CrossMark

Applied Psycholinguistics **38** (2017), 601–631
doi:10.1017/S0142716416000345

Visual salience effects on speaker choices: Direct or indirect influences on linguistic processing?

INÉS ANTÓN-MÉNDEZ
University of New England, Australia

Received: November 22, 2015 Accepted for publication: August 7, 2016

ADDRESS FOR CORRESPONDENCE
Inés Antón-Méndez, Linguistics (E11), University of New England, Armidale, New South Wales 2351, Australia. E-mail: iantonm2@une.edu.au

ABSTRACT
The effect of visual salience on speakers' choices is investigated by contrasting the effects of both visual and linguistic manipulations on picture descriptions and eye movements. Two-character pictures were used, which can be described in one of two complementary ways (e.g., a cop chasing a robber can be described either from a chasing or from a fleeing perspective), and using simple actives or other alternative syntactic structures (e.g., "a robber is being chased by a cop"). The pictures were preceded by a verb priming one of the two perspectives and/or a preview of one of the two characters. The results show that the visual manipulation affects looks to the characters regardless of which perspective had been linguistically primed, but it only affects verbal descriptions in the absence of a linguistic prime. Linguistically priming one of the perspectives, in contrast, has a reliable effect on both looks to the characters and verbal descriptions. These results suggest that visual salience does not influence linguistic choices directly.

> Language is the means of getting an idea from my brain into yours without surgery.
> Mark Amidon

When speakers talk, they are normally trying to give linguistic form to a notion they have in their heads that they wish to implant in their interlocutors' heads. Insofar as speakers can and do describe things that they are seeing either at that very moment or with their minds' eyes, it is logical to think that visual characteristics will have an influence on speaker choices regarding most obviously the content, but also the final form of their descriptive utterances. Vision has been found to influence language in many ways. For example, a character who is looked at first for perceptual reasons is more likely to be mentioned first (e.g., Gleitman, January, Nappa, & Trueswell, 2007; Myachykov, Garrod, & Scheepers, 2012; Myachykov, Thompson, Garrod, & Scheepers, 2011); looks to characters in an image have been consistently found to predict mention of the fixated characters

© Cambridge University Press 2016 0142-7164/16

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

(e.g., Coco & Keller, 2012; Coco, Malcolm, & Keller, 2014; Griffin, 2001) and their properties (Haskell, 2005; Meyer, van der Meulen, & Brooks, 2004; van der Meulen, 2001); and visual content influences choice of referring expression (Fukumura, van Gompel, & Pickering, 2010). However, speakers are not always describing a scene, and even when they are, some of the information conveyed is not straightfordwardly attributable to something the speakers perceive or have a perceptual memory of. Moreover, the information derived from the visual system is unlikely to be in a format that can be directly transferred to the linguistic system. Some kind of translation is required (Jackendoff, 1987). In short, the blueprints for sentences are likely to come from a conceptual representation of the perceived event, rather than from the perceived event itself (Bunger, Papafragou, & Trueswell, 2013; Vogels, Krahmer, & Maes, 2013).

Although most researchers will probably agree with the previous statement, some explanations of perceptual influences on language production seem to either disregard the existence of an intermediate level of event conceptualization or confound it with sentence preparation under the implicit assumption that what you see is what you say (Ibbotson, Lieven, & Tomasello, 2013; Montag & MacDonald, 2013; Myachykov et al., 2012; Myachykov, Thompson, et al., 2011), rather than determining what you *may want* to say. More specifically, looks to elements in a scene are linked directly with the building blocks of sentence production rather than with the way the scene is apprehended (see, e.g., Myachykov, Posner, & Tomlin, 2007, p. 464, their figure 2; Myachykov, Thompson, Garrod, et al., 2011). In consequence, the fact that speakers' gazes are drawn to a specific object and this object is then chosen as the syntactic subject of the final description is taken to mean that the starting point in sentence production is a lexical item rather than a syntactic frame (Gleitman et al., 2007; for a detailed description of the "syntactic planning starting point" debate, see Bock, Irwin, & Davidson, 2004).

In truth, however, even when perceptual salience correlates with choice of starting point in a sentence, it is possible that the perceptual properties are influencing how the event is conceptualized rather than linguistic planning directly (Vogels et al., 2013). Cueing a participant in a scene may be one way of anchoring the way the event is interpreted, which could make the cued character then become the topic of the linguistic message, what Bock and Ferreira (2014) call the "aboutee," and thus be the first mentioned noun by virtue of its conferred conceptual status. This would be similar to the special status conferred by other conceptual properties such as animacy, which also influences linguistic choices (Antón-Méndez, Gerfen, & Ramos, 2016; Branigan, Pickering, & Tanaka, 2008; Brawley, 2012; Hickmann, Taranne, & Bonnet, 2009; Prat-Sala & Branigan, 2000).

To determine which of the two mechanisms, scene conceptualization or lexical activation, is responsible for the observed effects of visual salience on linguistic output, it would be necessary to pitch the visual manipulation against another manipulation thought to affect at least one of these two levels of processing: for example, a linguistic prime. Myachykov et al. (2012) conducted such an investigation. In their Experiments 1 and 2, the researchers looked at the proportion of active descriptions of depicted transitive events (e.g., a cowboy punching a boxer) after a linguistic prime consisting of either a matching or a nonmatching verb (e.g., punch vs. kiss), and a visual prime in the form of a dot directing

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

attention to the location of one of the two characters prior to the image appearing on screen. They found that visually cueing the agent of the event (e.g., the cowboy) increased the proportion of active descriptions but only when the linguistic prime was the nonmatching irrelevant verb (e.g., kiss). For the authors, this interaction points to the effect of perceptual manipulations on syntactic choice over and above the effect of a verbal prime. There is an alternative interpretation of these results, however, since there was an imbalance between the two kinds of priming in Myachykov et al.'s design: the visual manipulation primed one or the other participants in an unambiguous event, but the linguistic one either primed or did not prime the event. The interaction could thus be due to the visual cue having an effect only in the absence of a "useful" linguistic one, while when there is a matching linguistic prime, the visual cue did not make any difference. To create a more balanced design, the scenes to be described should allow linguistic priming in both directions, something that can best be achieved by the use of perspective predicates (Gleitman et al., 2007).

To investigate whether perceptual manipulations influence sentence production directly through lexical access, the following experiments were designed to look for the effects that visual salience would have on sentence production if it influenced linguistic processing directly by determining order or mention or argument assignment. Specifically, the experiments looked for the effects of visual priming on scenes that can be described from two complementary perspectives (e.g., a chasing event is also a fleeing event), in the presence of a linguistic prime. The linguistic priming consisted of a verb associated with one of the two perspectives of the event to be described (e.g., either "chase" or "flee"), which should influence descriptions by biasing the speaker toward the associated conceptualization of the event (i.e., reading the word "chase" should make the reader think in terms of chasing and interpret the coming scene from that perspective); and the visual priming consisted of a preview of one of the two characters involved in the event (e.g., the robber or the cop in a chase scene being shown ahead of the full scene), which draws the eyes to first land on the intended character when the complete scene appears.

PREDICTIONS

A scene depicting a cop chasing a robber who is fleeing (see Figure 1) can be described in one of two main ways: as a regular active sentence from a chasing perspective (e.g., "the cop is chasing the robber") or from a fleeing perspective (e.g., "the robber is fleeing from the cop"). In addition, speakers may also choose to describe the scene using either a passive sentence (e.g., "the robber is being chased by the cop"), or a sentence where the direct object has been placed in a more prominent position (e.g., "the robber, the cop is chasing him"). Which of these forms gets chosen should depend on whether the speaker is already primed to think about or construe the scene as a chasing or fleeing event after encountering one of the two verbs. Assuming incremental sentence production (see review of lexicalist frameworks in Bock et al., 2004; Kempen & Hoenkamp, 1987; Levelt, 1989), characters who are more visually salient and therefore are made more accessible should tend to dictate how to describe the scene in which they appear

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

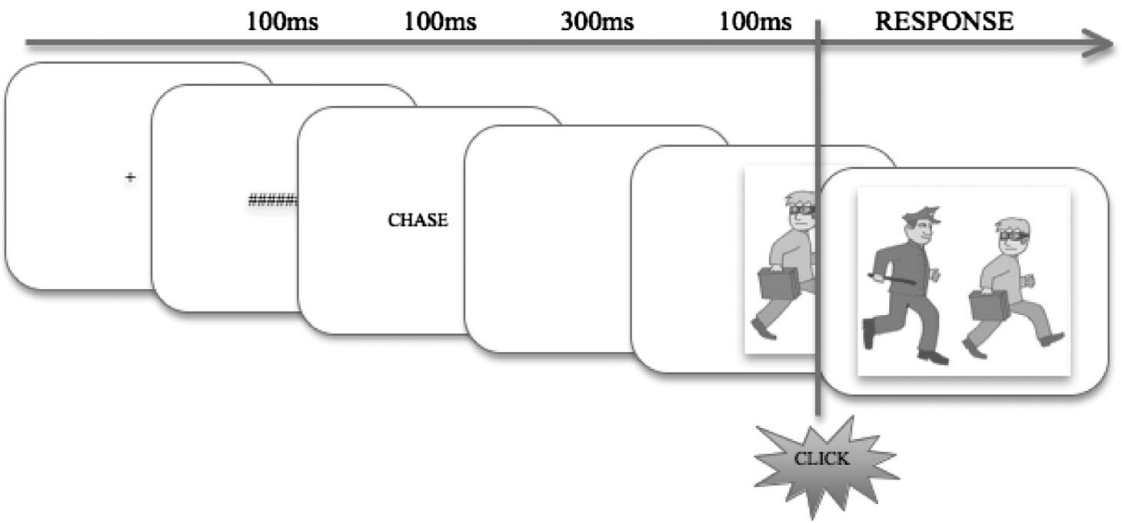


Figure 1. Outline of the experimental procedure for both experiments.

by being processed earlier and, thus, determining the sentence's starting point, as has been found (e.g., Gleitman et al., 2007; Prat-Sala & Branigan, 2000).

