Journal of Psycholinguistic Research (2020) 49:415–434
https://doi.org/10.1007/s10936-020-09691-x

Modulating "Surprise" with Syntax: A Study on Negative Sentences and Eye-Movement Recording

Matteo Greco[1] · **Paolo Canal**[1] · **Valentina Bambini**[1] · **Andrea Moro**[1]

Published online: 8 February 2020

Abstract
This work focuses on a particular case of negative sentences, the Surprise Negation sentences (SNEGs). SNEGs belong to the class of expletive negation sentences, i.e., they are affirmative in meaning but involve a clausal negation. A clear example is offered by Italian: '*E **non** mi è scesa dal treno Maria?!*' (let. 'and *not* CLITIC.to_me is got off-the train Mary' = 'The surprise was that Maria got off the train!'). From a theoretical point of view, the interpretation of SNEGs as affirmative can be derived from their specific syntactic and semantic structure. Here we offer an implementation of the visual world paradigm to test how SNEGs are interpreted. Participants listened to affirmative, negative or expletive negative clauses while four objects (two relevant—either expected or unexpected—and two unrelated) were shown on the screen and their eye movements were recorded. Growth Curve Analysis showed that the fixation patterns to the relevant objects were very similar for affirmative and expletive negative sentences, while striking differences were observed between negative and affirmative sentences. These results showed that negation does play a different role in the mental representation of a sentence, depending on its syntactic derivation. Moreover, we also found that, compared to affirmative sentences, SNEGs require higher processing efforts due to both their syntactic complexity and pragmatic integration, with slower response time and lower accuracy.

Keywords Negation · Expletive negation · Eye movements · Surprise effect · Visual world paradigm · Syntax–Pragmatics interface

✉ Matteo Greco
matteo.greco@iusspavia.it

Paolo Canal
paolo.canal@iusspavia.it

Valentina Bambini
valentina.bambini@iusspavia.it

Andrea Moro
andrea.moro@iusspavia.it

1 Center for Neurocognition, Epistemology and Theoretical Syntax (NETS), Scuola Universitaria Superiore IUSS Pavia (University School for Advanced Studies IUSS Pavia), Piazza della Vittoria 15, 27100 Pavia, Italy

Introduction

Sentential negation is a one-place complement-set operator (Horn 1989; Speranza and Horn 2012; Delfitto and Fiorin 2014) occurring universally in human languages; however, its role is not always clear and predictable. Consider, for instance, the following Italian sentence:

(1)	*Luca*	*è*	*più*	*alto*	*di*	*quanto*	***non***	*sia*	*Gianni*
	Luke	is	more	tall	of	how-much	*not*	is	John
	'Luke is taller than John.'								

In this case the occurrence of the negative marker *non* ("not") does not reverse the truth-value conditions of the proposition, offering an example of what it is called *expletive negation* (EN). Roughly speaking, EN consists in the use of negative markers that do not contribute to the polarity of the proposition in which they occur (see Jespersen 1917; Greco 2017, 2019a, b for a more detailed discussion). In the present study we will address two questions raised by the EN phenomenon: (1) how is negation processed when it does not deny a sentence, as it is supposed to do? (2) Are EN sentences elaborated as affirmative clauses, according to their semantic value, or as negative clauses, according to their morphological shape?

No experimental studies in the psycho-linguistic literature focused on EN and, therefore, we will consider the literature on standard negation to model the empirical approach to EN. One crucial piece of evidence on the elaboration of negative sentences is that they yield a longer processing time and higher error rates in their interpretation than the non-negative ones, regardless of the specific experimental task (Wason 1961; Clark and Chase 1972; Carpenter and Just 1975; Trabasso et al. 1971; Carpenter et al. 1999; Hasegawa et al. 2002; Kaup et al. 2007). Moreover, it has also been shown that negation is incorporated into the discourse model only at a late stage, even after thousands of milliseconds from the end of the sentence in which it occurs (Hasson and Glucksberg 2006; Giora 2006; Kaup et al. 2005, 2006; Lüdtke et al. 2008; Scappini et al. 2015).

The longer processing time, the higher error rates, and the late integration have been taken as the consequences of two principles: negative sentences require additional processing compared to the non-negative ones (Carpenter and Just 1975; Carpenter et al. 1999; Hasegawa et al. 2002; Dale and Duran 2011); linguistic negation reduces the accessibility of the information taken under its scope (MacDonald and Just 1989; Kaup 2001; Tettamanti et al. 2008; Liuzza et al. 2011; Bartoli et al. 2013). However, some intermediate positions have been advanced proposing that specific linguistic and discourse contexts can reduce both its inhibitory power (Kaup 1997; Giora et al. 2007) and its higher processing costs (Wason 1965; Villiers and Flusberg 1975; Glenberg et al. 1999; Lüdtke and Kaup 2006; Ferguson et al. 2008).

Among others, Orenes et al. (2014) showed that individuals understand negation differently depending on whether the discourse context provides one or multiple alternative objects to the negated one: only when it is one, rather than when they are multiple, negation shows its inhibitory power. In their study eye movements were tracked while participants listened to negative and affirmative sentences (e.g., *The figure is/is not red*) that were presented after a verbal context that was either binary (e.g., *The figure could be red or green*) or multiple (e.g., *The figure could be red or green or yellow or blue*) while four pictures with different colors were displayed on the screen. When the context was binary, participants shifted their gaze toward the alternative color (e.g., green) with a delay of 1340 ms after the end of the negative sentence (confirming the late integration of negation), but when the context was multiple they kept looking at the negated argument (e.g., red), similarly to what happened in the affirmative sentences. The linguistic context can affect, therefore, the specific parameters usually associated to the elaboration of negation (longer processing time, higher error rates and late integration) eventually canceling or reducing the processing differences between negative and affirmative sentences. In particular, to warrant a clear processing difference between affirmative and negative sentences we used discourse information to make the context binary, like the one used in Orenes et al. (2014), in which participants evaluated the meaning of the target sentences dealing with two alternatives. Since ENs are set up by a negative marker—like the Italian *non* ("not") in (1)—even thought their meaning is affirmative, our general aim is to clarify whether they are elaborated either as affirmative clauses, according to their semantic value, or as negative clauses, according to their morphological shape.

Crucially, different instances of EN structures display syntactic and semantic dissimilarities, not allowing them to be treated as a unitary phenomenon (Greco 2019a); therefore, we will focus on one single EN structure in Italian, i.e., Surprise Negation sentences (SNEGs) (Greco and Moro 2015; Greco 2019a, b), which is exemplified by the following sentence:

(2)	*E*	***non***	*mi*	*è*	*scesa*	*dal*	*treno*	*Maria?!*
	and	*not*	CLITIC.to-me	is	got	off-the	train	Mary
	'The surprise was that Mary got off the train!'[1]							

The sentence is affirmative regardless of the occurrence of the negative marker *not* and, therefore, it is considered an EN clause. Pragmatically, SNEGs are limited to a restricted context in which speakers are struck by an unexpected event—hence, the label "Surprise"—and they are mostly used in oral conversations. Since SNEGs display a marked intonation blending the acoustic features pertaining to both questions and exclamatives, they show the combined diacritic "?!".

