

Franc Lanko Marušič
Petra Mišmaš
Rok Žaucer¹

Looking for Cognitive Foundations of Functional Sequences

Abstract

With the multiplication of various functional projections, syntactic structures became very complex entities. Approaches like Cartography (e.g. Cinque and Rizzi 2008) went one step further than most other approaches, proposing that each sentence comprises of a number of universal, strictly ordered functional projections. In the noun phrase, the strictly ordered functional projections are said to be responsible not only for the relative order of numerals, demonstratives and nouns (cf. Cinque 2005), but also for the universal order of various types of adjectives (cf. Hetzron 1978; Sproat and Shih 1991; Cinque 1994; Scott 2002, etc.). Cinque and Rizzi (2008) discuss possible origins of the many hierarchies of functional projections and suggest that they might derive from general cognition. If cognition and its restrictions are behind the hierarchy of functional projections, then the order of projections hosting adjectives should be reflected in various non-linguistic cognitive processes. We designed several experiments to test this hypothesis. Our experiments did not confirm our hypothesis; but as we have also identified problems in the design of our experiments, our results do not warrant a clear rejection of the hypothesis either.

Keywords

noun phrase, adjective ordering restrictions, functional hierarchy, experimental syntax, cognitive foundations of syntax

¹ This work was supported by the Slovenian Research Agency through the research project J6-7282 (PI Rok Žaucer) and research program P2-0382 (PI Franc Marušič).

1. Background

Sentences are composed of lexical words that are linked together with the help of grammatical/function words. Whereas the number of different lexical phrases is fixed by the number of different lexical categories (verbs head the VP, nouns head the NP, etc.), the list of different functional categories, and in particular functional syntactic projections/phrases, is much longer and still growing. Just like the sheer identity of functional projections, the question of their relative order in the syntactic structure has also not been resolved. The main functional projections making up the core spine of the entire sentence are widely accepted to be ordered in the hierarchy CP > TP > vP, but these are far from being the only functional projections that have been proposed. An extreme position in this respect is represented by the cartographic approach to syntax (Cinque and Rizzi 2008). Rizzi (1997) and Cinque (1999) proposed that each of these three core functional projections of the clausal spine should itself actually be seen as consisting of a number of projections, altogether forming a sequence of 35 separate, strictly ordered projections. The situation is similar in the noun phrase domain, which cartographic research has argued to consist not only of the sequence DP > NP but of a much more elaborate sequence.

Cartography focuses on the functional make-up of syntactic structures and looks for the smallest amount of syntactic material that minimally increases a syntactic unit, i.e. features, and the order in which they are introduced into the structure. In the center of the cartographic approach is the assumption that lexical heads are dominated by hierarchically ordered functional projections. The head of each functional projection hosts one feature, which is related to one (morphosyntactic) property (Cinque and Rizzi 2008). Cartography “draws” precise maps of these functional hierarchies.

Evidence for the relative order within sequences of functional projections typically comes from crosslinguistic restrictions on the order of various elements. Such observations are in the center of the cartographic approach. Cinque (1999) and Rizzi (1997) developed their view of a more elaborate landscape of the CP-, TP-, and vP-domains describing crosslinguistic similarities in the word order of adverbs in the IP/vP area and the order of different elements in the CP, respectively. But many word-order generalizations have been made also for noun-phrase internal elements. For example, Greenberg (1963) described an ordering restriction on cooccurring demonstratives, numerals and adjectives and proposed a generalization stating that in languages with prenominal demonstratives, numerals and adjectives appear in the order Dem > Num > A > N. Another old observation suggests that attributive adjectives crosslinguistically come in a uniform order (cf. Hetzron 1978; Sproat and Shih 1991, etc.).

When charting the hierarchies of functional projections, researchers typically focus on the number of functional projections and their order. In this

paper, however, we build on previous studies of the hierarchy of functional projections that host adjectives in the noun phrase to explore another fundamental question asked by cartography: What is the source of the functional hierarchies? We address this question by exploring whether we can find traces of the proposed hierarchy outside of language, and more specifically, whether the universal hierarchy of functional projections is related to general cognition. Asking a similar question about the universal sequence *demonstrative* > *numeral* > *adjective* > *noun* (Greenberg's 1963, generalization 20), Culbertson and Adger (2014) argue that it is rooted in general cognition. Even more related to the topic of this paper are the findings of Scontras et al. (2017) and Kotowski and Hartl (2019), who argue that the universal sequence of adjectives has origins in more general cognitive principles. To motivate the approach that we are taking to address the above question, we need to have a closer look at the functional make-up of the noun phrase.

2. The hierarchy of functional projections in the noun phrase

As mentioned above, the fundamental assumption we make in this paper is that languages share uniform sequences of functional projections (following Rizzi 1997; Cinque 1999; Cinque and Rizzi 2008, etc.). Furthermore, we assume that these functional projections come on top of lexical heads and appear in a fixed universal order. The specific sequence we intend to investigate is the one that has been proposed to exist inside the noun phrase, the lexical core of which is, naturally, the noun. The functional projections above N^0 host various adnominal elements, including (but not limited to) determiners, quantifiers, demonstratives, numerals, and attributive adjectives. We will focus on the part of the sequence that is responsible for attributive adjectives.

Crosslinguistically, attributive adjectives typically appear in a fixed word order (Hetzron 1978; Dixon 1982; Sproat and Shih 1991, etc.). For example, English (1a) and Slovenian (2a) are acceptable, whereas English (1b) and Slovenian (2b) are not.

- | | | | | |
|-----|----|---|------|---|
| (1) | a. | a huge green book | b. * | a green huge book |
| (2) | a. | ogromna zelena knjiga
huge green book
'huge green book' | b. * | zelena ogromna knjiga
green huge book
'green huge book' |

For the cartographic model, this is a direct result of the hierarchy of functional projections that host attributive adjectives. This idea, developed from Cinque's (1999) proposal for adverbs in the IP, is in the center of Scott's (2002) proposal for adjective ordering restrictions (AOR), which are taken to be a direct

consequence of the hierarchy of functional projections. Importantly, when determining the order of adjectives to chart the landscape of functional projections in the NP, adjectives orders should be compared in neutral intonation, i.e. without a comma intonation, focus, etc. The functional field of the NP has been charted to varying degrees by several authors,² with Scott (2002) advancing what is perhaps the most detailed proposal for a universal hierarchy of functional projections which host attributive adjectives, a hierarchy which also includes several classes of adjectives which had previously not been considered. As in all similar approaches to AORs, adjectives fall into separate classes based on their meaning: for example, *young*, *old*, *green*, *ancient*, *modern*, *elderly*, etc., fall into the AGE class, adjectives like *long* and *short* into the LENGTH class, etc. The “map” of the functional projections that host attributive adjectives therefore includes the ordering of these classes. Finally, under this view, adjectives themselves appear in the specifiers of functional projections nested between the DP and the NP. The order of these projections is shown in (3) below:

- (3) DETERMINER > ORDINAL NUMBER > CARDINAL NUMBER > SUBJECTIVE COMMENT > ? EVIDENTIAL > SIZE > LENGTH > HEIGHT > SPEED > ? DEPTH > WIDTH > WEIGHT > TEMPERATURE > ? WETNESS > AGE > SHAPE > COLOR > NATIONALITY/ ORIGIN > MATERIAL > COMPOUND ELEMENT > NP (Scott 2002: 114, [47]).