In Experiment 1, the scenes to be described are preceded by either a linguistic or a visual prime, but not both, to test the predicted influences of each type of prime on description choice. In particular, a linguistic prime that is known to always match the scene to be described should create an expectation of what the coming scene will be about and, thus, shape the linguistic intention of how to describe it (Myachykov et al., 2012). Therefore, a linguistic prime on its own should influence how the scene is construed (as a chasing or a fleeing event) and how often it is described from the associated perspective; that is, there should be more chase descriptions when "chase" is the prime than when "flee" is the prime. In the case of visual primes, they are also expected to influence the choice of descriptions by either (a) making the associated lexical item more accessible to start with (Gleitman et al., 2007; Myachykov et al., 2007, p. 469) and thus more likely to be mentioned early; or (b) promoting the salient entity to a syntactically prominent position such as subject (Tanaka, Branigan, McLean, & Pickering, 2011; Tomlin, 1997). In the case of active sentences, both possibilities would result in the same kind of description: one in which the visually primed character is expected to be the subject. Therefore, under conditions that do not promote alternative sentence structures, more chase descriptions would be expected after visually priming the cop than after visually priming the robber.

Furthermore, when each prime type is acting independently, we do not expect a difference in the time needed to start speaking depending on the priming direction (i.e., whether the linguistic prime is "chase" vs. "flee," or whether the cop or the robber is visually primed) as there is no reason to believe that one of the priming directions carries any advantage over the other in terms of lexical access or facilitating sentence preparation.

In Experiment 2, the two types of primes are combined to answer the main research question: whether perceptual salience has a direct effect on linguistic processing by determining a sentence's starting point. If the visual prime affects lexical accessibility while the linguistic prime acts by promoting one of the two construals of the scene, the two influences could be accommodated in the eventual verbal description: a referent made more accessible could be processed first and should, at any rate, affect the time needed for sentence preparation (Smith & Wheeldon, 1999; Wheeldon, Ohlson, Ashby, & Gator, 2013). This means that, at the very least, visually priming the cop should result in shorter sentence preparation times when the speaker is preparing a congruent chase description due to the linguistic prime favoring a chasing construal, but in longer preparation times when the speaker is preparing an incongruent flee description because the linguistic prime favors a fleeing construal.[1] Therefore, a statistical interaction would be expected for sentence preparation times because the direction of priming of the visual cue would be associated with opposite effects depending on the direction of the linguistic prime.

An effect of visual salience on lexical accessibility could also affect the form of the description instead of (or in addition to) how long it takes to start talking. For example, if the linguistic prime induces a construal of chasing while the lexical item for "robber" is more accessible, the speaker could take advantage of the

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

increased accessibility of this lexical item by fronting the object (e.g., "the robber, the cop is chasing him"), or uttering a passive sentence (i.e., "the robber is being chased by a cop"). In this case, visually priming the robber should result in more alternative syntactic structures than visually priming the cop when the linguistic prime favors a chasing description, and in fewer such structures when the linguistic prime favors a fleeing description. Therefore, as for sentence preparation times, a statistical interaction would be expected with respect to the number of alternative syntactic structures. Such an effect would mean that visual salience affects order of mention (first ready, first out) and possibly even syntactic assignment (first one ready must be the subject). In turn, this would be evidence of sentence production proceeding in a very piecemeal fashion, something that has been termed radical incrementality (Ferreira & Swets, 2002). Object fronting, however, is very unusual in English, so while this language would allow us to see if visual salience affects syntactic function assignment by examining passive production, it would not be as useful in looking at the potential effects on order of mention independently of syntactic function. Spanish, in contrast, has more flexibility in word order and, therefore, more options for speakers to alter the order of mention leaving all else unchanged (e.g., "*al ladrón lo persigue un policía*," "the robber, the cop chases"). Spanish offers, thus, an additional opportunity to see the independent effects of visual and linguistic priming on event description, specifically on order of mention, should these exist.

Finally, in addition to considering description choices and voice onset times (VOTs), eye movements during the interval of sentence preparation were also analyzed. These data were intended, first, to confirm that the visual manipulation was effective and, second, to explore the effect of the linguistic prime on looking patterns. Furthermore, to the extent that visual features of the scene can impact speaker choices, we would expect to see some parallelism between looking patterns and description choices. That is, we would expect the impact of the manipulations on looking patterns to be reflected on description choices.

Throughout the article, the term *construal* is used to refer to how a scene is conceptualized and described. This is strongly related but different from the term *perspective*, which is used to refer to the actual verbs used as primes. Thus, a linguistic cue priming a certain perspective (e.g., "chase" primes a chasing perspective) could result in a verbal description that indicates a compatible construal of the scene (i.e., as a chasing event; e.g., "the robber is being pursued").

EXPERIMENT 1

The total number of conditions (nine) that would have resulted from crossing three levels (no priming plus priming one of the two alternatives) of two different dimensions (visual and linguistic priming), together with the limited number of picturable perspective predicates, necessitated the breakdown of the study into two experiments.

Experiment 1 included the simple, uncrossed, linguistic and visual conditions to determine the effect of each type of manipulation, linguistic and visual, on their own. Based on previous studies, the prediction was that both linguistic and visual manipulations would have an effect on looks to the characters and choice of

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

verbal description. However, there is no empirical or logical reason to believe that priming one character instead of another, or priming one verb instead of another, will have an effect on sentence preparation times. In addition, Experiment 1 also included a nonverbal baseline condition for eye-movement data to rule out visual biases unrelated to the manipulations.

Method

Participants. Participants in this experiment were 44 English-speaking students (M age = 21.3, SD = 4.60; 42.5% male) from the University of California, San Diego, and 45 Spanish-speaking students (M age = 23.8, SD = 5.54; 22.5% male) from either the University of Oviedo or the Autonomous University of Madrid (Spain). American participants were paid US $5, and Spanish participants were paid €5 for their participation. All the participants included were native speakers of either English or Spanish, and all considered themselves monolingual.

In the two experiments reported here, acceptable eye-tracking data loss per participant was set at 20% of all samples. In Experiment 1, this threshold resulted in four English-speaking and three Spanish-speaking participants being excluded from the analyses. Two more Spanish participants were excluded: one because of eye-tracker failure, and one due to voice recording failure.

Materials. Experimental stimuli consisted of 30 colored drawings made by the author depicting perspective predicates (see Figure 1 for an example), that is, events involving two characters doing complementary actions (e.g., selling and buying, chasing and fleeing).[2] The experimental drawings came in two versions that were mirror images of each other.

The experimental drawings were selected from a set of 33 given to 24 native monolingual English speakers (34% males, 100% aged between 18 and 29) and 24 native monolingual Spanish speakers (44% males, 74% aged between 18 and 29, 26% aged between 30 and 59) to describe in their own native language. In this norming study, the drawings were presented one at a time using the survey program Qualtrics (Qualtrics LLC, 2005). Participants were asked to type a sentence describing each scene in a blank window underneath the drawing.

From this set of 33, 30 drawings, which were found to be successful in eliciting the intended sorts of responses, were chosen. For the selected pictures, intended perspective descriptions were given on average 73% of the time (SD = 17%) in English and 67% (SD = 16%) in Spanish. In most cases, speakers showed a clear preference for one of the two scene construals of the event. On average, one of the alternatives was chosen 87% of the time (SD = 13%) by English speakers, and 90% of the time (SD = 12%) by Spanish speakers. This alternative was labeled the "dominant construal" of the event, as opposed to the secondary, or "subordinate construal" of the event (e.g., a description of the scene as a chasing-event indicated a dominant construal of the event, and a description of the scene as a fleeing-event indicated a subordinate construal of the event). Only one of the items had an even distribution of preferences in English and relatively even in Spanish (the construal preferentially shown in Spanish was considered the dominant construal in this case). Paired t tests confirmed the differences between dominant and subordinate

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

construals in both languages (in English, $t = 11.26$, $p = .000$; in Spanish, $t = 12.65$, $p = .000$), and the similarities of dominant-to-subordinate ratios across languages ($t = –1.79$, $p = .084$).

This norming study also provided the specific verbs that each drawing would be paired with as linguistic primes: one for each of the two complementary construals (e.g., "chase" or "flee"; "perseguir" or "huir"). These were selected to be one-word verbs that had been used by norming study participants in their descriptions (or close synonyms if the actual verbs used were not appropriate for the design).

There were also 30 filler drawings,[3] 15 depicting a transitive action and 15 an intransitive one. These drawings contained only one character (e.g., a man chopping wood or a child sleeping). Mirroring the experimental conditions, two-fifths of them were paired with a verb that described the depicted event, two-fifths were preceded by a preview of the character, and one-fifth had no primes. Finally, there were also 4 practice drawings. See Appendix A for a list of linguistic and visual primes.

Design and procedure. The stimuli were presented using E-Prime software, version 2.0. The first part of the experiment presented 6 experimental and 6 filler drawings, all of them unprimed, for the participant to watch at leisure, without having to describe the scene. Participants were unaware that the experiment would later require verbal descriptions. They were told they first had to examine the drawings carefully one by one while their eye movements were monitored, and later they would see them all together and would have to tell the experimenter which one(s) they had liked best and worst. This served as the nonverbal eye-movement baseline intended to find out whether the scenes themselves induced any perceptual biases.

In the second part of the experiment, there were four conditions according to whether the drawings were preceded by a linguistic prime consisting of one of the two perspective verbs (a verb denoting the *preferred perspective* pointing toward the dominant construal, or a verb denoting the *dispreferred perspective* pointing toward the subordinate construal), or a visual prime consisting of a preview of one of the characters in the drawing to be described (the agent in the dominant construal was labeled the *preferred agent*, and the agent in the subordinate construal was labeled the *dispreferred agent*). This operationalization of perceptual salience is arguably more naturalistic than the more inconspicuous attention capture method used in other studies (e.g., Gleitman et al., 2007; Kuchinsky, 2009; Myachykov, Garrod, & Scheepers, 2011), although the two types of attention capture do not seem to differ in their linguistic effects (Myachykov, Thompson, Garrod, et al., 2011). In any case, because the aim is to see how something that catches our visual attention affects further linguistic processing, it is irrelevant whether the visual attention is captured by means of a location marker or a preview of the character, as long as the attention *is* captured.

