

## **Widening agreement processing: a matter of time, features and distance**

Nicoletta Biondo<sup>a1</sup>, Francesco Vespignani<sup>a</sup>, Luigi Rizzi<sup>b,c</sup> and Simona Mancini<sup>d</sup>

*<sup>a</sup>Department of Psychology and Cognitive Science, University of Trento, Rovereto (TN), Italy*

*<sup>b</sup>Department of Linguistics, University of Geneva, Geneva, Switzerland*

*<sup>c</sup>Department of Social, Political and Cognitive Sciences, University of Siena, Siena, Italy*

*<sup>d</sup>Basque Center on Cognition, Brain and Language, Donostia-San Sebastián, Spain*

Correspondence should be addressed to Dr Nicoletta Biondo, Basque Center on Cognition, Brain and Language (BCBL), Paseo Mikeletegi 69, 20009 Donostia-San Sebastián, Spain. E-mail: n.biondo@bcbl.eu.

---

<sup>1</sup> The corresponding author has changed affiliations during the peer-review process. New affiliations are: Basque Center on Cognition, Brain and Language, Donostia-San Sebastián, Spain; Fondazione ONLUS Marica De Vincenzi, Rovereto (TN), Italy; Department of Psychology and Cognitive Science, University of Trento, Rovereto (TN), Italy.

## **Widening agreement processing: a matter of time, features and distance**

### **Abstract**

Existing psycholinguistic models typically describe agreement relations as monolithic phenomena amounting to mechanisms that check mere feature consistency. This eye-tracking study aimed at widening this perspective by investigating the time spent reading subject-verb (number, person) and adverb-verb (tense) violations on an inflected verb during sentence comprehension in Spanish. Results suggest that (i) distinct processing mechanisms underlie the analysis of subject-verb and adverb-verb relations, (ii) the parser is sensitive to the different interpretive properties that characterize the person, number and tense features encoded in the verb (i.e. anchoring to discourse for person and tense interpretation, as opposed to anchoring to cardinality information for number), and (iii) the (local, distal) position of the agreement controller with respect to the verb affects the interpretation of these dependencies. An account is proposed that capitalizes on the importance of enriching current sentence processing formalizations using a feature and relation-based approach.

### **Keywords**

agreement, person, tense, number, eye movements

## Introduction

Processing of verb inflection in morphologically rich languages requires checking feature consistency between verbs and other elements in the sentence (e.g. subjects and adverbs), and interpreting a wide set of information (e.g. person, number and tense). For example, in sentences like (1) in Spanish, the verb (*fui*, was) shares number and person features with the subject (*yo*, I). The “singular” feature value and the “first” person feature value inform the reader that the subject represents a single entity, and is also the speaker of the utterance. Moreover, the verb shares temporal features with the deictic adverb (*ayer*, yesterday) indicating that the act of going to the concert happened in the past, namely the day before the time of utterance.

(1)  $Ayer_{PAST} (yo)_{SG/1ST} fui_{SG/1ST/PAST}$  a un concierto de música jazz.

Yesterday I went to a jazz music concert.

In other words, each time comprehenders read the verb of this sentence, their language system needs to process number, person and tense features on the verb, and verify feature consistency with other constituents in the sentence (i.e. subject, adverb), which can be located at a different distance from the verb. For example, in (1) the subject is adjacent to the verb while the adverb is distally located from the verb.

An open issue is whether unique or different mechanisms underlie the processing of verb inflection and its features during sentence comprehension. How and when does the language system deal with verb inflectional features and their underlying relations? Does the distance between the verb and its related constituents play any role? To answer these questions, this study will explore the different impact of number, person and tense verb violations on sentence parsing

in Spanish, using an eye-tracking paradigm.

### *Inflectional features from a linguistic perspective*

Although number, person and tense features are all morphologically realized on the verb, these features differ in two important aspects, namely in the type of constituent the verb inflection interacts with, and in the intrinsic interpretive properties each feature specification entails.

Firstly, both the subject determiner phrase (DP) and verb inflection are *obligatory* components of the clausal structure, having an obligatory requirement of match in person and number features (the Extended Projection Principle, or EPP in Chomsky, 1981) so that subject-verb agreement can be established. This requirement is reached through a mandatory feature checking procedure that must take place both from the vantage point of the inflectional constituent, which must have its features valued by the subject, and of the subject DP, which must have its case licensed (i.e. nominative), even when the subject is omitted<sup>i</sup> (as in null-subject languages like Spanish). In contrast, the adverb-verb relation, called *adverb-verb tense agreement* or *temporal concord*<sup>ii</sup>, is very different from this viewpoint, since the adverb is an *optional* constituent. Indeed, adverbial DPs such as “last year, next year” bear temporal features for past and future (Alexiadou, 1997; Enc, 1987) but verb inflection does not need the presence of the adverbial to express a temporal value, as it carries an “interpretable feature” (in the minimalist terminology and formalization of Chomsky 1995, 2000) by itself. In other words, although temporal adverbs can provide detailed information about the time interval in which the event takes place, verb Tense features can provide the temporal location of the event in an independent way. Moreover, adverbial DPs do not have a case that needs to be structurally licensed. Hence, for adverb-verb agreement to be

established no formal feature checking is assumed to take place, unlike subject-verb agreement.

This contrast suggests a different status for number and person subject-verb agreement compared to tense adverb-verb agreement in terms of type and optionality of the verb-related constituent.

Secondly, the features involved in the two relations - number, person and tense - are characterized by different interpretive, or anchoring, requirements. The number feature relates to the cardinality of the subject, that is how many entities the constituent entails (singular or plural), while person expresses the role of the subject in the speech act (1<sup>st</sup> - the speaker, 2<sup>nd</sup> – the addressee, 3<sup>rd</sup> – neither the speaker nor the addressee). The tense feature expresses the time in which the event takes place relatively to the time of utterance ‘now’ (past, present, future). It has been suggested that the assignment of a speech participant role and the interpretation of the speech time expressed by person and tense requires linking the morpho-syntactic representation of these features to a clause-peripheral position in the complementizer (CP) zone, providing specifications connected to the discourse representation of the sentence (Bianchi 2003, Sigurðsson 2004, 2016; see also Speas & Tenny, 2003), a structural layer where the speech participants and the speech time are encoded. By contrast, the cardinality of number requires a clause-internal anchoring, as this property is expressed by the subject DP itself, located in a specifier position in the functional structure of the clause, the IP (Mancini, Molinaro, Rizzi & Carreiras, 2011; Mancini, Molinaro & Carreiras, 2013). This contrast suggests that person and tense should show some similarities and should pattern differently from number.

### ***Syntactic relations and features from a psycholinguistic perspective***

Existing models of sentence parsing (e.g. Bornkessel & Schlesewsky, 2006; Friederici, 2002; 2011; Gibson, 1998; Lewis, Vasishth & Van Dyke, 2006; Hagoort 2003, 2013), including those more informed by linguistic theories, appear to be largely underspecified with respect to

the impact that the establishment of different syntactic relations and the analysis of different features may have during comprehension.

One exception is represented by the Construal model (Frazier & Clifton, 1996). Although the Construal model has not received so much attention over the last decade, this model is specifically relevant to the issues addressed in the current article since, to our knowledge, it is the only model which tries to formalize the different cognitive mechanisms implied during the processing of optional constituents, such as temporal adverbs. Indeed it posits a fundamental distinction between primary (e.g. subject-predicate) and non-primary (e.g. adjunct, relative clauses attachment) relations. Primary relations, such as subject-verb agreement, are assumed to be processed using attachment mechanisms as described by classical syntax-first models, i.e. by positing initial reliance on purely syntactic criteria, while semantic-pragmatic factors are taken into account at subsequent stages. In contrast, the analysis of adverb-verb agreement would fall within the domain of non-primary relations, given the adjunct nature of the adverb. The Construal model claims that non-primary phrases are processed following a different parsing routine, that is, association or *construal*. The construal principle differs from traditional minimal attachment in that semantic or non-structural factors may affect the attachment processing. In other words, when a non-primary phrase such as temporal adverb is processed, it is associated to the current thematic domain, namely to the extended projection of the first theta assigner (i.e. the verb) which is encountered within the sentence, and it is interpreted based on structural and non-structural information.