The Present Study

The goal of this study is to investigate how EN is interpreted focusing on the SNEG case and measuring eye-movements in a visual world paradigm. According to the analysis proposed by Greco (2019b), our hypothesis is that SNEGs are mainly affirmative because negation does not affect their predicative core. More specifically, the adopted theoretical analysis argues that Italian negative marker "*non*" ("not") has a twofold interpretation encoded in syntax: (1) when it is merged in the predicative area (Belletti 1990; Zanuttini 1996, 1997), it gives the standard negative reading reversing the truth-value conditions of a sentence—as in the negative sentence '*Luca* ***non*** *è arrivato*' (let. Luke *not* is arrived = Luke has not arrived); (2) when it is merged in the complementizer area and the predicative one is already completed and inaccessible for further semantic and syntactic operations (see the *theory of phase* in Chomsky 2000, 2008, 2013), it gives the expletive

[1] In order to represent the "surprising" flavor of SNEGs, we used an inverse copular sentence construction. See Moro (1997) for a more detailed discussion.

reading shown in SNEGs (see sentence in 2). Moreover, according to Greco (2019a, b), SNEGs have some unique features distinguishing them from pure declarative sentences, such as the strong pragmatic interpretation of surprise, the interrogative-exclamative prosody, the focalization of the whole predicate, etc., which need to be correctly processed as well as integrated in the discourse context.

Based on Orenes et al. (2014, 2016), we expect that in binary contexts (1) individuals will show an increased rate of fixations on the mentioned object in both affirmative and SNEG sentences but not in negative ones; (2) individuals will show an increased rate of fixations on the alternative (not-mentioned) object after thousands of milliseconds only in negative sentences due to the integration of negation. Moreover, based on studies in experimental pragmatics (Schumacher 2012; Masia et al. 2017), we expect that (3) individuals will require a longer processing time to interpret SNEGs than declarative sentences (either Affirmative or Negative) since SNEGs display a higher syntactic complexity and the "surprising" information they carry needs to be pragmatically integrated in the discourse context. To verify these predictions, we manipulated the polarity of the sentences (affirmative, negative, and SNEG) and adopted a visual-world paradigm. The basic assumption of this methodology is that the comprehension of a sentence is mediated by its conceptual representation (Salverda et al. 2011). Eyes move toward the object denoted by a word when the linguistic input matches the visual input (Cooper 1974; Tanenhaus et al. 1995; Allopenna et al. 1998; Altmann and Kamide 2009; Kronmüller et al. 2017). Often, the expectations based on the conceptual representation of the previous discourse are sufficient to trigger eye-movements, before hearing the target word (we will refer to this behavior as to the *anticipatory effect*) (Altmann and Kamide 1999, 2007; Kamide et al. 2003; Altmann 2004).

Unlike many visual-world experiments (Altmann and Kamide 1999, 2007; Kamide et al. 2003; Altmann 2004; Orenes et al. 2014) using stylized drawings or black and white images, we used colored photos representing everyday-life objects and animals in order to increase the ecological validity of the visual stimuli (Viggiano et al. 2004) and their perceptual saliency (Ostergaard and Davidoff 1985; Brodie et al. 1991; Brodeur et al. 2010, 2014).

Since SNEGs are limited to a specific context in which speakers are struck by an unexpected fact (see above), in constructing the stimuli, we needed everyday-life situations that could be surprising, or unexpected. We tied the notion of expectedness to that of typicality (Rosch and Mervis 1975; Ashcraft 1978; Holmes and Ellis 2006; Maxfield et al. 2014): the more typical an event is for a given situation or for a given semantic category, the more expected its realization is, and vice versa (Mccloskey 1980; Hampton and Gardiner 1983; Pérez et al. 2015). For example, if a *soup* is considered a more atypical snack than an apple, then the sentence "The boy eats a soup" will be considered more unexpected than the sentence "The boy eats an apple" in a context in which the hearer knows that that boy is eating a snack. This is the reason why in this visual-world experiment all sets of visually displayed objects depicted atypical elements allowing for interpreting the unexpected value of the sentence, as SNEGs require.

We thus presented short everyday-life stories (e.g., *Laura invited some friends to her home. When they arrived, she showed them her pet*) introducing a specific semantic category (e.g., *pet*) followed by a target sentence that could be either affirmative (*e.g., La ragazza ha mostrato un serpente* translated as 'The girl showed a snake'), negative (e.g., *La ragazza non ha mostrato un serpente* translated as 'The girl did not show a snake') or SNEG (e.g., *La ragazza non ha mostrato un serpente?!* translated as 'The surprise was that the girl showed a snake!'), presented via headphones while four objects were shown on the screen. Two of the objects presented on the screen belonged to the semantic category introduced by the story (e.g., a *snake* and a *dog* as pets) and two did not (e.g., a *backpack* and an *air-conditioning device*). By doing so we provided a binary discourse context in which the mentioned object (e.g., *the snake*) co-occurred with an alternative one (e.g., *the dog*) belonging to the same semantic category. Such a context was required to distinguish the processing of affirmative and negative sentences (Orenes et al. 2014). One of the two objects related to the story was highly typical for the category under discussion (e.g., *the dog*) and, therefore, highly expected; the other one was highly atypical (e.g., *the snake*) and, therefore, highly unexpected. Such a configuration allowed us to create a natural context in which subjects could also evaluate SNEGs.

Materials and Methods

Thirty-four native Italian speakers with a university education participated in the eye-movement experiment (thirteen males and twenty-one females; 24.43 mean age). The recruitment process was voluntary. All participants had uncorrected vision or wore soft contact lenses. The Ethics Committee of the "Dipartimento di Scienze del Sistema Nervoso e del Comportamento" of the University of Pavia approved this study and all participants signed the informed consent.

Materials

Sixty experimental trials were presented to the participants divided in three conditions. Each trial consisted of an introductory context, describing a everyday-life scenario, and a target sentence—either affirmative, negative or SNEG—both presented via headphones. For each trial, four colored photos were shown on the screen. The introductory context was composed by two complex sentences:

> Part A: *Laura ha invitato delle amiche a giocare a casa sua* (translated as *Laura invited some friends to her home*)
> Part B: *Quando sono arrivate ha mostrato loro il suo animale domestico* (translated as *When they arrived, she showed them her pet*).

All verbs in the stories were used in the past tense. While Part A presented a situation, Part B included a direct object introducing a specific semantic category (e.g., *pets*).

The three experimental conditions were generated by creating three different target sentences for each context. The target sentence was either affirmative (e.g., *La ragazza ha mostrato un serpente* translated as 'The girl showed a snake'), negative (e.g., *La ragazza non ha mostrato un serpente* translated as *'The girl did not show a snake'*), or SNEG (*e.g., La ragazza non ha mostrato un serpente?!* translated as 'The surprise was that the girl showed a snake!'), and it displayed the same past-tensed verb of the introductory context (e.g., *ha mostrato* translated as 'showed'). Every participant was presented with 20 experimental trials per condition. A further 40 fillers were used. They were structurally identical to the experimental items with the notable difference that the target sentences always referred to the *expected* object (e.g., the *dog* in the previous example).

We then created six different experimental lists according to a Latin square design: in each list one third of the stories was presented in the Negative, SNEG, and Affirmative conditions, so that across lists each story was presented in different conditions to a similar number of participants. Also the response hand was counter-balanced, deriving the 6

Fig. 1 Example of photos accompanying the target sentence, either affirmative 'The girl showed a snake', negative 'The girl did not show a snake', or SNEG 'The surprise was that the girl showed a snake!'

different lists. Each list consisted of 60 experimental trials (20 affirmative, 20 negative and 20 SNEG) and 40 fillers (30 affirmative and 10 negative) for a total of 100 items. Trials were pseudo-randomized in order to avoid the occurrence of two, or more, successive SNEG sentences. Linguistic materials were recorded by a native Italian female speaker at a sampling frequency of 96 kHz and 32 bit. The onset and the offset time of each word forming the target sentence of each trial were manually marked using PRAAT software (Boersma 2001).

