eScholarship
University of California

Why bother? What our eyes tell about linking during the comprehension of psych verb (non) causative constructions

Carolina A. Gattei, Área de Educación, Escuela de Gobierno and Laboratorio de Neurociencia Universidad Torcuato Di Tella, AR, carogattei@utdt.edu

Federico Alvarez, Facultad de Filosofía y Letras, Universidad de Buenos Aires, AR, federicoalvarez.puan@gmail.com

Luis A. París, Instituto de Ciencias Humanas, Sociales y Ambientales, CONICET, AR, paris@mendoza-conicet.gob.ar

Alejandro J. Wainselboim, Instituto de Ciencias Humanas, Sociales y Ambientales, CONICET, AR, awainselboim@mendoza-conicet.gob.ar

Yamila Sevilla, Instituto de Lingüística Facultad de Filosofa y Letras, Universidad de Buenos Aires, AR, ysevilla@filo.uba.ar

Diego Shalom, Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, and IFIBA, CONICET, AR, diego@df.uba.ar

We present an eyetracking study that investigates how linking is achieved during real-time comprehension of Spanish sentences with causative psych verbs and alternative case marking. This group of verbs lead to verbs' argument structures that require direct or inverse syntax-to-semantics linking according to the type of case marking assigned to their object. The study aimed at disentangling whether processing inverse linking was more costly than direct linking, and exploring how incremental argument interpretation takes place when lexemes that accept several case markings are used. Results showed that during incremental comprehension, inverse linking is more difficult than direct linking, irrespective of word order. As for argument interpretation, the current study partially replicated the results of previous studies conducted in this language using different verb types. Findings are discussed under the light of different psycholinguistic models addressing case marking processing and incremental linking.

Introduction

Linking is the process by which the human language processing system maps syntactic function of constituents to semantic information provided by the thematic roles established in the lexico-semantic structure of the verb. The result of this process allows speakers to understand *'who is doing what to whom'* in a given utterance. Thus, accurate argument interpretation requires that linking is achieved correctly. To this end, during incremental processing speakers have to determine which participant of the sentence bears the most active thematic role, and which one bears the least active one. This process is known as *argument hierarchization* (Bornkessel et al., 2005). While different types of linguistic information may serve as cues for establishing the correct linking between syntax and semantics (e.g. word order, morphological case marking, animacy and definiteness of the constituents being the most important ones), the weight each cue imposes over argument interpretation is language-specific (MacWhinney et al., 1984). For instance, while English is a language that heavily relies on word order for argument interpretation, speakers of languages with a less rigid word order (e.g. German, Finnish, Russian) may use morphological case marking as a more informative linguistic cue (see Lamers and De Swart, 2012: for a collection of works that focus on this issue).

Several studies have inquired about how language-specific characteristics may affect the way linking is achieved during online processing, introducing evidence for psycholinguistic models that attempt to explain how this process is carried out (see Bornkessel and Schlesewsky, 2006; Bader and Bayer, 2006: for different models that focus on linking). Of particular interest for the current study, in a group of comprehension studies conducted in Spanish, Gattei et al. (2015a; 2015b; 2017) have shown that in this language, morphosyntactic information provided by case marking and word order, along with arguments' span (the amount of available arguments previous to the verb) trigger predictions about the type of event that will be described, and thus syntax-to-semantics linking of the type of verb predicted. Results of these studies show that when reading sentences with different word orders (SVO as in (1)a and (1)c and OVS as in (1)b and (1)d), the appearance of a verb that is incongruent with the type of event predicted leads to higher reading times and higher error rates (Gattei et al., 2015a; 2017), differential neural correlates (Gattei et al., 2015b) and higher amount of incoming regressions to the verb and the arguments regions (Gattei et al., 2017). The authors interpreted the outcome as a reflection of higher processing cost when the sentence word order did not reflect the order of arguments in the lexico-semantic structure of the verb, which positions the most prominent argument of the structure in first position.

(1)   a.   María   le   responde   a Ana.
            Mary$_{[NOM]}$ clitic$_{[DAT]}$ respond$_{[3rd-SG]}$ to Ana$_{[DAT]}$
            'Mary responds to Ana'

b. A Ana le responde María.
To Ana$_{[DAT]}$ clitic$_{[DAT]}$ respond$_{[3rd–SG]}$ Mary$_{[NOM]}$
'Mary responds to Ana.'

c. María le encanta a Ana.
Mary$_{[NOM]}$ clitic$_{[DAT]}$ love$_{[3rd–SG]}$ to Ana$_{[DAT]}$
'Ana loves Mary.'

d. A Ana le encanta María.
To Ana$_{[DAT]}$ clitic$_{[DAT]}$ love$_{[3rd–SG]}$ Mary$_{[NOM]}$
'Ana loves Mary.'

Furthermore, the authors found higher reading times for sentences that required a non-canonical linking between syntax and semantics (Gattei et al., 2015a: i.e. when the nominative-marked constituent is linked to the lowest-ranked thematic role) than for sentences with direct linking (i.e. the highest constituent in the syntactic hierarchy, is linked to the highest thematic role). This effect persisted even for the response times and accuracy rates of comprehension questions in both Gattei et al. (2015b) and Gattei et al. (2017). A question that remains unanswered relates to the role of the verb classes used in these experiments. The authors of the above mentioned studies aimed at maintaining the syntactic structure of sentences constant and used two-argument sentences with a nominative and a dative-marked object. They evaluated the interpretation of events that belong to different semantic classes that corresponded either to the class of activity verbs (e.g. 'gritar', *'to yell'*, 'mentir', *'to lie'*), or to the class of Object Experiencer psychological verbs (hereafter ObjExp psych verbs, e.g. 'gustar', *'to like'*, 'encantar', *'to love'*, and see Dröge et al., 2014; Bornkessel et al., 2003; 2005: for similar examples in German and Italian respectively).

Prior studies have shown that verb processing may differ according to several factors, including verb type (Pulvermüller et al., 2001), type of subcategorization and possible thematic roles (Shetreet et al., 2007), and semantic complexity (Brennan and Pylkkänen, 2010). Thus, it is unclear whether the verb type effects found in Gattei et al. (2015a) and Gattei et al. (2017) reflect greater difficulties related to linking complexity or if other factors may play a role.

The current study aims at responding to this question in order to shed light on how linking is achieved during real-time comprehension when the same class of verbs is used. It also goes a step further from other studies that investigated linking by using sentences with a special class of Spanish ObjExp psych verbs: Psych verbs which can bear both accusative or dative case marking.

The use of this subclass of psych verbs may provide further evidence on how a grammatical feature such as case marking serves as a valid cue for sentence overall interpretation in this language (see MacWhinney et al., 1984: for an experimental example of how different syntactic and semantic features modulate interpretation according to language-specific characteristics).

The study is organized as follows: we first introduce a few theoretical considerations about this class of verbs and discuss how they differentiate from psych verbs used in previous psycholinguistic studies. Then, we elaborate on the hypotheses and predictions for online comprehension of Spanish sentences with different word order. We finally test these hypotheses by means of a sentence comprehension eye-tracking study and an acceptability judgement task, and discuss the results under the light of the previous evidence on this topic and different psycholinguistic models that aim at explaining how linking between syntax and semantics is achieved during on-going processing.

Theoretical considerations about Spanish psych verbs with double case marking

The theoretical categorization and distinction of the different classes of psychological verbs (hereafter psych verbs) have been of great interest for the formal studies about the syntax-to-semantics linking, in particular, since the appearance of the influential work by Belletti and Rizzi (1988) about psych verbs and Θ Theory in Italian. In general terms, the authors classify Italian psych verbs in three classes. In the first class of verbs (the so-called *'temere'* class), the Experiencer takes nominative case, while in the second class (the *'preoccupare'* class), the Experiencer bares accusative case. Finally, the third class of psych verbs (the *'piacere'-like* class) assigns dative case to their Experiencer.

Spanish exhibits the same case-marking pattern for its first and third psych verbs classes, with verbs like 'amar' (*'to love'*) and 'gustar' (*'to like'*) respectively. However, it is still part of a controversial discussion what the most probable case marking is for the different exemplars that belong to the second class of psych verbs. For instance, the verb 'molestar' (*to bother*) allows accusative, dative and nominative case marking for the constituent related to its Experiencer, as it can be read in (2).

(2)   a.   Los   perros   lo   molestan   siempre   que   llega   tarde.
          The   dogs$_{[NOM.3rd.pl]}$ clitic$_{[ACC]}$ bother$_{[3rd.pl]}$ always   that   arrives$_{[3rd.sg]}$ late
          'The dogs bother (harass) him when he comes late.'

      b.   Los   perros   le   molestan   (*siempre   que   llega   tarde).
          The   dogs$_{[NOM.3rd.pl]}$ clitic$_{[DAT]}$ bother$_{[3rd.pl]}$ always   that   arrives$_{[3rd.sg]}$ late
          'Dogs bother him (*when he comes late).'
          (Treviño, 1990: 20b and 21b)

    c.   Él   se   molesta   con los perros siempre que llega   tarde.
       He$_{[NOM.3rd.sg]}$ clitic$_{[refl]}$ bother$_{[3rd.sg]}$ with the dogs   always   that arrives$_{[3rd.sg]}$ late
       'He gets annoyed with the dogs when he comes late.'

In these sentences, the Experiencer is the event role performed by a masculine NP ("him" /"he"). This same role is realized by an accusative pronominal clitic in (2a), a dative pronominal clitic in (2b) and as nominative subject in (2c), entailing different types of linking between syntactic constituents and semantic arguments. Furthermore, the Experiencer in each example presents a different status regarding its hierarchization against the other arguments of the sentence, thus affecting the way linking is achieved. (2a) represents direct linking since the highest ranked participant (e.g. the Agent) is linked to the subject while the lowest ranked one (e.g. the Experiencer) is linked to the direct object. Conversely, (2b) represents inverse linking because the highest-ranked participant of the event (e.g. the Experiencer) is not linked to subject (nor direct object) and, in consequence, it is marked dative morphology. The alternating nature of linking allows for these realizations but the fundamental question is whether these different couplings convey different meanings.