For cartographers, then, AORs fall out as a natural consequence of this proposed hierarchy. Concretely, the acceptability of (4a) in contrast to (4b) below is a consequence of the existence of this hierarchy: since the functional projection that hosts adjectives for age dominates the projection that hosts adjectives for color, the acceptable linear word order is the one in which the adjective *old* precedes the adjective *red*.

- (4) a. an old red book b. *a red old book

As Scott (2002: 110) notes, the hierarchy of FPs in the DP is not language specific but is rather “almost certainly” universal with respect to hierarchical direct modification, though perhaps not with respect to the specific fine-grained order of FPs. Scott (2002) provides supporting data from English, Finnish, Ibibio, Malayalam, Welsh, Swedish, Serbo-Croat and German. At least parts of this hierarchy have also been proposed as holding in Italian (Cinque 1994; Ramaglia 2011) and Russian (Pereltsvaig 2007). We will not go into the details

² For example, Dixon (1982) arrives at the order in (i), Sproat and Shih (1991) at the order of adjectives in (ii), and Cinque (1994) proposes the structure in (iii) for object denoting nominals:

- (i) VALUE > DIMENSION > PHYSICAL PROPERTY > SPEED > HUMAN PROPENSITY > AGE > COLOR
(ii) QUALITY > SIZE > SHAPE > COLOR > PROVENANCE
(iii) POSSESSIVE > ORDINAL NUM > QUALITY > SIZE > SHAPE > COLOR > NATIONALITY (i.e. PROVENANCE)

of Scott's hierarchy or try to find further crosslinguistic confirmation for the entire hierarchy, but we do want to mention that the same hierarchy has (at least partially) been found to hold also in Slovenian. The same order was in fact described for Slovenian already in traditional grammars, for example Toporišič (2000):

- (5) SUBJECTIVE PROPERTY (*lovely*) > OBJECTIVE PROPERTY (*small/beautiful/wild ...*) > AGE (*young/old/new/modern...*) > COLOR (*red*) > PARTICIPLES (*broken*) > SOURCE/NATIONALITY (*German*)

And a subset of this order was empirically confirmed for Slovenian by Plesničar (2017), who focuses on adjectives for size, shape, color, and origin and shows that these appear in the order in (6).

- (6) SIZE > SHAPE (>) COLOR > ORIGIN

However, Plesničar (2017) also notes that in Slovenian adjectives for shape and color are not completely irreversible. This is not completely unexpected, Pereltsvaig (2007) similarly found that different orders are often found equally acceptable for pairs of adjectives that are members of groups that are close in the hierarchy. Also important is the fact that cross-linguistic variations in word order of adjectives are not excluded, as Scott (2002) himself notes, and can be a consequence of, for example, movement (Cinque 2005).³

To sum up, given these cross-linguistic data, we assume a universal hierarchy of functional projections hosting adjectives inside the noun phrase. Given the existence of a sequence of functional projections, one question that needs to be addressed is where the hierarchy of functional projections comes from.⁴ Naturally, our question could be restated in a theory that does not accept the existence of functional projections, namely, we would be asking what extragrammatical properties influence the universal adjective order. This is the approach of Scontras et al. (2017) and Kotowski and Hartl (2019). Given our starting assumptions, we will keep talking about the functional hierarchy, but we are really interested in finding extragrammatical basis that ultimately result in cross-linguistically fixed order of adjectives.

³ Other factors also play a role, for example, Pereltsvaig (2007: 65) observed that English speakers preferred shorter adjectives to come before long adjectives (length is measured in the number of syllables). Kotowski and Hartl (2019) discuss temporariness and weight as potential factors and conclude that they do not reliably predict adjective order, although this need not mean that they never contribute to ordering preferences with adjective sequences.

⁴ Other questions, naturally, can also be asked. One is the question of the exhaustiveness of the currently proposed classes: are there other classes of adjectives in the NP (e.g. a class of adjectives for taste, smell, etc.), and if yes, where in the hierarchy is their place. Another is what types of movement can take place in the NP.

2.1. The source of universal hierarchy of functional projections

Different proposals have been made in the past with respect to the source of the universal hierarchies, and this section will briefly review them. A somewhat radical suggestion is that all of these hierarchies are simply part of Universal Grammar, as proposed by Scott (2002), who notes that “[c]onjectures as to the psycholinguistic motivation for AOR need not be posed: AOR fall out as a direct consequence of UG” (Scott 2002: 97). But of course, this only changes the question to ‘What is the source of this hierarchy in UG?’ Cinque and Rizzi (2008) note that it seems implausible that hierarchies would be part of a UG that is completely isolated from other cognitive systems, just as it seems implausible that they are solely a cultural property. So even if they are part of UG, they might be somehow related and originating in other cognitive systems.

Recently, Scontras et al. (2017) argued that AORs are a result of the properties adjectives denote, whereby less subjective adjectives are placed closer to the noun than more subjective adjectives. It seems, therefore, that AORs are claimed to be based on an extralinguistic criterion. More concretely, Scontras et al. (2017) found a correlation between the distance of the adjective from the noun and the perceived average subjectivity of the adjective group it belongs to. Note, however, that given this finding, we can also expect – everything else being equal (length, morphological makeup, etc.) – that not all adjectives *within* a group will pattern in exactly the same way, as there can be different levels of subjectivity also within a group. Consider, for example, *boiling* and *warm*. While both seem to belong to the same group of ‘temperature’, *boiling* clearly has a less subjective meaning than *warm*. If it is subjectivity that determines adjective positioning, one would expect *boiling* to appear closer to the noun than other, more subjective temperature adjectives. Similarly, the suffix *-ish* increases the subjectivity of the adjective it attaches to, with *reddish* thus being more subjective than the *red*. Given Scontras et al. (2017) finding, we would then expect to find *reddish* further away from the noun than the less subjective color adjective *red*. Given our intuition about the order of Slovenian adjectives, none of these predictions are borne out, as shown in (7) and (8) respectively.