While the target sentence was auditorily presented, four colored photos (size = 450 × 450 pixels; resolution = 72 pixel/inch) were presented on the screen, dislocated in the four quadrants of the display (counterbalanced quadrant rotation both across and within participants). A single set of four photos was associated with each trial (Fig. 1). Two of the photos belonged to the semantic category named in the introductory context (*related photos*; e.g., *the snake* and *the dog*), two did not (*distractors*; e.g., *the backpack* and *the air-conditioning*). One of the related photos represented the highly expected element of the semantic category (e.g., the *dog* as a pet); the other represented the unexpected element, which was mentioned in target sentence (e.g., the *snake*).

In addition to the photos display, a written yes–no question appeared on the screen at the end of each trial. The question referred to the content of the target sentence (e.g., *Did the girl show a snake?*) and not to the photos in order to not influence the participants' eye-movements and stimulating their attention on the sentence content. The polarity of the answers was balanced within and among lists. Importantly, photos were rated for name agreement and ease of recognition, while target sentences were rated for their "expectedness" value (see Appendix).

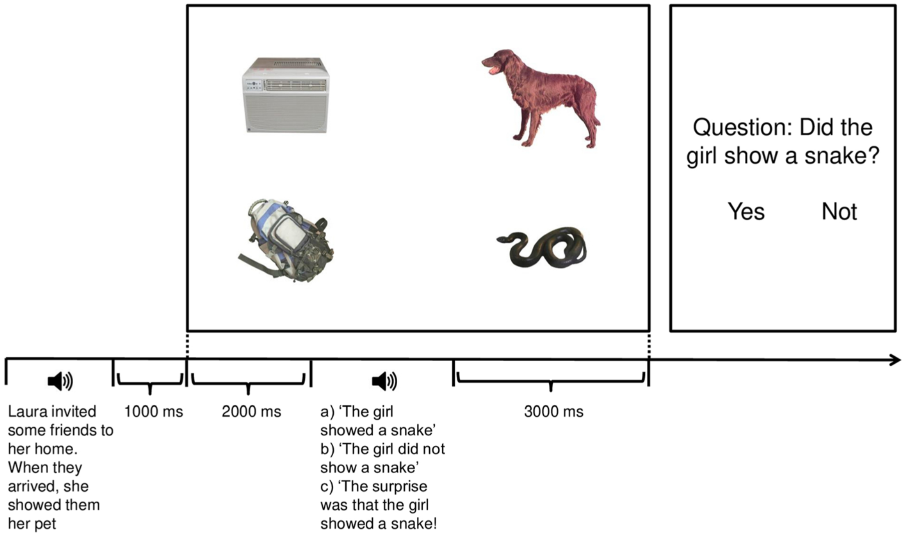

Fig. 2 Experiment time-line (the sentences are translated in English from Italian)

Apparatus and Procedure

Participants were seated in a quiet room in front of a 17-inch display at the distance of 56 cm from their eyes. Eye-movements were recorded at a rate of 1000 Hz with an SR Research EyeLink II eye-tracker, using a desktop-mount configuration. Head movements were restricted using a chin-rest to ensure the most precise estimation of gaze position. The whole experiment was implemented in the SR Research Experiment Builder. There were two blocks of trials. Calibration and validation of the recording apparatus were carried out at the beginning of each block and repeated when needed. Participants were told that the experiment aimed to investigate what happens when people listen to sentences while they simultaneously look at pictures and that their eye-movements would be recorded. They were not informed that the experiment focused on negation. More specifically, participants were told that they would hear some stories via headphone and see some photos on a screen; at the end of each trial they were instructed to answer to a comprehension question by pressing the button corresponding to *yes* or *no* on a RB-540 Cedrus keyboard with no time pressure.

Each trial started with a central fixation dot for drift correction (see Fig. 2), which was replaced by a fixation cross remaining on the screen during the time participants listened to part A and part B. A blank screen of 1000 ms followed it and, after that, the display with the four colored photos appeared. After 2000 ms participants listened to the target sentence. Eye-movements were recorded from the offset of the introductory context (before photos presentation) and ended 3000 ms after the offset of the target sentences, when also the photos were removed from the screen. Finally, the comprehension question was displayed. The pre-sentence period guaranteed a suitable time both to correctly review the display and to express anticipatory effects (Kamide et al. 2003; Altmann 2004; Altmann and Kamide 2007); the time period at the end guaranteed enough time to integrate negation (Kaup et al. 2005, 2006; Lüdtke et al. 2008; Orenes et al. 2014). The experiment had a total duration of approximately 40 min. A training session of five trials preceded it.

It may be worth underlining that the task was not active, unlike the ones used in some eye-tracker experiments (see the discussion in Altmann 2004; Orenes et al. 2014), in which participants touched or did something on the screen. The aims of the comprehension task were both to help participants to stay focused during the experiment and to measure their accuracy in comprehension.

Data Analysis and Statistics

The aim of behavioral analysis was to test how well and how quickly the three types of questions were understood by the sample of participants (after removal of all observations with RTs above or below 3sd from single subject means). We used lme4 package in R (Bates et al. 2015) to analyze the responses to the comprehension questions: for Accuracy, generalized linear mixed models (using logit function, Jaeger 2008) were used, while for Reaction Times, linear mixed models were used. In both cases, the dependent variables (probability of correct answers, or the time taken to answer) were modeled from the single trials, and using the maximal random effects structure (Barr et al. 2013), thus accounting for by-subject and by-item random intercepts and by-subject and by-item random slopes for the effect of the Sentence Type.

Considering the eye movements data, the output of the recordings consisted of x and y coordinates of the gaze position. For each sampling point we determined whether the eye was fixating one of the four Area of Interest (AOI), corresponding to the Expected, the Unexpected and the two Unrelated objects and re-coded fixations accordingly. Two different analyses are reported, closely following the approach proposed in Orenes et al. (2014) in order to keep the results highly comparable across the studies. The two analyses were carried out on different dependent variables and with different statistical approaches. In the Anticipatory Analysis we investigated a period of interest ranging from the onset of the pictures to the onset of the target sentences (the first 2000 ms after the fixation cross); in the Contingent Analysis we investigated a period of interest during the auditory presentation of the sentence, ranging from the 200 ms before the offset of the second word to 3500 ms later. The choice of subjecting data to two different kinds of analysis is due to the different questions that the two analyses aim at answering.

The aim of the Anticipatory Analysis is to replicate a relatedness effect (e.g., Kamide et al. 2003), answering the question of whether the previous context actually induced fixations towards contextually compatible objects or not, before hearing the critical sentences. The dependent variable consisted of the (arc-sin transformed) single-subject proportions of looks to each object type. Null hypothesis testing was carried out dividing the time period into smaller time bins (25-ms wide), and testing the differences in fixation proportion between object types (the average of the two Unrelated conditions vs. Expected, and Expected vs. Unexpected) with a series of (paired) *t* tests, using FDR adjustment for multiple comparisons ($n = 80$).