The literature regarding object case-marking alternation -with particular emphasis on causative constructions- has proposed that transitivity (Perlmutter, 1978; Comrie, 1976; Gibson and Raposo, 1986) and semantic / pragmatic factors (Hyman and Zimmer, 1976; Dorel, 1980; Reed, 1992) play an important role on the selection of arguments' case codification. As for the 'molestar-class' psych verbs, several authors have tried to theoretically explain which aspects of syntax and semantics motivate different linking patterns (Treviño, 1990; Franco, 1992; Ackerman and Moore, 1999; Honrubia, 2015; Mendivil, 2002; Fernández-Soriano, 1999; Cuervo, 1999; Masullo, 1992). Although there may be important differences among the assumptions that each approach entails, a number of commonalities may be established. Most importantly, in all accounts linking between syntax and semantics plays an essential role in capturing the semantic differences in the event's interpretation, either if linking is explained in terms of mapping the syntactic constituents with their specific Θ roles, proto-roles or generalized-semantic roles. Second, most accounts emphasize the dichotomic interpretation of the 'molestar-class' of event as being [+ causative] / [– stative] when accompanied with an accusative-marked object or [– causative] / [+ stative] when the object bears dative case. Following Van Valin and LaPolla (1997), events described by verbs similar to 'molestar' can be interpreted either as a causative state or as a state, entailing two different types of linking between syntax and semantics, as it may be observed in **Figure 1**. Linking for the accusative-marked verb is direct, while verbs with dative case-marking show inverse linking.
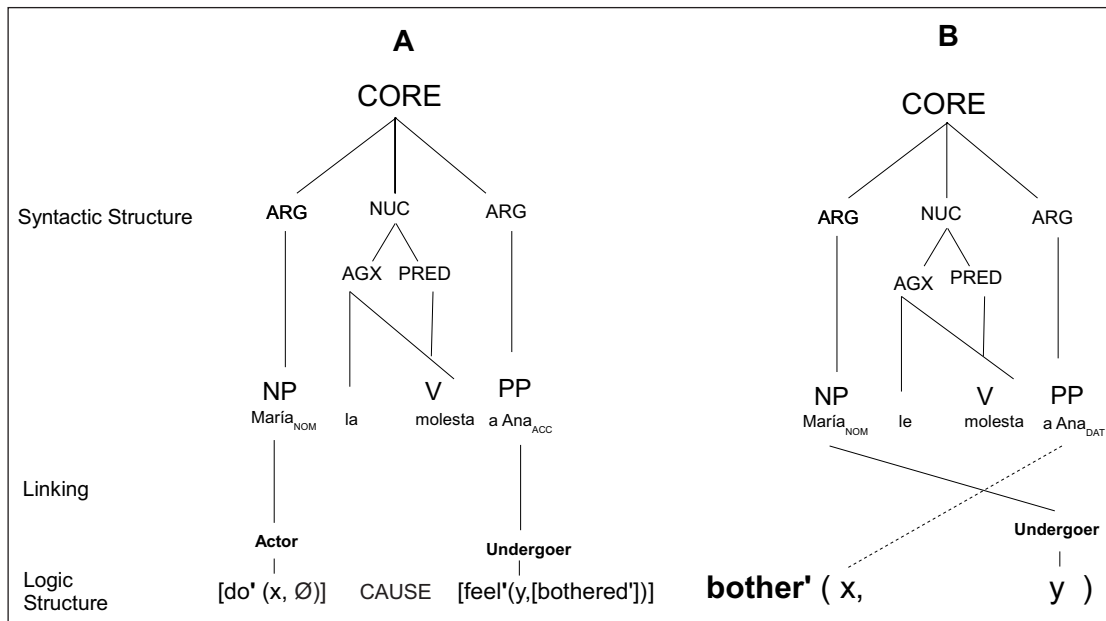
Figure 1: An illustration of the relation between syntactic templates and arguments' linking to generalized semantic roles for sentences with the same verb but different case-marking encoding (accusative vs dative) according to Role and Reference Grammar. Both sentences show the same syntactic template, but differ on the type of linking between constituents and the arguments of the lexico-semantic structure of the verb. While in sentence A the nominative-marked constituent is linked to the Actor (Agent) of a Causative State, in sentence B, it is linked to the Undergoer (Experiencer) of a State. Notice that the dative argument is directly associated to the lexico-semantic structure of the verb. CORE: the minimal phrase structure domain of the clause within which the nucleus and its arguments are realized; ARG: obligatory argument (noun phrase [NP] or prepositional phrase [PP]); NUC: nucleus (essentially equivalent to head, i.e., the element subcategorizing for the arguments); V: verb; AGX: agreement index (i.e. dependent of the NUCLEUS, it receives the agreement specifications of all core argument positions present in the Argument Structure); ACC: Accusative; NOM: Nominative; DAT: Dative.

In the following section, we attempt to explain the consequences of such characterization of the 'molestar'-type of verbs for incremental sentence comprehension

The current study: hypotheses and predictions

Recall that the purpose of the current study is to provide further evidence about the role of linking type (i.e. direct or inverse) for sentence comprehension. Previous studies about Spanish sentence comprehension (and also about Italian and German) resorted to the use of distinct verb classes (e.g. ObjExp psych verbs vs activity verbs Bornkessel et al., 2003; Dröge et al., 2014; Gattei et al., 2015b; a; 2017). Thus, we conducted an eye-tracking experiment with sentences that included the same subclass of psych verbs (i.e. 'molestar' type) but which differed in their objects' case marking. As it was explained in the previous section, a key feature of this subclass

of verbs is that it accepts both accusative and dative case marking, giving rise to distinct types of events, namely a causative psych state with direct linking or a state with inverse linking respectively. Take for instance sentences in **Table 1**.

Sentences (a) and (b) are SVO sentences, while sentences (c) and (d) are OVS sentences. Word order in these examples is marked via the case particle 'a', which is obligatory for dative-marked objects and most animate accusative-marked objects (Leonetti, 2004; Fábregas, 2013). Furthermore, these sentences use the periphrastic future construction in Spanish, consisting of the inflected verb 'va' ('go'), which takes the agreement features of the sentence, the particle 'a' and the non-finite main verb. The use of this construction allows for clitic postverbal attachment 'la/lo' or 'le' for accusative or dative respectively), only possible in verbal phrases using infinitive and gerund forms, and imperative affirmative commands.

The correlate of postverbal clitic attachment is that, during incremental reading, the correct interpretation of arguments only takes place once verb is read. Hence, with the use of this structure it is possible to test not only differences related to linking type (direct versus inverse), but also online argument hierarchization.

If linking type indeed leads to differential processing cost, we should see a greater difficulty to comprehend sentences with dative case marking (i.e. sentences (b) and (d)) than those ones with accusative case marking (i.e. sentences (a) and (c)). If differences previously found between sentences with Activity verbs and sentences with ObjExp psych verbs were related to differences in the verbs' semantic class, we should not see any case effect. Following previous evidence in this language, we expect that difficulty is reflected as lower accuracy rates and longer response times in the comprehension task.

As for possible linking effects during incremental comprehension, we expect any differences related to case marking to be seen at the verb or later regions in the sentence, once readers can disambiguate case marking. More specifically, we expect a modulation of late eye-movement measures, which are considered to reflect processing stages related to semantic and syntactic integration (see Clifton et al., 2007; Vasishth et al., 2013: for a review on this discussion).

Regarding argument incremental interpretation and hierarchization, the outcome of previous studies using the eye-tracking method shows that readers spend further time at the verb and later regions and regress more to initial regions of the sentence when the initial argument of the sentence is less prominent than the second argument of the sentence (Gattei et al., 2017; 2021). In the current experiment, this effect should be seen in sentences that require a reversal of thematic role structure (i.e. sentences (b) and (c) in **Table 1**). Since the effects expected involve the integration of syntactic and semantic information, we anticipate an overall modulation of late eye-movement measures, in contrast to early eye-movement measures, usually related to the integration of lexical information. **Table 2** shows the outcome of previous studies, expected

	Condition	Word Order	Case	Linking Type	Syntax – semantics alignment
a.	María va a molestarla a Ana. María.3rd.sg goes.3rd.sg to bother.infin.ACC.3rd.sg.fem to Ana.ACC.3rd.sg.fem. *María is going to bother Ana.*	SVO	ACC	Direct	S >> O Agent >> Experiencer
b.	María va a molestarle a Ana. María.3rd.sg goes.3rd.sg to bother.infin.DAT.3rd.sg. to Ana.DAT.3rd.sg.fem. *Ana is going to feel bothered by María.*	SVO	DAT	Inverse	S >> O Theme << Experiencer
c.	A María va a molestarla Ana. To María.3rd.sg goes.3rd.sg to bother.infin.ACC.3rd.sg.fem Ana.ACC.3rd.sg.fem. *Ana is going to bother María.*	OVS	ACC	Direct	O << S Experiencer << Agent
d.	A María va a molestarle Ana. To María.3rd.sg goes.3rd.sg to bother.infin.DAT.3rd.sg Ana.DAT.3rd.sg.fem. *María is going to feel bothered by Ana.*	OVS	DAT	Inverse	O << S Experiencer >> Theme

Table 1: Critical sentences used in the current eye-tracking study. Conditions differ in Word Order and Case, entailing different linking types between the Experiencer and the syntactic constituents of the sentences and different semantic argument structures. SVO = Subject-Verb-Object, OVS = Object-Verb-Subject; ACC = Accusative; DAT = Dative; >> = 'higher than'; << = 'lower than'.

Effect type	Effects found in previous studies in Spanish	Measures affected in previous studies in Spanish		Regions affected in previous studies in Spanish	Effects predicted for current study	Results found in current study		
		Overall comprehension	*Incremental reading*			*Overall Comprehension*	*Incremental Reading*	
							Early eye-movement measures	Late eye-movement measures
Argument structure Reinterpretation	Interaction between Verb Type and Word Order	Question accuracy (Gattei et al., 2015a; 2021) Total Reading Time (Gattei et al., 2021)	Reading Time (SPR Gattei et al., 2015a) Late eye-movement measures (Gattei et al., 2017; 2021) Early and Late eye-movement measures (Gattei et al., 2017; 2021)	NP after verb and spill-over regions Regions prior to the verb (TIR and TFT) Verb (RPD, TFT, TIR) NP after veb (RPD; RBRC,TFT, TIR) Spill-over regions (FFD, FPRT, RBRT, RPD, RBRC, TFT)	Interaction between Case x Word Order (SVO DAT >>SVO ACC OVS ACC >>OVS DAT)	Accuracy SVO DAT >>OVS DAT OVS ACC >>OVS DAT Response Time No interaction Total Reading Time SVO DAT >>SVO ACC SVO DAT >>OVS DAT OVS ACC = OVS DAT	No interaction	Dissambiguating verb (TIR) NP after verb (RPD, TFT)

Effect type	Effects found in previous studies in Spanish	Measures affected in previous studies in Spanish		Regions affected in previous studies in Spanish	Effects predicted for current study	Results found in current study		
		Overall comprehension	*Incremental reading*			*Overall Comprehension*	*Incremental Reading*	
							Early eye-movement measures	Late eye-movement measures
Linking effect	Verb Type effect (ObjExp > Act)	Question accuracy (Gattei et al., 2015b; 2017) Response Time (Gattei et al., 2015b; 2017; 2021)	Reading Time (SPR Gattei et al., 2015a) Late eye-movement measures Early eye-movement measures	NP after verb NP after verb (TIR) NP after verb (FPRT, RBRT)	Case effect (DAT > >ACC)	Accuracy No Case effect Response Time No Case Effect	No Case effect	Dissambiguating verb (TIR) NP after verb (RPD, TFT, TIR)

Table 2: Results of previous studies that analysed linking and argument interpretation in Spanish during incremental comprehension and predictions and results of the current study. SVO = Subject-Verb-Object; OVS = Object-Verb-Subject; ObjExp = Object Experiencer psych verbs; Act = Activity verbs; SPR = Self-paced reading; TIR = Total Incoming Regressions; FPRT = First Pass Reading Time; RBRT = Right-Bounded Reading Time; RPD = Regression Path Duration; TFT = Total Fixation Time; RBRC = Right-bounded Regression Count.

effects and results for the current experiment. In the following section, we present an eye tracking experiment that will test these hypotheses out.

