



ARTICLE



<https://doi.org/10.1057/s41599-022-01440-w>

OPEN

# Determining the cognitive biases behind a viral linguistic universal: the order of multiple adjectives

Evelina Leivada <sup>1</sup>✉

When people are asked to create a phrase with the elements {blue, earrings, beautiful}, they produce ‘beautiful blue earrings’. Several theories have been proposed about the origins of this universal tendency to order multiple adjectives in a specific way: an innate universal hierarchy with designated positions for each category of adjectives, sensitivity to the definiteness of the adjectival denotation, availability and psychological closeness of the adjective attributes to the speaker, the encoding of subjective vs. objective properties, and the adjective’s phonological weight. Although these theories have strong descriptive power, they often focus on what happens at the phenotypic level without explaining what cognitive needs trigger this behavior. Through a timed task that measures acceptability in ‘Adjective-Adjective-Noun’ sequences that either comply with the universal order or violate it, we adduce evidence for the high acceptability of the violating orders, whose processing did not take longer than that of the compliant orders, as should have happened if the former were non-canonical. The results suggest that ordering preferences exist but are not invariable, as one would expect if a strong linguistic universal was involved. We track the origin of adjective ordering preferences to the synergistic interplay of three cognitive biases: Zipf’s Law, Intolerance of Ambiguity, and Novel Information Bias. Last, we show that the linguistic manifestation of these preferences is sensitive to the statistical distribution of the input data, resulting to variation even among speakers of the same language.

<sup>1</sup>Universitat Rovira i Virgili, Tarragona, Spain. ✉email: [evelina.leivada@urv.cat](mailto:evelina.leivada@urv.cat)

## Introduction

Adjectives are ordered in a specific way across typologically different languages. This property of human language amounts to the most documented linguistic universal in digital media, repeatedly going viral every year since 2014 (Waldman, 2014; Hanson, 2015; Dowling, 2016; Horobin, 2016; Batchelor, 2017; Nordquist, 2018; Gutoskey, 2019; Saba, 2020). The linguistic formulation of the universal posits that different classes of adjectives appear in a so-called universal order that determines the distance between an adjective and the modified noun (Hetzron, 1978; Dixon, 1982; Sproat and Shih, 1991; Cinque, 1994; Scott, 2002; Laenzlinger, 2005; Teodorescu, 2006; Alexiadou et al., 2007; Panayidou, 2013). This universal order predicts that only (1a) is well-formed, with the rest of the orderings (1b–f) being ungrammatical in the absence of a special intonation or other licensing conditions that can legitimize their use (1g).

- (1) a. A set of beautiful blue porcelain earrings. [target order]  
 b. A set of beautiful porcelain blue earrings.  
 c. A set of blue beautiful porcelain earrings.  
 d. A set of blue porcelain beautiful earrings.  
 e. A set of porcelain blue beautiful earrings.  
 f. A set of porcelain beautiful blue earrings.  
 g. I asked for a set of BLUE beautiful porcelain earrings, not white.

Contrastive focus in (1g) legitimizes the deviation from the target order (1a), which shows that the adjective that denotes a subjective comment ("beautiful") is placed before the adjective that denotes color, which in turn precedes the material adjective. (1a) is the only order that is compliant with the universal hierarchy (2). The orders in (1b–f) have received various characterizations in the literature, such as ungrammatical (Bever, 1970), awkward (Teodorescu, 2006), semantically incorrect (Kemmerer et al., 2009), odd (Scontras et al., 2017), or marked/non-canonical (Smirnova et al., 2019).

- (2) Subjective > Comment > Evidential > Size > Length > Height > Speed > Depth > Width > Temperature > Wetness > Age > Shape > Color > Nationality/Origin > Material (adapted from Scott, 2002: p. 114).

This richness of views that present the structures that violate (2) as syntactically ill-formed, semantically/pragmatically odd, marked, or simply less preferred reflects the debated nature of adjective ordering preferences<sup>1</sup> (AOPs) as pertaining to core syntax, to the syntax-semantics interface, or to more general cognitive principles; a topic that is still under intense debate (Kotowski and Härtl, 2019).

According to the first theory (henceforth, the syntactic origin theory, SOT), an innate, Universal Grammar-encoded syntactic hierarchy with designated positions for adjective classes (2) is responsible for the grammaticality of (1a) and the ungrammaticality of (1b–f) (Cinque, 1994; Cinque, 2010; Scott, 2002; Panayidou, 2013). This syntactic hierarchy is part of a larger spine of syntactic positions (Demonstratives > Numerals > Adjectives > Nouns) that underlies the ordering of elements in the nominal domain (Cinque, 2005; Alexiadou et al., 2007; Alexiadou, 2014). This larger syntactic configuration is also available in Universal Grammar and makes predictions for the behavior of different types of adjectival modifications (Alexiadou, 2001).

According to the second theory (henceforth, the multifactorial cognitive origin theory, mCOT), the order is not so rigid, such that one can talk about ordering *preferences*, but not about ungrammatical or unattested orders. Under the assumptions of the second theory, AOPs are the outcome of one or more factors such as (i) the encoding of properties that are noun-inherent

(Whorf, 1945), objective (Hetzron, 1978; Scontras et al., 2017), absolute (Sproat and Shih, 1991), or object-oriented (Stavrou, 1999), (ii) the adjective's phonological weight/length (Wulff, 2003; Kotowski, 2016; Scontras et al., 2017), and (iii) the noun-specific frequency as well as the collocability and/or idiomaticity of certain combinations of nominals and adjectives (Wulff, 2003; Bouchard, 2005; Svenonius, 2008; Hahn et al., 2018).

The general tendencies are the following: Adjectives that denote noun-inherent, objective properties, which are object-oriented, and as such less likely to cause disagreement, are placed closer to the noun. A second contributing factor is noun-specificity. Certain adjectives occur more frequently with certain nouns, and this high specificity, or else noun-specific frequency, impacts ordering, placing the adjectives that have a high mutual information with the noun closer to it (Wulff, 2003; Hahn et al., 2018).<sup>2</sup> When adjectives are relatively interchangeable in terms of order (e.g., color and shape), a third factor becomes relevant: the morphophonologically lengthier adjective is placed closer to the noun (Wulff, 2003; Kotowski, 2016). The presence of a third tendency, which is predicated on the lax application of the other two, already suggests that more than one factors are at play behind what is considered to be the universal, unmarked order (2), such that a comprehensive answer to the question of why adjectives are ordered the way they are across languages is unlikely to make reference to only one cause (Wulff, 2003; Kotowski and Härtl, 2019).

The various proposals that fall under the second theory are grouped together because they have a similar reasoning. They all propose that one or more factors (i.e., inherentness, subjectivity, encoding of speaker-oriented vs. object-oriented properties) are behind the attested AOPs. Recent experimental studies that provide evidence for AOPs underscore the need for finding an explanation as to *why* these factors play a role in adjective ordering (Scontras et al., 2017; Fukumura, 2018). Undoubtedly, if human language consistently deploys a strategy, some function must be served.<sup>3</sup> The cross-linguistic preference for a specific order over others entails that this order is likely to lead to communicative success, possibly by facilitating referent identification (Fukumura, 2018; Franke et al., 2019; Scontras et al., 2019). What is missing from this picture is the overall underlying etiology: the *cognitive need(s)* that lead speakers/signers to prefer one ordering over others, sculpting cross-linguistic preferences accordingly. In Scott's (2002) words, almost "all writers claim that AO[P]s can be adequately accounted for using broad "psychological" criteria, yet none of them are able to provide a convincing argument—which is, moreover, consistent with the data—for a psychological basis to AO[P]s". More recent studies have re-affirmed that a clear answer to the question of what factors drive AOPs remains elusive (Trotzke and Wittenberg, 2019). Filling this knowledge gap requires (i) connecting linguistic structures observed at the phenotypic level to their cognitive underpinnings and (ii) bringing into the picture environmental triggers that may result to variation within or across linguistic communities. These are the main objectives of the present work. More concretely, the addressed research questions are:

- (I) What cognitive needs are subserved by AOPs?
- (II) Is there interspeaker variation in the attested preferences, and if yes, how can it be reconciled with the notion of a strong linguistic universal that should not allow for variability in its phenotypic realizations?

Research question (I) embeds the phenotypic behavior into the bigger picture, asking what triggers AOPs with the aim to explain their mosaic nature. Research question (II) imports a critical and hitherto missing comparative perspective: Experiments on AOPs

report insights from a single, often monolingual population, without offering any explicit comparisons of people with different developmental trajectories (e.g., monolinguals, early/late bilinguals, second language learners etc). Such a comparative perspective has the potential to reveal the ways in which aspects of the environment may influence the phenotypic manifestations of innate cognitive needs.

Methods

**Task.** A timed acceptability task was used to collect two types of responses: (i) acceptability judgments on a 3-point Likert scale ('correct', 'neither correct nor wrong', 'wrong') and (ii) reaction times (RTs). RTs are informative about the possible existence of an unmarked/preferred order, because a comparison across orders (i.e., unmarked/baseline vs. marked/less preferred) should reveal an extra processing cost in the latter (Erdocia et al., 2009). This cost occurs because the marked stimulus deviates from the expectations that the cognitive parser forms about upcoming input, based on its knowledge about what is the most frequently encountered, unmarked option (Imamura et al., 2016).