In the Contingent Analysis we adopted a growth-curve analysis approach that allows for investigating the time course of eye gaze across the screen, not requiring the selection of ad-hoc time windows of analysis, but rather modeling the “curve” depicted by the probability of looks towards a region of interest over time, and as influenced by a set of independent variables (Mirman et al. 2008). The time onset (t0) of this analysis was 200 ms before the offset of the 2nd word of each sentence (e.g., *ragazzo/a* translated as boy/girl), which for Negative and SNEG sentences immediately preceded the presentation of the particle *non* (e.g., L*a ragazza* ***non*** *ha mostrato un serpente* (?!)). This was necessary since the time interval between picture onset and the critical time point was variable across sentences. The input matrix consisted of the single trial data and the time variable was segmented into a series of 50 ms time-bins. The dependent variable was categorical (0,1) and its value depended on whether, at any moment in each single trial, a participant was fixating a particular object, or not. The logic behind the growth curve analysis is to first model the probability of fixations as an effect of Time (level-1 model) using orthogonal power polynomials (here we used first, second and third order), as a mean to capture the possibly complex shape of the curve (not only linear effects of time, but also quadratic, and cubic terms are evaluated). At a second stage (level-2 models) the effect of the independent variable(s) (Type of sentence: Affirmative, Negative and SNEG) over time is investigated, by evaluating the main effect of Type of sentence or the interaction between Type of sentence and the different terms (linear, quadratic, and cubic) of the variable Time. A likelihood ratio test (LRT) is then performed to evaluate whether level-2 models, and in particular which of the level-2 models better explains the variance in the data (the models Log Likelihood, the Chi square statistic for the LRT, and significance level are reported). For instance, an increase in likelihood associated to the interaction between Type of Sentence and the linear term of Time would suggest that differences between conditions linearly changed within the time period. Concerning the random effects structure (which caused no convergence issue), we allowed random intercepts for participants and items, and by-subject random slope for Type of Sentence (in interaction with the linear, and quadratic terms of time). We used repeated contrasts for Type of Sentence, comparing Negative with Affirmative and Affirmative with SNEG (Schad et al. 2020). In the results section we reported the results of the LRTs in tabular form, whereas all significant parameters are described in the text. Value of the parameters are expressed in percentages only when effects concerned a contrast between levels of Sentence Type, and are expressed in changes in coefficients when they refer to effects of the Time terms or interactions involving Time terms.

Results

Behavioral Data

Response accuracy and latency were both analyzed with mixed models testing the effect of Sentence type. Response accuracy to the three conditions differed: the probability estimates of correctly answering Affirmative (99.97%) and Negative (99.96%) sentences were very high, while the accuracy was lower with SNEGs (91.88%, $t=-4.7$, $p<0.001$), compared to Affirmative. The pattern for the RTs was similar in that questions on SNEGs required longer response times (+236 ms, $t=3.06$, $p<0.01$) than questions on Affirmative sentences, while Negative sentences did not differ (+55 ms, $t=1.35$, ns) from Affirmative ones (see Fig. 3).

Eye Movement Results

Anticipatory Analysis

The Relatedness effect was very robust. Compared to the average of the two distractors, fixations to Expected and Unexpected objects were more likely than fixations to the distractors as soon as after 275 ms and lasted until 2000 ms. Differences between Unexpected

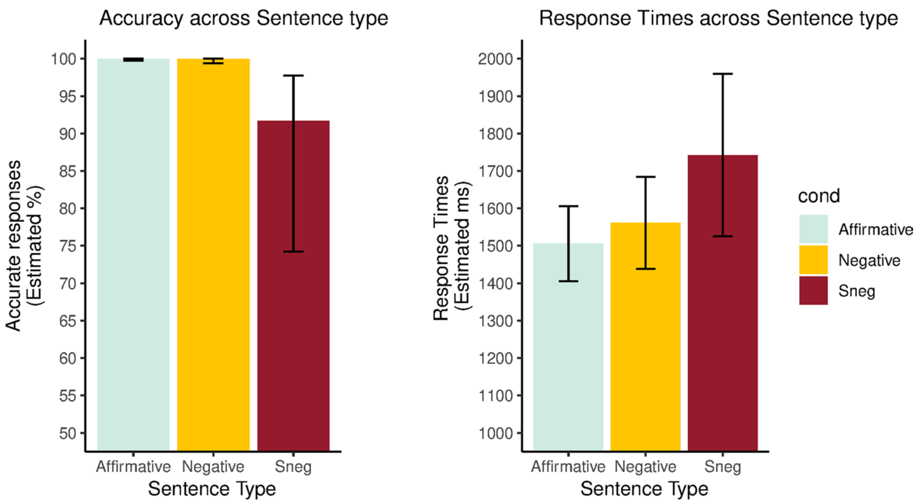

Fig. 3 Behavioral results: accuracy (on the left) and response times (on the right) to the comprehension questions following the three sentence types (affirmative, negative, and SNEG). The values are the estimated means of the models and the error bar represents 95% confidence intervals

and Expected objects were not very strong but participants looked at the unexpected ones slightly more often, and the difference was marginally significant ($0.05 < ps < 0.1$) in between 1150 and 1400 ms (see Fig. 4).

Contingent Analysis

The likelihood ratio tests (see Table 1) revealed the best fitting models were those allowing for the interaction between Sentence Type and the cubic term of Time in the looks to the Expected object, and the one allowing for the quadratic term for the looks to the Unexpected. Interestingly, the contribution of the different interaction terms differed in the analysis of two relevant objects. Below we inspect the models and describe, in detail, all the relevant parameters.

Contingent Expected Looking at the terms describing the time development of fixations independently of Sentence Type, the model shows that looks to the Expected object linearly increase over time (linear coefficient: +0.98, $t=2.58$, $p=0.009$) and show a general fall-rise pattern (quadratic coefficient: +0.33, $t=1.78$, $p=0.082$) (see Fig. 5). When considering the differences due to Sentence Type, the analysis revealed a significant effect on the intercept, with Negative being looked at less often (mean 28.0%, $t=-2.25$, $p=0.024$), and SNEG being looked at more often (mean = 29.8%, $t=3.38$, $p<0.001$) than Affirmative (mean = 28.7%). This effect on the intercept shows that on average (i.e., across all time points), Expected objects are looked at less often with Negative and more often with SNEGs, as compared to Affirmative sentences. However, interactions with the different terms of time also emerged, making these differences less “relevant”.

The differences between Negative and Affirmative are further described by two significant interactions. Negative and Affirmative differed for the contribution of quadratic term (+0.81, $t=1.73$, $p=0.082$), and the cubic term (−0.62, $t=-5.00$, $p<0.001$). The curve for the Negative sentences has a more pronounced fall-rise shape (the slightly

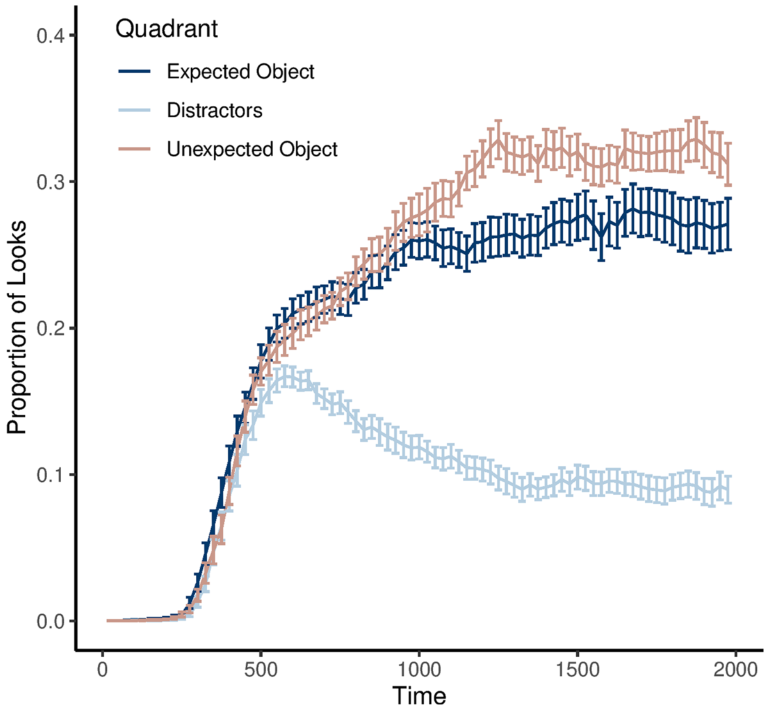

Fig. 4 Fixation proportion to the distractors, expected and unexpected objects during the 2000 ms before the target sentences. The value 0 on the horizontal axis represents the onset of the display with the photos. Light blue represents distractors, green represents the unexpected object and red represents the expected one. Error bars depict standard error of the mean

more positive quadratic coefficient) that further sums to a more pronounced fall-rise-fall shape (the more negative cubic coefficient). The interplay of these two coefficients describes the U shape of Negative sentences occurring in the first part of the epoch.