Methods

Participants

Sample size was calculated using the 'simr' R package (Green and MacLeod, 2016). Details of sample size estimation may be found at Appendix D. The power curve function of the package estimated that, considering the number of items we counted on, a sample of 50 participants would suffice to encounter effects of similar size (0.25) and similar power (70%; 95% CI: 55.39, 82.14) as those found for the interaction between Verb Type and Word Order in the accuracy model of Gattei et al. (2017) in a new group of participants. Following this recommendation, fifty-eight native Spanish speakers (33 women; age range 18 to 36 years; $M = 24.4$, $SE = .67$) participated in the current sentence reading experiment. All participants had normal or corrected-to-normal vision and had no history of prior neurological disease, drug or alcohol abuse, psychiatric disorders, developmental speech/language disorders, or learning disabilities. All of them provided written consent prior to the study. Eye movement data from 52 participants was entered for the final analysis. Data from five of the remaining participants was discarded due to software issues (N = 1), errors of the experimenter (N = 2) or calibration problems (N = 2). The data from one participant was discarded because after the experiment was completed she acknowledged to have been treated for dyslexia when she was a child.

Stimuli

The experimental sentences for this study were constructed on the basis of a list of 20 Spanish state psych verbs that allowed both accusative and dative case marking. Verbs were selected according to the Role and Reference Grammar characterization of state and causative state predicates (Van Valin and LaPolla, 1997; Van Valin Jr, 2005). Stimuli followed a 2 × 2 factorial design that manipulated Case (Accusative vs Dative) and Word Order (SVO vs. OVS).

For verb selection, we followed strict criteria of inclusion: First, we found around 60 verbs that belonged to the subclass of psych verbs here studied. We made sure that all verbs could be used with all accusative, dative and reflexive case markings (as suggested by most literature regarding this subclass of verbs), by conducting a thorough search of these constructions in "Corpus del Español", created by Davies (2002). The search included the appearance of each verb with singular and plural third person clitics, feminine and masculine clitics (in case of accusative clitics) and the reflexive clitic. Regular expressions were used in order to account for all occurrences in which the clitic was previous to the verb (e.g. "le molesta", "les molestó") or attached to it (e.g. "molestarle", "molestándole"). This search resulted in a total of 55 verbs.

Second, we considered aspectual characteristics of these verbs. For instance, Marín and McNally (2011) and Marín (2014) analyses of reflexive Spanish psych verbs propose different aspectual tests that could separate this class of psych verbs in two subgroups: verbs that denote a change of state with a very punctual beginning boundary (such as "molestarse", *'to feel bothered'*), and verbs with a non-punctual beginning boundary (like "aburrirse", *'to feel bored'*). We made sure that all verbs could pass all tests proposed by these authors and that half of them belonged to the first subclass and that half belonged to the second subclass. The rationale behind this decision was to minimize the preference for a dative or an accusative reading, since we had the intuition that the "molestarse" subclass of verbs was usually better combined with accusative case marking and verbs like "aburrirse" showed a preference for dative case marking. This step reduced the list of verbs to almost half. The final steps were excluding verbs that are not used in Argentina (e.g. "enfadarse", *'to get angry'*), and to use LEXESP to match frequency values of both "molestarse" and "aburrirse" subclass of verbs, in order to make sure that, although participants would read each verb once, there were no additional effects of aspect that would lead to our results.

Verbs were framed between two proper names, counterbalanced in gender. In SVO sentences, the second proper name was preceded by preposition 'a', thus giving rise to the sentence object, and in OVS sentences the preposition preceded the first proper name. In each sentence, the gender of the participants was kept constant (they were either both feminine or both masculine) since accusative clitics inflect in gender and we wanted to avoid that arguments' interpretation could be inferred from the clitic's gender.

Differently to Gattei et al. (2017), the verbs chosen were not inflected in the present tense but in the periphrastic future tense, formed by the verb 'go' in the present tense, the particle 'a' and the main verb in the infinitive form. This construction is similar to the 'going to + -ing' form in English and is the most used form of future tense in Buenos Aires Spanish. This form allows the attachment of case clitics at the end of the infinitive verb. Consequently, the type of event denoted by the verb (a causative psych state or a psych state) is only revealed once the verb and the clitic are encountered.

Additionally, an adverbial adjunct was included at the beginning of the sentence so that during the eye-tracking study the initial eye-fixations did not occur at the first word of the initial argument (the name in SVO conditions and the preposition 'a' in OVS conditions (see Kliegl et al., 2006: for a discussion on this issue). The 80 resulting sentences were divided into four lists of 20 sentences each (five sentences per condition). Participants saw each verb only once.

Twenty prepositional or adverbial phrases were added at the end of each set of sentences in order to avoid the critical region including 'wrap-up effects' during the eye-tracking study (Just et al., 1982). These phrases could modify both of the NPs included in the sentence and they were semantically neutral so that they did not facilitate a semantic interpretation of the event over the

other. **Table 3** shows the experimental conditions used in this experiment and subsequently used for the eye-tracking study.

Condition	Example	Comprehension Question
(a) ACC SVO	Indefectiblemente, María va a molestar**la** a Ana durante el viaje.	
(b) DAT SVO	Indefectiblemente, María va a molestar**le** a Ana durante el viaje.	
	Inevitably, María is going to bother Ana during the trip.	¿María se va a molestar?
(c) ACC OVS	Indefectiblemente, a María va a molestar**la** Ana durante el viaje.	*Is María going to feel bothered?*
(d) DAT OVS	Indefectiblemente, a María va a molestar**le** Ana durante el viaje.	
	Inevitably, Ana is going to bother María during the trip.	

Table 3: Critical sentences used in the current eyetracking experiment. ACC = Accusative case; DAT = Dative case; SVO = Subject-Verb-Object OVS = Object-Verb-Subject. Bold letters show the object clitics attached to the infinitive verb in their accusative ('la') or dative ('le') form.

Sixty filler sentences with different semantic and syntactic complexity were added so that participants did not notice the main purpose of the study. Finally, a set of three practice trials were included at the beginning of the experiment, and one question for each practice item, critical trial and filler sentence was prepared to test comprehension. Questions were formulated in such way that participants had to judge whether it correctly described the content of the preceding experimental sentence or not. Half of the questions required the answer "yes" and half of them required the answer "no". In the case of critical items, half of the questions involved the subject of the sentence and half of them asked about the object. A complete list of the experiment materials may be found at Appendix A of the Supplementary File.

Equipment

Participants were seated in front of a 19-inch screen (Samsung SyncMaster 997MB, 1024 × 768 pixels resolution, 100 Hz refresh rate) at a viewing distance of 65 cm. A chinrest aligned with the center of the screen prevented head movements. An EyeLink 1000 eye-tracker (SR Research Ltd.) was used to record gaze locations of both eyes during reading at a sampling rate of 1 kHz. Nominal average accuracy was 0.5° and space resolution was 0.01° root mean square (RMS), as given by the manufacturer. The participant's gaze was calibrated with a standard 13-point grid

for both eyes. All recordings and calibration were binocular. Only left eye data were used for the analysis.

All eye movements were labelled as fixations, saccades and blinks by the eye-tracker software using the default thresholds for cognitive experiments (30°/sec for velocity, 8000°/sec for acceleration, and 0.1° for motion) (Cornelissen et al., 2002). Stimuli presentation was developed using Matlab (http://www.mathworks.com/, Massachusetts, United States) and Psychophysics Toolbox Version 3 .

Procedure

All sentences were displayed on a single line and were presented in Courier New Bold font. At a distance of 65 cm, each letter subtended 0.44° of visual angle laterally. Subjects were instructed to read the sentences at their own rate. No instructions were given to suppress eye blinks. Before the eye-tracking experiment began, they had a practice session of three sentences. Then, trials were presented in randomized order. At the beginning of each trial, a dot appeared at the left edge of the screen and after participants fixated on this dot, the sentence appeared. The first letter of the sentence was located at the position of the dot. Participants were instructed to look at a second dot at the bottom right corner of the screen to indicate they had finished reading. The total reading time of each trial was measured starting from when participants triggered the appearance of the sentence by fixating on the left dot until they fixated on the bottom right dot and the sentence disappeared.

Comprehension questions appeared after every critical stimulus and filler sentence. Participants responded by mouse-clicking on one of two possible answers displayed horizontally. Half of the times, the correct answer was positioned over the left. The order in which it appeared at this position was randomized. Response time was measured starting from the appearance of the question until participants clicked on one of the possible responses. A calibration procedure was performed at the beginning of the eye-tracking experiment. Experimental sessions lasted approximately 30 minutes.

Data analysis

Eye movement data was screened for blinks and track losses. Fixations shorter than 50 ms and longer than 1000 ms were removed from the analysis. After this screening process, fixations were assigned to their respective word. Eye-tracking measures were computed using em2 package for R language for statistical computing (Logacev and Vasishth, 2013: version 3.0.2).

For the purpose of analysis, we divided the sentences into ten regions that consisted of the first ten words of each sentence, as shown in **Table 4**. Note that in order to facilitate statistical analysis and visual presentation of the results, we aligned the critical regions that comprised

the proper names (regions 3 and 8), and the verb phrase (regions 4 to 6). The region of the preposition has been labeled as (7) in subject-initial sentences, and (2) in object-initial sentences. The regions 'PP1' and 'PP2' (regions 9 and 10) correspond to the first and the second word of the prepositional phrase following the second noun phrase.

	1	2	3	4	5	6	7	8	9	10
SVO	Indefectiblemente,		María	va	a	molestarla|molestarle	a	Ana	PP1	PP2
OVS	Indefectiblemente,	a	María	va	a	molestarla|molestarle		Ana	PP1	PP2

Table 4: Regions of interest used for the statistical analysis of the current eye-tracking experiment according to Word Order. SVO = Subject-Verb-Object; OVS = Object-Verb-Subject; PP1 = First word of the Prepositional Phrase; PP2 = Second word of the Prepositional Phrase.

For each fixated word, we computed the following measures: (1) *First Fixation Duration* (FFD; the duration of the first fixation on the word); (2) *First Pass Reading Time* (FPRT; the sum of all fixation durations on the word before any other word was fixated); (3) *Right-Bounded Reading Time* (RBRT; the sum of all first-pass fixation durations on the word before another word to the right is fixated); (4) *Regression Path Duration* (RPD; also known as *go-past time,* it is the sum of all first-pass fixation durations on the word and all preceding words in the time period between the first fixation on the word up to the point where the reader leaves the critical region with a progressive saccade; (5) *Right-Bounded Regression Count* (RBRC; the number of regressions from the word before any word further to the right has been fixated); (6) *Total Fixation Time* (TFT; the sum of all fixations durations on a word); and (7) *Total Incoming Regressions* (TIR; the number of regressions to a specific word). Measures 1–3 are typically considered early measures, whereas measures 4–7 are considered late. Since most predictions are focused on the modulation of late eye-movement measures and for the purpose of improving reading flow, we only report the results of late eye-movement measures, but present all statistical analysis as part of Appendix C.