The task consists of two orders and three conditions. All test items have the same syntactic structure, featuring two adjectives and a Spelke object in the object position (e.g., 'I bought a square black table'). Each of the three conditions includes one of the following adjective pairs: 1. size adjective-nationality adjective, 2. shape adjective-color adjective, 3. subjective comment adjective-material adjective. Each condition has two orders with three test structures per order (18 test structures in total). In the congruent order, the size adjective precedes the nationality adjective, complying with (2). In the incongruent order, the nationality adjective precedes the size adjective, violating (2). The design of the task was based on Stowe and Kaan (2006). The task was implemented in Ibex Farm (Drummond, 2013). The task and the full dataset are available at <https://repositori.urv.cat/fourrepopublic/search/item/PC%3A3607>.

**Participants.** All participants were neurotypical adults, capable of providing informed consent. All participants provided written informed consent prior to their involvement in the study, in accordance with the Declaration of Helsinki. Regarding ethical approval, the Norwegian Center of Research Data reviewed and approved the study protocol (approval number: 55775/3/LH).

The task was administered to  $n = 139$  bilingual speakers of Greek and a Germanic language, mainly Norwegian, Swedish, Danish, English, or German. All bilingual participants stated they speak Greek as their native language, and at least one Germanic language, with varying degrees of proficiency ranging from good to near native. The original sample involved 167 participants, but 28 participants were excluded on the basis of the following pre-established criteria: (i) providing a series of automatic responses (i.e., RTs below 600 ms), (ii) not completing the task, (iii) non-native knowledge of Greek, (iv) reception of speech-pathology treatment, and (v) presence of neurological disorders. Criteria (iii)-(v) were assessed on the basis of self-report.

All participants were recruited online, through invitations posted on social media platforms, and completed the research online, in Ibex Farm. The language of testing was Greek. In previous work (Leivada and Westergaard, 2019), this task was run to two Greek-speaking populations that are different from the bilingual population tested in this work: monolingual speakers of Standard Greek and bidialectal speakers of Standard and Cypriot Greek. The results showed the high acceptability of the structures that deviate from the unmarked order (2), while not finding evidence for an extra processing cost associated with them. In the present work, the task is administered to speakers of Greek who

grew up as monolingual and were consistently exposed to a different language only upon relocating to another country/linguistic community as adults. It is very likely that, unlike the (monolingual) participants of previous experiments on AOPs, the participants of this experiment have been recently made aware of what is the prescriptively correct way of ordering adjectives, and this may affect their perception of a strong linguistic universal in ways that are yet to be determined.

At the time of testing, all participants had resided for a minimum of 4 years outside Greece (mean: 11.8 years, SD: 9.8), mainly in Scandinavia, UK, or Germany. Further information about the participants' length of residence in their L2/n communities is given in the Supplementary Information. Table 1 presents the participant demographics.

Results

The results were analyzed using jamovi, version 1.8 (the jamovi project 2021; R Core Team, 2021). RTs showed a skewed right tail; the standard logarithm ( $RT' = \log_{10}(RT)$ ) was applied to normalize them, and then the classical 3 SD filter was used to detect outliers. As a result, 15/1251 outliers have been removed from the congruent order (1.19%) and 13/1251 from the incongruent order (1.03%). In total, the results include 2474 data points for the on-line measure (reaction times) and 2502 for the off-line measure (acceptability judgments).

Starting from the acceptability judgments, the results show that participants accept both the congruent and the incongruent sentences as correct (Fig. 1). The latter is the predominant answer, followed by 'neither correct, nor wrong', and then by 'wrong', and this pattern is observed in both orders. Splitting for condition leaves this pattern unaltered (Fig. 2), although a more pronounced difference between the sentences that comply with the hierarchy in (2) and those that violate it can be observed. To test these differences statistically, a generalized linear model analysis was run with the categorical dependent variable being treated as multinomial. The model showed the effect of order on

Table 1 Participant demographics.

Gender	Female	74
	Male	65
Education	Secondary	18
	Tertiary	121
Handedness	R	121
	L	18

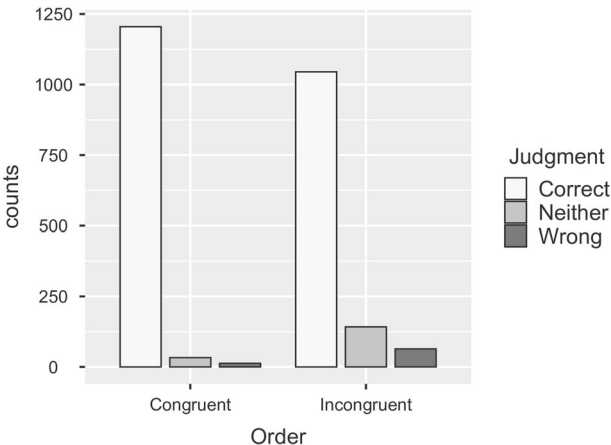
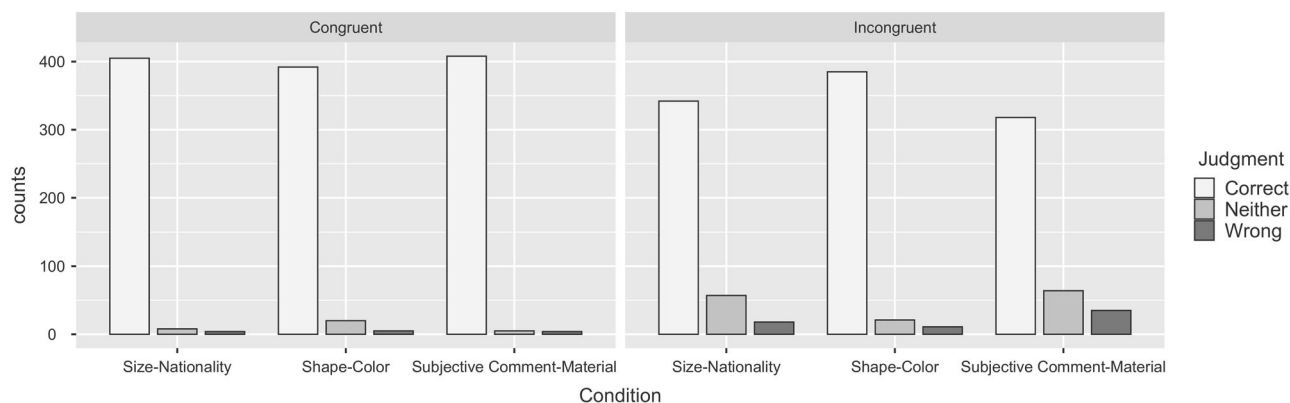
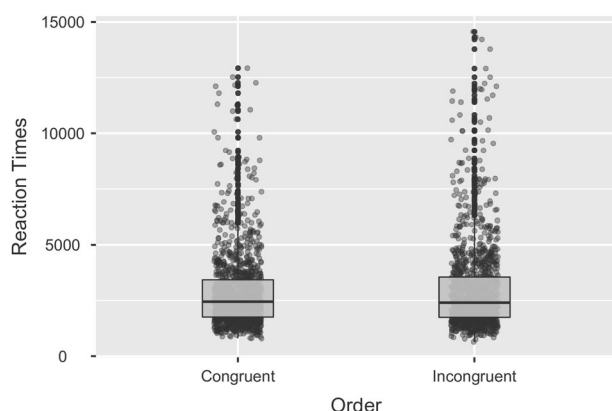


Fig. 1 Judgments split by order. The bars show raw scores.



**Fig. 2 Judgments split by order and condition.** The bars show raw scores.



**Fig. 3 RTs split by order.** The y-axis shows inverse-transformed ms.

acceptability judgments to be statistically significant ( $\chi^2 = 100.38$ ,  $p < 0.001$ ), the effect of condition to not be significant ( $\chi^2 = 2.22$ ,  $p = 0.694$ ), and the interaction order\*condition to be significant ( $\chi^2 = 36$ ,  $p < 0.001$ ).

Within conditions, the effect of order varies: the difference between congruent and incongruent orders is significant in the conditions ‘size-nationality’ ( $\chi^2 = 56.6$ ,  $p < 0.001$ ) and ‘subjective comment-material’ ( $\chi^2 = 99.2$ ,  $p < 0.001$ ), but not in the condition ‘shape-color’ ( $\chi^2 = 2.39$ ,  $p = 0.302$ ). This finding agrees with the results of previous experiments (Adam and Schecker, 2011; Leivada and Westergaard, 2019) and is due to the *distance effect*: the closest two adjective classes are in (2), the more interchangeable their members are in terms of order.<sup>4</sup> Importantly, this finding does not translate into evidence in favor of SOT, because the distance effect is compatible also with proposals that ground ordering preferences on cognitive notions (Scontras et al., 2017; Leivada and Westergaard, 2019).

Turning to the on-line measure, Fig. 3 shows the RTs split for order. The aim is to shed light on whether the incongruent, hierarchy-violating sentences are marked or dispreferred and, as such, incur an extra processing cost compared to their hierarchy-compliant, unmarked counterparts. Figure 4 brings condition into the picture. Both figures show that the obtained medians are quite similar across orders and conditions. Treating the dependent variable as continuous, a generalized linear model showed that processing times do not differ significantly when comparing congruent and incongruent sentences (effect of order:  $\chi^2 = 2.837$ ,  $p = 0.092$ ). The effect of condition and the interaction order\*condition do not reach the significance threshold either ( $\chi^2 = 0.111$ ,  $p = 0.946$  and  $\chi^2 = 2.052$ ,  $p = 0.358$  for condition and order\*condition, respectively).