Contingent Unexpected When looking at the parameters of the model, two terms of Time represent the shape of the curve, independently of differences across Sentence Type. Overall, the time development of fixations has a clear U shape (quadratic term = +0.94, $t = 3.70$, $p < 0.001$), while the negative cubic term (− 0.10, $t = -1.79$, $p = 0.073$) suggests that the curve reaches the lower end during the first half of the epoch (see Fig. 6). Sentence Type has a significant effect on the intercept, with SNEG being looked at (mean = 24.9%) less often than Affirmative (mean = 26.1%, $t = 3.83$, $p < 0.001$) and Negative sentences (26.2%, $t = -4.16$, $p < 0.001$). The differences between Negative and Affirmative are further characterized by a significant difference in the quadratic term of Time (− 1.21, $t = -2.58$, $p = 0.009$): crucially, the difference in these two parameters show that, compared to Affirmative, the curve for Negative sentences has a less pronounced fall-rise shape. Importantly, no significant difference in the Time coefficients were observed in the comparison between SNEGs and Affirmative.

Table 1 Results of the likelihood ratio tests

	Log likelihood	Likelihood ratio test (χ^2)	*p* value
(A) Looks to expected			
Level-1 model	−75,589		
Level-2 models			
Sentence type (intercept)	−75,574	28.46	<0.001***
Sentence type on linear term	−75,573	2.37	0.31
Sentence type on quadratic term	−75,571	4.51	0.10^
Sentence type on cubic term	−75,558	25.34	<0.001***
(B) Looks to unexpected			
Level-1 model	−72,430		
Level-2 models			
Sentence type (intercept)	−72,421	19.11	<0.001***
Sentence type on linear term	−72,420	0.72	0.69
Sentence type on quadratic term	−72,417	6.13	0.046*
Sentence type on cubic term	−72,417	0.12	0.93

For both looks to the expected (A) and unexpected (B) objects, a measure of goodness of fit (log likelihood), the test (LRT) to evaluate the likelihood change (χ^2) of models of increasing complexity and its associated *p* value are reported. Level-1 model only includes the four terms of time, while in the level-2 models the effect of sentence type on the intercept or the different time terms is reported

General Discussion

The aim of the present visual-world eye-tracking experiment was to shed light on the processing of ENs focusing on a specific instance of them, i.e., Surprise Negation Sentences. This puzzling topic has never been addressed before in the experimental literature. We investigated the hypothesis that SNEGs are elaborated as affirmative clauses, despite the occurrence of the negative marker *not*, since negation does not syntactically affect their predicative core (Greco 2019b).

More specifically, our predictions were that in the proper linguistic context, i.e., when participants evaluate the meaning of the negative sentences dealing with two alternative objects (in a binary context, see Orenes et al. 2014), individuals should (1) show an increased probability of fixating the mentioned object in both affirmative and SNEG sentences but not in negative ones, and (2) show an increased probability of fixating the alternative (not-mentioned) object after thousands of milliseconds in negative sentences due to the late integration of negation but not in affirmative and SNEGs sentences. Moreover, based on studies in experimental pragmatics (Schumacher 2012; Masia et al. 2017), we expected that (3) individuals would interpret SNEGs with longer processing time than declarative sentences (either Affirmative or Negative), since SNEGs display a higher syntactic complexity and the "surprising" information that they carry needs to be pragmatically integrated in the discourse context.

Our conditions were generated by three types of target sentences—affirmative (e.g., *The girl showed a snake*), negative (e.g., *The girl did not show a snake*) and SNEG (e.g., '*La ragazza **non** ha mostrato un serpente?!*' translated as 'The surprise was that the girl showed a snake!')—which followed a short everyday-life story (e.g., *Laura invited*

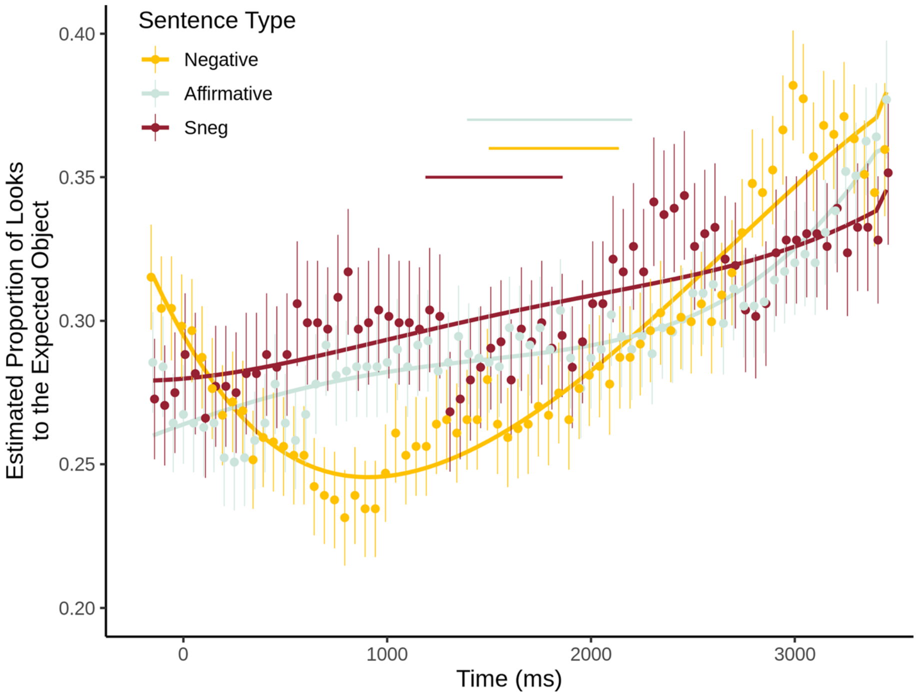

Fig. 5 Fixation proportion to expected object during the target sentences; in grey for the affirmative sentences, in yellow for the negative sentences and in dark red for the SNEGs. The offset of the 2nd word is represented on the horizontal axis by 0. The three horizontal segments present when, on average, the sentence object was pronounced. Error bars depict standard error of the mean

some friends to her home. When they arrived, she showed them her pet) while four photos were presented on the screen: two belonged to the semantic category introduced by the story, corresponding either to the highly unexpected one (e.g., a *snake*)—mentioned in the target—sentence or to a highly expected one (e.g., a *dog*); the other two were distractors (not members of that category, such as a *backpack* and an *air-conditioning unit*). Because only two objects were compatible with the discourse context, this configuration provided a binary visual context in which the differences between affirmative and negative clauses could emerge (Orenes et al. 2014).