While the Construal model predicts a difference in the processing of the subject-verb and adverb-verb relation, it does not consider potential interpretive differences among the features encoded in verb morphology. Recent accounts based on experimental data, however, raised the

hypothesis of a feature-sensitive language processing system. Carminati (2005) introduced the Feature Hierarchy Hypothesis, in which the processing of each agreement feature (e.g. number and person) varies depending on its “cognitive salience”, which is defined on the basis of the cross-linguistic occurrence of each feature (Greenberg, 1963). Person is considered more salient than number since the former can occur across languages independently from number, while the latter occurs in a language if person also does. However, this approach has been called into question by some experimental studies showing larger electrophysiological effects in later stages of processing for less salient features such as gender (Barber & Carreiras, 2005; Molinaro, Vespignani & Job, 2008). Moreover, the concept of “cognitive salience” is rather generic in the field of sentence processing and should be detailed in terms of stages of processing and/or in terms of type of cognitive resources that are at play in this hierarchy.

A typological perspective also underlies the implicit distinction between person and number drawn by Bornkessel & Schlesewsky (2006; see also Bornkessel-Schlesewsky & Schlesewsky 2009) in their extended Argument Dependency Model (eADM). In their account, which focuses on thematic role assignment, person (together with definiteness and animacy) provides the parser with reliable cues to determine who is the actor and who is the patient in a sentence (a mechanism called “compute prominence” in their formalization), with 1<sup>st</sup> person arguments more reliably mapping to actor roles compared to 2<sup>nd</sup> and 3<sup>rd</sup> person ones (Person Hierarchy,  $1^{st} < 2^{nd} < 3^{rd}$ , see Silverstein, 1975). In contrast, number information does not provide the processor with strong cues to agent and patienthood, and is therefore not used to draw argument hierarchies, but only to establish formal relations between arguments and predicates.

Finally, unlike the Feature Hierarchy Hypothesis and the eADM, Mancini et al. (2013) propose a differential impact of number and person features based on their distinct anchoring to the deictic context. In a self-paced reading study in Italian, Mancini, Postiglione, Laudanna & Rizzi (2014) showed that the presence of a person anomaly, as in “\**Il giornalista*<sub>3.sg</sub> *scrivo*<sub>1.sg</sub> *un articolo interessante*” (The journalist<sub>3rd.sg</sub> write<sub>1st.sg</sub> an interesting article) generated longer reading times compared to number anomalies as in “\**Il giornalista*<sub>3.sg</sub> *scrivono*<sub>3.pl</sub> *un articolo interessante* (The journalist<sub>3rd.sg</sub> write<sub>3rd.pl</sub> an interesting article). They interpreted these data as the result of the impossibility to assign a discourse role to the subject argument in case of person mismatch (i.e. to establish whether the writing event is told from the perspective of the speaker, as implied by the 1<sup>st</sup> person verb, or whether it involves a 3<sup>rd</sup> person party, as implied by the subject), as opposed to a faster resolution based on the subject’s number in case of number mismatch. Evidence of the differences between the processing of number and person features has been also found in ERP and functional magnetic resonance studies (Mancini, Molinaro, Rizzi & Carreiras, 2011; Mancini, Quiñones, Molinaro, Hernandez-Cabrera & Carreiras, 2017; Zawiszewski, Santesteban & Laka, 2016; but see Silva-Pereyra & Carreiras, 2007 for different results).

It should be noticed that feature-based approaches have mainly dealt with the processing of subject-verb agreement features, while no formalization or predictions have been provided on the processing of tense and temporal agreement. Experimental evidence on the processing of adverb-verb tense violations is rather sparse and mainly comes from ERPs studies (Steinhauer & Ullman 2002, Baggio 2008, Qiu & Zhou 2012), in which heterogeneous experimental material was tested (i.e. adverb-verb violations were tested in different languages and at different linear and structural distance), leading to inconsistent results. However, some evidence for different



mechanisms during the processing of tense and number verb violations was found by De Vincenzi and colleagues (unpublished) in a study conducted in Italian: number violations gave rise to a behavioural cost (self-paced reading, Exp.1) at the target region (the verb) and at the following one, while tense violations caused parsing costs only at the post-target region. ERP results (Exp. 2) further confirmed the number and tense dissociation. However, it is not clear whether this processing dissociation is due to the optionality of the verb-related constituent (i.e. obligatory subject vs. optional adverb) or to the discourse-related properties of the feature under computation (i.e. non-deictic number vs. deictic tense) since number and tense features differ in both aspects. In this respect, the comparison of number, person and tense features appears to be of crucial importance, because it allows us to understand whether this difference is due to the type of relation underlying verb inflection and its related constituent, to the different interpretive properties of each feature, or to the interplay of both factors. In fact, number and person pattern together with respect to the mandatory verb-related constituent (i.e. subject) but they differ in the anchoring mechanism. Conversely, person and tense pattern together in the (external) anchoring mechanism but they differ in the obligatoriness of the verb-related constituent.

Crucially, the comparison of person, number and tense is limited to one study. In an ERP study in French, Fonteneau, Frauenfelder & Rizzi (1998) investigated number, person and tense violations, in sentences such as in (2), and reported a differentiation in the ERP pattern elicited by number/person violations (posterior negativity around 300ms followed by a P600) compared to tense violations (frontal positivity and posterior negativity around 450ms). However, person/number and tense violations were compared with different baselines (i.e. the subject and the adverb were located in linearly distinct positions with respect to the verb), which may have contributed to the differences between the three conditions.

(2) Number disagreement: *En été l'herbe pousseront facilement.*

In summer the grass<sub>(3SG)</sub> will grow<sub>(3PL)</sub> easily.

Person disagreement: *Dans un an les travailleurs gagnerez de l'argent.*

In one year the workers<sub>(3PL)</sub> will earn<sub>(2PL)</sub> money.

Tense disagreement: *Demain l'étudiant lisait le livre.*

Tomorrow<sub>(FUT)</sub> the student was reading<sub>(PAST)</sub> the book.

To sum up, current mainstream models of parsing are largely underspecified with respect to possible differences either in the processing of different features or in the relations expressed by verb inflection, and the experimental studies described above only provide indirect evidence for a different computation of verb inflection when dealing with person, number and tense violations. This is of paramount importance, because such models fail to account for parsing mechanisms at work in morphologically rich languages such as Romance languages, in which verbs systematically and regularly encode tense, number and person information.

### **The current study**

The current study sets out to investigate the mechanisms involved in the processing of verb inflection violations using an eye-tracking paradigm. Thanks to the potentially ecological presentation of the materials, eye tracking allows us to make more specific predictions concerning reading mechanisms, compared to other behavioural methods, such as self-paced reading, or to electrophysiological paradigms, as it permits the analysis of the time spent in the critical region (generally characterized by longer fixations during the first reading of the wrong

constituent) and of the regressive saccades to previous parts of the sentence that are triggered to repair/reanalyse the critical constituent.

We will address three issues. Specifically, we are interested in determining whether the processing of verb inflection changes as a function of the type of constituent (or controller) the verb interacts with (i.e. an adverb or a subject DP), and/or of the different anchoring properties that characterize number, person and tense features encoded in the verb. Thirdly, we aim to assess whether the position of the subject and of the adverb with respect to the verb differentially impacts the processing of person, number and tense violations, as illustrated in 3a-3d.

#### Local adverb/subject-verb (dis-)agreement

(3.a.) Los viajeros cansados *mañana a mediodía* regresarán/\*regresaron a casa.