The data analysis was conducted in the R programming environment (R Core Team, 2021). For measures comprising reading or response time (i.e. Comprehension Task Response Time, FFD, FPRT, RBRT, and TFT) a linear mixed-effects model was used. Linear mixed-effects models are available in the package *lme4* (Bates et al., 2014; Pinheiro and Bates, 2000). For the accuracy measure, the data was fit to a generalized linear mixed-effects model with a binomial function. This function is adequate for analysing data measured on a dichotomous scale, namely "Correct" and "Incorrect" responses. Count data (RBRC and TIR), on the other hand, was analysed with a generalized mixed-effects model with Poisson link function, which is appropriate for counts of events in a fixed time window (Baayen, 2008: p. 322).

Case and Word Order were used as fixed effects and Subject, and Item as random effects for the models. Log Frequency and inverse length of each word were included as control factors

in every region except for regions 2, 4, 5 and 7 (preposition 'a' and verb 'va' which were equal for all conditions and sentences), the rationale being that these two variables may explain a significant part of the variability in reading times and amount of fixations on these regions Kliegl et al. (2004); Just and Carpenter (1980); Rayner and Well (1996). Data of names' approximate frequency was collected by conducting an advanced Google search in the domain of the Facebook website from Argentina.

A maximal random-effects structure was included whenever it was possible in both LMMs and GLMMs, as linear mixed-effects models that do not consider random intercepts and slopes involve the risk of Type I error inflation (Barr et al., 2013). When models either did not converge or the correlation between variance components could not be estimated, the random effects structure was simplified by removing the correlations. Family-wise error correction was applied for comparison among regions only. As suggested by (Von der Malsburg and Angele, 2017), we considered that regions' independence from each other made comparisons between multiple regions prone to produce Type I errors, so we applied Bonferroni correction and lowered the alpha threshold to $0.05/10 = 0.005$. However, we decided to not correct for multiple reading measures, in contrast with the suggestions by Von der Malsburg and Angele (2017) and considered worth discussing any effects or interaction that would reach significance in at least two of the aforementioned measures. As stated by these authors, the probability of finding false positives with the two-measures method is slightly lower than for the Bonferroni correction. Note, however, that this may render our results less conservative from the point of view of the article by von der Malsburg & Angele (2017). For all the models presented in the study, covariates that involved reading times were scaled and centred.

For measures regarding the comprehension task (percentage of correct answers, reaction times and total reading times), residuals (i.e. deviations of the observations from the sample mean) exceeding 2.5 standard deviations were eliminated, in order to avoid that extreme responses would affect the results. In all cases, model comparison procedures showed that trimming improved the model at a level of significance of $\alpha = .05$. In the comprehension task, the amount of removed data points was 18 in the accuracy analysis, 21 in the RT analysis and 22 in the total reading time analysis. This was equivalent to 1.77%, 2.06% and 2.16% of the critical items data respectively.

Finally, we used an orthogonal contrast coding to test the interaction between case and word order. For the case contrast, sentences with accusative case were coded as –1 and sentences with dative case were coded as 1. For the word order contrast, SVO sentences were coded as –1 and OVS sentences were coded as 1. Whenever we found a significant interaction between these factors, we used Tukey's multiple comparisons' test. Relevant pairwise comparisons for the hypotheses tested were those involving the same word order (e.g. SVO.DAT – SVO.ACC and OVS. DAT – OVS.ACC), which would show a preference for a specific case marking in each word order,

or those involving the same case marking (e.g. OVS.ACC – SVO.ACC and SVO.DAT – OVS.DAT), which would show a preferred word order for a specific case marking.

Results

Comprehension Task
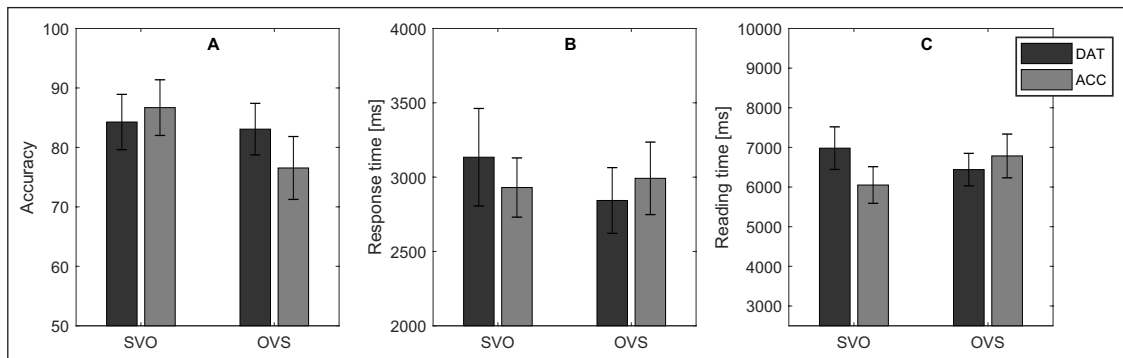
Question accuracy

Mean accuracy for all comprehension questions was 91% ($\pm$ .44%). This indicates that participants were paying attention to the content of the sentences. Mean accuracy of critical sentences was 82% ($\pm$ 1%). **Figure 2A** shows mean accuracy according to condition. Differences in accuracy according to case and word order were analysed with a generalized linear mixed-effects model. The analysis revealed a significant effect of word order. On average, participants responded questions about SVO sentences significantly better ($M = 85\%$, $SE = 2\%$) than questions about OVS sentences ($M = 80\%$, $SE = 2\%$, $\beta = -.278$, $SE = .094$, $t = -2.974$, $p < .01$). Besides, a significant interaction between case and word order was found; $\beta = .229$, $SE = .094$, $z = 2.446$, $p < .05$. Resolving this interaction revealed that when reading sentences in OVS word order, participants responded significantly better when the main verb included a dative clitic than when it included an accusative clitic, $\beta = .702$, $SE = .243$, $z = 2.882$, $p < .05$. This difference was not significant for SVO conditions, $p = .98$ However, when comparing conditions with a verb bearing accusative case marking, participants responded significantly better when the sentence followed the SVO word order than when it followed the OVS one, $\beta = 1.013$, $SE = .162$, $z = 3.861$, $p < .001$.

Response times

Total mean response time (RT) for critical items was 2974 ms ($\pm$ 62 ms). **Figure 2B** shows mean RTs according to condition. Analysis of differences in RTs between case and word order revealed no effects of case or word order, nor significant interaction between case and word order.

Total reading time

Mean total reading time was 6565 ms ($SE = 123$ ms). **Figure 2C** shows the average total reading time for the critical sentences used in the current eye-tracking experiment. The statistical analysis revealed an interaction between case and word order; $\beta = -0.045$, $SE = 0.011$, $t = -4.096$. Resolving this interaction showed that participants spent significantly longer time reading SVO sentences when the verb included a dative clitic than when they included an accusative clitic; $\beta = 0.126$, $SE = 0.034$, $z = 4.025$, $p < 0.001$. Besides, participants spent significantly longer time reading sentences with a dative-marked verb when they followed the SVO word order than when they followed the OVS order; $\beta = 0.112$, $SE = 0.031$, $z = 3.592$, $p < 0.01$. No significant differences between OVS conditions were found, $p = .29$.

Figure 2: Percentage of accurate answers (panel A), mean response times for the comprehension question (panel B, and mean total reading time for the critical sentences (panel C) in the current eye-tracking study according to case marking (Accusative vs. Dative) and word order (SVO vs. OVS). Error bars correspond to 95 % Confidence Intervals. ACC = Accusative-marked verb; DAT = Dative-marked verb; SVO = Subject-Verb-Object; OVS = Object-Verb-Subject.

Eye-tracking measures

Following Kliegl et al. (2006), we excluded the first word of all sentences for all reading duration analysis since the first word was positioned at the location of the fixation spot that triggered the appearance of the sentence. We also refrained from inspecting most reading measures for Region 2 (case marking preposition 'a') since the model could not be fit with any of the fixed variables. On the one hand, Region 2 only comprises sentences with OVS word order. On the other hand, case marking is only relevant for participants once they have read the clitic on Region 6. Thus, including these variables or the interaction between them did not make sense when fitting the model, and as it was explained at the data analysis section, control variables (i.e. frequency, and length) were kept fixed both at this region and Region 7. We thus inspected the reading measures FFD, FPRT, RBRT, RBRC and RPD from Region 3 onwards. However, at Region 1 and 2, we did analyse the probability of regressing to the first and second words (TIR) from other regions and the total fixation time.

Figure 3 summarizes the contrast between sentences with accusative case (causative psych states) and dative case (psych states) according to both word orders (SVO and OVS) for late eye-movement measures. Positive values mean that reading time and regression counts are higher for sentences with dative-marked verbs than for sentences with accusative-marked verbs. A positive black line (SVO conditions) and a negative grey dashed line (OVS conditions) correspond to an interaction between Case and Word Order. This representation makes the interaction visually clear. The asterisks show the regions where the interaction was significant. **Table 5** shows mean values and standard errors for each condition at all regions and measures in which case marking yielded a significant difference.
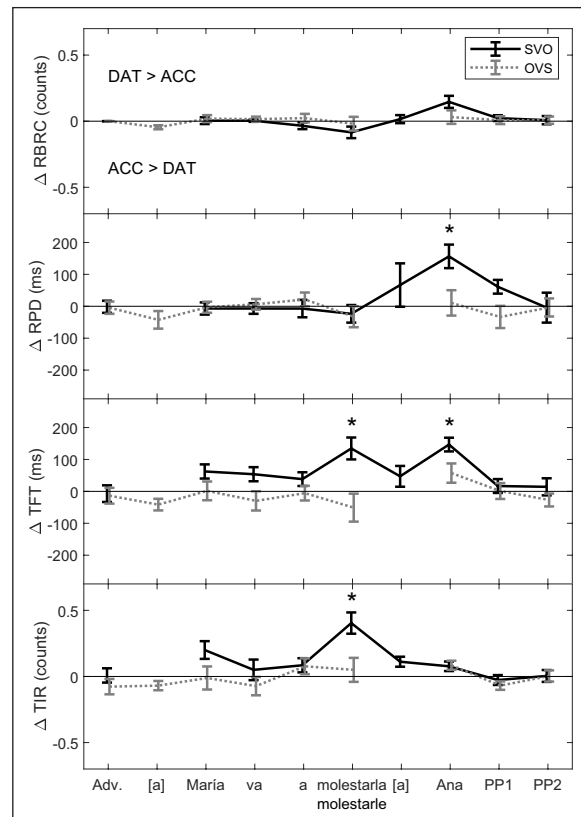
Figure 3: The Figure shows the difference (Δ) in mean fixation times (ms) and the amount of regressive saccades (counts) between conditions with dative and conditions with accusative case marking according to the sentence word order. Error bars correspond to Standard Error of the Mean. Eye-tracking measures: TFT = Total Fixation Time; RPD = Regression Path Duration; RBRC = Right-Bounded Regression Count; TIR = Total Incoming Regressions. Word Order: SVO = Subject-Verb-Object; OVS = Object-Verb-Subject. The asterisk shows that the interaction between Word Order and Verb Type was significant.

	TIR				TFT			
	Region 6		Region 8		Region 6		Region 8	
Condition	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
SVO.ACC	0.665	0.057	0.15	0.025	957	40	462	22
OVS.ACC	0.792	0.063	0.181	0.028	1080	46	539	27
SVO.DAT	1.069	0.068	0.227	0.03	1227	52	623	30
OVS.DAT	0.842	0.668	0.265	0.031	1099	54	597	31

Table 5: Mean values and standard error for each condition for all regions and measures in which case marking yielded a significant difference. TIR = Total Incoming Regressions; TFT = Total Fixation Time.