The robust relatedness effect observed in the Anticipatory Analysis indicates the appropriateness of the manipulation that we designed. The two objects related to the semantic category introduced by the short story were indeed looked at more often than the distractors as soon as after 300 ms, and before that the target sentence was presented (Fig. 4). This is in line with the anticipatory effects found and widely discussed in the literature (Kamide et al. 2003; Altmann 2004; Altmann and Kamide 2007), where participants' gaze was reported to fixate objects compatible with the verb, before their presentation. Interestingly, the proportion of looks toward the related objects showed a marginally significant effect of typicality in a lag of time between 1150 and 1400 ms for which the unexpected objects were looked at slightly more often than the expected ones. This result suggests that after a first visual investigation of the display, the objects that are consistent with the semantic category that is introduced by the discourse are looked at more often, and further

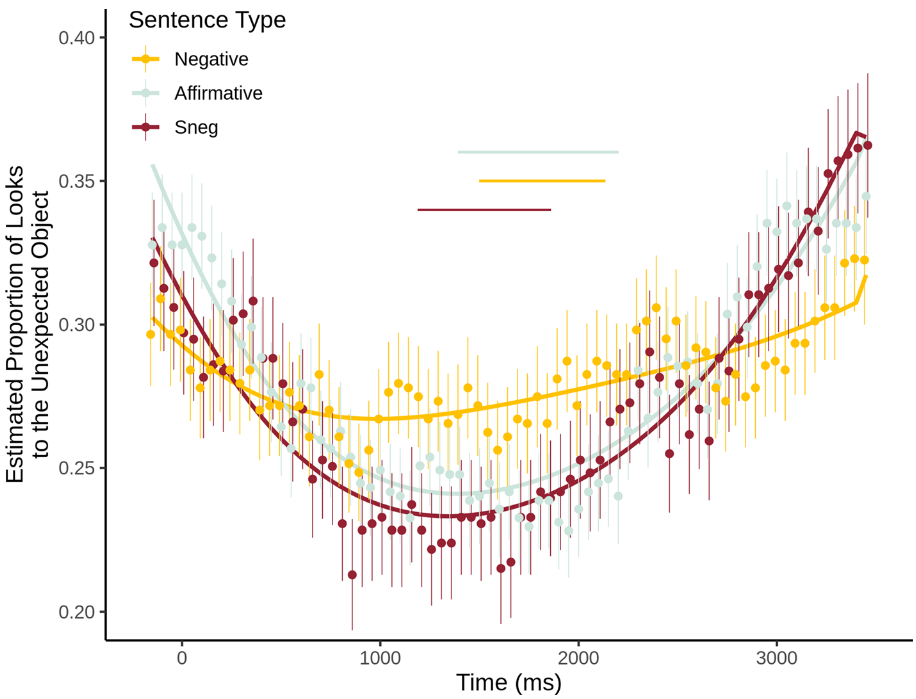

Fig. 6 Fixation proportion to unexpected object during the target sentences; in grey for the affirmative sentences, in yellow for the negative sentences and in dark red for the SNEGs. The offset of the 2nd word is represented on the horizontal axis by 0. The three horizontal segments present when, on average, the sentence object was pronounced. Error bars depict standard error of the mean

that objects that are less typical exemplars of that category are looked at even more likely, possibly because the unexpected element captures the listeners' visual attention, in keeping with what Boland (2005) and Pérez et al. (2015) showed.

Coming to our predictions, the effect of the Sentence type was investigated in the Contingent Analysis, looking at fixations towards the unexpected (Fig. 6) and the expected objects (Fig. 5). Let us start from the former: the quadratic term showed a less positive coefficient in negative sentences compared to the affirmative ones showing that the curve was more flat (the fall-rise shape was less pronounced). Crucially, the analysis did not manifest differences between SNEGs and affirmative in the time development of fixations to the Unexpected object. Data are compatible with the idea that participants kept looking at the unexpected object after hearing the negative particle in negative, but not in SNEG or in affirmative sentences. It is worth recalling that the unexpected object (e.g., *the snake*) was the mentioned one in all the target sentences and, coherently, the subjects' attention increased on it on the second half of the time epoch (the inflection point of the fall-rise shape occurs during to the mention of the object) for both affirmative (e.g., *The girl showed a snake*) and SNEG cases (e.g., *The surprise was that the girl showed a snake!*) because it realizes the actual meaning of the proposition, whereas this does not occur in negative sentences (e.g., *The girl did not show a snake*) because of negation. This result confirmed our first prediction and it is in line with the results in Orenes et al. (2014, 2016). Interestingly, the less pronounced fall-rise curve in negative sentences suggests that negation plays a role in the elaboration of the sentence earlier than we expected, since the difference appears soon after hearing *non* and before the object was mentioned.

Nonetheless, the differences in the elaboration of negative and SNEG sentences further emerge when we consider the probability of fixation to the expected object. Both the quadratic and cubic functions of time shows that participants move their gaze *away* from the expected (not-mentioned) object (e.g., *the dog*) to elsewhere on the screen, at the beginning of the negative sentences, and then they come back later on. This suggests that participants build a representation of the most likely scenario early on (the pet the girl will show is a dog), and negate it soon after the *non*; then, when they build the effective meaning of the sentence listening to it (*The girl did not show a snake*) they turn back to the *dog* (late integration). This moving away from—turning back to the expected object does not happen with SNEG and affirmative sentences, confirming our second prediction. Again, this difference appears before the object was mentioned and suggesting that negation plays a role in the elaboration of the sentence earlier than we expected.

On the other hand, behavioral data suggest a unique characterization of SNEGs versus Affirmative and Negative sentences. Statistical analysis showed that the probability of correctly resolve the task was higher after affirmative and negative target sentences (respectively, 99.97% and 99.96%) than after SNEGs (91.88%). This pattern is further mirrored by the slower response time associated with SNEGs and it can be seen as the consequence of the interaction of a set of independent factors. First, according to Greco (2019b), SNEGs have some specific features distinguishing them from the declarative sentences, such as a complex syntactic structure—involving the focalization of the whole predicate—, an interrogative-exclamative prosody and the "surprising" value that they carry. Such a "complexity" may alone cause the higher effort required by SNEGs. Consider, for instance, the interrogative-exclamative prosody of SNEGs. The F_0 in SNEGs has an initial high value (%H) and ends with a raising pattern (H%) (Greco 2019b) whereas declaratives start lower and end with a falling pattern (L–L%) (Avesani 1990; Sorianello 2011). A growing amount of studies shows that prosody influences the processing of unfolding utterances (Brown et al. 2011; Falé et al. 2016), suggesting that the differences in the fixation proportion of declarative and SNEG sentences may lay on their differences in prosody. Consider now the fact that SNEGs need to be integrated in the discourse context in a way that is absent in the negative and affirmative clauses: they all require the elaboration of the lexical meaning, but SNEGs also require a particular awareness of the context and of the speaker's intentions. In fact, the surprise effect characterizing SNEGs arises because new (and surprising) information needs to be accommodated in the mental model of discourse and, according to Masia et al. (2017), this elicits higher efforts "due the mismatch between the information packaging and the actual discourse representation" (p. 43) (see also Schumacher 2012; Domaneschi et al. 2018). Moreover, according to Tubau et al. (2017), expletive negation is often misunderstood and confused in languages that display the same negative marker for both expletive and standard negation—and Italian falls under this class of languages—possibly causing the higher error rate we found in SNEGs.

Crucially, the slower response time and the lower accuracy does not contradict the hypothesis that SNEGs are processed as affirmative sentences, as the eye-movements suggest, and certainly does not suggest a similarity with negative sentences, which do not show any of these higher efforts.

To conclude, the present study investigated for the first time the elaboration of EN focusing on an Italian construction, i.e., Surprise Negation. Surprise Negation clauses are affirmative in meaning even though they display a negative marker, as in the sentence '*E* ***non*** *mi è scesa dal treno Maria?!*' (let. 'and *not* CLITIC.to_me is got off-the train

Mary' = 'The surprise was that Maria got off the train!'). Following the analysis in Greco (2019b), we showed that speakers elaborate SNEGs as affirmative clauses, according to their syntactic and semantic derivation, and not as negative ones following their morphological shape. Moreover, we also found that surprise negation requires a major effort to be elaborated. In general, the evidence reported in this work suggests that negation plays a different role in the mental representation of a sentence depending on both its syntactic derivation and its pragmatic integration in the discourse context. These findings prompt future studies to consider the impact of the syntactic derivation when defining a pragmatic phenomenon, and, vice versa, the impact of the pragmatic context on the processing of a syntactic structure.