The tired travelers tomorrow at noon<sub>(FUT)</sub> will go<sub>(FUT)</sub> /\*went<sub>(PAST)</sub> back home.

(3.b) *Mañana a mediodía el viajero cansado* regresará/\*regresarán/\*regresarás a casa

Tomorrow at noon the tired traveler<sub>(3SG)</sub> will go<sub>(3SG)</sub> /\*will go<sub>(3PL)</sub> /\*will go<sub>(2SG)</sub> back home

#### Distal adverb/subject-verb (dis-)agreement

(3.c) *Mañana a mediodía* los viajeros cansados regresarán/\*regresaron a casa.

Tomorrow at noon<sub>(FUT)</sub> the tired travelers will go<sub>(FUT)</sub> /\*went<sub>(PAST)</sub> back home

(3.d) *El viajero cansado* mañana a mediodía regresará/\*regresarán/\*regresarás a casa.

The tired traveler<sub>(3SG)</sub> tomorrow at noon will go<sub>(3SG)</sub>/\*will go<sub>(3PL)</sub>/\*will go<sub>(2SG)</sub> back home.

In doing so, we aim to unveil important aspects of the language architecture. As pointed out by Kaan (2002), the presence of intervening material between two elements of a relation (e.g. subject-verb agreement) can affect different aspects of parsing: the way subject features are tracked before encountering the verb; the way the information coming from the verb is integrated with the information coming from the subject and the way an inconsistency is repaired when there is a mismatch in features between the subject and the verb. Crucially for the current study, the manipulation of subject-verb and adverb-verb violations in different configurations can be informative about the interplay of morphosyntactic and discourse information during sentence parsing. Indeed, models of sentence processing differ in the way different types of linguistic information are analysed during sentence comprehension.

Finally, through the manipulation of the linear distance between the verb and its related constituent we also overcome important limitations found in previous studies (e.g. Fonteneau et al. 1998), by providing a thorough comparison of the two relations and their features.

Previous eye-tracking studies on agreement processing report a heterogeneous scenario, with some studies pointing to the parser's early sensitivity to subject-verb manipulations, arising as early as during the first pass of fixations in the verb region (Deutsch, 1998; Deutsch & Bentin, 2001; Mancini, Molinaro, Davidson, Avilés & Carreiras, 2014), while others reporting later effects (Pearlmutter, Garnsey & Bock, 1999). As for adverb-verb agreement, to our knowledge only a few eye tracking studies are available in the literature. These studies mainly investigated the processing of ambiguous sentences such as "The carpenter *sanded/will sand* the shelves he

*attached/will attach* onto the kitchen wall *yesterday morning*, according to the foreman”, in which the deictic temporal adverb could refer to the verb (i.e. sanded) of the main clause and/or to the verb (i.e. attached) of the embedded clause (van Gompel, Pickering, Pearson & Liversedge 2005, Experiment 3; see also Altmann, van Nice, Garnham & Henstra, 1998). Data from these studies showed that the attachment of the temporal adverb and its match with the (main and/or embedded) verb is considered by the parser from early (first-pass) measures on. However, drawing a parallel with the current study would be inaccurate, given the substantial differences between the paradigm adopted by these studies (i.e. target adverb, ambiguous grammatical sentences) and ours (i.e. target verb, ungrammatical sentences). We will thus build our hypotheses for adverb-verb processing based on the existing literature on subject-verb agreement, predicting possible differences between the two relations as a function of optionality of the controller and anchoring to the deictic context.

The difference between subjects and adverbs, in terms of their optionality/obligatoriness in the relation they engage with the verb, permits drawing precise processing hypotheses. Specifically, if the processing of (person, number and tense) verb violations is modulated by the optionality of the controller and of the copying/checking operations, the two relations should elicit different reading patterns. Given that the subject is a fundamental and mandatory element of the sentence that triggers a formal feature-checking/valuing operation, a subject-verb agreement mismatch should force the parser to immediately detect and repair any inconsistency related to this relation. We thus expect immediate and sustained effects, from early measures, such as first-pass, to later measures (total reading times, regressions into target region) for both number and person mismatches compared to the correct agreement condition. On the contrary, the adverb’s optionality and the lack of an agreement-like checking operation could delay the

detection of a temporal inconsistency to later stages of processing (total reading time, regressions), leading to an asymmetry in the detection of subject- and adverb-verb anomalies. This pattern of effects would be in line with previous findings on the difference between the detection of subject-verb agreement violations and adverb-verb tense violations (De Vincenzi et al. unpub, Fonteneau et al. 1998). Alternatively, if the optionality of the constituents does not play a role during sentence comprehension, both person and number subject-verb mismatches and tense adverb-verb mismatches are expected to yield similar effects from early stages of reading measures.

A different set of hypotheses follows from the assumption that the three features under study can differently impact the processing of verb inflection because of their different anchoring positions. Under this assumption, we should expect a different mismatch effect for person compared to number mainly in late measures (i.e. total reading time, regressions to the verb), as a result of the different anchoring operation of the former feature compared to the latter (Bianchi 2006, Sigurðsson 2013, Mancini et al. 2013). In this respect, of relevance is the study by Mancini et al. (2014b, Experiment 4), which investigated the time course of discourse anchoring mechanisms during person agreement processing. By manipulating feature consistency and discourse plausibility between subject and verb, this study revealed an important dissociation between early stages of reading, which were sensitive to whether subject and verb shared the same person feature, irrespective of discourse plausibility, and later stages, when the parser showed sensitivity only to discourse implausible sentences. Moreover, if the discourse anchoring operations triggered by person and tense are similar (i.e. anchoring to an external position) but the two features differ in the formal feature checking procedure, we should also expect person and tense violations to yield similar total reading times and regression patterns at verb position,

but different mismatch effects in early measures, namely larger first-pass for person violations compared to the control condition and no mismatch effects for tense violations in early measures.

Finally, a third set of hypotheses can be formulated for the effect of distance in the processing of subject-verb and adverb-verb violations. The models accounting for the role of distance during sentence parsing are mainly memory-based models (e.g. Gibson, 1998; Grodner & Gibson, 2005; Lewis & Vasishth, 2005; Lewis et al., 2006). Some of these models, also called storage-based models (e.g. Gibson, 1998; Grodner & Gibson, 2005), assume that during parsing, constituents are incrementally stored in memory to be available for syntactic integration purposes, with storage costs increasing as a function of distance. Other models, also called cue-based retrieval models (e.g. Lewis & Vasishth, 2005; Lewis et al., 2006), assume that each parsed chunk fluctuates in the memory space and has an activation level which decays over time. In other words, all these models predict that the processing of a dependency becomes more difficult when the two related constituents are distally located in the sentence, either because of memory storage and integration costs or because of the difficulty in the reactivation of a constituent that has decayed over time. However, memory-based models do not make any prediction concerning potential differences between different types of relations or features. Previous experimental evidence on subject-verb agreement processing suggests possible effects of distance on verb inflection processing. An eye-tracking study in Hebrew (Deutsch, 1998) reports larger penalty for number agreement violations when the subject is located just before the verb compared to the condition in which the two constituents are distally located, both in early and late measures (but see Rispens & de Amesti 2016 for different ERP results). In line with this, we can expect less disruption when the subject is more distally located from the verb compared to when they are in a local configuration.

No previous studies have tested the effect of controller-target distance during the processing of tense violations. However, assuming that the optionality of the adverb can leave the parser more flexibility in the processing of the adverb-verb temporal relation, we can hypothesize that the detection of tense violations in a distal configuration can lead to shallower parsing costs in late measures, with respect to the local configuration.