We now provide the analysis of regions of interest for late eye-movement measures mentioned in the 'Data analysis' section. For each region, we first present the analysis of the interaction between Case and Word Order, since it is the contrast of interest of the current study. We then provide the relevant results of the multiple comparisons test whenever was needed. Finally, we report main effects of Case and Word Order. Appendix B shows the final converging models for each measure at each region. A table with all statistical coefficients including control factors (Length and Frequency), standard errors and t and p values for all measures, and the figures that depict the interaction between variables for early eye-movement measures may be found at Appendix C.

Regions 1 (Adverb) and 2 (Case marking preposition 'a')

Analysis of the probability of regressions into these regions ('TIR') and total fixation time ('TFT') revealed no interaction between case and word order.

Region 3 (First proper name)

Analysis of this region showed that TIR was affected by word order, showing a significantly higher probability of incoming regressions for object-marked constituents than for subject-marked constituents (OVS: $M = 0.95$, $SE = 0.05$; SVO: $M = 0.65$, $SE = 0.04$, $\beta = 0.198$, $SE = 0.041$, $z = 4.791$, $p < 0.0001$) .

Word order also modulated Total Fixation Time ('TFT'). Participants fixated for significantly longer time at this region when the sentence began with the object-marked constituent than when it began with the subject constituent (OVS: $M = 594$ ms, $SE = 21$ ms; SVO: $M = 482$ ms, $SE = 17$ ms; $\beta = 0.086$, $SE = 0.020$, $t = 4.361$, $p < 0.0001$).

Region 4 (Auxiliary verb 'ir')

A main effect of word order was found for TFT measure, as well as for Regression Path Duration ('RPD'). Participants took significantly longer to read this region and regressed for significantly longer time before continuing reading when they were reading an object-initial sentence (TFT: $M = 424$ ms, $SE = 17$ ms ; RPD: $M = 248$, ms, $SE = 12$ ms) than when they were reading a subject-initial sentence, TFT: $M = 355$ ms, $SE = 14$; $\beta = 0.074$, $SE = 0.023$, $t = 3.212$, $p < 0.005$; RPD: $M = 200$ ms, $SE = 6$; $\beta = 0.091$, $SE = 0.030$, $t = 3.016$, $p = 0.005$.

Region 5 (Auxiliary particle 'a')

Analysis of this region revealed no interaction between case and word order nor main effects of any of these two factors.

Region 6 (Disambiguating verb)

Analysis of this region showed a significant interaction between case and word order for TFT and TIR (TFT: $\beta = -0.064$, $SE = 0.017$, $t = -3.866$, $p = 0.0001$; TIR: $\beta = -0.109$, $SE = 0.034$, $z = -3.192$, $p = 0.001$). Post-hoc analyses of these measures revealed that, participants fixated for significantly longer time on this region and regressed significantly more times to this region in subject-initial sentences with a dative-marked clitic than in the other three conditions (TFT: SVO. DAT-SVO.ACC: $\beta = 0.185$, $SE = 0.068$, $z = 2.707$, $p < 0.05$; SVO.DAT-OVS.DAT: $\beta = 0.165$, $SE = 0.047$, $z = 3.486$, $p < 0.005$; SVO.DAT-OVS.ACC: $\beta = 0.185$, $SE = 0.068$, $z = 2.707$, $p < 0.05$; TIR: SVO.DAT-SVO.ACC: $\beta = 0.500$, $SE = 0.123$, $z = 4.059$, $p < 0.001$; SVO.DAT-OVS. DAT: $\beta = 0.259$, $SE = 0.090$, $z = 2.881$, $p < 0.05$; SVO.DAT-OVS.ACC: $\beta = 0.324$, $SE = 0.120$, $z = 2.713$, $p < 0.05$). The probability of regressing to this region was also higher for sentences with a dative-marked verb ($M = 0.96$, $SE = 0.05$) than for sentences with an accusative-marked verb ($M = 0.73$, $SE = 0.04$; $\beta = 0.141$, $SE = 0.051$, $z = 2.749$, $p = 0.005$).

Region 7 (Case marking preposition 'a')

Analysis of this region showed no significant interaction between case and word order nor main effects of any of these factors.

Region 8 (Second proper name)

Analysis of this region showed an interaction between case and word order for RPD and TFT (RPD: $\beta = -0.067$, $SE = 0.024$, $t = -2.766$ $p = 0.005$; TFT: $\beta = -0.055$; $SE = 0.018$; $t = -3.034$, $p < 0.005$). Resolving these interactions revealed that the regression path duration was significantly longer for SVO sentences with dative-marked verbs than for the other three conditions (SVO.DAT-SVO.ACC: $\beta = 0.300$, $SE = 0.068$, $z = 4.360$, $p < 0.001$; SVO.DAT-OVS. DAT: $\beta = 0.205$, $SE = 0.069$, $z = 2.975$, $p < 0.05$; SVO.DAT-OVS.ACC: $\beta = 0.237$, $SE = 0.069$, $z = 3.458$, $p < 0.05$). Planned comparisons of TFT showed that, in SVO sentences, participants fixated on this region for significantly longer time when the main verb included a dative clitic than when it included an accusative clitic, $\beta = 0.285$, $SE = 0.052$, $z = 5.521$, $p < 0.001$. Although no significant differences were found between OVS conditions, in sentences with an accusative-marked verb participants fixated for significantly longer time on this region when the sentence was object-initial than when it was subject initial, $\beta = 0.160$, $SE = 0.051$, $z = 3.126$, $p < 0.05$. Finally, the analysis also showed a main effect of case for RPD, TFT and TIR. On average, participants spent significantly longer time fixating on this region when the main verb included a dative clitic ($M = 618$ ms, $SE = 22$ ms) than when it included an accusative clitic ($M = 503$ ms, $SE = 18$ ms $\beta = 0.087$, $SE = 0.018$, $t = 4.819$, $p < 0.005$). The probability of regressing to this region (TIR) was also significantly higher for sentences with dative-marked verbs ($M = 0.24$, $SE = 0.02$) than for sentences with accusative-marked verbs ($M = 0.16$, $SE = 0.02$; $\beta = $

0.199, $SE = 0.069$, $z = 2.895$, $p < 0.005$). Likewise, the regression path duration at this region was significantly higher for sentences with dative-marked verbs ($M = 626$ ms, $SE = 34$ ms) than for sentences with accusative-marked verbs ($M = 504$ ms, $SE = 26$ ms; $\beta = 0.083$, $SE = 0.024$, $t = 3.428$, $p < 0.005$).

Regions 9 (First word of the Spill-over region) and 10 (Second word of the Spill-over Region)

Analysis of these regions showed no interaction between case and word order nor main effects.

Discussion

The general purpose of the current study was to shed light on the role of linking for online sentence processing. The experiment conducted involved the use of Spanish sentences that included the same verb but which accepted alternative case marking. The rationale behind the use of these verbs was to test whether keeping semantic content of the verb constant while manipulating case marking would lead to differences in overall sentence interpretation due to differences in linking type. While accusative case marking entailed direct linking, the use of dative case marking conveyed an interpretation where the highest-ranked thematic role (the 'Experiencer') was linked to the lowest-ranked syntactic constituent, the Object. Sentences were also manipulated regarding their word order (SVO vs OVS) in order to evaluate whether argument incremental interpretation played any role during online processing of these sentences as evidenced by previous works on this language. Results of the eyetracking reading task showed that participants regressed significantly more times to regions 6 (dissambiguating verb) and 8 (second NP) when the verb included a dative clitic than when it included an accusative clitic. They also took longer time to read region 8 and spent longer time regressing to previous regions before continuing reading when the sentences entailed an inverse linking type. As shown at **Table 5**, case marking affected reading at both SVO and OVS sentences and, consistently, it modulated sentences' total reading time as well. As for incremental argument interpretation, the eye-tracking measures of the current study showed an interaction between Case and Word Order for some of the late eye-movement measures. In particular, once participants found the main verb, they regressed significantly more to previous regions and fixated for significantly longer time at the verb and second NP when the sentence was in SVO word order and included a dative clitic (sentence (a) in **Table 1** than in the other three conditions. Recall that in previous studies in this language, Gattei et al. (2017; 2021) found that this difficulty took place every time the participants read a sentence where the first argument was the least prominent one, for both SVO and OVS sentences. The implications of the current results will be discussed at the General Discussion section under the light of different psycholinguistic models that address how incremental linking takes place.

Results of the comprehension task showed an effect of word order and an interaction between word order and case marking for accuracy rates, showing higher accuracy for questions about SVO sentences in general and, in the case of OVS sentences, showing better comprehension when the sentence included a dative clitic than when it included an accusative clitic. For accusative-marked sentences, better performance appeared when participants had to respond to questions about SVO sentences than for questions about OVS sentences.

Following (Gattei et al., 2017), we wondered whether results of the comprehension task could be driven by sentences' individual acceptability. Previous studies have shown an inverse correlation between the rate of acceptability associated with a specific sentence and the cognitive cost related to its comprehension, reflected in higher reading times, higher response times and higher error rates in comprehension tasks (Ditman et al., 2007; Wagers et al., 2009). A way to disentangle this possible confound was by running an acceptability judgement task on the sentences used in Experiment 1 and re-analysing the data collected at the comprehension task with acceptability scores as a control factor in the linear models used. We present this procedure in Experiment 2.

Experiment 2: Acceptability judgement task

All stimuli used in the eye-tracking experiment were used in an acceptability judgement task in order to disentangle whether difficulty that arose from the use of accusative or dative clitics in this group of verbs in sentences with different word order was related to overall sentence acceptability.

Hypotheses and Predictions

Following previous work on this issue, we expected that the current acceptability judgement task showed a rating pattern that matched prior evidence found in several comprehension tasks about this type of Spanish sentences (Gattei et al., 2015a; b). Namely, the prediction is that sentences where a reversal of argument prominence status is required should be rated as less acceptable than sentences that do not require such re-ranking. In other words, most acceptable sentences will be those in which the most prominent argument is in first position (sentences (a) and (d) in **Table 3**), whereas the least acceptable ones will be those in which the first position is occupied by the least prominent argument. Thus, we expected SVO sentences with accusative case being more acceptable than those with dative case, and OVS sentences with dative case being more acceptable than their accusative counterparts.

Material and methods
Participants

One hundred and thirty-nine native speakers of Buenos Aires Spanish (100 females) participated in the experiment voluntarily. Subjects ranged in age from 18 to 53 years old (M = 30.16 years

old; *SE* = 0.71). Participants reported no history of prior neurological disease, drug or alcohol abuse, psychiatric disorders, developmental speech/language disorders, nor learning disabilities. All participants had normal or corrected to normal vision. All of them provided written consent prior to the study.

Stimuli

Critical and filler sentences used for Experiment 1 were used. Twenty of the filler sentences were modified so that they would become unacceptable sentences in Spanish, and the syntactic constituents of 20 other filler sentences were scrambled so that they would become acceptable but not very common sentences, the rationale being that acceptability differed among sentences and participants were obliged to think about their response.