Acknowledgements This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' Contribution Theoretical linguistic framework: AM. Linguistic stimuli design: MG. Study concept and design: AM, MG, PC, VB. Experimental and stimuli design: MG, PC, VB. Data collection: MG. Data analysis and interpretation: MG, PC. Manuscript writing: MG, PC. Manuscript review: MG, PC, VB.

Compliance with Ethical Standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Appendix: Rating of the Materials

There were two goals in the rating study: testing the recognition of the photos and testing the expectedness/unexpectedness value of the target sentences. Twenty-seven native Italian speakers (ten males and seventeen females; 26.17 mean age) voluntarily participated to the rating by filling on-line questionnaires (SurveyMonkey® software). Most of the photos were extracted from BMB-corpus (Bates et al. 2015), COGPSY-corpus (Viggiano et al. 2004) and BOSS-corpus (Brodeur et al. 2010, 2014). We integrated them with new ones taken from the Internet. We chose photos with a frontal point of view and we used an open-source graphic-editor software, GNU Image Manipulation Program, to make them comparable to the ones in the corpora: we cut out the object from the scene background which was turned in a white color; we adjusted brightness, colors, and contrast. Even though the familiarity and the visual complexity of photos are cross-linguistically stable variables (Cfr. Pompeia et al. 2003), the name agreement is not (Sanfeliu and Fernandez 1996; Alario and Ferrand 1999; Pompeia et al. 2003) and, therefore, participants were asked to freely name each of them (naming task) and judge how easy it was to recognize the depicted object on a five-point Likert scale (1 = very difficult; 5 = very easy) (photos recognition task). Photos were presented one at a time for a total of 200 items. Participants were also asked to judge how expected a sentence was as a conclusion of a real-life story on a five-point Likert scale (1 = unexpected; 5 = very expected) (expectedness valuation task). 200 target sentences (either expected or unexpected) were prepared for a total of 100 real-life stories. Stimuli were organized in two lists according to a Latin square design so that each participant was presented with each story (with equal numbers of expected or unexpected target sentences) and each sentence was judged by the same number of participants. We selected all the items that satisfied the following criteria:

- *Photos* (1) More than 70% of participants agreed on the name of the depicted object (we considered correct all the narrow synonyms, such as *croissant/brioche*, *computer/laptop*, etc.); (2) scores > 4 in the photo recognition task.
- *Sentences* Score > 4 in the expectedness valuation task for the expected sentences and < 2.5 for the unexpected sentences. These cut-off points were chosen to make sure that the "most typical" among unexpected items were excluded from the materials, and also to have a sufficient number of items per each condition.

As a consequence, twenty-one photos and eight target sentences were excluded from the original set of materials. They were substituted and new ratings were collected. The final materials gained an overall name agreement of 95% (SD 0.08). Photos were judged to be recognized fairly easily [4.79, SD 0.25], whereas the average expectedness was 4.50 (SD 0.32) for the expected sentences and 1.79 (SD 0.43) for the unexpected ones.

The final set of materials was comprised of 100 stories (60 critical and 40 fillers), 220 target sentences (180 target in the three conditions and the 40 fillers) and 400 photos (200 were the photos that underwent the rating study and represented typical and atypical objects; other 200 photos formed the set of unrelated objects taken from the corpora without any rating).

References

Alario, F. X., & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers, 31,* 531–552.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419–439.

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm'. *Cognition, 93,* 79–87.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73,* 247–264.

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57,* 502–518.

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye-movements and mental representation. *Cognition, 111,* 55–71.

Ashcraft, M. H. (1978). Property norms for typical and atypical items from 17 categories: A description and discussion. *Memory & Cognition, 6*(3), 227–232.

Avesani, C. (1990). A contribution to the synthesis of Italian intonation. In *Proceedings of the 1st international conference on spoken language processing*, Kobe (pp. 1–13).

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*(3), 255–278.

Bartoli, E., Tettamanti, A., Farronato, P., Caporizzo, A., Moro, A., & Gatti, R. (2013). The disembodiment effect of negation: Negating action-related sentences attenuates their interference on congruent upper limb movements. *Journal of Neurophysiology, 109*(7), 1782–1792.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Belletti, A. (1990). *Generalized verb movement*. Torino: Rosenberg & Sellier.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341–345.

Boland, J. E. (2005). Visual arguments. *Cognition, 95,* 237–274.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The BOSS, a new set of 538 normalized photos of objects to be used as ecological stimuli in vision and memory paradigms. *PLoS ONE, 5,* e10773.

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of standardized stimuli (BOSS) phase II: 930 new normative photos. *PLoS ONE, 9*(9), e106953.

Brodie, E., Wallace, A., & Sharrat, B. (1991). Effect of surface characteristics and style of production on naming and verification of pictorial stimuli. *American Journal of Psychology, 104,* 517–545.

Brown, M., Salverda, A. M., Dilley, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin & Review, 18*(6), 1189–1196.

Carpenter, P., & Just, M. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review, 82*(1), 45–73.

Carpenter, P., Just, M., Keller, T., Eddy, W. F., & Thulborn, K. R. (1999). Time course of fMRI-activation in language and spatial networks during sentence comprehension. *NeuroImage, 10*(2), 216–224.

Chomsky, N. (2000). Minimalist Inquiry: The framework. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik* (pp. 89–155). Cambridge, MA: The MIT Press.

Chomsky, N. (2008). On phases. In R. Freidin, C. P. Otero, & M. L. Zubizarreta (Eds.), *Foundational issues in linguistic theory* (pp. 133–166). Cambridge, MA: MIT Press.

Chomsky, N. (2013). Problems of projection. *Lingua, 130,* 33–49.

Clark, H., & Chase, W. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3,* 472–517.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6,* 84–107.

Dale, R., & Duran, N. D. (2011). The cognitive dynamics of negated sentence verification. *Cognitive Science, 35*(5), 983–996.

Delfitto, D., & Fiorin, G. (2014). Negation in exclamatives. *Studia Lingüística, 68*(3), 284–327.

Domaneschi, F., Canal, P., Masia, V., Vallauri, E. L., & Bambini, V. (2018). N400 and P600 modulation in presupposition accommodation: The effect of different trigger types. *Journal of Neurolinguistics, 45,* 13–35.

Falé, I., Costa, A., & Luegi, P. (2016). Reading aloud: Eye movements and prosody. *Speech Prosody*. https://doi.org/10.21437/SpeechProsody.2016-169

Ferguson, H. J., Sanford, A. J., & Leuthold, H. (2008). Eye-movements and ERPs reveal the time course of processing negation and remitting counterfactual worlds. *Brain Research, 1236,* 113–125.

Giora, R. (2006). Anything negatives can do affirmatives can do just as well, except for some metaphors. *Journal of Pragmatics, 38*(7), 981–1014.

Giora, R., Fein, O., Aschkenazi, K., & Alkabets-Zlozover, I. (2007). Negation in context: A functional approach to suppression. *Discourse Processes, 43*(2), 153–172.

Glenberg, A. M., Robertson, D. A., Jennifer, L. J., & Johnson-Glenberg, M. C. (1999). Not propositions. *Cognitive Systems Research, 1*(1), 19–33.

Greco, M. (2017). *Surprise negation sentences: Expletive negation and the left periphery*. Ph.D. dissertation, University Vita-Salute San Raffaele—IUSS Pavia, Milano-Pavia

Greco, M. (2019a). Is expletive negation a unitary phenomenon? *Lingue e Linguaggio, 1,* 25–58.

Greco, M. (2019b). On the syntax of surprise negation sentences: A case study on expletive negation. *Natural Language & Linguistic Theory*. https://doi.org/10.1007/s11049-019-09459-6.