Similarly, the lack of a large body of empirical data on the comparison among features makes it difficult to predict how and whether distance can impact feature consistency verification and anchoring mechanisms, which are triggered by verb-inflection processing. Nevertheless, based on previous findings (Mancini et al. 2014a,b), we can plausibly hypothesize that if the distance of the controller affects formal checking operations, differences between subject-verb and adverb-verb processing are expected to arise during the first pass through the verb region, as a result of the obligatoriness of checking operations for the former but not for the latter relation. In contrast, if controller-target distance impacts discourse anchoring mechanisms, later reading stages are more likely to be impacted. Under this scenario, distance is expected to impact differently subject-verb person (and adverb-verb tense) disagreement compared to subject-verb number disagreement. A summary table of the predictions for each factor that can affect the processing of number, person and tense violations (i.e. optionality of the controller, feature anchoring properties and linear distance) is provided in Table 1.

(Table 1 here)



## ***Method***

### *Participants*

Forty-eight participants (35 female, 19-33 years, mean age = 23 years, SD= 2.5 years) took part in this experiment in exchange for a small payment. They were all native speakers of Spanish and had normal or corrected-to-normal vision.

### *Design and materials*

The experimental material consisted of 120 experimental stimuli, in which both subject-verb and adverb-verb relations were manipulated (Type of Relation factor), as illustrated in Table 2. Within each relation, Person, Number and Tense congruence was manipulated, giving rise to correct and incorrect number, person and tense stimuli (Condition factor). In addition, the distance between subject/adverb and verbs was also manipulated to create local and distal relations (Configuration factor).

Each sentence contained an animate subject (e.g. *el viajero cansado*), a temporal adverb (e.g. *mañana a mediodía*) and a simple past or future verb (in equal proportions) followed by a direct (or indirect) object. The subject was always a lexical DP, sometimes followed by a modifier to balance, across items, the length (in characters) of the constituents preceding the target verb (Subject phrase: mean = 14.54, SD= 3.69; Adverb phrase: mean = 14.39, SD= 3.63). The temporal adverbs were all deictic so they encoded a specific temporal information depending on the context, namely the time of utterance (e.g. “yesterday” denotes the day before the time of utterance that is “now”). We used deictic temporal adverbs of different forms to avoid participant habituation to the critical constituents (e.g. *ayer/ mañana por la tarde, hace/en dos meses, el año próximo/pasado*).

Controller-target local and distal configurations are both grammatical and frequent in Spanish, bearing very similar meaning<sup>iii</sup>. It must be noted that the distal configuration in the two relations is equivalent both in terms of linear distance and of structural distance. There are optional constituents, such as prepositional phrases, which are structurally embedded within the noun phrase they refer to (e.g. [The man [with the red scarf]]) and they generally provide some additional information about the noun phrase they are embedded in. However, temporal adverbs do not fall into this category of constituents. Indeed, temporal adverbs and subject noun phrases are two different phrases that cannot be embedded into each other. The adverb cannot provide additional information about the subject noun phrase and vice versa. Critically for the distal conditions tested in the current study, the adverb cannot be embedded within the subject phrase in the subject-verb distal conditions (e.g. in English: [The man] [yesterday] [went [to the supermarket]]), and the subject noun phrase cannot be embedded within the adverb phrase in the adverb-verb distal conditions (e.g. in English [Yesterday] [the man] [went [to the supermarket]]). The same number of words and the same number of constituents separate the two critical constituents across conditions and this guarantees the same linear and a structural distance between the controller and the target across conditions. Moreover, at the processing level this may make a difference<sup>iv</sup>. In the local configuration, the controller is processed just before parsing the verb, so its analysis could still be in progress during the processing of the verb. On the other hand, in the distal configuration the analysis of the controller has probably been completed, since a different constituent was also parsed before encountering the verb.

We also note that the stimuli used for subject-verb manipulations all contain third person singular subjects (*el viajero cansado*, the tired traveller), while third person plural subjects (*los viajeros cansados*, the tired travellers) are used for adverb-verb tense manipulations. This choice

was motivated by two methodological reasons. On the one hand, person violations in Spanish are only possible with singular subjects (due to the presence of unagreement patterns with 3<sup>rd</sup> person plural subjects, see Mancini et al. 2011, 2014b), and thus the use of singular subjects allowed us to use the same baseline (“*el viajero cansado*”, the tired traveller) across the three subject-verb agreement conditions<sup>v</sup>. On the other hand, the use of plural-marked subjects and verbs for the adverb-verb manipulations was meant to balance the length of tense correct and incorrect verbs.

(Table 2 here)

Seventy-two filler sentences of different nature were included (number violations with plural subjects, sentences containing unagreement patterns and historical present tense), to balance the proportion of correct and incorrect sentences and vary the type of agreement and tense manipulation. The experimental material was randomly assigned to different lists according to a Latin Square design, so that each subject could see only one version of each sentence. Thus, each subject read 12 sentences in each of the ten experimental conditions in addition to 72 filler sentences, making a total of 192 items.

Before conducting the eye tracking experiment, an offline naturalness judgment task was administered to a different group of 24 participants. Participants were asked to read each sentence and to evaluate its naturalness within a 7-point Likert scale. Results (see Table 3) show that correct sentences were clearly rated as more natural than violations. With respect to the correct version, a trend is evident to consider adverb-initial sentences as more natural (in line with the time frame hypothesis outlined in footnote 3).

(Table 3 about here)

## *Procedure*

Eye-movements were recorded using an SR Eye-Link 1000 machine interfaced with a 19'' CRT Viewsonic monitor (60cm from participants' eyes) in which stimuli were displayed via Experiment Builder Software (SR Research, Ontario, Canada). Participants had binocular vision while movements were measured, but only the right eye was tracked. The experimental room was slightly dimmed to provide a favourable viewing environment. A chin rest bar and a forehead restraint were provided for each participant to minimize head movements. Before the experiment, and whenever necessary during the experiment, the experimenter calibrated the eye-tracker asking participants to fixate 13 positions indicated by a red dot, linearly distributed along the bottom, central and top line of the screen.

Participants initiated each trial by fixating on a red dot on the left side of the screen, where the first word of the sentence would have appeared. Once a fixation in the target region reached a stable value, the entire sentence was displayed. All sentences were presented in a 20-point font (Times New Roman). Participants ended the presentation of each sentence by pressing one of the buttons of the response box. Twenty-five percent of the trials were followed by a comprehension question concerning the content of the sentence just read (e.g. *¿Es su casa el destino de este viaje agotador?*, Is home the final destination of this tiring trip?). Participants answered by pressing either one of the two buttons placed on a response box, corresponding respectively to YES and NO. The experimental session was preceded by 5 practice trials to familiarize the participant with the procedure. Testing sessions lasted approximately 1 hour, including practice, calibration, breaks and debriefing.

### *Data analysis*

Sentences were divided into 5 regions: the adverb/the subject, the verb, the object and the end of the sentence (e.g. prepositional phrases, adverbs, indirect objects). Eye-movements were analyzed at the target region V (the verb in all conditions), at the pre-target region V-1 (the subject or the adverb) and at the post-target region V+1 (the object in all conditions).

We report five measures for each region of interest. As for early latency measures, we analyzed *first pass reading time* which was calculated by summing all fixations on an area of interest before leaving it (either to the left or the right) and *go-past time* (the time spent in reading an area before moving to the right, including any time spent re-reading previous parts of the sentence), as opposed to late measures such as *total time* (the sum of all fixations on an area). In addition to these latency measures, we also report the *probability of regression in and out*. The former represents the probability that a regression is made into a specific area after reading subsequent regions in a sentence (e.g. the probability of rereading the verb region after reading the object region). The latter deals with the probability of exiting a specific region to read previous parts of the sentence (e.g. the probability of exiting the verb region to reread the subject region).