Stimuli were counterbalanced across 10 lists: 2 drawing orientations (mirror-inverted versions of the same drawing) × 5 conditions (nonverbal baseline, or paired with either one of the linguistic primes or one of the visual primes). Each condition was represented by six experimental items in each list. All participants saw all drawings only once in an individually randomized order (except that

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

the stimuli in the nonverbal baseline were always presented at the beginning, as explained).

The procedure depicted in Figure 1[4] was as follows. A cross-hair appeared in the middle of the screen and remained there until the participant had fixated it for a minimum of 100 ms. Afterward, a row of hashes was presented in the middle of the screen for 100 ms. For items with a linguistic prime, this followed the hashes for 100 ms, long enough for the participant to be fully aware of it. For items with no linguistic prime, the screen was left blank for 100 ms. This was followed by an interval of 300 ms. Next, for items with a visual prime, the prime was presented for 100 ms (or a blank screen if there was no visual prime). An audible click signaled the coming into view of the to be described scene, which participants were asked to describe in terms of who was doing what to whom without thinking too much about it. The resulting interval between linguistic prime and scene onset provided enough time for the participant to process the priming word at a semantic level before encountering the stimulus (cf. N400; Kutas & Hillyard, 1980), thus ensuring the prime could elicit the intended linguistic expectation of what the scene would be about. Nevertheless, participants were told that, although some of the scenes would be preceded by a word, which would always be related to it, they did not need to use it in their descriptions. It was stressed that they should say whatever came to mind first. The presentation of the visual prime right before the stimulus matches previous studies (see, e.g., Gleitman et al., 2007; Kuchinsky & Bock, 2010; Myachykov et al., 2012). This ensures that the elicited saccadic movement, which takes between 100 and 200 ms (Fischer & Ramsperger, 1984), will result in a fixation on the cued position within the stimulus itself. The experiment was self-paced.

Data coding. The experiment produced three types of data: verbal descriptions, VOTs, and eye movements. Verbal descriptions were recorded with a 96-kHz digital recorder. The ELAN program was used for transcribing. The English descriptions were transcribed by a research assistant and coded partly by a research assistant and partly by the author. The Spanish descriptions were entirely transcribed and coded by the author. The coding was later rechecked in its entirety by the author to ensure consistency and correct mistakes. In those sentences coded by the research assistant and rechecked by the author (74% of the English data), there were 11% coding discrepancies, which were resolved by the author in accordance with the following coding guidelines. Sentences were coded with regard to the construal of the event reflected by the verb used in the description and to their syntactic structure. Regarding construal, the description was coded as "dominant construal" if the verb used was the preferred perspective verb or a synonym (e.g., chase, pursue, follow, or run after); as "subordinate construal" if the verb named the dispreferred perspective (e.g., flee, run away from, or escape); or as "something else" if it did not name either perspective. Regarding syntactic structure, the sentence could be a regular "active" (e.g., "a thief is trying to escape the long arm of the law"); could be a regular "passive" (e.g., "the white-collar criminal is being chased by a policeman"); contain an object placed in a more prominent position by dislocation or altered word order ("prominent object"; e.g., "a robber it is that the cop is chasing"); or be coded as "something else" if it did not describe the scene

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

as a perspective predicate (e.g., "a crime has been committed"), did not contain a verb (e.g., "a chase"), or the participant misinterpreted the scene.

It should be noted that the alternative syntactic structures (i.e., passives and prominent objects) were only consistently possible in both languages for the dominant event construals (see Appendix A). Because some arguments are indirect, rather than direct objects for many of the dispreferred perspective verbs, descriptions of the subordinate event construal did not all allow passivization or, in Spanish, object dislocation. Therefore, analyses of syntactic choice were restricted to dominant construal descriptions (other analyses did include all responses).

VOTs were manually measured in ELAN either by a research assistant or by the author. VOTs measured the interval between when the scene to be described appeared on the screen (marked by the click on the sound spectrograph) and the first word uttered. Descriptions started after some kind of nonverbal physiological sound (e.g., sniffle, cough, laugh) were excluded from VOT analyses, and so were all cases where there were disfluencies on the initial noun phrase, or where the initial word was not part of the final description (i.e., self-corrections). Together with missing data points (e.g., undetectable clicks or voice onsets due to noise in the spectrograph), this resulted in the exclusion of approximately 22% of the utterances in the English data set (evenly distributed across conditions) and about 15% of the utterances in the Spanish data set (also quite evenly distributed across conditions).

Participants' eye movements were tracked with a portable Tobii X-30 taking samples at 30 Hz. The drawings were divided into two areas of interest, each containing one of the characters in the event: the preferred agent or the dispreferred agent. The absence of background ensured that gazes would fall on one or the other character, as there was little else to look at. The target for the eye movement data and analyses was the preferred agent, that is, the character who would be the agent in the dominant construal.

Analyses. In the analyses of VOT and verbal descriptions, untransformed response data were modeled with linear mixed-effects logistic regression (Jaeger, 2008) using lme4 in R version 3.1.3 (Bates, Maechler, Bolker, & Walker, 2014; R Core Team, 2015) and lmerTest to calculate the resulting coefficients' degrees of freedom (Satterthwaite approximation) and associated p values (Kuznetsova, Brockhoff, & Christensen, 2014). For the analyses on verbal descriptions, the dependent variable was binomial (dominant construal descriptions out of all descriptions, i.e., dominant and subordinate construal descriptions), so a GLMER with a binomial link was used. Full models initially included (unless otherwise specified) the three fixed factors of linguistic priming (preferred or dispreferred perspective, or prime absent), visual priming (preferred or dispreferred agent, or prime absent), and language (English or Spanish). These fixed factors were contrast coded and centered as follows: linguistic priming with preferred perspective = 0.5, with dispreferred perspective = –0.5, absent = 0; visual priming with preferred agent = 0.5, with dispreferred agent = –0.5, absent = 0; Spanish = 0.5; and English = –0.5.

A maximal random effect structure including random intercepts and slopes over subjects and items was always attempted first (Barr, Levy, Scheepers, & Tily,

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

2013). However, nonconvergence issues usually forced a gradual simplification of the random effects, which was implemented by first eliminating the higher order factors and then the random slopes of language and priming as required (Barr et al., 2013). The models were subjected to systematic reduction of the fixed effects and subsequent chi-square comparison of models' fits. This allows us to find the simplest model with the highest explanatory value as fixed factors that do not affect the results can be excluded from the model without significantly affecting the model's fit.

For the eye movement data, growth curve analyses (Mirman, Dixon, & Magnuson, 2008) were carried out on the empirical logit (elogit) transformed dependent variable of looks to target (the preferred agent) for data collapsed by either subjects or items and aggregated in 100-ms time bins starting from the time the whole scene came into view. For these analyses, only trials associated with valid descriptions in the active voice were included; that is, passives and unintended descriptions were excluded.

As there were no a priori predictions regarding the specific time course of eye movements, the intervals of interest were determined from visual inspection of the participants' scan patterns and based on clear curve inflection points. Two intervals of interest emerged: from onset of the complete scene to be described (stimulus onset) to 400 ms, and from 400 to 1200 ms. The overall time course of fixations was modeled with a first- (linear) and a second- (quadratic) orthogonal polynomial and always a maximal random effect structure consisting of random effects of time terms over subjects and items, respectively, as well as over subject-by-linguistic priming and subject-by-visual priming (in the subjects' analyses), or item-by-linguistic priming, item-by-visual priming, and item-by-language priming (in the items' analyses). In these analyses, a significant intercept for a given main effect or interaction indicates an overall difference of total looks to the target due to the main effect or interaction; an effect of the linear term gives an indication of whether fixations rose or fell; and an effect of the quadratic term indicates a difference across the time window in the dynamics of eye movements directed at the target. As with the analyses of verbal descriptions, a procedure of systematic fixed effect reduction was followed to arrive at the simplest model with the best fit. In order to make the Results section easier to follow, only the simplest models with the best fit are reported.

Results and discussion

Verbal descriptions. Figure 2 shows the number of verbal descriptions reflecting the dominant or subordinate construal for each priming condition, with the dominant construal responses broken down into actives (e.g., "the cop is running after the thief" or the "the thief is running from the cop") and passives (e.g., "a robber is being chased by a cop"). No prominent object sentences were produced in English and only two in Spanish: one in the dispreferred linguistic priming condition and one in the dispreferred visual priming condition. All the subordinate construal descriptions were regular actives as expected (see Data coding section).

The simplest model accounting for choice of dominant construal (e.g., a chase description such as "the cop is pursuing a robber") with the best fit included

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

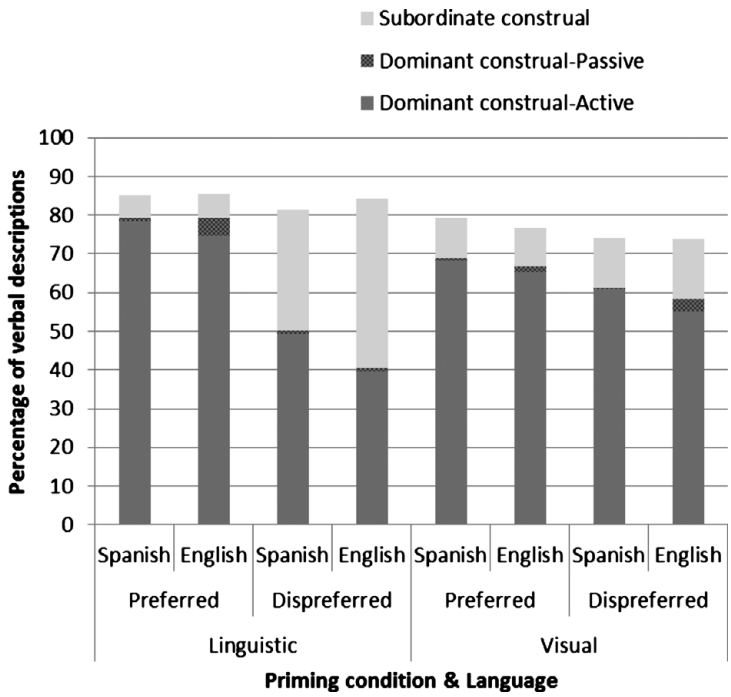


Figure 2. Percentage of verbal descriptions of different types in Experiment 1, according to language and priming condition. Maximum number of responses per condition = 240.

the fixed effects of linguistic and visual priming. Linguistically priming the dispreferred perspective (e.g., "flee" was the prime) resulted in significantly fewer dominant construal descriptions (estimate = 2.26, $SE = 0.32$, $p < .001$) and visually priming the dispreferred agent (e.g., presenting the robber first) also decreased dominant construal descriptions (estimate = 0.46, $SE = 0.16$, $p = .003$). It seems that both priming manipulations were effective in driving construal of the event for the purpose of description, more so when the prime was linguistic (a verb) than when it was visual (a salient character).