- (7) a. ledeno-mrzlo staro nemško pivo b. hladno staro nemško pivo
 ice-cold old German beer cold old German beer
- (8) a. kvadraten rdečkast lesen zaboj b. kvadraten rdeč lesen zaboj
 square reddish wooden crate square red wooden crate

Both pairs of adjectives behave alike and seem to be placed in the same position within the sequence of adjectives as the less/more subjective element of the pair.⁵

⁵ Note that we do not find the reverse orders of (7a) and (8a) completely impossible, but we need to stress that the reverse orders of the ones presented in (7) and (8) are equally available for

Of course, we are basing these subjectivity rankings on our intuitions, but this is essentially how Scontras et al. were measuring ‘subjectivity,’ too: subjects had to report the perceived subjectivity of individual adjectives. Similarly to Scontras et al.’s subjectivity, other components of adjective meaning had been proposed to be the factor determining adjective ordering: based on previous studies, Martin (1969), for example, suggested that definiteness of denotation (i.e. the stability of adjective meaning independently of the noun) plays a crucial role in adjective ordering – see Scontras et al. (2017) for an overview of Martin (1969) and similar proposals.

There is another line of proposals which, in one way or another, explicitly place the root of functional hierarchies in general cognition. For example, Cinque and Rizzi (2008: 53) mention that elements which are included in the hierarchy are essentially a “linguistic crystallization of a particular set of cognitive categories among the many more that simply do not find a grammatical encoding in UG.” For example, adjectives for color (such as *red*) are simply how we encode one variety of the concept of color. Similarly, we also find accounts which mention that general cognition is the most prominent source of the hierarchy since language reflects how concepts are composed, as per Ramchand and Svenonius (2014), who argue for cognition as the source of the functional sequence on the basis of the cognitive tendency to understand experience in terms of events, situations and propositions (the composition of which mirrors the syntax of the clause). A similar proposal is mentioned in Kotowski and Hartl (2019), who mention Eichinger (1991) as associating adjective order preferences with memory capacities and a strategy of the cognitive system to reduce complexity.⁶

Finally, a different role is ascribed to the meaning of adjectives by Teodorescu (2006), who observes that adjective ordering restrictions do not apply when reversing the order of two adjectives results in structures that are truth-conditionally distinct (see also Svenonius 1994 for a similar observation). In (9), for example, the two word orders have the same interpretation, and the order in (9a) is preferred because of AOR; in (10), on the other hand, the two variants contain different structures with different interpretations, so the order of the two (operator) adjectives appears not to be restricted by AOR.

(9) a. vegetarian Russian lawyer b.# Russian vegetarian lawyer (Teodorescu 2006: (1))

(10) a. a famous former actor b. a former famous actor (Teodorescu 2006: (12))

both elements of the pair. Thus, to the extent that *a reddish square wooden crate* sounds possible in Slovenian, so does *a red square wooden crate*.

⁶ Here we could add Champollion (2006), who attempts a game-theoretic account of adjective ordering restrictions, although he eventually suggests that AORs are potentially cognitively imposed and mutually known to all speakers.

Truswell (2009) mirrors this observation and argues, based on data collected from the internet with Google, that the meaning of adjectives plays a role in AORs in the sense that subjective adjectives dominate intersective adjectives, although he does also claim that the order of intersective adjectives is essentially free, which is not expected given the results in Scontras et al. (2017) or Pereltsvaig (2007) and similar empirical studies.

In this paper we will also focus on general cognition as a possible source of the universal hierarchy of functional projections that host adjectives. Given the large number of different types of adjectives and the lack of consensus as to what is the correct typology of adjectives, we limit ourselves to three classes of adjectives whose existence seems widely accepted: color, shape, and size. Our basic assumption is that if what determines the hierarchy of functional projections and the adjective ordering restrictions is cognition and its restrictions, then we can expect to find a reflection of the specific order of projections hosting adjectives in non-linguistic cognitive processes. In what follows, we report on a series of experiments that we designed to test this prediction, discussing both the results themselves and the methodological issues that we detected.

3. Experiments

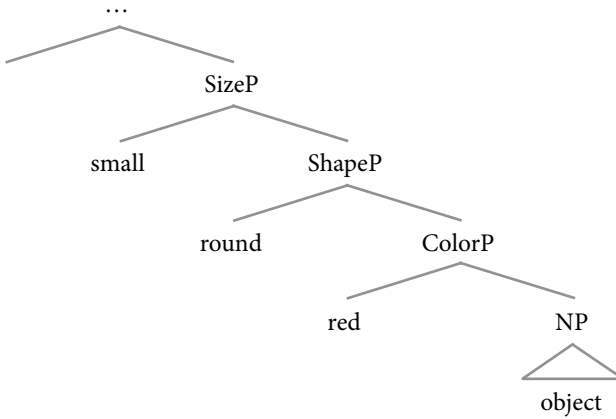
The goal of the experiments which we will discuss in this part of the paper was to investigate if the universal order of adjectives (and the hierarchy of functional projections hosting them) is rooted in nonlinguistic cognitive domains. If this is the case and what is behind the hierarchy is general cognition and its restrictions, then the order of projections hosting adjectives should also be reflected in non-linguistic cognitive processes. To test this prediction we focus on SIZE, SHAPE, and COLOR, that is, concepts which are grammatically encoded in adjectives for size, shape, and color.

The adjectives for size, shape, and color are widely claimed to come in the order presented with Slovenian data in (11) and (12) (Sproat and Shih 1991; Cinque 1994; Scott 2002, a.o.).

- (11) a. velik rdeč balon vs. *rdeč velik balon size > color
big red balloon red big balloon
- b. velik okrogel balon vs. *okrogel velik balon size > shape
big round balloon round big balloon
- c. okrogel rdeč balon vs. ?rdeč okrogel balon shape > color
round red balloon red round balloon
- (12) size > shape > color
velik okrogel rdeč balon
big round red balloon

Following Scott (2002) (also Cinque 1994), the observed order of these three classes of adjectives is a consequence of the ordering of functional projections in the DP, as shown in (13).