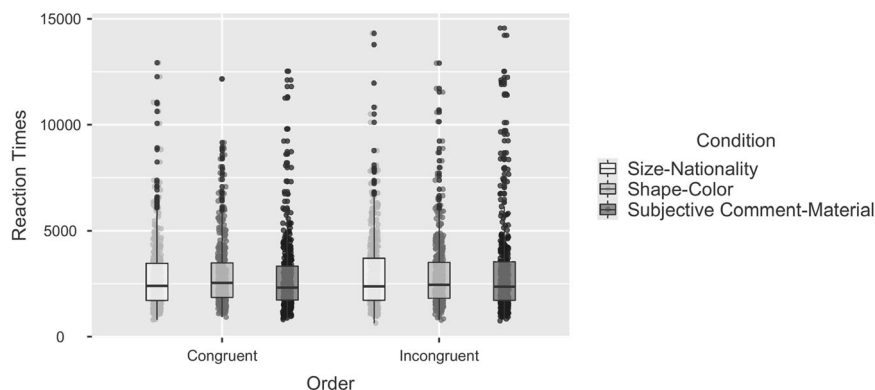
Bringing together the two measures, acceptability judgments and RTs, Fig. 5 shows that for both orders, the judgment ‘correct’ is the one that is associated with the shortest decision times. More importantly, unlike what one would expect if the incongruent order was marked and legitimized only under special licensing conditions, Fig. 5 shows that it is the congruent order that is associated with slightly longer RTs, when the sentences are accepted as ‘correct’. According to a generalized linear model, the effect of judgment on RTs is significant ( $\chi^2 = 339$ ,  $p < 0.001$ ). Figure 6 presents a more direct comparison of acceptability judgments and their associated RTs in the two orders.

Having presented the results of the dataset produced by this experiment, the next aim is to determine the effect of language group, by comparing this dataset with that of other populations that completed the same task. Taking the group of  $n = 140$  monolingual speakers of Greek presented in Leivada and Westergaard (2019) as the monolingual comparison group, we compare the two datasets in order to examine the possible effect of developmental trajectory. Recall that although both groups grew up as monolinguals in Greece, the bilinguals tested in the present study relocated as adults to a linguistic community where their L1 is not societally present. Figure 7 shows a first comparison of the two groups:  $n = 140$  monolingual speakers of Greek and  $n = 139$  bilingual speakers of Greek and a Germanic language.

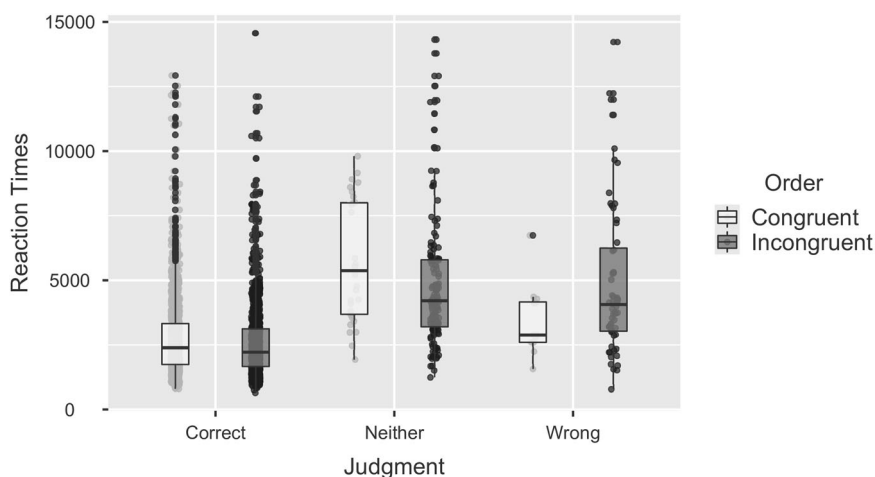
To evaluate the difference in judgments between the two groups, a generalized linear model analysis was run with the categorical dependent variable being treated as multinomial. The effect of language group on acceptability judgments was not found to reach significance ( $\chi^2 = 0.075$ ,  $p = 0.963$ ), but the effect of the interaction language group\*order is marginally significant ( $\chi^2 = 6.459$ ,  $p = 0.040$ ). However, these two results are not very informative about the potential differences between the two groups, because half of the sample on which they are based consists of stimuli from the congruent order, for which the predominant answer in both groups is the target answer ‘correct’, as expected. In other words, there is no room for differences in this order. For this reason, a separate analysis of accuracy in the two groups was conducted, targeting the incongruent order alone. Treating accuracy as a two-level variable (accurate vs. inaccurate), the acceptability judgments given to the incongruent stimuli were classified as accurate if they matched the target answer (which for this order is ‘wrong’) and as inaccurate if otherwise. Figure 8 shows the behavior of the two groups. Figure 9 shows the distribution of judgments across conditions in the two groups.

A generalized linear model examined the significance of the effect of group and condition on accuracy in the incongruent order, as well as their interaction, treating accuracy as logistic. The effect of language group on accuracy is significant ( $\chi^2 = 7.48$ ,  $p = 0.006$ ), the effect of condition is significant too ( $\chi^2 = 43.40$ ,  $p < 0.001$ ),

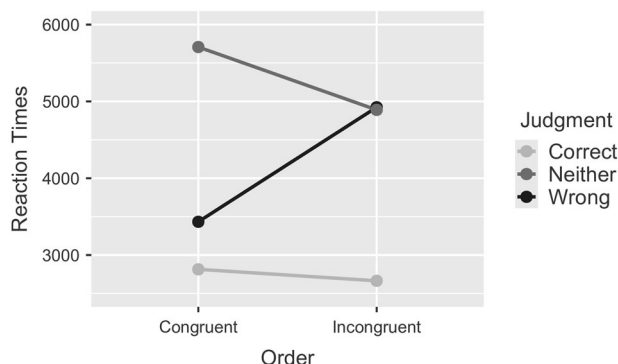




**Fig. 4 RTs split by order and condition.** The y-axis shows inverse-transformed ms.



**Fig. 5 Judgments and RTs split by order.** They-axis shows inverse-transformed ms.



**Fig. 6 A direct comparison of judgments and reaction times across orders.** The y-axis shows inverse-transformed ms.

while the interaction language group\*condition is not significant ( $\chi^2 = 4.53$ ,  $p = 0.104$ ). Post-hoc tests with Bonferroni correction confirmed the effect of language group ( $p = 0.014$ ) and the effect of condition (condition 'shape-color' vs. condition 'subjective comment-material':  $p < 0.001$ , condition 'size-nationality' vs. condition 'shape-color':  $p = 0.011$ , condition 'size-nationality' vs. condition 'subjective comment-material':  $p = 0.003$ ).

The second important finding showed in Fig. 8 is the degree to which the behavior of the participants deviates from what all theories describe as the ill-formed, odd, or marked order. Both SOT and mCOT predict that the incongruent stimuli employed in this study should be rejected or treated as marked respectively;

however, the participants of the present experiment accepted them as well-formed, without taking extra time to process them compared to the congruent stimuli.

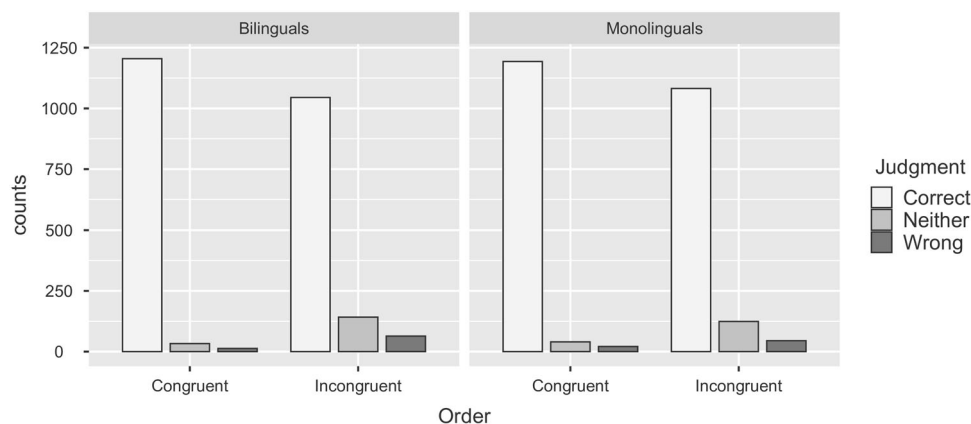
## Discussion

The research questions (RQ) behind this experiment are the following:

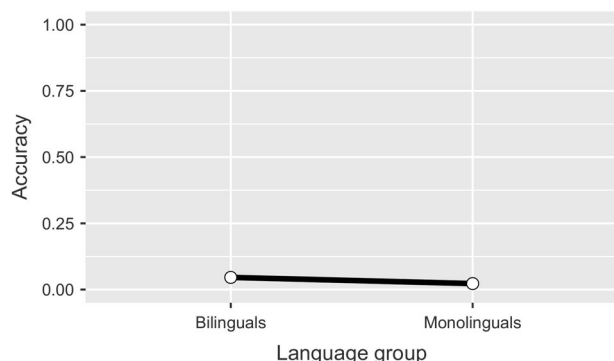
- (I) What cognitive needs are subserved by AOPs?
- (II) Is there interspeaker variation in the attested preferences?

Given that the time component of the experiment did not adduce evidence for an extra processing cost for the incongruent orders—as should have happened if this was a marked order that the parser either disprefers or legitimizes only under special licensing conditions—the conclusion is that all the orders in (1), that have been variably described as ungrammatical, odd, awkward, marked, or semantically incorrect, are grammatically well-formed and highly acceptable. At the same time, the congruent orders elicited a higher degree of acceptability than the incongruent ones. This finding, in combination with the results from other experiments that found robust ordering preferences (e.g., Scontras et al., 2017), begs the question: If the parser consistently likes some orders more than others, some cognitive needs must be subserved by such preferences. Which are these needs?