Greco, M., & Moro, A. (2015). Surprise negation (SNEG) sentences. *Poster session presented at the Götttingen Summer School on Negation*. Göttingen: University of Göttingen.

Hampton, J., & Gardiner, M. M. (1983). Measures of internal category structure. A correlational analysis of normative data. *British Journal of Psychology, 74*(4), 491–516.

Hasegawa, M., Carpenter, P. A., & Just, M. A. (2002). An fMRI study of bilingual sentence comprehension and workload. *NeuroImage, 15*(3), 647–660.

Hasson, U., & Glucksberg, S. (2006). Does understanding negation entail affirmation? An examination of negated metaphors. *Journal of Pragmatics, 38*(7), 1015–1032.

Holmes, S. J., & Ellis, A. W. (2006). Age of acquisition and typicality effects in three object processing tasks. *Visual Cognition, 13*(7–8), 884–910.

Horn, L. (1989). *A natural history of negation*. Chicago: University of Chicago Press.
Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446.
Jespersen, O. (1917). *Negation in English and other languages*. Copenhagen: A. F. Høst.
Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49,* 133–156.
Kaup, B. (1997). The processing of negatives during discourse comprehension. In M. B. Shafto & P. Langley (Eds.), *Proceedings of the 19th conference of the cognitive science society* (pp. 370–375). Mahwah: Lawrence Erlbaum.
Kaup, B. (2001). Negation and its impact on the accessibility of text information. *Memory & Cognition, 29*(7), 960–967.
Kaup, B., Lüdtke, J., & Zwaan, R. A. (2005). Effects of negation, truth value, and delay on picture recognition after reading affirmative and negative sentences. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th annual conference of the cognitive science society* (pp. 1114–1119). Mahwah: Lawrence Erlbaum.
Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics, 38,* 1033–1050.
Kaup, B., Lüdtke, J., & Zwaan, R. A. (2007). The experiential view of language comprehension: How is negation represented? In F. Schmalhofer & C. A. Perfetti (Eds.), *Higher level language processes in the brain. Inference and comprehension processes* (pp. 255–288). London: Lawrence Erlbaum Associates.
Kronmüller, E., Noveck, I., Rivera, N., Jaume-Guazzini, F., & Barr, D. (2017). The positive side of a negative reference: The delay between linguistic processing and common ground. *Royal Society Open Science, 4,* 160827.
Liuzza, M. T., Candidi, M., & Aglioti, S. M. (2011). Do not resonate with actions: Sentence polarity modulates cortico-spinal excitability during action-related sentence reading. *PLoS ONE, 6*(2), e16855.
Lüdtke, J., Friedrich, C. K., De Filippis, M., & Kaup, B. (2008). ERP correlates of negation in a sentence-picture-verification paradigm. *Journal of Cognitive Neuroscience, 20*(8), 1355–1370.
Lüdtke, J., & Kaup, B. (2006). Context effects when reading negative and affirmative sentences. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th annual conference of the cognitive science society* (pp. 1735–1740). Mahwah: Lawrence Erlbaum Associates.
Macdonald, M., & Just, M. (1989). Changes in activation levels with negation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(4), 633–642.
Masia, V., Canal, P., Ricci, I., Lombardi Valluri, E., & Bambini, V. (2017). Presupposition of new information as a pragmatic garden path: Evidence from event-related brain potentials. *Journal of Neurolinguistics, 42,* 31–48.
Maxfield, J. T., Stalder, W. D., & Zelinsky, G. J. (2014). Effects of target typicality on categorical search. *Journal of Vision, 14*(12), 1–11.
Mccloskey, J. (1980). The stimulus familiarity problem in semantic memory research. *Journal of Verbal Learning and Verbal Behavior, 19,* 485–502.
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*(4), 475–494.
Moro, A. (1997). *The raising of predicates*. Cambridge: Cambridge University Press.
Orenes, I., Beltrán, D., & Santamaría, C. (2014). How negation is understood: Evidence from the visual world paradigm. *Journal of Memory and Language, 74,* 36–45.
Orenes, I., Moxey, L., Scheepers, C., & Santamaría, C. (2016). Negation in context: Evidence from the visual world paradigm. *The Quarterly Journal of Experimental Psychology, 69*(6), 1082–1092.
Ostergaard, A. L., & Davidoff, J. B. (1985). Some effects of color on naming and recognition of objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 579–587.
Pérez, A., Joseph, S. L., Bajo, T., & Nation, K. (2015). Evaluation and revision of inferential comprehension in narrative texts: An eye movement study. *Language, Cognition and Neuroscience, 31*(4), 549–566.
Pompeia, S., Miranda, M. C., & Bueno, O. F. (2003). Brazilian standardized norms for a set of pictures are comparable with those obtained internationally. *Arquivos de Neuro-Psiquiatria, 61,* 916–919.
Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605.
Salverda, A. P., Brown, M., & Tanenhaus, M. K. (2011). A goal-based perspective on eye movements in visual world studies. *Acta Psychologica, 137*(2), 172–180.

Sanfeliu, M. C., & Fernandez, A. (1996). A set of 254 Snodgrass Vanderwart pictures standardized for Spanish: Norms for name agreement, image agreement, familiarity, and visual complexity. *Behavior Research Methods, Instruments, & Computers, 28,* 537–555.

Scappini, M., Delfitto, D., Marzi, C. A., Vespignani, F., & Savazzi, S. (2015). On the non-incremental processing of negation: A pragmatically licensed sentence-picture verification study with Italian (dyslexic) adults. *Cahiers de linguistique française, 32,* 45–58.

Schad, D. J., Hohenstein, S., Vasishth, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038.

Schumacher, P. B. (2012). Context in neurolinguistics: Time-course data from electrophysiology. In R. Finkbeiner, J. Meibauer, & P. B. Schumacher (Eds.), *What is a context? Linguistic approaches and challenges* (pp. 33–53). Amsterdam: John Benjamins.

Sorianello, P. (2011). Aspetti pragmatici e prosodici dell'atto esclamativo. *Studi Linguistici e Filologici Online (SLIFO), 9,* 287–332.

Speranza, J. L., & Horn, L. (2012). A brief history of negation. In D. M. Gabbay, F. J. Pelletier, & J. Woods (Eds.), *Logic: A history of its central concepts* (pp. 127–174). Amsterdam: Elsevier.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632–1634.

Tettamanti, M., Manenti, R., Della Rosa, P., Falini, A., Perani, D., Cappa, S., et al. (2008). Negation in the brain: Modulating action representation. *NeuroImage, 43*(2), 358–367.

Trabasso, T., Rollins, H., & Shaughnessy, E. (1971). Storage and verification stages in processing concepts. *Cognitive Psychology, 2*(3), 239–289.

Tubau, S., Déprez, V., Borràs-Comes, J., & Espinal, M. T. (2017). How speakers interpret the negative markers no and no… pas in Catalan. *Probus: International Journal of Romance Linguistics*. https://doi.org/10.1515/probus-2017-0008.

Viggiano, M., Vannucci, M., & Righi, S. (2004). A new standardized set of ecological pictures for experimental and clinical research on visual object processing. *Cortex, 40,* 491–509.

Villiers, J., & Flusberg, H. (1975). Some facts one simply cannot deny. *Journal of Child Language, 2*(2), 279–286.

Wason, P. (1961). Response to affirmative and negative binary statements. *British Journal of Psychology, 52*(2), 133–142.

Wason, P. (1965). The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior, 4*(1), 7–11.

Zanuttini, R. (1996). On the relevance of tense for sentential negation. In A. Belletti & L. Rizzi (Eds.), *Parameters and functional heads. Essays in comparative syntax* (pp. 181–207). Oxford: Oxford University.

Zanuttini, R. (1997). *Negation and clausal structure. A comparative study of Romance languages*. Oxford: Oxford U. Press.