Data for each region of interest were analyzed separately. Individual fixations that were shorter than 80 milliseconds (ms) and longer than 800ms were considered outlier and removed. For the latency measures, values that were higher or lower than 2.5 standard deviations around the mean (separately for each subject and current word, across conditions) were also removed using the recursive procedure with moving criterion developed by Van Selst & Jolicoeur (1994). Overall outlier removal procedures led to 2.7% of removals for the subject-verb relation and 2.9% of removal for the adverb-verb relation in first-pass reading time, 3.1% and 2.9% in go-

past time respectively for subject-verb agreement and adverb-verb agreement, and 2.7% for both agreement relations in total time.

The analysis was carried out fitting linear mixed-effect models (Baayen, Davidson, & Bates, 2008) to our data, using both item and subject as random variables.

Different analyses were run to investigate the effects of a violation depending on the optionality of the controller (optional for adverb-verb and obligatory for subject-verb agreement) and on the anchoring properties of the features under investigation. As a consequence, different fixed effect factors were included in the models depending on the theoretical question we wanted to address.

Firstly, a linear mixed effect model was built including the factors *relation type* (adverb-verb, subject verb), *grammaticality* (match and mismatch, consisting of the average reading times for number and person violations), *configuration* (local, distal), as well as their interaction. This was to assess whether reading times differ during the processing of the verb depending on the type of relation under investigation (relation type) and during the processing of a correct and incorrect sentences (grammaticality). Moreover, we also wanted to test whether the position of the controller (configuration) had any significant impact during the processing of the verb.

To evaluate the effect of different feature anchoring properties during the processing of the verb, a second set of analyses was conducted. In particular, separate comparisons were carried out for person vs. number and for person vs. tense. In the former case, linear mixed effect models were built considering the factor *condition* with three levels, namely control, number (mismatch) and person (mismatch), the factor *configuration* with two levels, local and distal, and the interaction between the two fixed effect factors. The comparison between number and person conditions was carried out by changing the reference level of the intercept from control to

number violations. In the latter case, linear mixed effect models were built considering *condition* (tense, person) *grammaticality* (match, mismatch), *configuration* (distal, local) and the interaction among these fixed effect factors. In each analysis, the best-fitting model was chosen by comparing models whose random-effects structure had a different degree of complexity. In other words, a first model that included only the by-subject random intercept was compared to a model including by-subject and by-item random intercepts, and subsequently, more complex models presenting by-subject and by-item random intercept and slopes were considered, to allow the slope of the fixed factors effect to vary across subject and items. For each pair of models in each dependent variable, the results of the likelihood ratio test were applied to evaluate whether the inclusion of additional random-effects parameters provided a better fit of the model to the data. More complex models were disregarded only if the p-value for the significance of the difference between two models was above .20. P-values for the estimated effects in each model were calculated using the R package *lmerTest* (Kuznetsova, Brockhoff & Christensen, 2015). For the analysis of all the dependent variables in the critical regions (pre-target, target and post-target) a best-fit model was adopted with by-Subject and by-Item random intercepts, unless a more complex model with a by-Subject random slope was required. For the analysis of the probability of regression measure, logistic mixed-effect models were employed (Jaeger, 2008).

For each dependent variable analysed, non-significant or marginally significant results will be reported only if relevant to the issues discussed in the study. Bar plots of mean reading times and probability of regressions in each sentence region are respectively illustrated in Figures 1a and 1b for subject-verb agreement manipulations, while Figure 2a and 2b represent adverb-verb temporal agreement manipulations. For each comparison, we report the intercept,

the estimated regression coefficient (Estimate), standard error (SE) and t/Wald's z values resulting from the linear mixed effects model analysis.

(Figures 1a, 1b and Figures 2a and 2b about here)

## ***Results***

All participants reached at least 75% accuracy on the comprehension questions.

### ***Comparing adverb-verb and subject-verb violations***

A summary of the main results found in this analysis are schematized in Table 4.

#### *Pre-target region*

The analysis of total reading times in the adverb region revealed a marginal effect of grammaticality for tense violations in the local configuration, with longer reading times for ungrammatical compared to grammatical trials (Intercept: 621.42 ms; Estimate: 30.91 ms, SE: 15.40,  $t = 2.01$ ,  $p = 0.05$ ). No grammaticality effects were found on the subject region for subject-verb violations.

#### *Target region*

*First pass.* A two-way interaction *relation type x grammaticality* (Estimate: -52.66 ms, SE: 14.75,  $t = -3.57$ ,  $p < 0.001$ ) and a three-way interaction *relation type x grammaticality x configuration* (Estimate: 45.04 ms, SE: 20.58,  $t = 2.19$ ,  $p < 0.05$ ) were found, showing that in this reading measure, the processing penalty generated by a violation varies as a function of the type of relation and configuration. In fact, when dealing with subject-verb agreement, incorrect



verbs were read more slowly than correct verbs (grammaticality effect) both in the local configuration (Intercept: 341.50 ms; Estimate: 57.93 ms, SE: 10.97,  $t = 5.28$ ,  $p < 0.0001$ ) and in the distal configuration (Intercept: 351.32 ms; Estimate: 33.89 ms, SE: 10.71,  $t = 3.17$ ,  $p < 0.01$ ). Conversely, in the adverb-verb conditions, incorrect verbs were read more slowly compared to correct ones in the distal configuration (Intercept: 345.06 ms; Estimate: 26.27 ms, SE: 9.88,  $t = 2.66$ ,  $p < 0.01$ ) but not in the local one (Intercept: 373.42 ms; Estimate: 5.27 ms, SE: 9.85,  $t = 0.53$ ,  $p = 0.59$ ).

*Go-past.* In the subject-verb agreement conditions, an effect of grammaticality was found both in the local configuration (Intercept: 406.19 ms; Estimate: 39.09 ms, SE: 13.76,  $t = 2.84$ ,  $p < 0.01$ ) and in the distal configuration (Intercept: 372.45 ms; Estimate: 48.19 ms, SE: 13.43,  $t = 3.59$ ,  $p < 0.001$ ), while in the adverb-verb conditions, a grammaticality effect was found only in the distal configuration (Intercept: 376.09 ms; Estimate: 41.44 ms, SE: 12.30,  $t = 3.37$ ,  $p < 0.01$ ) and not in the local one (Intercept: 400.06 ms; Estimate: 15.14 ms, SE: 12.25,  $t = 1.24$ ,  $p = 0.21$ ). However, no significant *relation type*  $\times$  *grammaticality*  $\times$  *configuration* interaction was found in this measure (Estimate: 17.21 ms, SE: 25.57,  $t = 0.67$ ,  $p = 0.5$ ).

*Total.* Longer reading times were found for incorrect verbs compared to the correct ones both in the subject-verb conditions (local: Intercept: 536.65 ms; Estimate: 78.14 ms, SE: 15.86,  $t = 4.93$ ,  $p < 0.0001$ ; distal: Intercept: 435.29 ms; Estimate: 55.24 ms, SE: 15.40,  $t = 3.59$ ,  $p < 0.001$ ) and in the adverb-verb conditions (local: Intercept: 452.73 ms; Estimate: 43.46 ms, SE: 14.03,  $t = 3.10$ ,  $p < 0.01$ ; distal: Intercept: 428.58 ms; Estimate: 64.06 ms, SE: 14.07,  $t = 4.55$ ,  $p < 0.0001$ ).

*Probability of regression out.* No significant interactions or grammaticality effects were found in this measure.

*Probability of regression in.* A higher probability of regression inside the verb region was found for incorrect verbs compared to the correct ones, both in the subject-verb conditions (local: Intercept: -2.38; Estimate: 0.90, SE: 0.17, Wald's  $z = 5.42$ ,  $p < 0.0001$ ; distal: Intercept: -2.14; Estimate: 0.41, SE: 0.16, Wald's  $z = 2.60$ ,  $p < 0.05$ ) and in the adverb-verb conditions (local: Intercept: -2.34; Estimate: 0.57, SE: 0.18, Wald's  $z = 3.09$ ,  $p < 0.01$ ; distal: Intercept: -2.08; Estimate: 0.38, SE: 0.17, Wald's  $z = 2.19$ ,  $p < 0.05$ ).