So far, this question has not been tackled in a multifactorial way that goes significantly beyond observations formed at the surface level (but see Kotowski and Härtl, 2019 for an exception). To explain this better, an important note is due with respect to the various notions put forth by the various subproposals within



**Fig. 7 Comparison of judgments across orders in two populations: monolinguals and bilinguals.** The bars show raw scores.



**Fig. 8 Accuracy in rejecting the incongruent stimuli.** The comparison is between the bilingual participants of the present experiment and the monolingual participants of Leivada & Westergaard (2019).

mCOT: Subjectivity, inherentness, noun-specific frequency, high collocability, and phonological weight/length are not explanations of the observed AOPs; they are observations of what happens at the phenotypic level. Put differently, the fact that an adjective A is statistically likely to occur given a noun N, or that less subjective adjectives tend to appear closer to the noun, are surface observations that must be explained from a cognitive point of view; they are not explanations of the origin of AOPs themselves. They tell us *what* happens at the phenotypic level, but not *why* it happens and what cognitive needs trigger this behavior. Therefore, the aim is to tackle RQ (I) by uncovering the cognitive underpinnings of AOPs, focusing on the parts of the hierarchy in (2) that have been examined in this study: 1. size adjective-nationality adjective, 2. shape adjective-color adjective, 3. subjective comment adjective-material adjective.

Starting with the condition ‘size-nationality’, nationality adjectives appear closest to the noun because of their special nature as sociopragmatic conventions. Succinctly put, nationality/origin adjectives form idiosyncratic concepts that encapsulate the pragmatic conventions of the linguistic community in which they are uttered. To give a concrete example, what is often referred to as Turkish coffee in Turkey, is called Greek coffee in Greece, Bosnian coffee in Bosnia, and Cypriot coffee in Cyprus. Essentially, these nationality adjectives do not refer to a property *inherent* to the described object (i.e., Turkish/Greek/Cypriot/Bosnian coffee refers to the same type of coffee); they rather express a fixed relation between the noun on which they are formed and the noun they modify in each of the respective languages/linguistic communities. This relation is fixed within, but not across languages/linguistic communities, something that

attests to its idiosyncratic, pragmatically determined nature. For this reason, nationality adjectives pose a challenge for subjectivity and inherentness accounts: A black cat is black in all linguistic communities, and it is unlikely that faultless disagreement will emerge over its blackness. The same cannot be claimed for the nationality adjective in ‘Turkish coffee’. Examples (3a-d) can all receive similar analyses to the one for Turkish coffee.

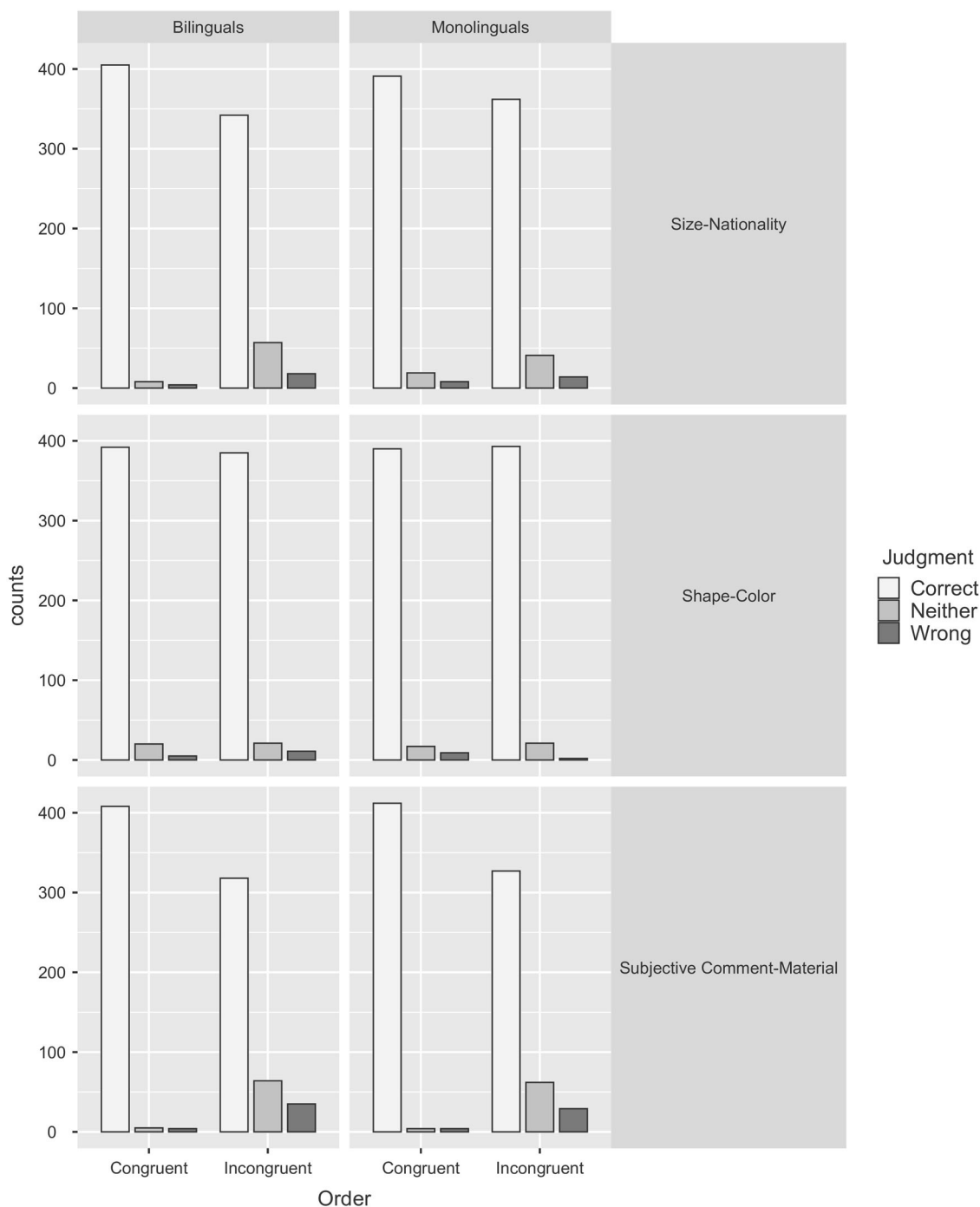
- (3) a. Turkish delights
- b. Russian salad
- c. Bavarian cream
- d. Italian dress

In (3a-d), the adjective expresses an idiosyncratic relation to the noun that is determined by pragmatic conditions and does not necessarily refer to origin. Turkish delights are called Greek delights in Greece, while they are not known as Turkish delights as Turkey. The Russian salad is not known as such in Russia and was not invented by a Russian. The Bavarian cream was neither conceived in Bavaria, nor by a Bavarian, while it is perfectly possible that a dress marketed by an Italian brand is manufactured in another country and designed by a non-Italian designer. In all these examples, there is *some* type of relation expressed between the nominal and the adjective, but this relation is neither semantically transparent, nor always the same: It is place of production in (3a), historical origin in (3b), first recipient in (3c), and origin of brand in (3d).

One could counter that ‘Spanish coast’ is cross-linguistically uncontroversial if one knows basic geography. Yet, even in this example, the relation between the nationality adjective and the noun is a matter of idiosyncratic convention and cross-linguistic variation exists (cf. Costa Brava being described as a ‘Spanish coast’ vs. a ‘Catalan coast’). To explain the idiosyncrasy, in order to classify an object as red, a specific condition must be met: the presence of redness. This condition is a salient property of the described object. A Spanish coast though may lack any salient property that can be construed as Spanishness. Unlike ‘red eyes’, ‘red car’, and ‘red soil’, which share the tangible property of redness, ‘Spanish eyes’, ‘Spanish car’, and ‘Spanish soil’ denote some relation to Spain or to an individual from Spain, but this relation is idiosyncratic and can be variably described as origin, place of production, or any link to the Spanish country, language, or culture.

Similar is the case with animate referents (4a-b), whereby an idiosyncratic relation is expressed, without the adjective denoting a quality inherent to the nominal.

- (4) a. Meghan Markle was a British duchess.
- b. Giannis Antetokounmpo is an American star.



**Fig. 9 Comparison of judgments across orders and conditions in the two populations: monolinguals and bilinguals.** The comparison is between the bilingual participants of the present experiment and the monolingual participants of Leivada & Westergaard (2019).

Meghan Markle was a British duchess, but she does not originate from Britain and, at least in early 2022, she does not have the British citizenship. Giannis Antetokounmpo is an American star ever since he plays in the NBA, however he is not American, but Greek, born to Nigerian parents, and stateless in the first years of his life. Evidently, there are contexts that legitimize the use of these nationality adjectives as denoting various types of relations. Crucially, the expressed relations are a matter of convention, as there is no *inherent* British or American quality in any of the referents in (4). As in (3), these adjectives express some relation between the country and the individual. Precisely because this relation is largely idiosyncratic and does not refer to an

inherent, objective property of the nominal, these adjectives are preferably placed closer to the noun, being part of a socio-pragmatic convention.

