistic data: a Bayesian approach. *Journal of Quantitative Linguistics* doi:10.1080/09296174.2018.1499457.
Kiparsky, Paul & Judith Tonhauser. 2012. Semantics of inflection. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics*, vol. 3, 2070–2097. Berlin: de Gruyter.
Križ, Manuel. 2017. Bare plurals, multiplicity, and homogeneity. Ms., Institut Jean Nicod.
Kruschke, John K. 2014. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Cambridge, MA: Academic Press 2nd edn.
Kruschke, John K. & Torrin M. Liddell. 2018. The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bullitin & Review* 25. 178–206. doi:10.3758/s13423-016-1221-4.
Martí, Luisa. 2020a. Dual number and the typology of the numeral-noun construction. Ms., Queen Mary University of London.
Martí, Luisa. 2020b. Inclusive plurals and the theory of number. *Linguistic Inquiry* 51(1). 37–74. doi:https://doi.org/10.1162/ling_a_00330.
Marty, Paul P. 2017. *Implicatures in the DP Domain*: Massachusetts Institute of Technology dissertation.
Marušič, Franc & Rok Žaucer. to appear. Case study: Slovenian dual. In Patricia Cabredo Hofherr & Jenny Doetjes (eds.), *The Oxford Handbook of Grammatical Number*, Oxford: Oxford University Press.
Marušič, Franc, Rok Žaucer, Vesna Plesničar, Tina Razboršek, Jessica Sullivan & David Barner. 2016. Does grammatical structure speed number word learning?: Evidence from learners of dual and non-dual dialects of Slovenian. *Plos ONE* .
Marušič, Franc, Rok Žaucer, Amanda Saksida, Jessica Sullivan, Dimitrios Skordos, Yiqiao Wang & David Barner. 2019. Children derive exact meanings pragmatically: Evidence from a dual morphology language. Ms.
Mayr, Clemens. 2015. Plural definite NPs presuppose multiplicity via embedded exhaustification. In *Proceedings of SALT 25*, 204–224.
McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press 2nd edn.
Noveck, Ira A. 2018. *Experimental Pragmatics*. Cambridge: Cambridge University Press.
Pearson, Hazel, Manizeh Khan & Jesse Snedeker. 2010. Even more evidence for the emptiness of plurality: An experimental investigation of plural interpretation as a species of implicature. In *Proceedings of SALT 20*, 489–508.
Percus, Orin. 2006. Antipresuppositions. Tech. rep. Japan Society for the Promotion of Science. Report of the Grant-in-Aid for Scientific Research (B), Project No. 15320052.
R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. https://www.R-project.org/.
van Rooij, Robert & Katrin Schulz. 2004. Exhaustive interpretation of complex sentences. *Journal of Logic, Language and Information* 13(4). 491–519.
Sauerland, Uli. 2003. A new semantics for number. In Robert B. Young & Yuping Zhou (eds.), *Proceedings of SALT 13*, 258–275. Ithaca, NY: Cornell Linguistics Club.
Sauerland, Uli. 2008. On the semantic markedness of phi-features. In Daniel Harbour, David

Adger & Susana Béjar (eds.), *Phi Theory: Phi-Features across Modules and Interfaces*, 57–82. Oxford: Oxford University Press.

Sauerland, Uli, Jan Anderssen & Kazuko Yatsushiro. 2005. The plural is semantically unmarked. In Stephan Kepser & Marga Reise (eds.), *Linguistic Evidence*, 409–430. Berlin: Mouton de Gruyter.

Spector, Benjamin. 2007. Aspects of the pragmatics of plural morphology: On higher-order implicatures. In Uli Sauerland & Penka Stateva (eds.), *Presuppositions and Implicatures in Compositional Semantics*, 243–281. New York: Palgrave-Macmillan.

Spector, Benjamin. 2013. Bare numerals and scalar implicatures. *Language and Linguistics Compass* 7(5). 273–294. doi:10.1111/lnc3.12018.

Spector, Benjamin & Yasutada Sudo. 2017. Presupposed ignorance and exhaustification: how scalar implicatures and presuppositions interact. *Linguistics and Philosophy* 40(5). 473–517. doi:https://doi.org/10.1007/s10988-017-9208-9.

Stalnaker, Robert. 1978. Assertion. In Peter Cole (ed.), *Syntax and Semantics 9: Pragmatics*, 315–332. New York: Academic Press.

Sudo, Yasutada. 2019. The plurality inference as a quantity implicature. Ms., UCL.

van Tiel, Bob, Emiel van Miltenburg, Natalia Zevakhina & Bart Geurts. 2016. Scalar diversity. *Journal of Semantics* 33(1). 137–175.

van Rooij, Robert. 2017. A fine-grained global analysis of implicatures. In Salvatore Pistoia-Reda & Filippo Domaneschi (eds.), *Linguistic and Psycholinguistic Approaches to Implicatures and Presuppositions*, 73–110. Palgrave McMillan.

Zweig, Eytan. 2009. Number-neutral bare plurals and the multiplicity implicature. *Linguistics and Philosophy* 32(4). 353–407. doi:10.1007/s10988-009-9064-3.