It is important to stress that speakers could not just be blindly using the verb provided in the prime, at least not all of the time, because that would have resulted in as many subordinate construal descriptions in the dispreferred linguistic priming condition (i.e., flee descriptions after "flee" had been primed) as dominant construal descriptions in the preferred linguistic priming condition (i.e., chase descriptions after "chase" had been primed). Instead, subordinate construal descriptions did not exceed 52% of valid descriptions in the dispreferred linguistic priming conditions ($N = 226$, see Figure 2), while dominant construal descriptions could amount to 93% of valid descriptions in the preferred linguistic priming conditions ($N = 220$). Furthermore, a count of cases where the description contained the exact verb used in the prime revealed that this was the case in 25% of the

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

responses in the dispreferred linguistic priming condition ($N = 480$) and 60% of the responses in the preferred linguistic priming condition ($N = 480$). This suggests that, as predicted, the linguistic primes are shaping the linguistic intention of participants rather than just activating the associated lexical items.

In line with previous results for similar types of scenes (Gleitman et al., 2007, Experiment 1), very few passives were produced (4% by English speakers and less than 1% by Spanish speakers) and virtually no prominent object constructions (only two, as noted above). This is not surprising as, in this experiment, there was little reason to expect an effect on syntactic structure: there were no differences in animacy between patients and agents (cf. Gleitman et al., 2007, Experiment 2; Prat-Sala & Branigan, 2000; Tomlin, 1997), there were no conflicting cues, and a salient dispreferred agent could be easily accommodated in an active description adopting the subordinate construal of the event (e.g., "the robber is fleeing from the police"). Nevertheless, given the base preferences for the dominant construal shown in the norming study, it is possible that visually priming a dispreferred agent (e.g., the robber) resulted in more descriptions where a base event construal preference is combined with a dispreferred agent by the use of a passive (e.g., "the robber is being chased by the police").

To look at this possibility, a mixed logistic regression model was fitted on passive production for the conditions with a visual prime only. The full model in this case included language and visual priming as fixed effects, but the simplest model with the best fit contained only the fixed effect of visual priming. As hypothesized, visually priming the dispreferred agent induced significantly more passive descriptions (estimate = –13.18, $SE = 0.001$, $p < .001$).

VOTs. VOTs for the different conditions can be seen in Figure 3. A mixed-effects linear regression was done on the VOT data with language, linguistic priming, and visual priming as fixed effects. The fit of the model containing only the intercept was not significantly different from that of other models with main effects. However, in the comparison with the model containing only the intercept, the model containing the fixed effect of linguistic priming did approach significance ($\chi^2 = 3.57$, $p = .059$) due to shorter VOTs in conditions with a preferred over those with a dispreferred linguistic prime (estimate = –9.95, $SE = 5.22$, $p = .058$). It seems that speakers of the two languages take equally long in preparing their descriptions and that the visual manipulation had no effect on VOTs, while the linguistic one may affect them.

Eye movements. To see whether the visual stimuli on their own induced looking preferences and also if English and Spanish speakers differed in how they scanned the scene in the absence of any linguistic task, a growth curve analysis was carried out on looks to the preferred agent comparing the nonverbal baseline condition (see Figure 4) with a hypothetical "expected" line representing random looks to either character. In both intervals of interest, the simplest models with the best fit were the models containing only the time terms indicating that, when not preparing for a verbal description, participants did not have a preference for any of the characters, and that speakers of both languages behaved similarly during these two time intervals. Although this result appears to contrast with the clear

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

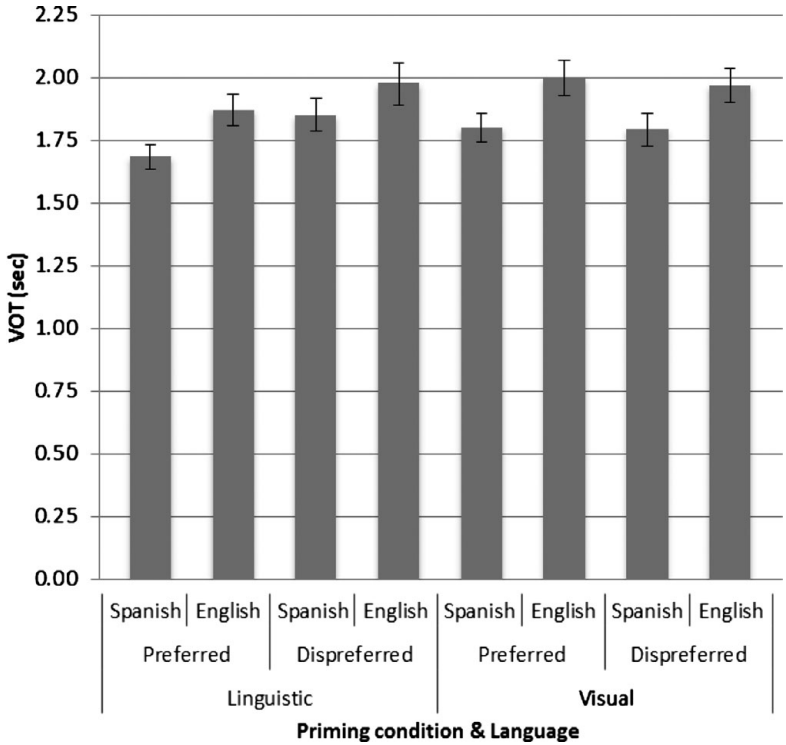


Figure 3. Voice onset times (s) and standard errors for each language and priming condition in Experiment 1.

preference to look at the patient in Griffin and Bock's (2000) study, it should be borne in mind that, unlike in the Griffin and Bock's study, the events depicted here did not have a unique, unambiguous agent or patient.

Given that there were no differences between the languages regarding verbal behavior, or baseline looking preferences, the eye movement data for the two speaker populations were merged for the following analyses.[5] Figure 4 shows proportions of looks to target (when a valid active response was produced) in the different conditions as a function of time. For convenience, the graphs plot proportion of looks to target, but the analyses are done on elogit transformed data, as explained in the Analyses section above.

With respect to the conditions where a verbal response was required, in the first interval of interest, up to 400 ms after stimulus onset, the simplest model with the best fit included the fixed effect of visual priming only. Visually priming the preferred agent, that is, the target, significantly increased total looks to this character (by subjects' intercept: estimate $= 0.80$, $SE = 0.05$, $p < .001$; by items' intercept: estimate $= 0.84$, $SE = 0.06$, $p < .001$), with looks to target increasing when it was primed and decreasing when the dispreferred agent was primed (by

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

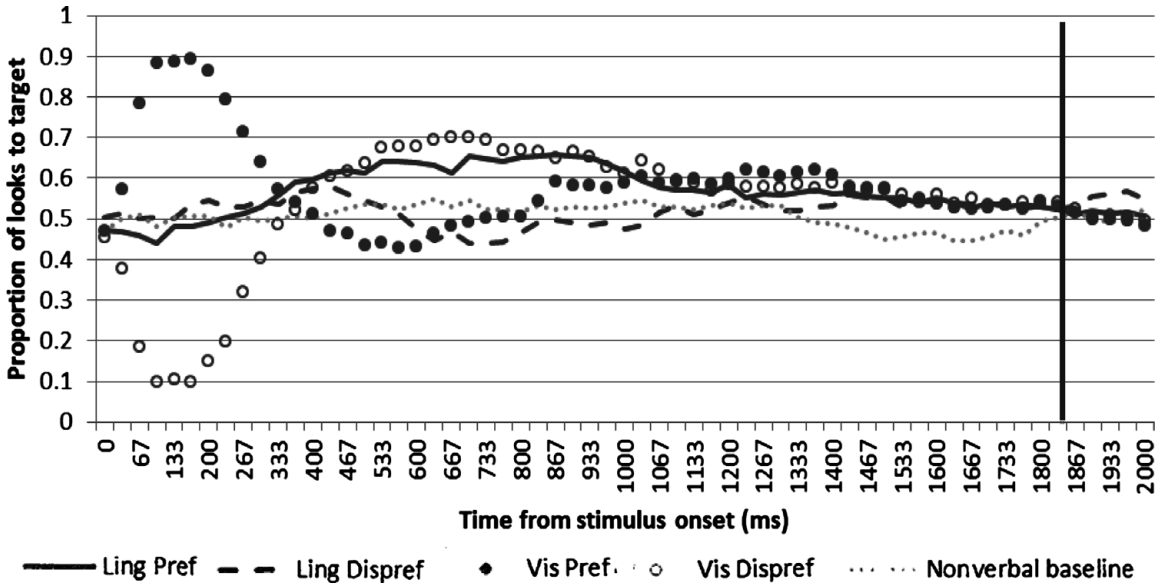


Figure 4. Proportion of looks to target (the preferred agent) in Experiment 1 as a function of time (ms) and priming condition. Only trials where the verbal description was an active valid response are included. The vertical line marks the average voice onset time.

Table 1. *Parameter estimates and significance values from the growth curve analysis of looks to target in Experiment 1*

	Subjects			Items		
	Est.	*SE*	*p*	Est.	*SE*	*p*
Linguistic priming						
Intercept	0.22	0.04	<.001	0.24	0.05	<.001
Linear term	0.04	0.07	.556	0.04	0.08	.578
Quadratic	−0.25	0.05	<.001	−0.28	0.07	<.001
Visual priming						
Intercept	−0.11	0.04	.014	−0.14	0.05	.004
Linear term	0.36	0.08	<.001	0.41	0.09	<.001
Quadratic	0.10	0.06	.114	0.09	0.06	.138

Note: The analyses were done on the agent of the preferred perspective according to linguistic priming and visual priming independently as a function of time in the second interval of interest (400–1200 ms after stimulus onset).

subjects' linear term: estimate = –0.26, $SE = 0.05$, $p < .001$; by items' linear term: estimate = –0.24, $SE = 0.07$, $p < .001$; by subjects' quadratic term: estimate = –0.99, $SE = 0.04$, $p < .001$; by items' quadratic term: estimate = –1.00, $SE = 0.06$, $p < .001$).