Procedure

Participants were tested individually through the Ibex Farm application (developed by Alex Drummond, http://spellout.net/ibexfarm). Trials were presented randomly according to a Latin square design. Thus, each participant saw 20 critical sentences, five sentences per condition. Participants were asked to judge on a 5-point scale how the sentence they had just read sounded to them. They were reminded that their answer should be driven according to whether they found the sentence acceptable (5 points), fairly acceptable (4), not that acceptable (3 points), doubtfully acceptable or unlikely to be acceptable (2 points), or not an acceptable sentence in Spanish (1 point). Although no time limit was set for each of the trials, participants were encouraged to rate sentences as fast as possible and to base their answers on their own intuition. Apart from this, half of the filler sentences included a comprehension question that participants had to respond with 'yes' or 'no' by clicking on the correct answer. The experimental session lasted between 15 to 20 minutes approximately.

Data analysis

Differences in the degree of acceptability of sentences were analysed with a cumulative linear mixed-effect model using the *ordinal* package (Christensen, 2019). This type of analysis is adequate for ordinal variables, in which data consists of ordered, but discrete categories (Veríssimo, 2021). The models included Case and Word Order as fixed factors, and Items and Subjects as random factors. Pairwise comparisons tests were run whenever a significant interaction took place using the *emmeans* (Lenth, 2022) R package.

For statistical reanalysis of the results of the comprehension task, we included mean acceptability rating of each item as a fixed factor in the linear model used for analysis of this task at Experiment 1.

Results

Acceptability Judgement Task

Results showed that mean accuracy for all items was 89%. This means that participants were paying attention to the sentences they were reading. Mean acceptability for all critical items was 2.88 points (SE = .02). **Table 6** shows mean values of acceptability ($\pm$SE) according to condition.

Condition	Mean	*SE*
ACC SVO	3.23	.05
ACC OVS	2.90	.05
DAT SVO	2.66	.05
DAT OVS	2.77	.05

Table 6: Mean acceptability and standard error (*SE*) of sentences used in Experiments 1 and 2. ACC = Accusative; DAT = Dative case; SVO = Subject-Verb-Object OVS = Object-Verb-Subject.

Statistical analysis shows that acceptability ratings were affected by case marking and word order. The outcome of the model shows that sentences with a dative clitic were rated significantly lower than sentences with an accusative clitic, $\beta = -.295$, $SE = .102$, $t = -2.888$, $p = .004$, and subject-initial sentences were rated higher than object-initial ones, $\beta = .64$, $SE = .103$, $t = 6.194$, $p < .001$. Besides, a significant interaction between case and word order was found; $\beta = -.869$, $SE = .015$, $t = -5.935$. Tukey post-hoc multiple comparisons revealed that differences in mean acceptability rating were significant for SVO sentences, in which participants rated sentences with accusative case marking significantly higher than sentences with dative case marking, $\beta = 1.165$, $SE = .011$, $z = 11.012$, $p < .001$. Difference between OVS conditions was also significant, with participants rating sentences with accusative case marking higher than dative-marked sentences, $\beta = .295$, $SE = .102$, $z = 2.888$, $p = .02$.

Question accuracy

Analyses of differences in accuracy according to case marking and word order revealed a main effect of word order; $\beta = -0.265$, $SE = 0.095$, $z = -2.791$, $p = 0.005$. The interaction between case and word order present at the analysis without mean rating as a fixed factor showed a tendency towards significance, $\beta = 0.183$, $SE = 0.10$, $z = 1.829$, $p = 0.06$. Tukey's multiple comparisons test revealed that this interaction followed the same pattern as in the analysis without acceptability rating as a fixed factor. When reading object-initial sentences, participants responded significantly better when the sentence included a dative clitic than when it included an accusative clitic; $\beta = 0.772$, $SE = 0.250$, $z = 3.085$, $p = 0.01$. When comparing accusative-

marked sentences, readers performed better after reading SVO sentences than when reading OVS ones; $\beta = 0.894$, $SE = 0.277$, $z = 3.230$, $p = 0.006$. No significant effects of acceptability rating were found.

Response time

Analyses of differences in response times revealed a tendency towards significance for word order; $\beta = -0.024$, $SE = 0.012$, $z = -1.925$, $p = 0.05$. On average, participants took longer time to respond to questions about SVO sentences ($M = 3036$ ms. $SE = 95$ ms.) than to respond questions about OVS sentences ($M = 2918$ ms. $SE = 82$ ms.) No main effects of case, mean rating nor significant interactions between fixed factors were found.

Total reading time

Analyses of differences in total reading time revealed a main effect of acceptability rating, showing that participants took longer time to read less acceptable sentences; $\beta = -0.126$, $SE = 0.062$, $z = -2.011$, $p = 0.04$. The interaction between word order and case marking was also significant; $\beta = -0.034$, $SE = 0.012$, $z = -2.845$, $p = 0.004$. Tukey's multiple comparisons test revealed a significant difference in reading time between dative-marked conditions, with SVO sentences conveying longer reading time than OVS sentences; $\beta = 0.099$, $SE = 0.032$, $z = 3.099$, $p = 0.01$.

Discussion

Experiment 2 was included in order to test whether there were significant differences in the acceptability of sentences that required a reversal of thematic arguments, and whether the results from the comprehension task of Experiment 1 mainly depended on sentence acceptability. Results of the current acceptability judgement task revealed that even in sentences that contain the same verb but differ in case marking, participants penalized those sentences in which the arguments' linear word order displayed an argument with less prominent status in first position. As a consequence, participants considered these sentences less acceptable. This result was particularly seen when SVO sentences were compared (participants preferred accusative case marking), and between accusative-marked sentences (participants preferred the SVO word order). Thus, data from the comprehension task was reanalysed by using mean acceptability rating of the sentences as a fixed factor in the statistical models used to analyze the results of Experiment 1. Results of this reanalysis task confirmed that the main effect of word order and the interaction between case marking and word order persisted for accuracy rates, with a small decrease in the statistical significance value of the interaction and the interactions' multiple comparisons tests. Mean acceptability was significant when considering total reading time as a dependent variable, showing longer reading time for sentences with lower acceptability ratings.

The interaction between word order and case marking still persisted for total reading time, but the multiple comparisons test revealed that the difference between SVO conditions disappeared and the difference between dative-marked conditions was still significant. Thus, results of this experiment show that even when sentence acceptability could modulate overall reading behavior -as seen in total reading time-, the effects of presenting a word order that introduces the lowest-ranked argument in the thematic hierarchy in first position still persist.

General discussion

The general purpose of the current study was to shed light on how linking is achieved during online processing. Substantial evidence from studies run in German (Bornkessel et al., 2005; Bornkessel-Schlesewsky and Schlesewsky, 2009; Bader and Bayer, 2006; Bornkessel et al., 2003; 2002; Frisch and Schlesewsky, 2001), English (Weckerly and Kutas, 1999), Chinese (Wang et al., 2009; 2012), and Italian (Droge et al., 2014) has shown that the parser does not wait until verbal information is available in order to make predictions about the argument structure of the sentence. These predictions are based on morphosyntactic (i.e. case morphology, word order) and semantic cues (i.e. animacy, definiteness) provided by the arguments of the sentence. Cross-linguistic evidence also suggests that the grade of relevance that each cue has for triggering these predictions may vary from language to language. Spanish is a language in which linking mainly depends on word order and case morphology. Previous studies in this language have shown that participants take longer time to read the second argument of an SVO or OVS sentence (Gattei et al., 2015a) and make longer and more regressions to previous regions (Gattei et al., 2017; 2021) when the order of semantic arguments does not reflect the order established by the argument structure of the verb. Evidence in this language has also shown that arguments' misinterpretation modulates neurophysiological correlates (Gattei et al., 2015b). Interestingly, these studies revealed a main effect of verb type when participants had to integrate the second argument of the sentence (Gattei et al., 2015a; 2017) and when they had to respond questions about 'who did / felt what to / for whom' (Gattei et al., 2015b; 2017), with sentences with ObjExp psych verbs being more difficult than sentences with activity verbs. The authors suggested that such difficulty was related to differences in linking, since activity verbs entailed a direct linking, and sentences with ObjExp psych verbs involved an inverse linking between syntax and semantics. Although this suggestion goes in line with evidence found in other languages (Thompson and Lee, 2009), it remained unclear from these studies whether such differences are indeed related to linking or whether they could respond to differences in verb class (Pulvermüller et al., 2001), type of subcategorization and possible thematic roles (Shetreet et al., 2007), or semantic complexity (Brennan and Pylkkanen, 2010) of the two classes of verbs used (Activity verbs vs ObjExp psych verbs). Thus, we studied the comprehension of sentences that included a class of verbs that allowed different case marking for identical lexemes (see Ferreira, 1994: and subsequent works

for studies addressing the choice of active or passive voice in sentence production through the comparison of different verb types). This case manipulation leads to different syntax-to-semantics linking, as shown in **Table 1**. On the one hand, accusative-marked sentences like (a) and (b) result in a direct linking between syntactic constituents and their semantic roles. On the other hand, dative-marked verbs like those in sentences (c) and (d) lead to sentences with inverse linking. The hypothesis of the current experiment was that if linking type played a role on the cognitive cost of sentence comprehension, differences according to case marking should have taken place.

On this regard, results of the current study show that during incremental reading, participants *did* regress significantly more times to regions 6 (dissambiguating verb) and 8 (second NP) when the verb included a dative clitic than when it included an accusative clitic. They also took longer time to read region 8 and spent longer time regressing to previous regions before continuing reading when the sentences entailed an inverse linking type. These results suggest that integrating verb and arguments information gives rise to a higher processing difficulty when linking between syntax and semantics is inverse than when it is direct, replicating similar findings on the self-paced reading and eye-tracking studies with different verb classes in Spanish (Gattei et al., 2015a; 2017; 2021). In other words, the current study managed to delink the verb type effect found in previous studies from the use of different verb classes and highlights the role of linking *per se*.

As for the results of the comprehension task, the current experiment did not find evidence showing that case (and hence linking type) alone modulated accuracy rates or reaction times as Verb Type did in Gattei et al. (2015a; 2015b; 2017). The fact that the linking effect found during incremental reading did not persist during the offline task could be explained by one main factor: differently to the questions formulated on the three previous experiments conducted in Spanish, comprehension questions on this experiment only focused on one of both participants of the event and used a reflexive version of the verb being tested so that participants had to respond to the same question for sentences with both case markings. For instance, for sentences like those in **Table 1**, the question was '¿María se va a molestar?' (*Is María going to feel bothered?*). In the previous experiments, questions included a relative clause and both participants were somehow mentioned. For example, for sentences with a verb like 'gustar' (*to like*) or 'gritar' (*to yell*) questions were formulated as '¿Es María quien le gusta / grita a alguien?' (*Is María who likes /* yells at someone?). Thus, we suggest that linking effects in response times and accuracy rates in questions of previous experiments could have been present mainly because participants had to process linking information from both the sentences and the questions arguments, which in turn included an embedded relative clause, leading to increasing parsing difficulty (see Murujosa et al., ress: for details on processing of relative clauses with psych predicates). In the current experiment, the questions were feasible to answer by only understanding if the participant mentioned was

the Experiencer or not, both in accusative and dative marked sentences, and questions did not include a relative clause. Hence, although the current results do not invalidate previous findings, they call for further investigation of task-related effects when testing both comprehension and linking effects simultaneously.