Turning to the condition 'shape-color', our results show that this is the condition with the smallest difference in terms of acceptability ratings across the congruent and the incongruent stimuli (Fig. 2). This finding illustrates the *distance effect*: the closer two adjective classes are in (2), the more interchangeable their members are in terms of order. The reason is not that (2) is an innate hierarchy that predicts a more rigid ordering among its distant components. The explanation we propose is that shape adjectives and color adjectives are variably ordered because the

*expectation value* assigned by the cognitive parser is the same for both these categories of adjectives.

It is a well-established fact that the parser, while processing the linguistic message, forms expectations about incoming stimuli. Words with low expectation value, as in the final word in “She spread the bread with socks” are known to elicit larger N400s (i.e., a negative-going deflection that peaks around 400 milliseconds post-stimulus onset) than expected words, showing that the cognitive parser reacts when encountering deviations from the expectations it has formed (Kutas and Hillyard, 1980). Going back to the tested condition, it has been found that shape and color are perceived almost simultaneously (Viviani and Aymoz, 2001 and references therein). Since the parser registers them at the same time, and even uses cues from the one to categorize the other (e.g., the prominent role of color in mediating shape, which has been found in non-human primates too; Lafer-Sousa and Conway, 2013), it assigns them the same expectation value and has no *strong* preference in ordering them in a specific way, hence their high interchangeability.

One explanation for any remaining weak preference for placing shape adjectives before color adjectives is that language reflects vision. Under this account, syntactic preferences may be mirroring the syntax found in the perception of a visual object (Pinna and Deiana, 2015). The visual object is a set of multiple properties, both explicit (i.e., readily visible) and implicit. Visual attributes like shape, material, and color are often placed in the foreground, with respect to other properties like illumination, density, or contouring (Pinna and Deiana, 2015). If shape is granted a slightly more prominent position than color in the visual object, the linguistic object may reflect this preference. At the same time, these perceptually induced biases do not seem to provide the full picture. Visual syntax may indeed have a role in the attested orderings, but other cognitive biases come into play. For instance, the production-driven *availability bias* posits that the most available adjectives are placed first to ease production (Fukumura, 2018 and references therein). This claim is supported by evidence from both within and outside the literature on adjective order, as it has been found that attributes that are more familiar or closer to the identity of the speaker tend to be mentioned first (Smirnova et al., 2019). Consequently, language does not always externalize the syntax of the visual object in a faithful or uniform fashion. In fact, linguistic syntax has devices for overwriting the input of visual syntax, based on the communicative needs it faces in different contexts. Focus fronting (1g) for adding emphasis to one aspect of the linguistic message is such an example. This means that driven by the need to subserve different communicative needs, speakers produce shape > color or color > shape, depending on which order is more likely to facilitate referent identification in a specific context, while also taking into account the needs of both the speaker (e.g., availability) and the addressee (e.g., discriminability; Danks and Glucksberg, 1971; Haywood et al., 2003; Fukumura, 2018). In other words, it seems that linguistic syntax plays a key role in AOPs: It is the interface that both mediates the cognitive needs subserved by AOPs and fine-tunes their externalization, deploying different strategies (e.g., fronting) to ensure that the linguistic message is saliently conveyed, satisfying context-specific communicative needs. Therefore, AOPs, as a determining factor of nominal syntax, both underlie adjective serialization and boil down to cognitive factors.

If we attempt to synthesize the overarching connection between variable adjective order and effective communication, it becomes clear that word order is mindful to both production- and perception-driven tendencies. Weighing all these tendencies when deciding which order to produce seems amenable to ultimately being attributable to a general cognitive bias: *Ambiguity*

*Intolerance*. According to this bias, the cognitive parser tends to treat ambiguous situations as undesirable (Frenkel-Brunswick, 1949; Tanaka et al., 2015). The linguistic manifestation of this bias amounts to the Gricean maxim of manner that suggests that one’s conversational contribution must be as clear and as orderly as possible, avoiding obscurity and ambiguity (Grice, 1957). Of course, ambiguity is pervasive both in human life and in human language, but equally pervasive are the strategies we can use to dissolve ambiguities when necessary. Specifically for adjective order, the proposal is that speakers can externalize shape > color or color > shape, placing first the property of the nominal they find most disambiguating in each context (Kemmerer et al., 2007). If the context poses no such need for disambiguation, other factors such as availability and visual syntax representation kick in. This proposal predicts that the incongruent order color > shape is perfectly acceptable both when color is the most appropriate discriminating factor among a set of qualities, but also when color is just one quality among many equally discriminatory ones. Evidence for the second scenario comes from the acceptability judgments presented in the previous section. Specifically for the condition ‘shape-color’, the two orders were found to be identical in terms of their acceptability in the *out-of-context* presentation of the stimuli, which was employed in the present experiment (Figs. 2 and 9). This absence of context is important, because it shows that the two types of adjectives can be ordered freely, even when there are no contextual needs that force the speaker to invert the usual order.

Turning to the third condition, ‘subjective comment-material’, this is the domain where the clearest difference between congruent and incongruent stimuli was found (Figs. 2 and 9). The explanation we propose for this finding is based on Scontras et al.’s (2017) results, and more specifically on their brief observation that as noun phrases are built semantically outward from the noun, the less subjective content is the one that enters earlier in the process (see also Scontras et al., 2019). Addressing the question of what drives this behavior, the answer is that the more subjective adjectives usually enter the computation last, due to a cognitive bias called *Novel Information Bias* (NIB).

NIB refers to the cognitive tendency to avoid tokenizing multiple, adjacent occurrences of the same type, because of a general bias to provide more attentional resources to novel information (Leivada, 2017). Consequently, this information is often granted a more prominent position in the linguistic message to facilitate clear and, to the degree possible, effortless identification. More prominent in this case means first in languages like English (i.e., Adj-Adj-Noun) or last in languages like Spanish (i.e., Noun-Adj-Adj). This happens because one of the key abilities of the parser is to aptly keep track of sequence edges (Endress et al., 2009; Ferry et al., 2015). Therefore, novel information with a low expectation value is typically placed in a saliently accessible position: at the edges of the Adj-Adj-Noun/Noun-Adj-Adj constellation.

To unpack the somewhat unclear notion of novelty, a concrete finding from corpus studies refers to how often an adjective collocates with a noun. For instance, Wulff’s (2003) results suggest that noun-specific frequency is a highly significant factor that mediates order: adjectives with high noun-specific frequency tend to appear closer to the noun in a multi-adjective string. Noun-specific frequency is proportional to expectation value—due to the high collocability of the adjective and the noun, the parser expects to see them together—and inversely proportional to novelty: The higher the expectation value is, the lower becomes the novelty. Put another way, a high expectation value entails a diminished degree of novelty that may surprise the addressee. The notions of novelty and surprise should be understood in this context as referring to attributes that the parser does not actively



Table 2 Frequency of constructions in the categories 'subjective comment + toy' and 'material + toy' based on COCA.			
Subjective comment adjective		Material adjective	
Nice	22	Plastic	73
Ugly	0	Wooden	18
Good	19	Metal	12
Bad	1	Rubber	12
Precious	3	Porcelain	1
Pretty	9	Stuffed	94
Silly	10	Furry	2
Cheap	12	Fluffy	4
Cute	10	Plush	40
Expensive	24	Magnetic	2
Total	110	Total	258

expect in a context, because it does not consider them default dimensions of information that often appear with the nominal. For example, in the context of talking about kittens, the adjective 'fluffy' has a higher expectation value and a lower degree of novelty compared to the adjective 'boring'.<sup>5</sup>

Let's illustrate what this means for the way subjective comment adjectives are ordered in relation to material adjectives more generally, through comparing 'nice toy' (subjective comment) to 'plastic toy' (material). If 'plastic' collocates with 'toy' more often than 'nice' does, the parser will not be surprised if it sees 'plastic' as close to this nominal as possible. 'Nice', on the other hand, denotes a subjective comment that is primarily informative about the speaker's perception of the object, not the object itself. If an adjective like 'nice' patterns with a large set of nouns, while an adjective like 'fluffy' or 'plastic' is compatible with a smaller set (i.e., specific Spelke objects), then 'fluffy' or 'plastic' have a high expectation value in the contexts of nouns denoting these specific Spelke objects. In other words, we expect to see 'fluffy' being mentioned in relation to kittens or 'plastic' in relation to toys, but our expectations about 'nice' are weaker because this adjective patterns with many nouns. When 'nice' and 'plastic' must be ordered in one construction that features 'toy', the parser recognizes that 'nice' is more generic and less noun-specific, and thus assigns it a lower expectation value in constructions that feature this noun. This claim relies on two premises: first, there are more constructions in the 'material + toy' category than the 'subjective comment + toy' category, and second, the parser is mindful of such differences in frequency. The second premise has already been established empirically (see Wulff, 2003). To shed light on whether the first premise holds, the Corpus of Contemporary American English (COCA) was searched. The first search featured one of the following subjective comment adjectives: {nice, ugly, good, bad, precious, pretty, silly, cheap, cute, expensive} + toy. The second search featured one of the following material adjectives: {plastic, wooden, metal, rubber, porcelain, stuffed, furry, fluffy, plush, magnetic} + toy. The results, given in Table 2, grant support to the first premise.