#### *Post-target region*

*First pass.* A marginal two-way interaction *relation type* x *grammaticality* (Estimate: 27.95 ms, SE: 15.81,  $t = 1.77$ ,  $p = 0.08$ ) and a significant three-way interaction *relation type* x *grammaticality* x *configuration* (Estimate: -51.31 ms, SE: 22.37,  $t = -2.29$ ,  $p < 0.05$ ) were found, showing that the cost of a violation differed as a function of the type of relation and configuration. Reading times on the post-target region were faster in the subject-verb control condition compared to the the subject-verb mismatch condition in the local configuration (Intercept: 501.74 ms; Estimate: -23.36 ms, SE: 10.33,  $t = -2.26$ ,  $p < 0.05$ ), while in the distal configuration no difference between correct and incorrect conditions was found (Intercept: 373.42 ms; Estimate: 5.27 ms, SE: 9.85,  $t = 0.53$ ,  $p = 0.59$ ). On the other hand, no difference between correct and incorrect adverb-verb conditions was found in either configuration (local: Intercept: 487.82 ms; Estimate: 4.58 ms, SE: 11.98,  $t = 0.38$ ,  $p = 0.70$ ; distal: Intercept: 489.55 ms; Estimate: -20.06 ms, SE: 12.03,  $t = -1.67$ ,  $p = 0.10$ ).

*Go-past.* In the subject-verb agreement conditions, there was an effect of grammaticality both in the local configuration (Intercept: 551.43 ms; Estimate: 33.32 ms, SE: 11.70,  $t = 2.85$ ,  $p < 0.01$ ) and in the distal configuration (Intercept: 541.69 ms; Estimate: 26.86 ms, SE: 11.69,  $t = 2.30$ ,  $p < 0.05$ ), while in the adverb-verb conditions no grammaticality effect was found in either configuration (local: Intercept: 542.27 ms; Estimate: 20.43 ms, SE: 13.48,  $t = 1.52$ ,  $p = 0.13$ ; distal: Intercept: 559.07 ms; Estimate: 0.16 ms, SE: 13.58,  $t = 0.01$ ,  $p = 0.99$ ).

*Total.* At the post-target region, longer reading times were found in the subject-verb mismatch condition compared to the control one in the distal configuration (Intercept: 587.72 ms; Estimate: 28.60 ms, SE: 12.70,  $t = 2.25$ ,  $p < 0.05$ ), while no effect of grammaticality was found in the local configuration (Intercept: 609.03 ms; Estimate: 12.72 ms, SE: 12.72,  $t = 1.00$ ,  $p = 0.32$ ).

Conversely, in the adverb-verb conditions, marginally longer reading times for the incorrect conditions were found in the local configuration (Intercept: 579.42 ms; Estimate: 26.54 ms, SE: 14.75,  $t = 1.80$ ,  $p = 0.07$ ) while no effect of grammaticality was found in the distal one (Intercept: 605.55 ms; Estimate: -1.01 ms, SE: 14.81,  $t = -0.07$ ,  $p = 0.95$ ).

*Probability of regression out.* No significant interactions were found in this measure, although a higher probability of regression out of the post-target region was found in both configurations of subject-verb agreement (local: Intercept: -2.37; Estimate: 0.65, SE: 0.16, Wald's  $z = 4.13$ ,  $p < 0.0001$ ; distal: Intercept: -2.27; Estimate: 0.34, SE: 0.16, Wald's  $z = 2.19$ ,  $p < 0.05$ ) but only in the local configuration of adverb-verb agreement (Intercept: -2.49; Estimate: 0.43, SE: 0.19, Wald's  $z = 2.32$ ,  $p < 0.05$ ).

*Probability of regression in.* No significant interactions or grammaticality effects were found in this measure.

(Table 4 here)

### ***Comparing number and person violations***

A summary of the main results is presented in Table 5.

#### *Pre-target region*

In this region, no significant differences among the control and the mismatch conditions were found.

#### *Target region*

*First-pass.* In the local configuration, number mismatches (Intercept: 338 ms; Estimate: 47.1 ms, SE: 12.13,  $t = 3.88$ ,  $p < 0.001$ ) and person mismatches (Intercept: 338 ms; Estimate: 62.67 ms, SE: 11.86,  $t = 5.29$ ,  $p < 0.001$ ) gave rise to longer reading times with respect to the control condition. Similarly, in the distal configuration longer first-pass reading times were found for number mismatches (Intercept: 345.42 ms; Estimate: 32.86 ms, SE: 11.84,  $t = 2.76$ ,  $p < 0.01$ ) and person mismatches (Intercept: 345.42 ms; Estimate: 35 ms, SE: 11.66,  $t = 3$ ,  $p < 0.01$ ) with respect to control. No differences in reading times and no interaction *condition x configuration* were found.

*Go-past.* Number mismatches led to marginally longer reading times compared to control (Intercept: 401.72 ms; Estimate: 25.37 ms, SE: 15.21,  $t = 1.67$ ,  $p = 0.1$ ) while person mismatches

gave rise to significantly longer reading times with respect to the control condition (Intercept: 401.72 ms; Estimate: 45.83 ms, SE: 14.83,  $t = 3.1$ ,  $p < 0.05$ ) in the local configuration. Similarly, longer go-past reading times both for number (Intercept: 368.50 ms; Estimate: 48.57 ms, SE: 14.85,  $t = 3.27$ ,  $p = 0.001$ ) and person (Intercept: 368.50 ms; Estimate: 43.14 ms, SE: 14.6,  $t = 2.96$ ,  $p < 0.05$ ) compared to control, emerged in the distal configuration. No differences in reading times and no interaction *condition x configuration* were found.

*Total.* Both number mismatches (Intercept: 425.72 ms; Estimate: 59.14 ms, SE: 17.55,  $t = 3.37$ ,  $p < 0.001$ ) and person mismatches (Intercept: 425.72 ms; Estimate: 99.66 ms, SE: 17.15,  $t = 5.81$ ,  $p < 0.001$ ) gave rise to longer reading times with respect to the control condition, when the subject and the verb were adjacent (local). Similarly, when the subject was located at the beginning of the sentence (distal configuration), longer total reading time for number mismatches (Intercept: 428 ms; Estimate: 50.35 ms, SE: 17.02,  $t = 2.96$ ,  $p < 0.01$ ) and person mismatches (Intercept: 428 ms; Estimate: 56.61 ms, SE: 16.74,  $t = 3.38$ ,  $p = 0.001$ ) were found compared to control. Moreover, in the local configuration person mismatches led to significantly longer reading times with respect to number mismatches (Intercept: 484.86 ms; Estimate: 40.51 ms, SE: 15.26,  $t = 2.65$ ,  $p = 0.01$ ). No interaction *condition x configuration* was found when comparing number and person violations, however the difference in reading times that showed up between number and person mismatches in the local configuration was no longer present in the distal configuration (Intercept: 478.35; Estimate: 6.26, SE: 15.07,  $t = 0.42$ ,  $p = 0.68$ ). In particular, person anomalies in the distal configuration led to significantly smaller reading times with respect to person anomalies in the local configuration (Intercept: 525.37; Estimate: -40.77, SE: 14.85,  $t = -2.75$ ,  $p < 0.01$ ).

*Probability of regression (out).* In the local configuration, only number mismatches significantly differed from the control condition (Intercept: -2.14; Estimate: -0.67, *SE*: 0.27, Wald's  $z = -2.49$ ,  $p = 0.01$ ) while no differences among the three experimental conditions were found in the distal configuration.