(13)



A substantial part of the linguistic literature claims that syntactic structure is built bottom-up (Chomsky 1995, 2000, etc.), which means that what lies lower in the syntactic structure enters the syntactic derivation sooner and stays in the derivation longer than what lies higher. This suggests that the lower elements might receive some sort of bias. Further, a large number of acquisition studies found that the acquisition of syntactic structure also proceeds bottom-up (Radford 1996; De Cat 2000; Hulk 2004; Vainikka and Young-Scholten 2011, a.o.), which also suggests that the lower syntactic elements are somehow more prominent compared to the elements from the higher functional projections.⁷ In other words, there seems to be some bias for the syntactic elements located in the lower functional projections over the elements from located in the higher functional projections. Consequently, we assume that if the syntactic hierarchies have some parallel in the organization of extralinguistic cognitive domains, such as the concepts of SIZE, SHAPE, and COLOR, we can also expect some sort of bias for the concepts that correspond to the properties denoted by adjectives expressed

⁷ For example, in acquisition studies of English, it has been found that the noun phrase is acquired gradually: when children start using nouns, they first use bare nouns, then they start using plural inflection and later still they start adding instances of more formal inflection, such as the possessive (’s) and the case of pronouns (he vs. him). This order of acquisition corresponds to the functional make up of the noun phrase, where the NP (where the noun is) is the lowest projection, the DP (where the possessive ’s is) the highest, and the plural inflection is somewhere in-between (cf. Radford 1996).

in the lower functional projections.⁸ Put differently, we predict that we will find a preference for COLOR over SHAPE (over SIZE) in certain non-linguistic tasks since COLOR concepts are, among these three groups, related to and expressed by adjectives that are located in functional projections closest to the lexical head. In a categorization task, for example, subjects' reactions should reveal COLOR as the most prominent out of the concepts of COLOR, SHAPE, and SIZE. If such a bias is detected, it could be taken as an argument for the claim that AORs derive from extralinguistic properties of general cognition and that (given our assumptions about the functional hierarchy hosting adjectives) also for the idea that universal hierarchies of functional projections are based on properties of general cognition.

3.1. Experiments 1, 2 and 3

We conducted a series of four experiments to try to test the hypothesis that there should be a prominence bias for the properties expressed in the lower functional projections. Even though we will focus mostly on the last experiment (Experiment 4), we will first report on Experiments 1, 2 and 3 since they provide relevant background for the last experiment. The basic idea behind all of the reported experiments is the same: when subjects are asked to memorize or categorize an object, they will do so on the basis of the relative prominence of the objects' relevant properties. In other words, they will memorize or categorize an object on the basis of its most prominent property. Therefore, by designing a task that tests which property out of color, shape and size determines subjects' memorization or categorization of an object, we will have a way of determining which of these three properties is more prominent in subjects' perception and/or memory.

Experiments 1 through 3 were all conducted with IxbyFarm (Drummond 2011), and all three used the same basic design. Participants were first presented with a picture of an object (visible for 2 seconds) on a computer screen. This picture was followed by a screen with three objects, each of which matched the initial picture in one property (either size, shape, or color). Participants had to click on the picture of the object which they thought best matched the initial picture of the object. Figure 1 below shows the design of Experiment 1, which is also the basic design of Experiments 2 and 3, although as will be seen below, there *were* also some important differences between the three experiments.

⁸ In a corpus study, Dolenc (2018) finds that these three types of adjectives are acquired in the reverse order of the order in which they appear in the noun phrase, which matches exactly the order one would expect assuming a bottom-up acquisition of syntactic structure.





1 st picture	3 possibilities		
	1 st possibility	2 nd possibility	3 rd possibility
			
(2 seconds)	SIZE	SHAPE	COLOR
	“Click on the image that best matches the previous picture.”		

Figure 1: Experiment 1

Experiment 1 consisted of practice examples, 27 test examples and 28 fillers (the order of test examples and fillers was randomized).⁹ The initial image/object was followed by the three images/objects from which the subjects had to choose the best match immediately. The objects used as stimuli were simple pictures of shapes (square, circle, and pentagon), the tested colors were green, red, and blue, and the three tested sizes were related to the area of the shape (there was a two-time increase in the area between sizes). All together this makes 27 different/unique stimuli (3 colors × 3 shapes × 3 sizes). All of the participants in Experiment 1 were native speakers of Slovenian (n = 37).

The majority of participants made their decision based on shape (80%), followed by color (17%) and size (3%).¹⁰ This means that our prediction (COLOR over SHAPE over SIZE) was not confirmed (at least with respect to the relative ordering of shape and color), given that shape turned out to be the most prominent of the three tested properties. However, using simple pictures of shapes (square, circle and pentagon) for our stimuli may very well have been problematic. The pictures represented objects that differed in color, shape, and size, but at the same time, shape is the only property of the three that could be used to classify these objects if we were to use only nouns. The objects used in this experiment have names that are defined with their shapes – in Slovenian, *kvadrat* ‘square’, *krog* ‘circle’ and *petkotnik* ‘pentagon’ – whereas nothing of the sort exists for color and size. There is no way of classifying these objects with nominal names describing their color or size. One might want to counter here with the existence of some terms derived from color adjectives, such as Slovenian *rdečina* ‘red color’, ‘state of being red’, ‘redness’ or *rdečost* ‘the state of being red’, ‘redness’, though

⁹ In Experiments 1 and 2 the fillers consisted of 14 examples in which one of the three possibilities completely matched the 1st picture and 14 examples in which none of the three possibilities matched the 1st picture in any of the three properties. In Experiments 3 and 4, the fillers consisted of examples with simple counting tasks. For example, participants were shown a picture with some objects on it and then had to choose between three pictures, each of which had a different number of objects on it: out of these, only 1 picture matched the initial picture in the number of objects.

¹⁰ We only report descriptive statistics for the initial experiments as we subsequently learned that they all had design issues, as will become clear lower down.

the use of these is rather limited and usually also includes an additional component, e.g. *rdečina* more commonly being ‘a red spot/area’ than really just ‘red color.’ This could mean that when subjects had to remember the object and compare it to the three objects on the next page, they were not really comparing three properties, but rather three classified objects, one of which matched the object on the first slide, except that it was of a different color and size. If we assume, as we have done, that when defining an object, properties/concepts encoded lower in the hierarchy of functional projections will be more prominent, then the fact that nouns are located in the lowest position (even if this is the lexical head) easily accounts for the prominence of shape in this experiment as the participants potentially thought of the 1st picture in Figure 1 as ‘a medium red circle’ and not ‘a medium round red object.’ To avoid this issue, we designed Experiments 2 and 3 with pictures of buttons (instead of pictures of unidentified objects that could then be identified as circles, squares, and pentagons).