To continue with the previous example, having established that 'plastic' has a higher noun-specific frequency value in relation to 'toy', it follows that the parser assigns to their co-occurrence a high expectation value. As a result, when the speaker/signer must choose between placing 'plastic' or 'nice' next to 'toy' in a multi-adjective string, 'plastic' wins as the adjective that has the higher expectation value, because the parser prefers to have the novel, less object-oriented, and less expected in the context, information placed at the outmost position to facilitate easy retrieving and orienting attention accordingly.<sup>6</sup> By extension, the slot closest to

the noun will host adjectives that do not need the special, outmost position, because such adjectives are already remembered easily enough upon the occurrence of the noun (Lockhart and Martin, 1969), probably due to the fact that they describe predictable properties (Eichinger, 1991). In other words, besides sociopragmatic conventions that form idiosyncratic concepts (e.g., Russian salad, Spanish coast, American star) and give rise to specific orders, higher-level ontological categories, such as saliently observable object-oriented attributes vs. speaker-oriented evaluative attributes, appear to come into play. High collocability may then convert an attribute and a nominal into a stereotype (e.g., 'fluffy kitten'), and this in turn further increases the noun-specific frequency (Posner, 1986). Bouchard (2005) also proposes concept iconicity as a general principle of serialization of adjectives: If the adjective is likely to form an idiomatic concept with the noun, it tends to be placed close to it (see also Kotowski, 2016 for a review).

Relations between higher-level ontological classes come into play *after* sociopragmatic conventions that form idiosyncratic concepts: If you show somebody 'a red Russian ball', its color is immediately more apparent and less likely to cause any disagreements than its origin (Scott, 2002). Therefore, according to the literature that puts forth the subjectivity rule (e.g., "the major rule is to place the more objective and undisputable qualifications closer to the noun, and the more subjective, opinion-like ones farther away" Hetzron, 1978: p. 178; see also Scontras et al., 2017), we should observe the reversed version of the pattern we observe: COLOR and not ORIGIN should be placed closest to the noun. This does not happen because idiosyncratic conventions take precedence over other relations between adjective classes.

The last factor that plays a key role in adjective order is phonological weight, also referred to as length in the literature. This factor posits that when two or more adjectives are freely ordered, the lengthier adjectives tend to appear closer to the noun (Wulff, 2003; Kotowski, 2016). Although the results of the present experiment are not directly informative about the weight factor, because this was controlled for in the experimental design (see Leivada & Westergaard, 2019), it is worth integrating it in the overall discussion of the cognitive underpinnings of AOPs. The reason is that this factor stands out from the rest for it does not refer to some semantic notion (e.g., subjectivity, inherentness, absoluteness) or some observation over the distribution of the data (e.g., lemma frequency, noun-specific frequency, degree of collocability). Briefly put, this factor, unlike any other, has been presented in the literature as a purely phonological one (Wulff, 2003). However, it is unclear why or even how the articulatory-motor interface can have a say that affects word order (i.e., syntax). It is equally unclear why the rule 'place the lengthier adjective closer to the noun' would be activated only with adjectives that are freely ordered (e.g., shape-color), and what prevents it from generalizing and applying more broadly, especially since it seems that all adjectives are freely ordered to varying degrees, and there are no flat-out rejected, unacceptable orders (cf. endnote 3 and Fig. 2).

Addressing these issues, the first step is to propose that weight/length is relevant across categories of adjectives, since the results of this experiment show that there are preferred orders, but not unacceptable or ungrammatical ones. Second, the way this factor has been presented in the literature brings forward an overlooked problem. In certain linguistic frameworks (e.g., the inverted-Y model in Minimalism and its precursors; Chomsky and Lasnik, 1977), it is hard to sustain the claim that phonology affects syntax, as the latter is taken to be "phonology-free" (Miller et al., 1997; Irurtzun, 2009). The solution to this problem lies in recognizing the cognitive needs of the parser. Unlike previous studies that presented this factor as a phonological one, we

propose that its phonological repercussions are only an epiphenomenon, and the effect itself boils down to a cognitive principle called the *Principle of Least Effort*.

It is well known that words that are used more frequently (e.g., ‘and’, ‘the’) tend to be shorter (*Zipf’s law of abbreviation*; Zipf, 1932; 1949). Zipf theorized that this pattern is the result of accommodating two competing needs: the pressure to take the path that entails the least effort (i.e., short words need less effort to produce) and the pressure for successful communication (i.e., short words are more susceptible to noise in the transmission of the linguistic message). The Principle of Least Effort was proposed as an explanation of the law of abbreviation. Language strives for optimizing form-meaning mappings under competing pressures, such that a ‘frequency-length-meaning’ relationship is formed: words that are used more frequently tend to be shorter and tend to have the most frequent meanings (see Kanwal et al., 2017 for a recent overview).

If some adjectives are more generic in meaning, it follows that they are used more frequently than others and are compatible with many nouns. For example, the evaluative adjective ‘good’ has a frequency of 1,130,305 in COCA, while the material adjective ‘plastic’ has a frequency of 43,844. It is likely that genericity/frequency of meaning is what makes some adjectives appear first in an Adj-Adj-Noun construction, while by virtue of Zipf’s law, it is frequency of use that makes them be shorter. Under our explanation, the weight/length factor has been mistakenly identified as an individual factor that pertains to phonology, according to the literature; it is a by-product of the Principle of Least Effort. Also, unlike many accounts of Zipf’s law that focus exclusively on the interaction between length and frequency, the explanation of AOPs put forth here suggests that *meaning* acts as the main determinant of the attested orders, affecting both frequency and length. As Piantadosi (2014) puts it, word meaning is the best causal force in shaping frequency. In his words, “[h] appy” is more frequent than “disillusioned” because the meaning of the former occurs more commonly in topics people like to discuss” (Piantadosi, 2014: np).

Another important finding is the effect of language group on providing the target acceptability judgment: ‘correct’ for the congruent stimuli and ‘wrong’ for the incongruent stimuli. The effect of language group on accuracy provides an answer to RQ (II), about the existence of interspeaker variation in the attested preferences. As the previous discussion suggested, AOPs are sensitive to the statistical distribution of the input data. According to Yang’s (2000) *Generalized Statistical Learning Hypothesis*, this is a general property of acquisition. The child learner can be parallelized to a generalized data processor, which approximates the target language based on the statistical distribution of the input data. If this is an accurate description of the process of extrapolating the target grammar, statistical learning can explain the obtained differences in accuracy across language groups (Figs. 7 and 8). The effect of language group (monolingual vs. bilingual) in providing the target judgment in the incongruent order was found to be significant, with bilinguals performing better than monolinguals. This finding is at odds with the long-entertained claim that a linguistic universal is behind AOPs. If AOPs boil down to an innate universal, how is it possible that different groups of informants *variably* accept sentences that flatly violate it as well-formed?

The answer has to do with the statistical distribution of the data. The observed interspeaker variation among speakers of the same language attests to the sensitivity to the statistical properties of the input. Although both groups were raised as monolinguals in Greece and were tested in their L1, bilinguals differ from monolinguals in having relocated in an L2 community as adults. This means that they were consistently exposed to a foreign

language and learned aspects of its grammar through learning the rules. Dictionaries and books of grammar that are used in foreign language learning contexts offer explicit instructions on adjective use (e.g., see the lemma ‘Adjectives: Order’ in the Cambridge Dictionary). In other words, it is very likely that bilinguals have received instructions about the prescriptively correct order of adjectives in their L2/n, and these rules mention a version of the hierarchy in (2). This explanation suggests that the multifactorial origin of AOPs relies on different cognitive biases, but in practice, the realization of different orders in language is also modulated by statistical learning over the input.

Last, addressing the scope of the obtained results, a question that arises is whether these favor one of the two proposals about the origin of AOPs, SOT, or mCOT, which have a syntactic and a cognitive orientation, respectively. This is not one of the two research questions of this experiment, and the experimental design cannot adduce evidence for either one of these theories. In fact, it can be argued that a theory does not have to choose between these two theories, under the premise that syntax reflects cognitive principles. In this sense, the obtained results cannot confirm or disconfirm any of the two theories, SOT or mCOT, and the results do not seem more probable under one of them, because it seems that the two theories together can explain them. There is, however, one important disagreement between SOT and mCOT, and in this respect the obtained results seem to tentatively favor the latter: *rigidity*. More specifically, the obtained results suggest that it is more meaningful to talk about ordering *preferences* than rigid hierarchies that predict that certain orders are ungrammatical (Cinque, 2014).<sup>7</sup>

As mentioned in the Introduction, many syntactic accounts within SOT accept that AOPs may be violated if the emphasis (on any adjective) shifts, if the scope relations change (Alexiadou et al., 2007 and references therein), or in cases of parallel modification (Ferris 1993), where a pause exists in between the adjectives (e.g., ‘a blue, beautiful, woollen, expensive jumper’). In other words, many syntactic accounts accept that the ordering restrictions are not absolute, recognizing some freedom in the ordering, especially in the presence of specific communicative needs (e.g., emphasis, contrastive focus). From this perspective, the variation in the orderings is not at odds with all syntactic accounts. At the same time, the underlying assumption in many SOT accounts is that, in the absence of any special licensing conditions like emphasis, there *is* an unmarked order that predicts a rigid spine of adjectives. Although the obtained results do not conclusively settle this matter, it seems that variation exists even in the default, unmarked setting (i.e., in the *out-of-context* presentation of the stimuli in the present experiment that does not call for emphasis or other special conditions that justify any deviation from the unmarked order), casting some doubt on claims about rigidity in what is typically viewed as the unmarked order.