In the second interval of interest, from 400 to 1200 ms, the full model containing the fixed effects of time, linguistic priming, and visual priming was significantly better than the reduced models (all ps $< .001$). The coefficients of the full model can be seen in Table 1. Linguistically priming the preferred perspective resulted in more overall looks to the target, and a more pronounced curve than priming the dispreferred perspective. Visually priming the preferred agent, however, resulted in fewer looks overall to the target and a different pattern of eye movements than when the dispreferred agent had been primed.

All in all, the linguistic manipulation had a robust effect in that (a) eye movements were directed to the agent of the primed perspective about 400 ms after stimulus onset and about 1500 ms before onset of the verbal response, and (b) verbal descriptions then tended to reflect the construal that matched the primed perspective. In addition, sentence preparation times may also have been affected by the linguistic priming. This effect could be because, although priming the linguistically dispreferred perspective increased the number of subordinate construal descriptions, it still resulted in a substantial amount of dominant construal descriptions, which would have resulted in a delay because the speaker would have had to revise his or her expectation/intention. It thus seems that the linguistic manipulation was effective in inducing an intention on the part of the speaker to construe the event in terms of the primed perspective, which made the speaker more likely to look for the right agent in the scene and produce a description from this character's point of view. The visual manipulation was very successful in driving looks to the primed character right after picture onset, and also had an effect on verbal descriptions in terms of event construal (replicating previous

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

findings; Gleitman et al., 2007) and syntactic structure. Nevertheless, although the visual primes had a more marked effect on looks to target, their effect on event construal was smaller than that of the linguistic manipulation. In addition, the early looks to the visually primed character were likely to be followed by "reactionary" looks to the second character in the scene (the character not primed) as a result of the inhibition of return effect (Klein, 2000). This is a tendency to shift attention toward novel visual areas after gazes had been directed to a particular point in the visual field.

EXPERIMENT 2

Experiment 2 addressed the research question directly: can visual salience interact with a linguistic intention and influence specific aspects of sentence preparation such as lexical accessibility, order of mention, or syntactic function assignment? This experiment included crossed linguistic and visual priming manipulations, as well as a control condition with no priming of any kind. The prediction is that the linguistic priming, like in Experiment 1, will influence event construal (e.g., whether the description is made from a chase or flee perspective), and that, if visual priming influences lexical accessibility, it will affect VOTs and/or syntactic choice (i.e., use of passive voice or prominent object constructions).

Method

Participants. Forty-three English-speaking students (M age = 21.0, SD = 4.25; 32.5% male) from University of California, San Diego, and 49 Spanish-speaking students (M age = 22.5, SD = 5.10; 30.0% male) from either the University of Oviedo or the Autonomous University of Madrid (Spain) participated in this experiment. American participants were paid US $5, and Spanish speakers were paid €5.

Three English-speaking and six Spanish-speaking participants were associated with eye-tracking data loss higher than the preset 20% threshold and were thus excluded from further analyses. Three more Spanish participants were excluded from the analyses: one because of missing voice recording, one because of data overwriting, and one because she had been raised bilingually with sign language. All the participants included in the analyses below were either native English or native Spanish speakers and considered themselves monolingual.

Materials. The same experimental, filler, and practice colored drawings used in Experiment 1 were used in this experiment. For the experimental items, each drawing was paired with the same linguistic and visual primes as in Experiment 1, calling attention to either the dominant or the subordinate construal of the experimental drawings (see Appendix A for a full list). For the filler items, in this case, one-fifth was presented with no primes as in the baseline condition, and the other four-fifths were paired with a verb that described the depicted event and preceded by a preview of the character.

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

Design and procedure. There were four conditions resulting from crossing the preferred linguistic prime with the preferred and dispreferred visual primes, and the dispreferred linguistic prime with the preferred and the dispreferred visual prime. There was also one control condition with no primes or any kind, which, unlike in Experiment 1, did require a verbal description.

Stimuli were counterbalanced across 10 lists: 2 drawing orientations × 2 linguistic priming conditions × 2 visual priming conditions, plus 2 drawing orientations × 1 unprimed control condition. As in Experiment 1, each condition was represented by six items in each list, and all participants saw all drawings only once in an individually randomized order.

The procedure was the same as for Experiment 1 (see Figure 1). The order of the two primes was not counterbalanced. Instead, the chosen ordering was intended to ensure that the linguistic prime had been semantically processed by the time the scene came into view. The reason for this is that the predictions are based on a potential visual effect on lexical accessibility and/or syntactic assignment combining with a prior expectation/intention generated by the linguistic prime. The interval needed for full processing of the linguistic cue was estimated based on the existence of a neurophysiological semantic effect at 400 ms, the N400 (Kutas & Hillyard, 1980). By presenting the linguistic prime 400 ms before the visual prime, we ensure that the first fixation on the cued character is done when the speaker already has formed an expectation of what the scene will be about.

Data coding. Data coding details can be found in the Data Coding section of Experiment 1. Of the 46% cross-checked English descriptions in Experiment 2, only 5% resulted in coding disagreements, which were resolved by the author according to the preestablished coding guidelines.

Inclusion criteria for VOT analyses resulted in the exclusion of about 20% of the English utterances (23% of the items in the unprimed control condition, between 18% and 20% in the primed conditions), and 18% of the Spanish utterances (22% in the unprimed control condition, between 16% and 18% in primed conditions).

Analyses. The same procedures detailed in the Analysis section of Experiment 1 were followed to analyze the data from Experiment 2. The only differences concern the nature of the fixed effects and the time intervals of interest for the eye-movement analyses. The three fixed factors analyzed here were linguistic priming condition (preferred or dispreferred perspective, or baseline), visual priming condition (preferred or dispreferred agent, or baseline), and language (English or Spanish). These fixed factors had been contrast coded and centered as follows: priming of preferred perspective/agent = 0.5, priming of dispreferred perspective/agent = –0.5, and no priming = 0; Spanish = 0.5; and English = –0.5. As already explained, there was no a priori reason to look for effects in any particular time frame, and therefore, the windows were chosen on the basis of clear points of inflection in the data. The first interval of interest was the same here as in Experiment 1, from picture onset to 400 ms, but the second one was longer, from 400 to 1500 ms. Again, in order to make the Results section easier to follow, only the simplest models with the best fit are reported.

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

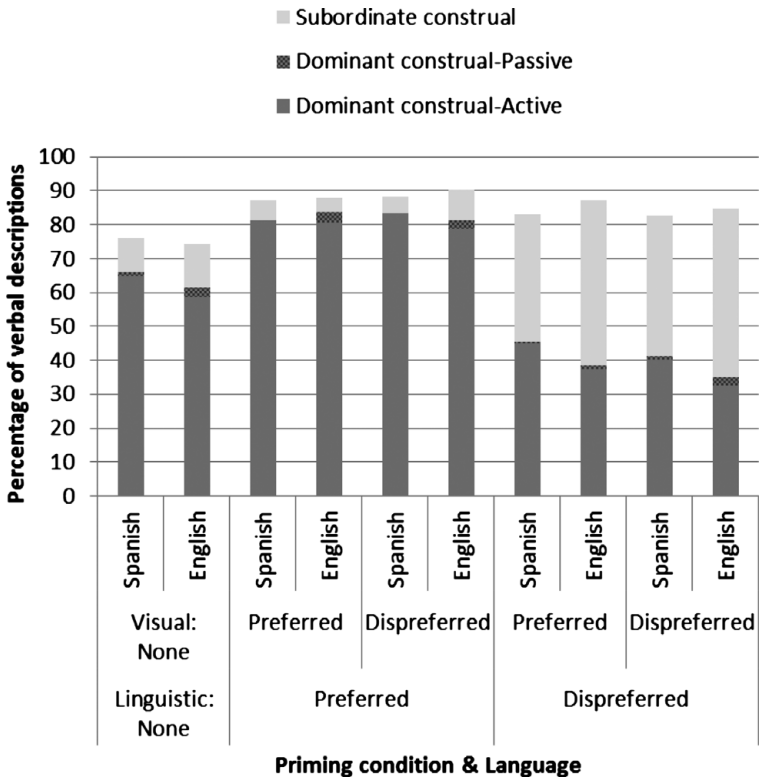


Figure 5. Percentage of verbal descriptions of different types in Experiment 2, according to language and priming condition. Maximum number of responses per condition = 240.

Results and discussion

Verbal descriptions. Figure 5 shows the number of verbal descriptions reflecting the dominant or subordinate construal for each priming condition, with the dominant construal responses broken down into actives and passives (no prominent object sentences were produced in either English or Spanish). All the subordinate construal descriptions were regular actives as expected (see Data coding section in Experiment 1).

The simplest model on dominant construal descriptions with the best fit included only the fixed effect of linguistic priming. Dominant construal descriptions were significantly increased when the corresponding preferred perspective had been linguistically primed (estimate = 2.57, $SE = 0.23$, $p < .001$). In Experiment 2, linguistic priming was, thus, also influencing construal choice. However, while visual priming did have an effect on construal choice in the absence of linguistic primes in Experiment 1, this effect appears not to be powerful enough to overcome the effect induced by an accompanying linguistic prime.