Another result that requires analysis is the effect of word order found at region 4 (Auxiliary verb 'va'). This effect was present for Regression Path Duration, Total Fixation Time and First Pass Reading Time (although not discussed at the Results section), showing longer reading times for OVS sentences than for SVO sentences. The appearance of this effect at the auxiliary verb region could be related to the fact that the OVS word order is not compatible with the lexical intransitive verb 'ir' (*'to go'*). It is possible that participants initially tried to integrate the lexical form of this verb without success and soon realized that it was part of a periphrastic future construction, leading to longer reading times and regression path duration.

At the same time, the use of sentences with verbs that accept alternative case marking (accusative or dative), allowed us to assess whether the effects related to incremental argument interpretation found in previous studies comparing sentences with different verb types could also be found when sentences only differed in one character denoting case assignment.

As it was briefly mentioned in the Introduction section, different psycholinguistic models have attempted to explain how case marking is processed and thus, how linking takes place. While it is not the aim of the current study to fully commit to one of them, in the following section we introduce their general assumptions and how they would account for the results of the current study.

On the one hand, the *Linking and Checking Model (LCM)* is a behavioral model that bases its assumptions and predictions on the Minimalist Program of the Generative Grammar (Bader and Bayer, 2006). In order to explain how the human sentence processing mechanism deals with morphological case ambiguous sentences, the authors of this model thoroughly describe a group of studies that involved word order alternation in German. The model proposes that comprehension and linking between syntactic to semantic information is mostly syntactically-driven, with morphological case marking as the most important cue for predicting what type of event the sentence will describe. The Linking and Checking algorithm that the model assumes proposes that in case of ambiguity, accusative case marking should be preferred over dative case marking. Crucially for the sentences used in the current experiment, the LCM predicts that in the case of OVS sentences, in which the first constituent of the sentence is an ambiguous noun phrase, the appearance of a dative clitic at the verb region should entail a higher processing difficulty than an accusative clitic.

On the other hand, the *extended Argument Dependency Model* proposes that during sentence processing the parser attempts to assign the most active semantic role possible to the first

constituent. In order to do so, the model assumes a prominence computation mechanism, that uses the available syntactic and semantic features of the language and the dependencies between them (Bornkessel and Schlesewsky, 2006). While this is a neurocognitive model that bases most of its assumptions and predictions on evidence from neurophysiological and brain imaging studies, it is still relevant for the current study as "one would expect to find at least a partial correspondence between eyetracking measures and different processing steps within the eADM." (Bornkessel and Schlesewsky, 2006: p. 811). According to this model, Spanish prominence computation would be sensitive to case marking, word order and animacy alternations. In the case of sentences like the ones analysed in the current study, the use of different clitics yields different verb argument structures and different configurations of their arguments' prominence status. For instance, accusative-marked sentences present a more-to-less prominent order of arguments in the case of SVO sentences and a less-to-more prominent structure in OVS sentences. Sentences with dative-marked clitics present the opposite pattern: a more-to-less prominent order of arguments in OVS sentences and a less-to-more prominent order in SVO sentences. Hence, the model predicts an interaction between word order and case marking, with SVO sentences with Dative case marking being more difficult than those with accusative case marking, and with OVS sentences with accusative case marking being more difficult than those with dative case marking.

As it may be seen in **Table 2**, the data collected in the current experiment partially supports the predictions of both models regarding case marking. In particular, results show that there was a significant interaction between case and word order at regions 6 and 8 (second NP) in late eye-movement measures TIR, TFT and RPD. However, this effect was not symmetric for both word orders. Subject-initial sentences with a dative clitic yielded more parsing difficulty than SVO sentences with accusative case marking but no differences were found between OVS conditions.

A third psycholinguistic model that is relevant for the current study is the *Competition Model*, proposed by MacWhinney et al. (1984); MacWhinney and Bates (1993); Bates et al. (1982). This model grounds most of its assumptions and predictions regarding linking on evidence from behavioral comprehension studies. It assumes that sentence processing is based on a direct form-to-function mapping that results from the interaction of several linguistic cues (e.g., word order, animacy, agreement, stress). The strength of each cue is language-specific, meaning that a cue may be a strong indicator for semantic interpretation in one language but a weak one in another. Cue validity is calculated according to the combination of cue applicability (how available a cue is) and cue reliability (how unambiguous a cue is) and, when considering what the most active argument of a sentence is, the different cues will interactively compete against each other to guide comprehension. This model shares certain similarities with the eADM in that they both try to explain cross-linguistic findings by resorting to a higher-level operation (i.e. prominence computation and cue validity respectively) that uses interacting information from different linguistic levels, and the outcome of overall sentence interpretation may be modeled according

to the relevance of each cue for a specific language. While this was not the focus of the current study, accuracy results of the comprehension task showed that both position of the noun and case marking are relevant cues for Spanish sentence interpretation, but that participants do not rely on them similarly for all sentences. When trying to identify if the participant mentioned in the question was the most active argument of the sentence previously read (i.e. the Agent for accusative-marked sentences or the Experiencer in dative-marked ones), the results showed an interaction between word order and case which depended on case marking for OVS sentences and on word order for accusative-marked ones. This pattern suggests that when facing OVS sentences, participants relied more on a cue like the position of the noun, prioritizing preverbal nouns than case-marking of the object and yielding worse results for accusative than for dative-marked sentences. No differences were found between SVO conditions, showing that for these conditions, participants also relied on case marking of the object for interpretation. These hypotheses are compatible with the results of the eyetracking measures. Recall that during incremental reading the opposite pattern was found: higher difficulties to parse SVO sentences when they included a dative clitic than when they included an accusative one, and no differences between OVS sentences. It is possible that in subject-initial sentences, a higher effort was put on correcting mispredictions about case assignment -as well as processing word order- showing similar difficulty for both SVO sentences, and thus no differences between conditions. Conversely, it is possible that in the case of OVS sentences participants relied more on word order and engaged less in the clitic's case marking, showing a higher rate of misinterpretation for those sentences in which OVS sentences is not the most frequent one (accusative-marked ones). On the other hand, interpretation was different for accusative than for dative-marked sentences. Participants showed worse performance at interpreting sentences with distinct word order in accusative-marked sentences but not in dative-marked ones, indicating that a noun being at preverbal position may have been a stronger cue for accusative-marked sentences than for their dative-counterparts. This pattern also complements the results of the eyetracking measures: during reading, participants regressed to the preverbal noun and main verb significantly more and spent more time reading these regions when the sentence included a dative than an accusative clitic. Once again, the longer time spent reading dative-marked sentences could have resulted in more accurate syntax-to-semantic linking and better performance, and differences regarding word order for accusative-marked sentences could have raised due to less reliance on noun position for these sentences. It is matter of future research to characterize whether this pattern of results is consistent with Spanish corpora analysis of cue reliability and applicability.

Conclusions

The current work aimed at providing further evidence on how linking is achieved during real-time comprehension when the same verbs class is used. It went a step further from other studies on the

topic by using a special class of psych verbs in Spanish, which can bear both accusative or dative case marking, leading to verbs' argument structures that require either direct or inverse syntax-to-semantics linking respectively. Together with the manipulation of word order, the sentences used in this experiment could disentangle whether processing inverse linking was more costly than direct linking. All in all, results showed that during incremental comprehension, inverse linking is more difficult than direct linking, irrespective of word order. Although the linking effect was not present at the comprehension task, it did take place at the sentence acceptability judgement task, with participants penalizing those sentences with indirect linking more than sentences with direct one. This dissociation of results, as well as the differences between the types of questions formulated for the current experiment and those from previous studies in the same language suggest that the type of questions used for the comprehension task (probably easier in terms of linking) could have led to a dissipation of the linking effect. Furthermore, results were briefly discussed under the light of different psycholinguistic models that address case processing and linking, providing a broader view on how the processing of this subclass of psych verbs could be informative for their own theoretical assumptions and experimental predictions.

Abbreviations

ACC = Accusative

Act = Activity

AGX = Agreement Index

DAT = Dative

eADM = extended Argument Dependency Model

fem = feminine

FFD = First Fixation Duration

FPRT = First Pass Reading Time

Infin = Infinitive

LCM = Linking and Checking Model

M = Mean

NOM = Nominative

ObjExp = Object Experiencer Psych Verbs

OVS = Object – Verb – Subject

pl = plural

REFL = Reflexive

RBRC = Right Bounded Regression Count

RPD = Regression Path Duration

SE = Standard Error

sg = Singular

SVO = Subject – Verb – Object

TFT = Total Fixation Time

TIR = Total Incoming Regressions

Data availability

The raw data, analyses' scripts and supplementary files supporting the conclusions of this article are available at the OSF repository.

Ethics and consent

The experiments described in this work were reviewed and approved by the Ethics Board of Facultad de Filosofía y letras, Universidad de Buenos Aires, protocol # 16. All participants provided their written informed consent to take part in this study.

Acknowledgements

Competing interests

The authors have no competing interests to declare.

Authors contributions

CAG, LP, AJW and DS conceptualized the experiment and decided on the methodology, contributed on writing and on data curation and analysis. FA took the lead on participants' recruitment and data collection. YS contributed with the revision of materials, writing and editing. Analyses' code was written by CAG and DS.

References

Ackerman, F., & Moore, J. (1999). Syntagmatic and paradigmatic dimensions of causee encodings. *Linguistics and Philosophy, 22*(1), 1–44. DOI: https://doi.org/10.1023/A:1005462027684

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511801686

Bader, M., & Bayer, J. (2006). *Case and linking in language comprehension: Evidence from German,* volume 34 *of Studies in Theoretical Psycholinguistics.* Springer Science & Business Media.

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. DOI: https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *ArXiv e-prints.* DOI: https://doi.org/10.18637/jss.v067.i01

Bates, E., McNew, S., MacWhinney, B., Devescovi, A., & Smith, S. (1982). Functional constraints on sentence processing: A cross-linguistic study. *Cognition, 11*(3), 245–299. DOI: https://doi.org/10.1016/0010-0277(82)90017-8

Belletti, A., & Rizzi, L. (1988). Psych-verbs and θ-theory. *Natural Language & Linguistic Theory, 6*(3), 291–352. DOI: https://doi.org/10.1007/BF00133902

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological review, 113*(4), 787. DOI: https://doi.org/10.1037/0033-295X.113.4.787

Bornkessel, I., Schlesewsky, M., & Friederici, A. D. (2002). Beyond syntax: Language-related positivities reflect the revision of hierarchies. *NeuroReport, 13*(3), 361–364. DOI: https://doi.org/10.1097/00001756-200203040-00022

Bornkessel, I., Schlesewsky, M., & Friederici, A. D. (2003). Eliciting thematic reanalysis effects: The role of syntax-independent information during parsing. *Language and Cognitive Processes*, *18*(3), 269–298. DOI: https://doi.org/10.1080/01690960244000018

Bornkessel, I., Zysset, S., Friederici, A. D., von Cramon, D. Y., & Schlesewsky, M. (2005). Who did what to whom? The neural basis of argument hierarchies during language comprehension. *Neuroimage*, *26*(1), 221–233. DOI: https://doi.org/10.1016/j.neuroimage.2005.01.032

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). *Processing syntax and morphology: A neurocognitive perspective.* Oxford University Press Oxford.