**Outlook.** A timed acceptability judgment task showed that deviations from what is often deemed as the universal order for adjectives are highly acceptable. The reaction times component showed that the acceptability of these deviating orders is not subject to long processing times, contrary to what should have happened if these sentences were marked and legitimized under special licensing conditions. Overall, three cognitive principles were identified as driving AOPs: the Ambiguity Intolerance, the Novel Information Bias, and the Principle of Least Effort. These principles explain why one order is deemed more natural than others, but crucially they do not predict the ungrammaticality of the other orders. Upon comparing two groups of speakers of the same language, who differ in terms of their developmental trajectory (i.e., monolinguals vs. late bilinguals), we observed significant variation in their acceptability judgments of the deviating

orders. This finding is unanticipated since the relevant literature has claimed that an innate universal is behind adjective ordering. Precisely because the obtained results point to the existence of adjective ordering preferences, and not adjective ordering restrictions that ban certain orders via an innate universal, we argue that these preferences are the outcome of the synergistic interplay of specific cognitive biases in terms of origin, but in terms of their manifestation in language, they are subject to statistical processing and sensitivity to the input, giving rise to interspeaker variation.

## Data availability

The full dataset is available at <https://repositori.urv.cat/fourrepublic/search/item/PC%3A3607>.

Received: 24 February 2022; Accepted: 14 November 2022;

Published online: 05 December 2022

## Notes

- 1 The term 'adjective ordering preferences' is used instead of 'adjective ordering restrictions' because no order is flat-out rejected. Recent experiments have found that orders that deviate from (2) may be less preferred than the compliant order (Scontras et al., 2017), but still, they are acceptable to varying degrees (Leivada and Westergaard, 2019).
- 2 Although this factor has been argued to impact ordering independently of subjectivity/inherentness (Hahn et al., 2018), the two converge on giving rise to the same order. The interaction of these two factors poses a chicken and egg question: Did a high degree of collocability (due to noun-specificity) result to certain ordering preferences or did subjectivity/inherentness-based ordering preferences result to a higher degree of collocability between certain nouns and adjectives? Similarly, if a causative relation between high collocability and AOPs is assumed, the observation that some nouns and adjectives have a high collocability is not informative about the direction of the effect. Is high collocability the origin of AOPs or did the frequency of collocability become high thanks to independently motivated AOPs that favored certain combinations of adjectives and nouns more than others?
- 3 Consistency here refers to the fact that the adjective ordering in (1a) is attested cross-linguistically. However, the strength of the preference for this order over others is not equally consistent. Many theoretical, corpus, and experimental studies have shown that adjective orderings that deviate from (1a) and (2) are both acceptable and productively used in spontaneous speech (e.g., 'red gorgeous color' in Wulff, 2003; *sabates noves blanques* 'white new shoes' in McNally and Boleda, 2004; *kokino megalò vivlio* 'red big book' in Alexiadou, 2014; *großen schönen Audi* 'big beautiful Audi' in Kotowski and Härtl, 2019; *malino omorfo fustani* 'woolen beautiful dress' in Leivada and Westergaard, 2019).
- 4 See section 'Discussion' for the cognitive origins of the distance effect.
- 5 Almost all English dictionaries offer 'cat/kitten/dog' as the prototypical examples of showing how 'fluffy' is used.
- 6 This claim agrees with the predictions of the *Subjective Distance Principle*, according to which attributes that are more familiar, or closer to the identity of the speaker are mentioned first (Smirnova et al., 2019). Also, the current proposal that connects adjective ordering with NIB is in agreement with evidence outside the domain of adjectives that also suggests that language has strategies to place the most unexpected/novel material first. Focus fronting (1g) is one example.
- 7 One possible counterargument is that the rigid order pertains only to direct modification adjectives (Cinque, 2014), such that the attested flexibility in the ordering preferences is not really at odds with SOT. To evaluate this counterargument, let us examine the criteria upon which an adjective can be classified as a direct modification adjective (Cinque, 2014): direct modification adjectives are (i) non-predicative, though they may also have predicative uses; (ii) closer to the noun; (iii) a closed functional class; and (iv) rigidly ordered. The first criterion is somewhat ambiguous in its definition. The adjectives employed in the present experiment have predicative uses. By this criterion alone, it is not clear whether they should be classified as direct modification adjectives or not. The predictions of the second criterion are challenged both by corpus studies and elicitation studies that have provided ample evidence (see endnote 3) that the position closer to the noun can be occupied interchangeably by different categories of adjectives. Although Cinque (2014) recognizes that both 'big brown dog' and 'brown big dog' are possible in English, he also argues that the former corresponds to the *natural* order. He further argues that this reversal should not be taken as showing that no rigid order exists in English among direct modification

adjectives. The third criterion yields unclear results. Adjectives import lexical information. A closed class, however, refers to a category of function words (i.e., determiners, complementizers) that does not readily accept new members. One cannot rule out the possibility that a speaker of a language that has only a few color adjectives, including red, blue, green and yellow, may describe an object as blue-green, thus coining a new lexical item in a way that does not work for true function words (i.e., creating a new determiner with transparent meaning by spontaneously merging two existent determiners). The last criterion is perhaps the most difficult to apply. SOT's claim is that only direct modification adjectives are rigidly ordered, and in order to determine whether an adjective is a direct modification one or not, we employ a criterion that says that these adjectives are rigidly ordered. The circular reasoning behind this last criterion brings along several important challenges about what counts as an explanation in linguistics (Leivada, 2020; Larson, 2021).

## References

- Adam S, Schecker M (2011) Position und Funktion: kognitive Aspekte der Abfolge attributiver Adjektive. In: Schmale G (ed.) *Das Adjektiv im heutigen Deutsch. Syntax, Semantik, Pragmatik*. Tübingen, Stauffenburg, pp. 157–172
- Alexiadou A (2001) Adjective syntax and noun raising: word order asymmetries in the DP as a result of adjective distribution. *Studia Linguist* 55:217–248
- Alexiadou A (2014) Multiple determiners and the structure of DPs. John Benjamins, Amsterdam
- Alexiadou A, Haegeman L, Stavrou M (2007) Noun phrase in the generative perspective. Mouton de Gruyter, Berlin
- Batchelor K (2017) Adjective order: The red big barn. BookHive. <https://www.bookhivecorp.com/blog/research/adjective-order-red-big-barn/>
- Bever TG (1970) The cognitive basis for linguistic structures. In: Hayes JR (ed.) *Cognition and the development of language*. Wiley, New York, pp. 279–362
- Bouchard D (2005) Sériation des adjectifs dans le SN et formation de concepts. *Recherches linguistiques de Vincennes* 34:125–142
- Chomsky N, Lasnik H (1977) Filters and control. *Linguist Inq* 8(3):425–504
- Cinque G (1994) On the evidence for partial N-movement in the Romance DP. In: Cinque G, Koster J, Pollock J-Y, Rizzi L, Zanuttini R (eds.) *Paths towards Universal Grammar. Studies in honor of Richard S. Kayne*. Georgetown University Press, Washington, DC, pp. 85–110
- Cinque G (2005) Deriving Greenberg's Universal 20 and its exceptions. *Linguist Inq* 36:315–332
- Cinque G (2010) *The syntax of adjectives. A comparative study*. MIT Press, Cambridge, MA
- Cinque G (2014) The semantic classification of adjectives. A view from syntax. *Stud Chin Linguist* 35(1):3–32
- Danks JH, Glucksberg S (1971) Psychological scaling of adjective orders. *J Verbal Learn Verbal Behav* 10:63–67
- Dixon RMW (1982) *Where have all the adjectives gone?: And other essays in semantics and syntax*. Mouton, Berlin
- Dowling T (2016) Order force: the old grammar rule we all obey without realising. *The Guardian*. <https://www.theguardian.com/commentisfree/2016/sep/13/sentence-order-adjectives-rule-elements-of-eloquence-dictionary>
- Drummond A (2013) Ibox Farm. Available at <http://spellout.net/ibexfarm>
- Eichinger LM (1991) Woran man sich halten kann: Grammatik und Gedächtnis. *Jahrbuch Deutsch als Fremdsprache* 17:203–220
- Endress AD, Nespor M, Mehler J (2009) Perceptual and memory constraints on language acquisition. *Trend Cogn Sci* 13(8):348–353
- Erdocia K, Laka I, Mestres-Missé A, Rodriguez-Fornells A (2009) Syntactic complexity and ambiguity resolution in a free word order language: behavioral and electrophysiological evidenceS from Basque. *Brain Lang* 109:1–17
- Ferris C (1993) *The meaning of syntax. A study in the adjectives in english*. Longman, London
- Ferry AL, Fló A, Brusini P, Cattarossi L, Macagno F, Nespor M, Mehler J (2015) On the edge of language acquisition: inherent constraints on encoding multisyllabic sequences in the neonate brain. *Dev Sci* 19:488–503. <https://doi.org/10.1111/desc.12323>
- Franken M, Scontras G, Simonić M (2019) Subjectivity-based adjective ordering maximizes communicative success. In: Goel A, Seifert C, Freksa C (eds.) *Proceedings of the 41st Annual Conference of the Cognitive Science Society*. Cognitive Science Society, Montreal, QB, pp. 344–350
- Frenkel-Brunswick E (1949) Intolerance of ambiguity as emotional and perceptual personality variable. *J Pers* 18:108–143
- Fukumura K (2018) Ordering adjectives in referential communication. *J Mem Lang* 101:37–50
- Grice HP (1957) Meaning. *Philos Rev* 62:377–388
- Gutoskey H (2019) The very particular grammar rule you probably never knew—but use every day. *Mental Floss*. <https://www.mentalfloss.com/article/597981/adjective-order-grammar-rule-you-probably-never-knew-about>
- Hahn M, Degen J, Goodman N, Jurafsky D, Futrell R (2018) An information-theoretic explanation of adjective ordering preference. *Proceedings of the 40th Annual Meeting of the Cognitive Science Society (CogSci)*, pp. 1–6