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

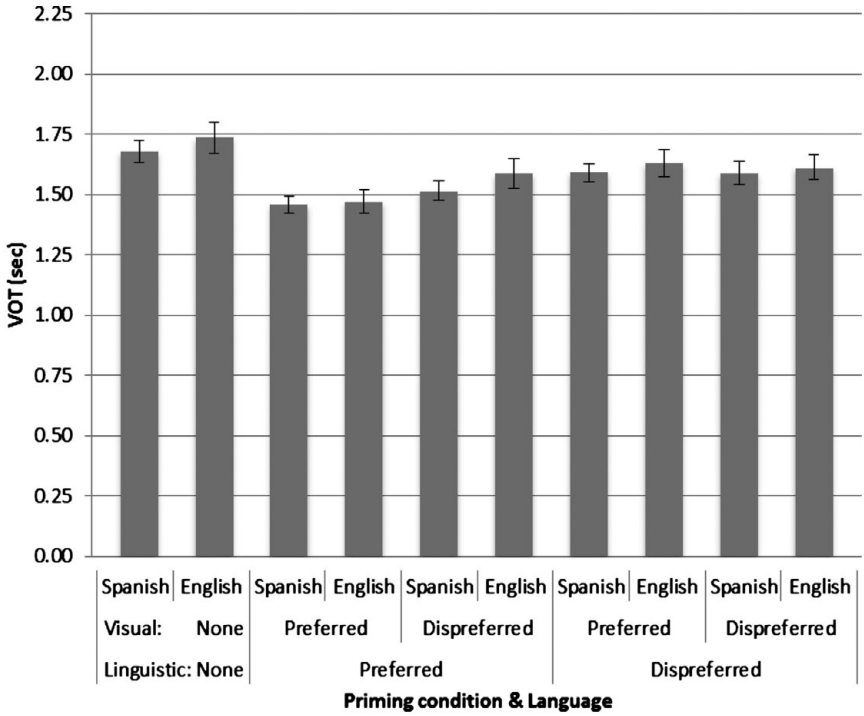


Figure 6. Voice onset times (s) and standard errors for each language and priming condition in Experiment 2.

As in Experiment 1, it was not the case that participants were merely repeating the verb token presented as the linguistic prime as the exact primes were used on average in 42% of the descriptions in dispreferred linguistic conditions and in 73% of the descriptions in preferred linguistic conditions.

To see if visual priming induced passive use, a mixed logistic regression model was fitted on passive production ($N = 34$) within dominant construal responses ($N = 1{,}480$). The best fit, simplest model included only the intercept, suggesting that, in this experiment where visual priming accompanied linguistic priming, making the dispreferred agent more salient does not increase passive use.

VOTs. VOTs for the different conditions can be seen in Figure 6. The simplest model with the best fit contained only the fixed effect of linguistic priming. Apparently, neither the speakers' native language nor the visual priming affected VOTs, and neither did visual priming interact with linguistic priming. However, linguistic priming resulted in shorter VOTs than no priming (estimate = –19.87, $SE = 5.81$, $p = .002$).

Eye movements. As with Experiment 1, given that there were no differences between the languages regarding verbal behavior, the eye movement data for the

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

two speaker populations were merged for the following analyses.[6] Figure 7 shows proportions of looks to the target (when a valid active response was produced) in the different conditions as a function of time. For convenience, the graphs plot proportion of looks to the target, but the analyses are done on elogit transformed data, as explained above.

For the first interval of interest, the simplest model that best explained looks to the target contained only the fixed effect of visual priming. Visually priming one of the characters had a clear effect on how and how much participants looked at the primed character: a preview of the preferred agent resulted in more looks to this character (by subjects' intercept: estimate = 1.01, $SE = 0.04$, $p < .001$; by items' intercept: estimate = 1.01, $SE = 0.06$, $p < .001$), with looks to target increasing or decreasing across time depending on the prime (by subjects' linear term: estimate = –0.15, $SE = 0.05$, $p = .001$; by items' linear term: estimate = –0.11, $SE = 0.05$, $p = .038$; by subjects' quadratic term: estimate = –1.11, $SE = 0.05$, $p < .001$; and by items' quadratic term: estimate = –1.01, $SE = 0.05$, $p < .001$).

The second interval of interest was somewhat longer in Experiment 2 than Experiment 1: from 400 to 1500 ms, or 1100 ms as opposed to 800 ms. This lengthening of the interval where the effects of the primes seem to be taking place may have been due to having two cues, linguistic and visual, imposing demands on eye fixations. However, tellingly, this did not translate into longer times to process the scene and start the description, as evidenced by the shorter VOTs in Experiment 2 in comparison with Experiment 1 (see above), and, within Experiment 2 in particular, in the conditions with primes in contrast with the verbal baseline condition. The best fit model for this interval was the full model including the fixed effects of linguistic and visual priming. The coefficients associated with it can be seen in Table 2.

Priming the preferred perspective linguistically resulted in more looks to the target than priming the dispreferred perspective, with the former increasing and the latter decreasing fixations to the target as a function of time up to about 700 ms, after which point the trends are inverted. Priming the preferred agent visually induced more looks to the nonprimed character in this time window. It is critical that there was no significant interaction between visual and linguistic priming either on the intercept or in any of the time terms.

To summarize, looks to the target in the initial 400 ms after stimulus onset are positively influenced by the visual manipulation. In the second interval of interest, from 400 to 1500 ms, both types of priming had an effect on looks to the target but in opposite directions as visual priming of one of the characters significantly reduced overall looks to that character during the second time interval as a result of the inhibition of return effect (Klein, 2000). There is no interaction between the two priming manipulations, suggesting the two types of priming act independently on eye movements, with the final result being a combination of the individual influences. Furthermore, and more crucially, despite the clear effects of both types of priming on eye movements, the eventual descriptions were only affected by linguistic priming in terms of construal of the event and sentence preparation times. To this, it should be added that, unlike in Experiment 1, visual priming did not influence syntactic choice in Experiment 2, where the speaker had a prior linguistic expectation/intention.

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

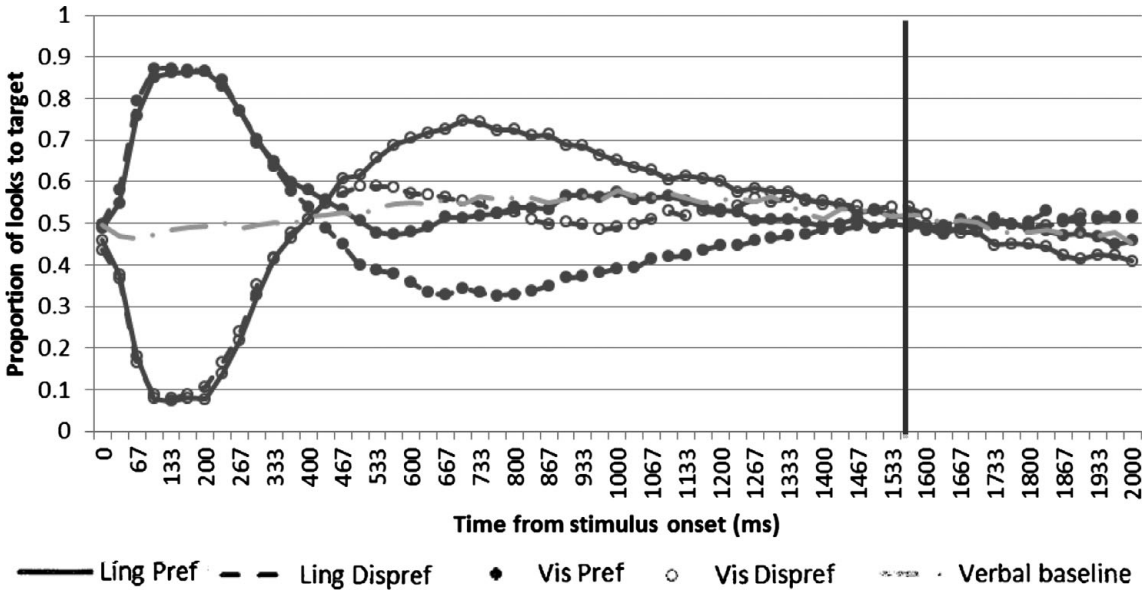


Figure 7. Proportion of looks to target (the preferred agent) in Experiment 2 as a function of time (ms) and priming condition. Only trials where the verbal description was an active valid response are included. The vertical line marks the average voice onset time.

Table 2. *Parameter estimates and significance values from the growth curve analysis of looks to targets in Experiment 2*

	Subjects			Items		
	Est.	*SE*	*p*	Est.	*SE*	*p*
Linguistic priming						
Intercept	0.27	0.03	<.001	0.27	0.04	<.001
Linear term	−0.20	0.07	.003	−0.18	0.08	.023
Quadratic	−0.41	0.06	<.001	−0.37	0.08	<.001
Visual priming						
Intercept	−0.22	0.03	<.001	−0.21	0.04	<.001
Linear term	0.17	0.07	.017	0.14	0.08	.096
Quadratic	0.14	0.05	.013	0.17	0.07	.015
Linguistic Priming × Visual Priming						
Intercept	−0.01	0.02	.520	0.02	0.02	.480
Linear term	0.04	0.07	.559	0.12	0.08	.137
Quadratic	0.01	0.07	.873	−0.05	0.08	.543

Note: The analyses were done on the agent of the preferred perspective according to linguistic and visual priming conditions as a function of time in the second interval of interest (400–1500 ms after stimulus onset).

GENERAL DISCUSSION

Experiment 1 showed that both types of manipulations, linguistic and visual, were independently having an effect on eye movements and on description choice, but their effects were not equivalent. Visually priming one of the characters exerted a strong effect on very early looks to the primed character, and it also increased the likelihood of the verbal description taking on the perspective of the primed character, a result that replicates previous findings (Gleitman et al., 2007). Linguistically priming one of the two event perspectives had a smaller but clear effect on later eye movements, as well as a more pronounced effect on description choice, indicating that the linguistic manipulation was successful in generating a linguistic intention. These results are not very surprising: a perceptual measure is most affected by a visual manipulation, and a measure of verbal production is primarily affected by a linguistic manipulation. Nevertheless, given previous findings of a good relationship between what people look at and what comes out of their mouths (Coco & Keller, 2012; Griffin, 2001; Griffin & Bock, 2000), it is perhaps noteworthy that there is so little correlation between the relative size of the effects in the two domains: looking and saying.

Against these individual visual and linguistic priming effects, in Experiment 2, when the visual manipulation was superposed to the linguistic one, the visual influence on verbal descriptions disappeared. In particular, the lack of an effect of visual priming on either VOTs or syntactic choice when there is a linguistic prime present is incompatible with the predictions based on the visual cue affecting lexical accessibility or argument assignment. It appears as if the influence of visual priming could not overcome the effect on event conceptualization of the

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

linguistic cue. In contrast, for eye movements, the independent linguistic and visual influences seen in Experiment 1 were replicated in Experiment 2 without signs of an interaction: when the two influences were present simultaneously, the final looks appeared to be the result of accommodating both of them in an additive manner.