*Probability of regression (in).* In the local configuration, a higher probability of regression into the target region was found for person mismatches compared to control (Intercept -2.55; Estimate: 1.14, *SE*: 0.23, Wald's  $z = 4.89$ ,  $p < 0.001$ ) and for number mismatches compared to control (Intercept: -2.55; Estimate: 0.97, *SE*: 0.24, Wald's  $z = 4.05$ ,  $p < 0.001$ ). Conversely, in the distal configuration the difference between control and person violations was only marginal (Intercept -2.09; Estimate: 0.38, *SE*: 0.21, Wald's  $z = 1.83$ ,  $p = 0.07$ ), while no difference between control and number violations showed up in the distal configuration. Finally, no differences in reading times and no interaction *configuration x condition* were found in this measure.

#### *Post-target region*

*First-pass.* In the local configuration, the number mismatch condition gave rise to shorter reading times compared to the control condition (Intercept: 496.17 ms; Estimate: -39.26 ms, *SE*: 12.25,  $t = -3.21$ ,  $p < 0.01$ ), while person mismatches did not differ from control. Also, reading times for number were faster than for person mismatches (Intercept: 456.91 ms; Estimate: 40.45 ms, *SE*: 12.20,  $t = 3.32$ ,  $p < 0.001$ ). In contrast, in the distal configuration the three conditions elicited equivalent reading times. An interaction *condition x configuration* was found when comparing number and person violations (Estimate: -37.35, *SE*: 17.24,  $t = -2.16$ ,  $p < 0.05$ ). In

particular, number violations were marginally faster in the local configuration than in the distal one<sup>vi</sup> (Estimate: 456.91 ms, Estimate: 24.03 ms, SE: 12.18,  $t = 1.97$ ,  $p = 0.05$ ).

*Go-past.* In both configurations, person mismatches showed longer reading times with respect to the control (local: Intercept: 548.56 ms; Estimate: 48.16 ms, SE: 13.54,  $t = 3.56$ ,  $p < 0.001$ ; distal: Intercept: 540.29 ms; Estimate: 44.63 ms, SE: 13.55,  $t = 3.29$ ,  $p = 0.001$ ), while no difference was found between the control condition and number violations in either configuration. No interaction *condition x configuration* was found when the reference level of the model was the number condition (Estimate: -6.90, SE: 19.17,  $t = -0.36$ ,  $p = 0.71$ ) although longer reading times were found for person mismatches compared to number mismatch conditions in both configurations (local: Intercept: 568.83 ms; Estimate: 27.88 ms, SE: 13.55,  $t = 2.06$ ,  $p < 0.05$ ; distal: Intercept: 550.13 ms; Estimate: 34.78 ms, SE: 13.56,  $t = 2.57$ ,  $p < 0.05$ ).

*Total.* In the local configuration, person anomalies elicited longer reading times compared to the control condition (Intercept: 606.6 ms; Estimate: 40.9 ms, SE: 14.8,  $t = 2.76$ ,  $p = 0.01$ ), while there was no significant difference between control and number mismatches. Moreover, person mismatch reading times were found to be significantly longer with respect to number mismatch (Intercept: 594.37 ms; Estimate: 53.14 ms, SE: 14.79,  $t = 3.59$ ,  $p < 0.001$ ). In the distal configuration, number mismatches only marginally differed from control (Intercept: 587.34 ms; Estimate: 25.30 ms, SE: 14.75,  $t = 1.72$ ,  $p = 0.09$ ) as well as from person mismatches (Intercept: 587.34 ms; Estimate: 28.43 ms, SE: 14.80,  $t = 1.92$ ,  $p = 0.06$ ). No difference between person and number mismatches emerged in the distal configuration (Intercept: 612.64 ms; Estimate: 3.14 ms, SE: 14.78,  $t = 0.21$ ,  $p = 0.83$ ). The different impact that person and number violations had on

this independent variable led to a significant *condition x configuration* interaction (Estimate: 50.01 ms, SE: 20.91,  $t = 2.39$ ,  $p < 0.05$ ).

*Probability of regression out.* In the local configuration, number mismatches led to higher probabilities of regression out of the post-target region with respect to the control condition (Intercept: -2.11; Estimate: 0.53, SE: 0.17, Wald's  $z = 3.10$ ,  $p < 0.01$ ). The same pattern was found for person mismatches compared to control both in the local configuration (Intercept -2.11; Estimate: 0.67, SE: 0.17, Wald's  $z = 3.97$ ,  $p < 0.001$ ) and in the distal configuration (Intercept: 0.13; Estimate: 0.06, SE: 0.02, Wald's  $z = 2.74$ ,  $p < 0.01$ ). No interaction *conditionxconfiguration* was found in this measure, although when the subject was sentence-initial a higher probability of regression out was found for person mismatches compared to number (Intercept: 0.15; Estimate: 0.05, SE: 0.02, Wald's  $z = 2.07$ ,  $p < 0.05$ ).

*Probability of regression in.* The comparison between number mismatch and control revealed a higher probability of regression in the post-target area for the former condition only in the local configuration (Intercept: -2.22; Estimate: -0.44, SE: 0.19, Wald's  $z = -2.27$ ,  $p < 0.05$ ). In contrast, person mismatches did not significantly differ from the control condition in either configuration. This led to a marginally significant interaction *conditionxconfiguration* (Estimate: -0.50, SE: 0.27, Wald's  $z = -1.81$ ,  $p = 0.07$ ).

(Table 5 here)

### ***Comparing tense and person violations***

A summary of the main results is presented in Table 6.



### *Pre-target region*

A marginal effect of grammaticality was found in total reading time for tense violations at the adverb region (Intercept: 621.42 ms; Estimate: 30.91 ms, SE: 15.40,  $t = 2.01$ ,  $p = 0.05$ ), while no grammaticality effects were found at the subject region for person violations.

### *Target region*

*First pass.* A two-way interaction *feature type x grammaticality* (Estimate: 61.63 ms, SE: 15.39,  $t = 4.01$ ,  $p < 0.0001$ ), and three-way interaction *feature type x grammaticality x configuration* (Estimate: -52.71 ms, SE: 21.56,  $t = -2.45$ ,  $p < 0.05$ ) were found in this measure, showing that the cost of a violation was different for the two features under computation, depending on the configuration between the two critical constituents. Longer reading times were found for tense-incorrect verbs compared to the correct ones (grammaticality effect) only in the distal configuration (Intercept: 342.32 ms; Estimate: 25.96 ms, SE: 10.05,  $t = 2.58$ ,  $p < 0.01$ ).

Conversely, person-incorrect verbs were read more slowly than correct verbs both in the local configuration (Intercept: 339.40 ms; Estimate: 65.56 ms, SE: 11.68,  $t = 5.61$ ,  $p < 0.0001$ ) and in the distal configuration (Intercept: 349.02 ms; Estimate: 34.88 ms, SE: 11.47,  $t = 3.04$ ,  $p < 0.01$ ).

*Go-past.* A marginal interaction *feature type x grammaticality* was found (Estimate: 34.64 ms, SE: 19.23,  $t = 1.80$ ,  $p = 0.07$ ) meaning that the effect of grammaticality is marginally different for the two features under computation. In particular, a grammaticality effect for tense violations was found only in the distal configuration (Intercept: 373.49 ms; Estimate: 41.06 ms, SE: 12.53,  $t = 3.28$ ,  $p < 0.01$ ), while a grammaticality effect for person violations was found both in the local configuration (Intercept: 406.19 ms; Estimate: 39.09 ms, SE: 13.76,  $t = 2.84$ ,  $p < 0.01$ ) and in the distal one (Intercept: 404.13 ms; Estimate: 48.21 ms, SE: 14.63,  $t = 3.30$ ,  $p < 0.01$ ).