Brennan, J., & Pylkkanen, L. (2010). Processing psych verbs: Behavioural and MEG measures of two different types of semantic complexity. *Language and Cognitive Processes*, *25*(6), 777–807. DOI: https://doi.org/10.1080/01690961003616840

Christensen, R. H. B. (2019). ordinal—regression models for ordinal data. R package version 2019.12-10. https://CRAN.R-project.org/package = ordinal.

Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. van Gompel (Ed.), *Eye Movements: A Window on Mind and Brain* (pp. 341–372). Elsevier Science. DOI: https://doi.org/10.1016/B978-008044980-7/50017-3

Comrie, B. (1976). The syntax of causative constructions: Cross-language similarities and divergences. *Syntax and Semantics*, *6*, 261–312.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers, 34*(4), 613–617. DOI: https://doi.org/10.3758/BF03195489

Cuervo, M. C. (1999). Quirky but not eccentric: dative subjects in Spanish. *MIT Working papers in Linguistics, 34*, 213–227.

Davies, M. (2002). Corpus del espanol: 100 million words, 1200s–1900s.

Ditman, T., Holcomb, P. J., & Kuperberg, G. R. (2007). An investigation of concurrent ERP and self-paced reading methodologies. *Psychophysiology*, *44*(6), 927–935. DOI: https://doi.org/10.1111/j.1469-8986.2007.00593.x

Dorel, M. A. (1980). The two verbs FAIRE in French expressions of causation. In F. H. Nuessel Jr (Ed.), *Contemporary studies in Romance languages, proceedings of the eighth annual symposium on Romance languages* (pp. 27–47).

Dröge, A., Maffongelli, L., & Bornkessel-Schlesewsky, I. (2014). Luigi piace a Laura? In A. Bachrach, I. Roy, & L. Stockall (Eds.), *Structuring the Argument: Multidisciplinary Research on verb argument structure,* volume 10 of *Language Faculties and Beyond* (pp. 83–118). John Benjamins Publishing Company. DOI: https://doi.org/10.1075/lfab.10.05dro

Fábregas, A. (2013). Differential object marking in Spanish: State of the art. *Borealis–An International Journal of Hispanic Linguistics, 2*(2), 1–80. DOI: https://doi.org/10.7557/1.2.2.2603

Fernández-Soriano, O. (1999). Two types of impersonal sentences in Spanish: Locative and dative subjects. *Syntax, 2*(2), 101–140. DOI: https://doi.org/10.1111/1467-9612.00017

Ferreira, F. (1994). Choice of passive voice is affected by verb type and animacy. *Journal of Memory and Language, 33*(6), 715–736. DOI: https://doi.org/10.1006/jmla.1994.1034

Franco, J. (1992). Towards a typology of psych verbs: Evidence from Spanish. *Anuario del Seminario de Filología Vasca "Julio de Urquijo"* (pp. 119–134).

Frisch, S., & Schlesewsky, M. (2001). The N400 reflects problems of thematic hierarchizing. *NeuroReport, 12*(15), 3391–3394. DOI: https://doi.org/10.1097/00001756-200110290-00048

Gattei, C. A., Dickey, M. W., Wainselboim, A. J., & Paris, L. (2015a). The thematic hierarchy in sentence comprehension: A study on the interaction between verb class and word order in Spanish. *The Quarterly Journal of Experimental Psychology, 68*(10), 1981–2007. DOI: https://doi.org/10.1080/17470218.2014.1000345

Gattei, C. A., París, L. A., & Shalom, D. E. (2021). Information structure and word order canonicity in the comprehension of Spanish texts: an eye-tracking study. *Frontiers in Psychology, 12*. DOI: https://doi.org/10.3389/fpsyg.2021.629724

Gattei, C. A., Sevilla, Y., Tabullo, A., Wainselboim, A., París, L., & Shalom, D. (2017). Prominence in Spanish sentence comprehension: An eye-tracking study. *Language, Cognition and Neuroscience, 33*(5), 587–607. DOI: https://doi.org/10.1080/23273798.2017.1397278

Gattei, C. A., Tabullo, Á., París, L., & Wainselboim, A. J. (2015b). The role of prominence in Spanish sentence comprehension: An ERP study. *Brain and Language, 150*, 22–35. DOI: https://doi.org/10.1016/j.bandl.2015.08.001

Gibson, J., & Raposo, E. (1986). Clause union, the stratal uniqueness law and the chômeur relation. *Natural Language & Linguistic Theory, 4*(3), 295–331. DOI: https://doi.org/10.1007/BF00133372

Green, P., & MacLeod, C. J. (2016). Simr: an r package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498. DOI: https://doi.org/10.1111/2041-210X.12504

Honrubia, J. L. C. (2015). Causativity and psychological verbs in Spanish. In *Verb classes and aspect* (pp. 110–130). John Benjamins. DOI: https://doi.org/10.1075/ivitra.9.06cif

Hyman, L. M., & Zimmer, K. E. (1976). Embedded topic in French. In C. N. Li (Ed.), *Subject and Topic* (pp. 189–211). New York, NY: Academic Press.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329. DOI: https://doi.org/10.1037/0033-295X.87.4.329

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*(2), 228. DOI: https://doi.org/10.1037/0096-3445.111.2.228

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology, 16*, 262–284. DOI: https://doi.org/10.1080/09541440340000213

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General, 135*(1), 12. DOI: https://doi.org/10.1037/0096-3445.135.1.12

Lamers, M., & De Swart, P. (2012). *Case, Word order and Prominence: Interacting Cues in Language Production and Comprehension* (vol. 40). Springer Science & Business Media.

Lenth, R. V. (2022). emmeans: Estimated marginal means, aka least-squares means. R package version 1.7.2.

Leonetti, M. (2004). Specificity and differential object marking in Spanish. *Catalan journal of linguistics*, *3*(1), 75–114. DOI: https://doi.org/10.5565/rev/catjl.106

Logacev, P., & Vasishth, S. (2013). em2: A package for computing reading time measures for psycholinguistics. R package version 0.9.

MacWhinney, B., & Bates, E. (1993). The crosslinguistic study of sentence processing. *Journal of Child Language*, *20*(2), 463–71. DOI: https://doi.org/10.1017/S0305000900008382

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in english, German, and Italian. *Journal of verbal learning and verbal behavior*, *23*(2), 127–150. DOI: https://doi.org/10.1016/S0022-5371(84)90093-8

Marín, R. (2014). Stativity and agentivity in Spanish psychological verbs. In *Workshop on the Syntax and Semantics of Experiencers. Berlin.*

Marín, R., & McNally, L. (2011). Inchoativity, change of state, and telicity: Evidence from Spanish reflexive psychological verbs. *Natural Language & Linguistic Theory*, *29*(2), 467–502. DOI: https://doi.org/10.1007/s11049-011-9127-3

Masullo, P. J. (1992). Quirky datives in Spanish and the non-nominative subject parameter. *MIT Working Papers in Linguistics*, *16*, 89–103.

Mendivil, J.-L. (2002). La estructura ergativa de 'gustar' y otros verbos de afección psíquica en español. In *Actas del V Congreso de Lingüística General. León* (pp. 2003–2014). Arco Libros.

Murujosa, M., Gattei, C. A., Shalom, D. E., & Sevilla, Y. (In press). Comprensión de oraciones relativas con predicados psicológicos en español. *Estudios de Linguistica Aplicada, 74.*

Perlmutter, D. M. (1978). Impersonal passives and the unaccusative hypothesis. In *annual meeting of the Berkeley Linguistics Society* (vol. 4, pp. 157–190). DOI: https://doi.org/10.3765/bls.v4i0.2198

Pinheiro, J. C., & Bates, D. M. (2000). Linear mixed-effects models: Basic concepts and examples. In J. C. Pinheiro, & D. M. Bates (Eds.), *Mixed-effects models in S and S-Plus* (pp. 3–56). Springer. DOI: https://doi.org/10.1007/978-1-4419-0318-1_1

Pulvermüller, F., Härle, M., & Hummel, F. (2001). Walking or talking?: Behavioral and neurophysiological correlates of action verb processing. *Brain and Language, 78*(2), 143–168. DOI: https://doi.org/10.1006/brln.2000.2390

R Core Team (2021). R: A language and environment for statistical computing.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*(4), 04–509. DOI: https://doi.org/10.3758/BF03214555

Reed, L. (1992). Remarks on word order in causative constructions. *Linguistic inquiry, 23*(1), 164–172.

Shetreet, E., Palti, D., Friedmann, N., & Hadar, U. (2007). Cortical representation of verb processing in sentence comprehension: Number of complements, subcategorization, and thematic frames. *Cerebral Cortex*, *17*(8), 1958–1969. DOI: https://doi.org/10.1093/cercor/bhl105

Thompson, C. K., & Lee, M. (2009). Psych verb production and comprehension in agrammatic Broca's aphasia. *Journal of Neurolinguistics*, *22*(4), 354–369. DOI: https://doi.org/10.1016/j.jneuroling.2008.11.003

Treviño, E. (1990). Non-canonical subjects in Spanish: evidence from causative and psych verbs. In *Romance Languages and Modern Linguistic Theory: Selected papers from the XX Linguistic Symposium on Romance Languages, University of Ottawa,* volume 91 of *Current Issues in Linguistic Theory* (pp. 309–324). John Benjamins, Ottawa.

Van Valin Jr, R. D. (2005). *Exploring the syntax-semantics interface.* Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511610578.001

Van Valin, R. D., & LaPolla, R. J. (1997). *Syntax: Structure, meaning, and function.* Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139166799

Vasishth, S., von der Malsburg, T., & Engelmann, F. (2013). What eye movements can tell us about sentence comprehension. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(2), 125–134. DOI: https://doi.org/10.1002/wcs.1209

Verissímo, J. a. (2021). Analysis of rating scales: A pervasive problem in bilingualism research and a solution with bayesian ordinal models. *Bilingualism: Language and Cognition*, *24*(5), 842–848. DOI: https://doi.org/10.1017/S1366728921000316

Von der Malsburg, T., & Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *Journal of memory and language*, *94,* 119–133. DOI: https://doi.org/10.1016/j.jml.2016.10.003

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*(2), 206–237. DOI: https://doi.org/10.1016/j.jml.2009.04.002

Wang, L., Schlesewsky, M., Bickel, B., & Bornkessel-Schlesewsky, I. (2009). Exploring the nature of the 'subject'-preference: Evidence from the online comprehension of simple sentences in Mandarin Chinese. *Language and Cognitive Processes*, 24(7–8): 1180–1226. DOI: https://doi.org/10.1080/01690960802159937

Wang, L., Schlesewsky, M., Philipp, M., & Bornkessel-Schlesewsky, I. (2012). The role of animacy in online argument interpretation in Mandarin Chinese. In M. Lamers, & P. de Swart (Eds.), *Case, Word Order and Prominence* (pp. 91–119). Springer. DOI: https://doi.org/10.1007/978-94-007-1463-2_5

Weckerly, J., & Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology*, *36*(5), 559–570. DOI: https://doi.org/10.1111/1469-8986.3650559