The effects of perceptual salience on sentence production

Perceptual salience has been generally assumed to induce a lexical starting point by facilitating lexical access and, as a consequence, order of mention and even syntactic function assignment (Gleitman et al., 2007; Kuchinsky, 2009; Myachykov et al., 2012; Myachykov, Garrod, et al., 2011). If that was the case, the results of Experiment 2 should have shown signs of it as having an overall expectation of what the event to be described will be about and a preexisting intention to proceed from that perspective should not preclude taking advantage of more accessible lexical items if they are compatible with the linguistic intention (Smith & Wheeldon, 1999). In particular, a more accessible lexical entry for the dispreferred agent should have resulted in an increase in subordinate event construals (e.g., "the robber is running from the cop") relative to other conditions, or in more dominant construals where this lexical item appeared earlier in the sentence (e.g., "the robber is being chased by the cop"), or in shorter VOTs when the active lexical item was compatible with the utterance being planned (i.e., when it was congruent with the linguistic intention established by the linguistic prime). None of these effects were found. Visual salience had an influence on description choices when on its own, but no effect when combined with a linguistic prime that predisposed speakers to construe the scene in a particular way.

However, if not via lexical activation, how does visual salience influence descriptions when it does, as in Experiment 1 here and in the cited research? Most likely visual salience affects descriptions by influencing how a scene is conceptualized. In particular, visual salience should make an entity more prominent and therefore more likely to be mentioned (Coco et al., 2014) and made the protagonist (Vogels et al., 2013), so that the description should then focus on what this entity is doing or experiencing. If visual salience is primarily affecting how a scene is construed, we would expect this to also influence downstream linguistic decisions and surface as differences in order of mention and syntactic choices because what we want to say determines how it should be said. That is, if we want to talk about the boxer and not the cowboy, we are likely to make it the topic of the sentence even if that implies resorting to a passive sentence when the boxer is the patient in the event (Myachykov et al., 2012). However, that effect should disappear when the conceptualization of the scene, how the scene is "read," does not hinge on what is most visually salient, as would be the case if there is a linguistic cue priming the speaker toward a particular event construal, and this cue is more powerful (Myachykov et al., 2012, their experiment 2, match conditions; Experiment 2 here).

The interpretation of perceptual cues as serving to promote a particular conceptualization of the scene rather than to provide a lexical starting point is supported by the findings of Coco and Keller (2009) and Vogels et al. (2013). Coco and Keller report effects of visual manipulations (clutter and number of animate characters in

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

the scene) on time to start a description even when a lexical starting point (a cue naming an animate character in the picture) had been provided that could readily be used to initiate a description; speakers were slower to initiate their descriptions when the scene was more cluttered and when there were two as opposed to only one animate character, that is, when the number of descriptive possibilities was increased. In conclusion, speakers do not start speaking before they know what they want to say, even if a lexical starting point has been served to them on a platter. Vogels et al. (2013) found that both visual salience (foregrounding) and linguistic salience influenced choice of sentential subject without interacting. They think that foregrounding served to make the foregrounded character the figure in the discourse, and they interpret their results as "consistent with the view that speakers primarily plan their utterances based on the global structure of an event" rather than "by taking into account the salience of individual elements" (p. 1345). Montag and MacDonald (2013) also find that their effects of visual salience on relative clause choice are better explained by how salience influences inspection times, which then impacts on speakers' communicative goals by allowing the speaker to notice fewer or more features of the scene.

What comes first: Seeing or planning to say?

The fact that the eye movements associated with the visual manipulation remained so stable across experiments despite the difference in consequences regarding verbal descriptions suggests there is a decoupling of perceptually guided eye movements and linguistic choice. The very dramatic effect on looks to the salient character found in the first 400 ms after stimulus onset was not followed by a commensurate increase in descriptions starting with that character in either experiment.

The expectation generated by the linguistic prime also had a clear and stable influence on eye movements: priming one of the perspectives increased looks to the character most compatible with the primed perspective from around 400 ms after picture onset (second time window). In the present experiments, furthermore, the linguistic priming effects on verbal descriptions were equally stable across experiments, with significantly more descriptions from the primed perspective and, thus, with the most fixated character in the second time window as the subject. The implication is that the linguistic expectation/intention is guiding gazes to characters rather than the other way around (Griffin & Bock, 2000). However, it is worth noting that, during the second time window in Experiment 2, the looks to the preferred character were strongly influenced by both the linguistic and the visual priming. This suggests that eye movements can be, but are not necessarily guided by, linguistic factors even when describing what is in front of our eyes (Coco et al., 2014).

The effect of native language

The results of the two experiments show that there is no difference between speakers of English and Spanish regarding description choice, production of sentence structures other than regular actives, or sentence preparation times. With respect to

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

choice of syntactic structure, Spanish greater word-order flexibility was expected to allow speakers to accommodate competing influences more easily than English if an accessible lexical entry was more likely to influence order of mention than syntactic function assignment, but this was not found. This replicates a similar minimal or nonexistent effect of perceptual salience effects on syntactic choice and word order in two scrambling languages, Russian and Finnish (Myachykov & Garrod, 2008; Myachykov, Thompson, Scheepers, & Garrod, 2011), which was attributed to the inability of the perceptual information to override the constraining power of these languages' case marking systems: initiating the description with a patient requires the speaker's commitment to a specific argument assignment, something that is not necessary in English. Myachykov and colleague's findings of visual salience effects being modulated by argument assignment requirements also supports the conclusion that visual salience is not simply promoting a character to be mentioned first: the whole event has to be understood before the speaker commits to a specific syntactic structure.

Limitations and future directions

The question of interest here was whether the impact of a visual cue on linguistic output is due to its effect on specific linguistic processes such as lexical access and argument assignment. In these experiments, this was studied by looking at the effect of a visual cue when a linguistic intention had been induced, since an accessible lemma should result in shorter sentence preparation times and/or specific syntactic choices. However, a visual cue could potentially have a different effect if the order of presentation of the primes was inverted. For example, encountering the robber before coming across the verb "chase" may result in significantly more passives in English because the character would already be prominent in the speaker's mind by the time the intention to talk about chasing is formed. If so, visual cueing may be able to influence argument assignment or topicalization given the right conditions. It would thus be useful to continue the investigation into the effects of visual characteristics on linguistic processing by inverting the order of the primes.

Conclusions

To conclude, the results of the present experiments strongly suggest that perceptual characteristics affect the content rather than the specific form of utterances, which is predominantly determined by the linguistic intention the speaker has in terms of what to communicate. This conclusion is in accordance with findings of perceptual salience on probability of mention of a character but not on choice of referring expression (Vogels et al., 2013), and the preeminence of linguistic and conceptual factors on syntactic choices in scene descriptions (Kuchinsky, 2009; Myachykov & Garrod, 2008; Myachykov et al., 2007; van de Velde, Meyer, & Konopka, 2014). What influences have been found of visual attention capture on linguistic choices during scene descriptions (Coco & Keller, 2009; Gleitman et al., 2007; Montag & MacDonald, 2013; Myachykov et al., 2012; Myachykov, Garrod, et al., 2011; Myachykov, Thompson, Garrod, et al., 2011; Tomlin, 1997) are unlikely to be

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

the result of a direct influence on linguistic processing. They are most likely due to the effect of those influences on event conceptualization. As Coco and Keller put it, "scene understanding has to happen before constraints supplied by visual attention are applied to sentence production" (Coco & Keller, 2015, p. 147). The present results support accounts of sentence production being based on a general sentence plan rather than being lexically incremental (i.e., Bock, Irwin, Davidson, & Levelt, 2003; Griffin & Bock, 2000; Kuchinsky, Bock, & Irwin, 2011).

APPENDIX A

Primes used in the study

	Linguistic Primes					
	English		Spanish		Visual Primes	
Item	Pref. Perspec.	Dispreff. Perspec.	Pref. Perspec.	Dispref. Perspec.	Pref. Agent	Dispref. Agent
101	feed	eat	alimentar	comer	man	cat
102	feed	eat	alimentar	comer	woman	man
103	chase	flee	perseguir	huir	policeman	robber
104	give	take	dar	tomar	nurse	child
105	feed	eat	alimentar	comer	bird	chick
106	give	beg	dar	recibir	woman	man
107	interview	answer	entrevistar	responder	journalist	woman
108	buy	sell	comprar	atender	girl	man
109	give	get	regalar	recibir	man	woman
110	feed	eat	dar	comer	mother	child
111	breastfeed	suck	amamantar	mamar	mother	baby
112	chase	flee	perseguir	huir	dog	man
113	pay	give	pagar	cobrar	woman	man
114	give	receive	dar	recibir	boy	girl
115	win	lose	ganar	perder	black boxer	white boxer
116	win	lose	ganar	perder	black runner	white runner
117	feed	suck	amamantar	mamar	dog	puppies
118	lend	borrow	prestar	pedir	boy	girl
119	indicate	ask	indicar	preguntar	boy	girl
120	give	receive	entregar	recibir	doctor	father
121	listen	play	escuchar	tocar	man	musician
122	speak	listen	hablar	escuchar	man	spectators
123	talk	listen	hablar	escuchar	woman	man
124	preach	listen	pregonar	escuchar	man	spectators
125	teach	learn	enseñar	aprender	teacher	student
126	buy	sell	comprar	vender	woman	man
127	play	listen	tocar	escuchar	cellist	man
128	send	receive	enviar	recibir	boy	girl
129	throw	catch	lanzar	coger	boy	dog
130	kick	stop	chutar	parar	player	keeper

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

ACKNOWLEDGMENTS

I thank the many people who helped me with various aspects of the research: Lila Gleitman and John Trueswell, who kindly shared their stimulus set with me; Fernando Cuetos, José Manuel Igoa, Tamar Gollan, and Vic Ferreira, who welcomed me to their labs and gave me space to work and the means to recruit participants; Noemí Arce, Beatriz Bermúdez, María González Nosti, and Mayra Murillo, who went out of their way to help me find enough participants; Cecilia Salcedo, who helped me run the experiments; Vicky Knox, who assisted me with transcription and coding of the data; and Alma Geesink-Anton, who so very graciously spent hours painstakingly measuring VOTs for her mother. I also thank Lucia Pozzan for her helpful comments on a previous draft of this manuscript.