3.1.1. Experiment 2

Experiment 2 was identical in design to Experiment 1 except that we used pictures of buttons for the stimuli. It was also run on IbexFarm (Drummond 2011), and it also had 27 test examples and 28 fillers. The choice of colors and sizes remained the same, but we made two minor modifications regarding the shapes we used. As pentagonal buttons seem unusual, we replaced the pentagonal button with a hexagonal one and made the square button slightly rounded. All buttons had two vertical holes that were proportional to the size of the button. Figure 2 below shows the design of Experiment 2. In the instructions, subjects were told they will be looking at pictures of buttons.





1 st picture	3 possibilities		
	1 st possibility	2 nd possibility	3 rd possibility
			
(2 seconds)	SIZE	SHAPE	COLOR
	“Click on the image that best matches the button/item on the previous picture.”		

Figure 2: An example of a stimulus from Experiment 2

Experiment 2 was only run on 11 subjects, and the general results were the same as in Experiment 1. Subjects predominantly chose objects/buttons that matched the initial stimulus in shape. Shape (61%) was followed by color

(33%) and size (6%).¹¹ Even though we used different stimuli for this experiment, this did not seem to have a big effect on the results.

Regardless of results and before going to collect the data on a bigger sample of subjects, we wanted to try and make the task a little more demanding.

3.1.2. Experiment 3

The basic design remained similar to that of Experiments 1 and 2. We made the task more difficult by separating the initial picture from the 3 possible answers by a 5-second gap during which two unrelated animated GIFs appeared. Other than that, the task stayed the same, with the participants needing to select which of the three pictures of buttons best matched the initial picture.





1 st picture	5-second GIF	3 possibilities		
		1 st possibility	2 nd possibility	3 rd possibility
				
		SIZE	SHAPE	COLOR
(2 seconds)	“Click on the image that best matches the button/item on the previous picture.”			

Figure 3: An example of a stimulus in Experiment 3

Again, Experiment 3 consisted of training examples, 27 test examples and 28 fillers (which involved counting tasks). The tested colors were green, red, and blue, the tested shapes were circle, square, and hexagon, and the tested sizes were related to the area of the shape (there was a two-time increase in the area between sizes). We tested 48 participants, the majority of which were speakers of Slovenian (34), followed by speakers of Serbo-Croatian (7), English (2), Czech (2), Bulgarian (1), and Dutch (1). Subjects were recruited through Facebook.

Despite the change in the stimuli (circle/square/pentagon to shaped buttons) and the additional memory-taxing 5-second pause, the results of Experiment 3 were similar to the results of Experiments 1 and 2: the majority of the participants chose their best matches on the basis of shape (58%), followed by color (30%) and finally size (11%). This means that our prediction that there would be a bias for color over shape over size is not confirmed in entirety. The

¹¹ See Marušič and Mišmaš (2018) for more details on Experiments 1 to 3.

results do, however, confirm one part of the prediction, namely, that size will be less important than color and shape.

3.1.3. Summing up Experiments 1, 2, and 3

Overall, Experiments 1, 2 and 3 yielded similar results despite the differences in the design. The results of the experiments did differ with respect to the share of responses for each property (Experiment 1 had as much as 80% of the matchings made on the basis of shape, while Experiment 3 had a considerably lower 58,31% matchings made on the basis of shape). Despite this difference, however, the results of all three experiments are consistent in revealing a clear bias for shape. They also all revealed color to be the second most prominent property, ahead of the least prominent property, size. The conclusion is thus clear: even though the results are in line with our prediction that size will be the least prominent property when memorizing an object, they do not confirm our prediction that there will be a preference for color over shape in a non-linguistic task. So, on the one hand, these results cannot be taken as an argument for the idea that universal hierarchies of functional projections are based on properties of general cognition. On the other hand, however, it is encouraging that the finding that size is the least prominent property when memorizing an object can be seen to be in line with the results of language-dependent tests in Scontras et al. (2017), who found, regarding the source of the hierarchy of adjectives, that adjective meaning, specifically its subjectivity, influences the position of the adjective, and adjectives for size are more subjective than adjectives for color and shape and therefore more removed from the noun.

In the course of analyzing results, we discovered several aspects of our experimental design that could have disqualified the validity of our results. Firstly, a relatively small number of participants were tested (37 participants in Experiment 1, 11 in Experiment 2, and 48 in Experiment 3), an issue that was avoided in further testing. Secondly, even though subjects were 'reminded' through the instructions that the objects are 'buttons' rather than just circles, squares and hexagons, it is difficult to be entirely sure that subjects do not still simply take them as circles, squares and hexagons, in which case we have not avoided the original problem that made us switch from circles, squares and pentagons to 'buttons' (see section 3.1 above). A much more important issue, however, is that it is not clear if or to what extent SIZE, SHAPE and COLOR are similar/comparable properties. Crucially, if these properties are not directly comparable, then the fact that participants in Experiments 1 through 3 typically chose their match on the basis of SHAPE might not have so much to do with SHAPE as a property, but rather with the fact that, for example, squares are more impactful (and consequently more memorable) than for example red as a color. To address this

issue, we designed and conducted a calibrating session, on which we report in the next section.

3.2. Calibrating

Experiments 1, 2 and 3 compared three differences on three dimensions, without knowing how comparable these differences are between these three dimensions. Issues in comparison of size, shape and color become obvious if, for example, we consider how we measure values related to these properties: while we can assess or measure size in terms of area or circumference and thus evaluate the difference between two sizes with an absolute number, we cannot do anything comparable with differences in color and shape. When it comes to color, differences may be measurable in terms of categories such as hue, brightness, saturation and gloss (see e.g. Schwartz and Krantz 2018), but differences in shape really cannot be assigned any comparable numerical value. We can describe some differences between specific shapes in terms of the number of angles or edges, but even shapes that have the same number of angles or edges can vary significantly among themselves (for example, square, rectangle, rhombus, trapezium and parallelogram are all quadrilaterals). Again, these ‘parameters’ are not necessarily comparable, so varying them does not necessarily mean that the results of their variations are comparable; therefore, the colors we used in our experiments (blue, green and red) were perhaps less different than the shapes we used (circle, square, and hexagon in Experiment 3), which would also mean that SHAPE was potentially easier to differentiate and memorize.