- Hanson D (2015) Ordering adjectives... Who knew? Crafting Connections <https://www.crafting-connections.com/2015/09/ordering-adjectives-who-knew.html>
- Haywood SL, Pickering MJ, Branigan HP (2003) Co-operation and co-ordination in the production of noun phrases. In: Alterman R, Kirsh D (eds.) Proceedings of the 25th Annual Conference of the Cognitive Science Society. Lawrence Erlbaum Associates, Mahwah, pp. 533–538
- Hetzron R (1978) On the relative order of adjectives. In: Seiler H (ed.) Language universals. Papers from the conference held at Gumpersbach/Cologne, Germany, October 3–8, 1976. Narr, Tübingen, pp. 165–184
- Horobin S (2016) Beware the bad big wolf: why you need to put your adjectives in the right order. The Conversation. <https://theconversation.com/beware-the-bad-big-wolf-why-you-need-to-put-your-adjectives-in-the-right-order-64982>
- Imamura S, Sato Y, Koizumi M (2016) The processing cost of scrambling and topicalization in Japanese. Front Psychol 7:536
- Irurtzun A (2009) Why Y: on the centrality of syntax in the architecture of grammar. Catalan J Linguist 8:141–160
- Kanwal J, Smith K, Culbertson J, Kirby S (2017) Zipf's law of abbreviation and the principle of least effort: language users optimise a miniature lexicon for efficient communication. Cognition 165:45–52
- Kemmerer D, Weber-Fox C, Price K, Zdzanzyk C, Way H (2007) Big brown dog or brown big dog? An electrophysiological study of semantic constraints on prenominal adjective order. Brain Lang 100(3):238–256. <https://doi.org/10.1016/j.bandl.2005.12.002>
- Kemmerer D, Tranel D, Zdzanzyk C (2009) Knowledge of the semantic constraints on adjective order can be selectively impaired. J Neurolinguistics 22:91–108. <https://doi.org/10.1016/j.jneuroling.2008.07.001>
- Kotowski S, Härtl H (2019) How real are adjective order constraints? Multiple prenominal adjectives at the grammatical interfaces. Linguistics 57:395–427. <https://doi.org/10.1515/ling-2019-0005>
- Kotowski S (2016) Adjectival modification and order restrictions. The influence of temporariness on prenominal word order. De Gruyter, Berlin
- Kutas M, Hillyard SA (1980) Reading senseless sentences: Brain potentials reflect semantic incongruity. Science 207:203–205
- Laenzlinger C (2005) French adjective ordering: perspectives on DP-internal movement types. Lingua 115:645–689
- Lafer-Sousa R, Conway BR (2013) Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. Nat Neurosci 16(12):1870–1878
- Larson RK (2021) Rethinking cartography. Language 97(2):245–268
- Leivada E (2017) What's in (a) Label? Neural origins and behavioral manifestations of Identity Avoidance in language and cognition. Biolinguistics 11(SI):221–250
- Leivada E (2020) Mid-level generalizations of generative linguistics: Experimental robustness, cognitive underpinnings and the interdisciplinarity paradox. Zeitschrift für Sprachwissenschaft 39(3):357–374. <https://doi.org/10.1515/zfs-2020-2021>
- Leivada E, Westergaard M (2019) Universal linguistic hierarchies are not innately wired. Evidence from multiple adjectives. PeerJ 7:e7438. <https://doi.org/10.7717/peerj.7438>
- Lockhart RS, Martin JE (1969) Adjective order and the recall of adjective-noun triples. J Verbal Learn Verbal Behav 8:272–275
- McNally L, Boleda G (2004) Relational adjectives as properties of kinds. In: O. Bonami O, Cabredo Hofherr P (eds.) Empirical issues in formal syntax and semantics 5, CSSP. pp. 179–196. [http://www.cssp.cnrs.fr/eiss5/index\\_en.html](http://www.cssp.cnrs.fr/eiss5/index_en.html)
- Miller PH, Pullum G, Zwicky AM (1997) The principle of phonology-free syntax: four apparent counterexamples in French. J Linguist 33:67–90
- Nordquist R (2018) Adjective order. ThoughtCo. <https://www.thoughtco.com/what-is-adjective-order-1688972>
- Panayidou F (2013) (In)flexibility in adjective ordering. Queen Mary University of London, Doctoral dissertation
- Piantadosi ST (2014) Zipf's word frequency law in natural language: a critical review and future directions. Psychon Bull Rev 21(5):1112–1130
- Pinna B, Deiana K (2015) Material properties from contours: new insights on object perception. Vision Res 115:280–301
- Posner R (1986) Iconicity in syntax. The natural order of attributes. In: Bouissac P, Herzfeld M, Posner R (eds.) Iconicity. Essays on the nature of culture. Stauffenburg Verlag, Tübingen, pp. 305–337
- R Core Team (2021) R: A Language and environment for statistical computing. (Version 4.0). (R packages retrieved from MRAN snapshot 2021-04-01). Retrieved from <https://cran.r-project.org>
- Saba W (2020) The Universal Adjective-Ordering 'Mystery'—not a Mystery. Medium. <https://medium.com/ontologik/the-universal-adjective-ordering-mystery-not-a-mystery-c10614e50761>
- Scontras G, Degen J, Goodman ND (2017) Subjectivity predicts adjective ordering preferences Open Mind: Discov Cogn Sci. 1:53–65
- Scontras G, Degen J, Goodman ND (2019) On the grammatical source of adjective ordering preferences. Semant Pragmat 12(7):1–21
- Scott GJ (2002) Stacked adjectival modification and the structure of nominal phrases. In: Cinque G (ed.) Functional structure in DP and IP. The cartography of syntactic structures. Oxford University Press, New York, pp. 91–120
- Smirnova A, Romero Sanchez R, Lenarsky A (2019) Contextual determinants of adjective order: beyond itsy bitsy teeny weeny yellow polka dot bikini. In: Goel AK, Seifert CM, Freksa C (eds.) Proceedings of the 41st Annual Conference of the Cognitive Science Society. Cognitive Science Society, Montreal, QB, pp. 2825–2831
- Sproat R, Shih C (1991) The cross-linguistic distribution of adjective ordering restrictions. In: Georgopoulos C, Ishihara R (eds.) Interdisciplinary approaches to language. Kluwer Academic Publishers, Dordrecht, pp. 565–593
- Stavrou M (1999) The position and serialization of APs in the DP. In: Alexiadou A, Horrocks G, Stavrou M (eds.) Studies in Greek Syntax. Kluwer Academic Publishers, Dordrecht, pp. 201–226
- Stowe LA, Kaan E (2006) Developing an experiment. Techniques and design. Unpublished MS
- Svenonius P (2008) The position of adjectives and other phrasal modifiers in the decomposition of DP. In: Kennedy C, McNally L (eds.) Adjectives and adverbs: syntax, semantics, and discourse. Oxford University Press, New York, pp. 16–42
- Tanaka Y, Fujino J, Ideno T, Okubo S, Takemura K, Miyata J, Kawada R, Fujimoto S, Kubota M, Sasamoto A, Hirose K, Takeuchi H, Fukuyama H, Murai T, Takahashi H (2015) Are ambiguity aversion and ambiguity intolerance identical? A neuroeconomics investigation. Front Psychol 5:1550
- Teodorescu A (2006) Adjective ordering restrictions revisited. In: Baumer D, Montero D, Scanlon M (eds.) Proceedings of the 25th WCCFL. Cascadia, Somerville, MA, pp. 399–407
- The Jamovi Project (2021) jamovi. (Version 1.8). Retrieved from <https://www.jamovi.org>
- Trotzke A, Wittenberg E (2019) Long-standing issues in adjective order and corpus evidence for a multifactorial approach. Linguistics 57(2-S):273–282
- Viviani P, Aymoz C (2001) Colour, form, and movement are not perceived simultaneously. Vision Res 41(22):2909–2918
- Waldman K (2014) The secret rules of adjective order. A long fascinating article—or is it a fascinating long article? Slate Magazine <https://slate.com/culture/2014/08/the-study-of-adjective-order-and-gssacpm.html>
- Whorf BL (1945) Grammatical categories. Language 21(1):1–11
- Wulff S (2003) A multifactorial corpus analysis of adjective order in English. Int J Corpus Linguist 8:245–282
- Yang C (2000) Knowing and learning in natural language. MIT, Doctoral dissertation
- Zipf GK (1932) Selected studies of the principle of relative frequency in language. Harvard University Press, Cambridge
- Zipf GK (1949) Human behavior and the principle of least effort: an introduction to human ecology. Addison-Wesley Press, Cambridge

## Acknowledgements

I thank the participants of this experiment for their willingness to devote their time to science. I am also grateful to Antonio Fábregas for his insights on the syntax of adjectives in Spanish. This work received support from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 746652 and from the Spanish Ministry of Science, Innovation and Universities under the Ramón y Cajal grant agreement no. RYC2018-025456-I. The funders had no role in the writing of the study and in the decision to submit the article for publication.

## Competing interests

The author declares no competing interests.

## Ethical approval

The Norwegian Center of Research Data reviewed and approved the study protocol (approval number: 55775/3/LH).

## Informed consent

All participants provided written informed consent prior to their involvement in the study, in accordance with the Declaration of Helsinki.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1057/s41599-022-01440-w>.

**Correspondence** and requests for materials should be addressed to Evelina Leivada.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.





**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022