NOTES

1. It may be that a linguistic cue primes the associated lemma as well as shaping the speaker's intention. However, as this lexical item being a verb cannot appear at the beginning of the sentence, this should not significantly affect sentence preparation times because the activated lemma would have to be buffered until it could be inserted in the sentence (Wheeldon et al., 2013). Therefore, the most noticeable contribution of the linguistic cue should be on scene construal and promoting one description perspective over another.
2. Of these, nine were based on stimuli used in Gleitman et al. (2007), which were kindly provided to the author. They were redrawn by the author in the same style as the others.
3. Of these, 10 were based on stimuli used in Gleitman et al. (2007).
4. It should be stressed that, in Experiment 1, the primes were not crossed and items could be preceded by no primes (nonverbal baseline), or either a linguistic or a visual prime, but not both. Figure 1 presents the procedure including both primes, but in Experiment 1, the reader should assume that the screens for one or the other (or both in the case of the baseline conditions) were blank.
5. A parallel set of analyses of eye movement data where the models included a fixed effect of language failed to identify any major effects of language on eye movements. The patterns for the two speaker populations were remarkably similar in shape and timing. Language did interact with the priming manipulations, showing some influence on the temporal dynamics of the eye movements, but not one that was relevant for the issues at hand in this study.
6. As in Experiment 1, the shape and timing of eye movements were very similar for the two speaker populations. A parallel set of analyses including the fixed effect of language in the models failed to identify any effects of language that had any relevance for the focus of the current paper.

REFERENCES

Antón-Méndez, I., Gerfen, C., & Ramos, M. (2016). Salience effects: L2 sentence production as a window on L1 speech planning. *Journal of Psycholinguistic Research*, *45*, 537–552. doi:10.1007/s10936-015-9361-7

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278. doi:10.1016/j.jml.2012.11.001

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *Journal of Statistical Software, 67*.

Bock, K., & Ferreira, V. S. (2014). Syntactically speaking. In M. Goldrick, V. Ferreira, & M. Miozzo (Eds.), *The Oxford handbook of language production* (pp. 21–46). Oxford: Oxford University Press.

Bock, K., Irwin, D. E., & Davidson, D. J. (2004). Putting first things first. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision and action: Eye movements and the visual world*. New York: Psychology Press.

Bock, K., Irwin, D. E., Davidson, D. J., & Levelt, W. J. M. (2003). Minding the clock. *Journal of Memory and Language*, *48*, 653–685.

Branigan, H. P., Pickering, M. J., & Tanaka, M. N. (2008). Contributions of animacy to grammatical function assignment and word order during production. *Lingua*, *118*, 172–189. doi:10.1016/j.lingua.2007.02.003

Brawley, H. (2012). *What informs event descriptions: Language, salience, and discourse in English and Japanese*. Unpubished manuscript, Ohio State University. Retrieved from https://etd.ohiolink.edu/ap/10?0::NO:10:P10_ACCESSION_NUM:osu1338275303

Bunger, A., Papafragou, A., & Trueswell, J. C. (2013). Event structure influences language production: Evidence from structural priming in motion event description. *Journal of Memory and Language*, *69*, 299–323. doi:10.1016/j.jml.2013.04.002

Coco, M. I., & Keller, F. (2009). *The impact of visual information on reference assignment in sentence production*. Paper presented at the 31st Annual Conference of the Cognitive Science Society, Amsterdam.

Coco, M. I., & Keller, F. (2012). Scan patterns predict sentence production in the cross-modal processing of visual scenes. *Cognitive Science*, *36*, 1204–1223. doi:10.1111/j.1551-6709.2012.01246.x

Coco, M. I., & Keller, F. (2015). Integrating mechanisms of visual guidance in naturalistic language production. *Cognitive Processing*, *16*, 131–150. doi:10.1007/s10339-014-0642-0

Coco, M. I., Malcolm, G. L., & Keller, F. (2014). The interplay of bottom-up and top-down mechanisms in visual guidance during object naming. *Quarterly Journal of Experimental Psychology*, *67*, 1096–1200. doi:10.1080/17470218.2013.844843

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*, 57–84. doi.org/10.1006/jmla.2001.2797

Fischer, B., & Ramsperger, E. (1984). Human express saccades: Extremely short reaction times of goal directed eye movements. *Experimental Brain Research*, *57*. doi:10.1007/BF0023114

Fukumura, K., van Gompel, R. P. G., & Pickering, M. J. (2010). The use of visual context during the production of referring expressions. *Quarterly Journal of Experimental Psychology, 63*, 1700–1715. doi:10.1080/17470210903490969

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, *57*, 544–569. doi:10.1016/j.jml.2007.01.007

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, *82*, B1–B14. doi:10.1016/S0010-0277(01)00138-X

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*, 274–279. doi:10.1111/1467-9280.00255

Haskell, T. R. (2005). *Speaker's eye movements during the production of adjectives*. Paper presented at the 18th Annual CUNY Sentence Processing Conference, Tucson, AZ.

Hickmann, M., Taranne, P., & Bonnet, P. (2009). Motion in first language acquisition: Manner and path in French and English child language. *Journal of Child Language*, *36*, 705–741. doi:10.1017/S0305000908009215

Ibbotson, P., Lieven, E. V. M., & Tomasello, M. (2013). The attention–grammar interface: Eye-gaze cues structural choice in children and adults. *Cognitive Linguistics*, *24*. doi:10.1515/cog-2013-0020

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

Jackendoff, R. (1987). On beyond zebra: The relation of linguistic and visual information. *Cognition, 26*, 89–114.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformational or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446. doi:10.1016/j.jml.2007.11.007

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*, 173–218.

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences*, *4*, 138–147. doi:10.1016/S1364-6613(00)01452-2

Kuchinsky, S. E. (2009). *From seeing to saying: Perceiving, planning, producing*. Ann Arbor, MI: ProQuest. Retrieved from http://eric.ed.gov/?id=ED526717

Kuchinsky, S. E., & Bock, K. (2010). *From seeing to saying: Perceiving, planning, producing*. Paper presented at the CUNY Sentence Processing Conference, New York.

Kuchinsky, S. E., Bock, K., & Irwin, D. E. (2011). Reversing the hands of time: Changing the mapping from seeing to saying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1–16. doi:10.1037/a0022637.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). *lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)*. Retrieved from http://CRAN.R-project.org/package=lmerTest

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Meyer, A., van der Meulen, F., & Brooks, A. (2004). Eye movements during speech planning: Talking about present and remembered objects. *Visual Cognition*, *11*, 553–576. doi:10.1080/13506280344000248

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*, 475–494. doi:10.1016/j.jml.2007.11.006

Montag, J. L., & MacDonald, M. C. (2013). Visual salience modulates structure choice in relative clause production. *Language and Speech*, *57*, 163–180. doi:10.1177/0023830913495656

Myachykov, A., & Garrod, S. (2008). Perception and word order in Russian and Finnish sentence production. In W. Ramm & C. Fabricius-Jansen (Eds.), *Linearisation and segmentation in discourse: Multidisciplinary approaches to discourse 2008* (pp. 87–97). Oslo: University of Oslo.

Myachykov, A., Garrod, S., & Scheepers, C. (2011). Perceptual priming of structural choice during English and Finnish sentence production. In R. K. Mishra & N. Srinivasan (Eds.), *Language and cognition: State of the art* (pp. 53–71). Munich: Lincom Europa.

Myachykov, A., Garrod, S., & Scheepers, C. (2012). Determinants of structural choice in visually situated sentence production. *Acta Psychologica*, *141*, 304–315. doi:10.1016/j.actpsy.2012.09.006

Myachykov, A., Posner, M. I., & Tomlin, R. S. (2007). A parallel interface for language and cognition in sentence production: Theory, method, and experimental evidence. *Linguistic Review*, *24*, 457–474. doi:10.1515/TLR.2007.020

Myachykov, A., Thompson, D., Garrod, S., & Scheepers, C. (2011). Referential and visual cues to structural choice in visually situated sentence production. *Frontiers in Psychology*, *2*, 1–9. doi:10.3389/fpsyg.2011.00396

Myachykov, A., Thompson, D., Scheepers, C., & Garrod, S. (2011). Visual attention and structural choice in sentence production across languages. *Language and Linguistics Compass, 5*, 95–107. doi:10.1111/j.1749-818X.2010.00265.x

Prat-Sala, M., & Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language*, *42*, 168–182. doi:10.1006/jmla.1999.2668

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press

Qualtrics LLC. (2005). Qualtrics [Computer software]. Retrieved from http://www.qualtrics.com/

R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Smith, M., & Wheeldon, L. R. (1999). High level processing scope in spoken sentence production. *Cognition, 73*, 205–246. doi:10.1016/S0010-0277(99)00053-0

Tanaka, M. N., Branigan, H. P., McLean, J. F., & Pickering, M. J. (2011). Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language*, *65*, 318–330. doi:10.1016/j.jml.2011.04.009

Tomlin, R. S. (1997). Mapping conceptual representations into linguistic representations: The role of attention in grammar. In J. Nuyts & E. Pederson (Eds.), *Language and conceptualization* (pp. 162–189). Cambridge: Cambridge University Press.

van der Meulen, F. F. (2001). *Moving eyes and naming objects*. Retrieved from http://repository.ubn.ru.nl/handle/2066/19013

van de Velde, M., Meyer, A. S., & Konopka, A. E. (2014). Message formulation and structural assembly: Describing "easy" and "hard" events with preferred and dispreferred syntactic structures. *Journal of Memory and Language*, *71*, 124–144. doi:10.1016/j.jml.2013.11.001

Vogels, J., Krahmer, E., & Maes, A. (2013). Who is where referred to how, and why? The influence of visual saliency on referent accessibility in spoken language production. *Language and Cognitive Processes, 28*, 1323–1349. doi:10.1080/01690965.2012.682072

Wheeldon, L. R., Ohlson, N., Ashby, A., & Gator, S. (2013). Lexical availability and grammatical encoding scope during spoken sentence production. *Quarterly Journal of Experimental Psychology, 66*, 37–41. doi:10.1080/17470218.2012.754913

https://doi.org/10.1017/S0142716416000345 Published online by Cambridge University Press