To address this issue, we tried to prepare a calibrated set of stimuli in order to: (i), determine which minimal pair per property participants perceive or recognize as different, and (ii), make these minimal pairs as comparable as possible between the three properties. To calibrate the three properties, we created:

- (a) color: we created 7 shades of three colors (red, green and blue), each shade 5 hue-points different;
- (b) shape: we created 7 versions for each tested shape which differed in curvature of angles and sides;
- (c) size: we tested 7 different sizes, where each size had a 5% bigger/smaller area for a randomly selected version of a shape from (b)

We tested 12 participants (3 male, 9 female; age 25 to 40), each participant took part in 3 tasks (1 per property). In all three tasks the participants were shown two objects and were then asked if they thought the objects they had seen were different or the same with respect to the property in question. The design used in calibrating is shown in Figure 4 for shape:


Screen 1	Screen 2
	The objects on the previous screen were a) the same b) different

Figure 4: Calibration of stimuli

We calculated the smallest difference for which there was 75% agreement among subjects that the two objects differ with respect to the relevant property, and then used this as stimuli in the new experiment, which we report on in the next section.

3.3. Experiment 4

The design of Experiment 4, shown in figure 5 below, was the same as that of Experiment 3: the participants were shown a picture, followed first by a 5-second animation (GIF) and then by a screen with 3 pictures, each of which matched the initial picture in one property (either size, shape or color). They were asked to choose the one which they thought best matched the initial picture. Differently from Experiment 3, however, Experiment 4 used the stimuli that we created on the basis of the calibrating session described in section 3.2 above. The use of such stimuli also meant that we avoided completely the issue which made us switch from squares, circles and pentagons to ‘buttons’ between Experiments 1 and 2 (see sections 3.1 and 3.1.3 above); to the best of our knowledge, there is no specific noun (or adjective) that would reflect the shape of the objects used in this experiment (nor are there specific names for each individual shade of blue and red used in the experiment).





1 st picture	5-sec. GIF	3 possibilities		
		1 st possibility	2 nd possibility	3 rd possibility
2 seconds				
		SIZE	SHAPE	COLOR
		“Click on the image that in your opinion best matches the previous picture.”		

Figure 5: An example of a stimulus from Experiment 4

Experiment 4 was also conducted online, and it also consisted of practice examples, 27 test examples and 28 control fillers (with the order of test examples and fillers randomized). We tested 160 participants, recruited through

Amazon Mechanical Turk (participants were compensated). Despite the fact that we have been assuming that the hierarchy of functional projections (and consequently, adjective order) is universal, we asked about the participants' first language. The majority of our participants were speakers of English (149, before exclusions), with some speakers of Spanish, Turkish, Indonesian, Italian or German (2), and Cantonese or Malayalam (1). 48 participants were excluded from the analysis, on the basis of: (i) reported color blindness (3), (ii) reported use of smart-phones (12), (iii) less than a 75% success rate at control fillers (33).¹² Following these exclusions, we are left with 112 participants. Note that even though we excluded almost a third of the participants, the number of participants included in the analysis was still much larger than in Experiments 1, 2 and 3.

3.3.1. Results

Like the results of Experiments 1 through 3, the results of Experiment 4 also do not confirm the predictions of our hypothesis. However, the results of Experiment 4 are markedly different from the results in Experiments 1 through 3.

Specifically, while in Experiment 3 (as well as 1 and 2) participants typically chose on the basis of shape, in Experiment 4, participants chose on the basis of color in 53,9% of cases, followed by size with a 26% share. Single factor ANOVA shows that the results for the three tested properties are statistically different ($p < .005$). This means that we found a preference for color, which is in line with our prediction (given that color is encoded by adjectives lowest in the hierarchy of adjectives). Contrary to the prediction, though, color, as already indicated, was not followed by shape but rather by size (26,5%): matching items were chosen on the basis of shape in only 19,6% of cases. Therefore, our prediction was again not confirmed in its entirety. Unfortunately, we have again identified some relevant problems in the design of our experiment.

3.3.2. Issues

Whereas running the experiment online and using Amazon Mechanical Turk for recruitment allowed us to fix the issue of the low number of participants, this setup also brought along some unwanted effects. Firstly, since Experiment 4 was conducted online, we could not account for different screen settings that participants might have used. For example, participants might

¹² Given that fillers included stimuli where a "correct" answer could easily be determined, we calculated the percentage of correct answers in choosing between the three options. If a subject failed to select the matching picture on more than 75% of the 28 controls, they were excluded from the analysis.

have zoomed in and thus enlarged the pictures, which could have favored choosing on the basis of size. Secondly, the quality of colors differs considerably between different types of screens, and the quality of colors we see on the computer screen also depends a lot on the lighting in the area where the computer is used, with artificial light reflections or direct sunlight on the computer screen basically making everything on the entire screen invisible. Thirdly, contrary to the predominance of color among subjects on the whole, we observed that there were subjects who consistently ignored COLOR but did not report color blindness. All of these issues can be avoided by testing participants in a controlled environment and by using additional screening for detecting color blindness (based on which participants would be excluded).

Furthermore, a post-hoc analysis also revealed some issues with the calibrating session, which could mean that the minimal differences for size, shape and color that we used were not completely comparable. Specifically, on the basis of the calibrating session we used a 10% increase in the area of the objects as a minimal difference between two sizes (which was expected to be comparable with the minimal pairs for shape and color). But we later discovered that a smaller increase would have worked. This means, importantly, that the 10% increase in size is not comparable to the minimal differences in color and shape that were used in the experiment, which may have influenced the results for size (note that in Experiments 1 through 3, size was the least prominent property). To address this issue a new calibration session is necessary (currently in progress) before the experiment is re-run.

4. Discussion

The four experiments we report on were conducted in order to test the hypothesis that we will find a preference for the properties expressed on the lower functional projections (i.e. COLOR > SHAPE > SIZE) in non-linguistic tasks such as memorizing or matching objects. If such a bias were detected, it could be taken as an argument for the idea that universal hierarchies of functional projections are based on properties of general cognition. At this point, the results of our experiments appear to be inconclusive. On the one hand, it is clear that none of our experiments confirmed the hypothesis in total; but on the other hand, it seems that we also cannot, on the basis of the results, completely reject the hypothesis.

Figure 6 shows a comparison of the results for Experiments 3 and 4 (we are leaving Experiments 1 and 2 aside at this point due to the differences in the design of the experiment, etc., see 3.1 to 3.1.3 above).

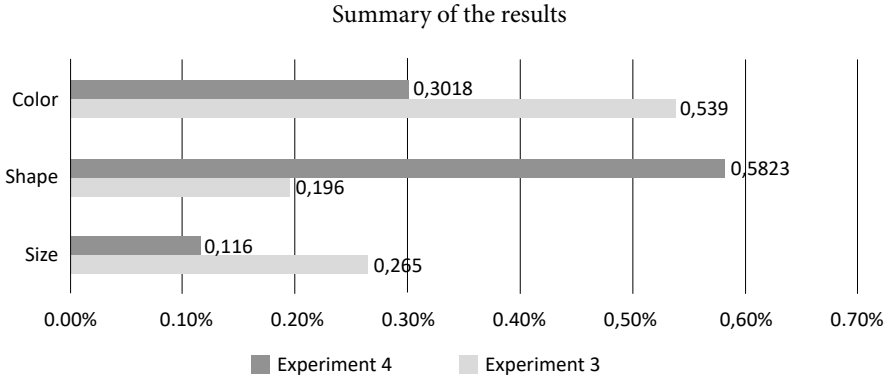


Figure 6: Summary of the results for Experiments 3 and 4

Given that the only significant design difference between these two experiments was in the stimuli, the comparison of the results suggests that the change in the materials between Experiments 3 and 4 had an effect. Importantly, the task in Experiment 4 revealed color as the most prominent property. This is in line with our starting hypothesis, as the concept COLOR is expressed by adjectives which come into the structure at the bottom of the hierarchy of functional projections that host adjectives, see the diagram in (13) in section 3 above. As the materials were modified in an attempt to correct identified problems in the design by calibrating the differences in the materials (from red, blue, and green circles, squares, and hexagons to calibrated versions of shades of objects with calibrated versions of objects with different curvatures, see 3.2. above), there is reason to take the results of Experiment 4 over those of Experiment 3.

In spite of this aspect of the results, our experimentation clearly has not so far provided evidence for confirming the hypothesis that cognition can be taken as the source of the functional hierarchy because size had precedence over shape in Experiment 4. But given that we also found reason to doubt the success of our calibration for size (see 3.3.2 above), we cannot have full confidence in the accuracy of these results. We therefore conclude that we are not yet in a position to reject the hypothesis either. In other words, it remains unclear what the source of the hierarchy of functional projections is, and we can exclude neither universal grammar as the source nor meaning of the adjectives.¹³

¹³ Note that our results also cannot be used, in any straightforward way, to support the opposite position of the one we were trying to experimentally verify, i.e. the position that advocates a modular architecture of cognition whereby syntactic structure is autonomous from general cognition. The reason for this is that the results are not consistent across our four experiments (they do not all yield one and the same order, which would be different from the one observed in language), and that despite the methodological improvements of Experiment 4 over previous

We should mention that the results of Experiment 4 (but not Experiment 3) may be in line with independent findings of some studies on the acquisition of concepts. For example, Baldwin (1989) finds that shape is used over color in generalizing object labels, and Diesendruck and Bloom (2003) find that shape has precedence in object naming and object matching over color and material. However, none of these studies calibrated differences of their materials; for example, Diesendruck and Bloom (2003) used objects created for the purpose of the study that ranged from something that could be described as vaguely resembling a turritella seashell, vaguely resembling a toy briefcase, vaguely resembling a blueberry (sphere with a round surface depression), etc., in undescribed colors that seem, based on the black-and-white photograph of the materials (Diesendruck and Bloom 2003: 171), to range from very bright to very dark and to include both single-color and multiple-color objects. So before we accept our experiments as supporting the conclusions of these studies, we need to perfect the calibration of our materials for size and re-run our experiment.

Another avenue to pursue in order to test our starting hypothesis is through an acquisition study with preschool children designed similarly to the experiments we report on in this article. Corpus investigations suggest that the acquisition of adjectives for color, shape and size in small children, as reflected by the rates of adjective use at different ages and their comparison to the rates of adjective use in adult speech, proceeds in line with the order that is predicted by the universal hierarchies of adjectives, assuming a bottom-up acquisition of syntactic structure (Dolenc 2018). This finding about language thus fails to find a parallel shape-bias that studies like Diesendruck and Bloom (2003) report for the acquisition of the related concepts. This could, naturally, mean simply that language acquisition is separate from concept acquisition and that the hierarchy of syntactic functional projections is independent of the cognitive domain of concepts. But as was already pointed out above, it seems that the design and material preparation in those concept acquisition studies was clearly suboptimal. So before endorsing such a conclusion on the basis of the results of such studies, an acquisition study that will use better designed materials and will at least only test the acquisition of individual compared concepts without forcing direct competition between them into the task is called for. We have designed a sorting task of this type and are currently testing preschool children with it.

References

- Baldwin Dare A. (1989). Priorities in children's expectations about object label reference: Form over color. *Child Development* 60(6), 1291–1306.

experiments, we do not yet seem to have achieved enough methodological soundness to clearly allow drawing such a conclusion.

- Champollion Lucas (2006). *A game-theoretic account of adjective ordering restrictions*. Ms., University of Pennsylvania. [URL: www.ling.upenn.edu/~champoll/adjective-ordering.pdf; accessed May 20, 2019].
- Chomsky Noam (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky Noam (2000). Minimalist inquiries: The framework. In *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, Roger Martin, David Michaels, Juan Uriagereka, and Samuel Jay Keyser (eds.), 89–155. Cambridge, MA: MIT Press.
- Cinque Guglielmo (1994). On the evidence for partial N-movement in the Romance DP. In *Paths towards Universal Grammar: Studies in Honor of Richard S. Kayne*, Guglielmo Cinque, Jan Koster, Jean-Yves Pollock, Luigi Rizzi, and Raffaella Zanuttini (eds.), 85–110. Washington, D.C.: Georgetown University Press.
- Cinque Guglielmo (1999). *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford University Press.
- Cinque Guglielmo (2005). Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry* 36(3), 315–332. DOI: 10.1162/0024389054396917.
- Cinque Guglielmo, Rizzi Luigi (2008). The cartography of syntactic structures. *Studies in Linguistics* 2, 43–95.
- Culbertson Jennifer, Adger David (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences of the United States of America* 111(16), 5842–5847. DOI: 10.1073/pnas.1320525111.
- De Cat Cecile (2000). Structure building and the acquisition of dislocations in child French. In *BUCLD 24: Proceedings of the 24th Annual Boston University Conference on Language Development*, S. Catherine Howell, Sarah A. Fish, and Thea Keith-Lucas (eds.), 242–252. Sommerville: Cascadilla Press.
- Diesendruck Gil, Bloom Paul (2003). How specific is the shape bias? *Child Development* 74/1: 168–178. DOI: 10.1111/1467-8624.00528.
- Dixon Robert Malcolm Ward (1982). *Where Have All the Adjectives Gone*. Berlin: De Gruyter.
- Dolenc Tina (2018). *Raba pridevnikov pri najmlajših govornih* [Adjective use among the youngest speakers]. Nova Gorica: University of Nova Gorica. BA thesis. [URL <http://repozitorij.ung.si/IzpisGradiva.php?id=4018&lang=eng>; accessed May 20, 2019].
- Drummond Alex (2011). *Ibex Farm*. [URL <http://spellout.net/ibexfarm>; accessed May 20, 2019].
- Eichinger Ludwig M. (1991). Woran man sich halten kann: Grammatik und Gedächtnis [What you can hold on to: grammar and memory]. In *Jahrbuch Deutsch als Fremdsprache* 17. München: iudicium, 203–220.
- Greenberg Joseph (1963). Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of Language*, Joseph Greenberg (ed.), 72–113. Cambridge: MIT Press.
- Hetzron Robert (1978). On the relative order of adjectives. In *Language Universals*, Hansjakob Seiler (ed.), 165–184. Tübingen: Narr.
- Hulk Aafke (2004). The acquisition of the French DP in a bilingual context. In *The Acquisition of French in Different Contexts. Focus on Functional Categories*, Philippe Prévost and Johanne Paradis (eds.), Amsterdam/Philadelphia: John Benjamins. 243–74.

- Kotowski Sven, Hartl Holden (2019). How real are adjective order constraints? Multiple prenominal adjectives at the grammatical interfaces. *Linguistics* 57(2), 395–427. DOI: <https://doi.org/10.1515/ling-2019-0005>.
- Marušič Franc, Mišmaš Petra (2018). Povezovanje hierarhije pridevnikov v samostalniški zvezi in kognicije [Relating the hierarchy of adjectives inside the noun phrase and cognition]. In *Škrabčevi dnevi 10: zbornik prispevkov s simpozija 2017*, Aleksandra Bizjak Končar and Helena Dobrovoljc (eds.), 99–110. Nova Gorica: University of Nova Gorica Press. [URL http://www.ung.si/media/storage/cms/attachments/2018/11/29/16/56/02/Skrabcevi_dnevi_10_11k18.pdf; accessed May 20, 2019].
- Martin James E. (1969). Semantic determinants of preferred adjective order. *Journal of Verbal Learning and Verbal Behavior* 8, 697–704. DOI: 10.1016/S0022-5371(69)80032-0.
- Pereltsvaig Asya (2007). On the universality of DP: A view from Russian. *Studia Linguistica* 61, 59–94. DOI: 10.1111/j.1467-9582.2007.00129.x.
- Plesničar Vesna (2017). Adjective ordering restrictions of attribute adjectives in Slovenian. Talk given at *12th Slavic Linguistics Society Annual Meeting*, Ljubljana, September 21–24, 2017.
- Radford Andrew (1996). Towards a structure-building model of acquisition. In *Generative Perspectives on Language Acquisition*, Harald Clahsen (ed.), 42–88. Philadelphia: Benjamins.
- Ramaglia Francesca (2011). *Adjectives at the Syntax-Semantics Interface*. München: LINCOM Europa.
- Ramchand Gillian C., Svenonius Peter (2014). Deriving the functional hierarchy. *Language Sciences* 46, 152–174. DOI: 10.1016/j.langsci.2014.06.013.
- Rizzi Luigi (1997). The fine structure of the left periphery. In *Elements of Grammar*, Liliane Haegeman (ed.), 281–337. Dordrecht: Kluwer Academic Publishers.
- Scontras Gregory, Degen Judith, Goodman Noah D. (2017). Subjectivity predicts adjective ordering preferences. *Open Mind: Discoveries in Cognitive Science* 1(1), 53–65. DOI: 10.1162/OPMI_a_00005.
- Scott Gary-John (2002). Stacked adjectival modification and the structure of nominal phrases. In *The Cartography of Syntactic Structures. Vol. 1.*, Guglielmo Cinque (ed.), 91–120. Oxford: Oxford University Press.
- Schwartz Bennett L., Krantz John H. (2018). *Sensation and Perception*. (Second Edition). Thousand Oaks, CA: SAGE Publishing.
- Sproat Richard, Shih Chilin (1991). The cross-linguistic distribution of adjective ordering restrictions. In *Interdisciplinary Approaches to Language: Essays in Honor of S.-Y. Kuroda*. Carol Georgopoulos and Roberta Ishihara (eds.), 565–593. Dordrecht: Kluwer Academic Publishers.
- Svenonius Peter (1994). On the structural location of the attributive adjective. In *Proceedings of the 12th West Coast Conference on Formal Linguistics*, Eric Duncan, Donka Farkas, and Philip Spaelti (eds.), 439–454. Stanford, CA: CSLI Publications.
- Teodorescu Alexandra (2006). Adjective ordering restrictions revisited. In *Proceedings of the 25th West Coast Conference on Formal Linguistics*, Donald Baumer, David Montero, and Michael Scanlon (eds.), 399–407. Somerville, MA: Cascadilla Proceedings Project.
- Toporišič Jože (2000). *Slovenska slovnica* [Slovenian Grammar]. 4th edition. Maribor: Obzorja.

- Truswell Robert (2009). Attributive adjectives and nominal templates. *Linguistic Inquiry* 40(3), 525–533. DOI: <https://doi.org/10.1162/ling.2009.40.3.525>.
- Vainikka Anne, Young-Scholten Martha (2011). *The Acquisition of German: Introducing Organic Grammar*. Berlin/Boston: De Gruyter Mouton.

Franc Lanko Marušič, Petra Mišmaš, Rok Žaucer
Center for Cognitive Science of Language
and
School of Humanities
University of Nova Gorica
Vipavska 13
SI-5000 Nova Gorica
Slovenia

Franc Lanko Marušič
franc.marusic(at)ung.si
ORCID: 0000-0002-0667-3236

Petra Mišmaš
petra.mismas(at)gmail.com
ORCID: 0000-0001-8659-875X

Rok Žaucer
rok.zaucer(at)ung.si
ORCID: 0000-0001-7771-6937