*Total.* A two-way interaction *feature type* x *grammaticality* (Estimate: 56.04 ms, SE: 22.00,  $t = 2.55$ ,  $p < 0.05$ ), and three-way interaction *feature type* x *grammaticality* x *configuration* (Estimate: -63.05 ms, SE: 30.59,  $t = -2.06$ ,  $p < 0.05$ ) were found in this measure, showing that the cost of a violation was different for the two features under computation, depending on the configuration between the two critical constituents. In particular, longer reading times were found for incorrect verbs compared to the correct ones in both configurations for tense violations (local: Intercept: 451.17 ms; Estimate: 41.09 ms, SE: 14.20,  $t = 2.89$ ,  $p < 0.01$ ; distal: Intercept: 425.75 ms; Estimate: 64.84 ms, SE: 14.24,  $t = 4.55$ ,  $p < 0.0001$ ) and for person violations (local: Intercept: 434.44 ms; Estimate: 97.13 ms, SE: 16.80,  $t = 5.78$ ,  $p < 0.0001$ ; distal: Intercept: 431.94 ms; Estimate: 57.83 ms, SE: 16.41,  $t = 3.52$ ,  $p < 0.001$ ). However, the grammaticality effect for person violations significantly decreased in the distal configuration compared to the local one (Intercept: 531.58 ms; Estimate: -41.81 ms, SE: 13.94,  $t = -3.00$ ,  $p < 0.01$ ) while a similar grammaticality effect was found for tense violations in either configuration.

*Probability of regression out.* No significant interactions or grammaticality effects were found in this measure.

*Probability of regression in.* No significant interactions were found in this measure.

Independently from the feature and the configuration under computation, a higher probability of regression inside the verb region was found for incorrect verbs compared to the correct ones both for tense violations (local: Intercept: -2.34; Estimate: 0.57, SE: 0.18, Wald's  $z = 3.10$ ,  $p < 0.01$ ; distal: Intercept: -2.08; Estimate: 0.38, SE: 0.17, Wald's  $z = 2.18$ ,  $p < 0.05$ ) and for person violations (local: Intercept: -2.38; Estimate: 0.99, SE: 0.18, Wald's  $z = 5.49$ ,  $p < 0.0001$ ; distal: Intercept: -2.15; Estimate: 0.50, SE: 0.17, Wald's  $z = 2.85$ ,  $p < 0.01$ ).

*Post-target region*

*First pass.* No significant interactions or grammaticality effects were found in this measure.

*Go-past.* No significant interactions were found and no grammaticality effect was found for tense violations in either configuration (local: Intercept: 542.11 ms; Estimate: 20.11 ms, SE: 13.48,  $t = 1.49$ ,  $p = 0.14$ ; distal: Intercept: 559.04 ms; Estimate: 0.07 ms, SE: 13.59,  $t = 0.01$ ,  $p = 0.1$ ), while there was an effect of grammaticality for person violations in both configurations (local: Intercept: 551.42 ms; Estimate: 48.94 ms, SE: 13.57,  $t = 3.61$ ,  $p < 0.001$ ; distal: Intercept: 541.96 ms; Estimate: 43.95 ms, SE: 13.60,  $t = 3.23$ ,  $p < 0.01$ ).

*Total.* Non-significant interactions were found in this measure. However, tense violations showed no grammaticality effects while person violations caused longer reading times compared to correct conditions in the local configuration (Intercept: 609.66 ms; Estimate: 38.89 ms, SE: 14.57,  $t = 2.67$ ,  $p < 0.05$ ) and in the distal configuration (Intercept: 587.73 ms; Estimate: 28.34 ms, SE: 14.58,  $t = 1.94$ ,  $p = 0.05$ ).

*Probability of regression out.* No significant interactions were found in this measure, although a higher probability of regression out of the post-target region was found in the incorrect conditions compared to the correct ones, only in the local configuration for tense violations (Intercept: -2.50; Estimate: 0.43, SE: 0.19, Wald's  $z = 2.33$ ,  $p < 0.05$ ) and in both configurations for person violations (local: Intercept: -2.38; Estimate: 0.72, SE: 0.17, Wald's  $z = 4.15$ ,  $p < 0.0001$ ; distal: Intercept: -2.28; Estimate: 0.52, SE: 0.17, Wald's  $z = 2.30$ ,  $p < 0.01$ ).

*Probability of regression in.* No significant interactions or grammaticality effects were found in this measure.

(Table 6 here)

## **Discussion**

The main goal of the current study was to investigate whether verb inflection was similarly or differently processed during sentence comprehension, depending on the different types of agreement relations the verb is engaged in, the different features involved during verb processing, and the linear/structural distance between the verb and its controller.

On the one hand, this data on number and person processing showed that the disruption of the subject-verb relation gives clear and sustained parsing costs at the target region and at the post-target region, both in the local and in the distal configuration, across all measures. Interestingly, the processing of a person mismatch resulted in larger costs compared to that of a number mismatch. On the other hand, the data on the adverb-verb processing showed that adverb-verb mismatches led to numerically smaller parsing costs with respect to subject-verb mismatches and these emerged consistently in the total reading time of the target region. In other words, in the (local) configuration in which the three (number, person and tense) violations were compared as strictly as possible, a processing difference was found both at the relational level (subject-verb, adverb-verb) and at the feature level (number, person and tense).

Moreover, the magnitude of this differential effect of verb inflection violations was also found to be sensitive to the linear/structural distance between the verb and its related constituents. Larger parsing costs appeared for person compared to number violations at the target region (in total reading time) and post-target region (in first-pass, total reading time) when the subject was adjacent to the verb. In contrast, in the distal configuration, the difference between person and number violation penalties was shallower (i.e. statistically emerged only in the go-past duration of the post-target region).

As for tense violations, while in the local configuration the mismatch effect showed up only in measures that include re-reading of the target region (total time), in the distal configuration the mismatch effect appeared from early measures on.

We will firstly discuss data from the local configuration (adjacent verb and controller), which allow us to discuss whether a relation-based approach (Frazier & Clifton 1996, Chomsky 1995, 2000) or a feature-based approach (Carminati, 2005; Mancini et al. 2013) is more suitable to describe these and other data available in the literature, without the interference of other relevant factors such as the distance between the target and the controller.

Subsequently, data from the distal configuration (distal verb and controller) will be discussed, providing a speculative implementation of the anchoring-approach (Mancini et al. 2013) and the factors affecting the discourse anchoring process during sentence comprehension.

### ***Optionality and anchoring in local relations***

When the subject and the verb were adjacent, we found immediate and sustained parsing costs in early and late measures for both number and person mismatches compared to the correct agreement condition. These data support previous accounts (Mancini et al. 2011, 2013, 2014a) claiming that number and person violations led to similar parsing costs in early stages of processing because of the disruption of the same morpho-syntactic feature checking mechanism. Furthermore, the immediate effect of the disruption is here interpreted as a function of the non-optionality of the controller in the subject-verb dependency. When the verb contains agreement features which are inconsistent with the features expressed by the mandatory subject, the parser immediately faces the incongruence, independently from the type of feature under computation. By contrast, there was a difference in the analysis of number and person anomalies in the

spillover region, with greater parsing difficulties for person anomalies. The data thus converge, in that different interpretive properties of the two features, namely the cardinality of the subject and the role of speech participants, are differently processed in later stages of processing, in line with the feature-anchoring approach proposed in Mancini et al. (2013). A discrepancy in defining the discourse role of the subject in the context of the utterance generates greater cognitive penalty compared to an inconsistency in the cardinality of the subject of the sentence, as already shown by previous studies (Mancini et al. 2014b).

Tense anomalies also caused longer reading times than grammatical control conditions in late measures (total time), with larger rereading costs on the verb and on the adjacent adverb. In other words, at the same stage in which person anomalies started to differ from number anomalies, tense anomalies also started showing a cost. This parallelism in the processing of person and tense anomalies in late measures is in line with theoretical claims proposing a similar implication of discourse-related information in the interpretation of the two features (Bianchi 2003, Sigurðsson 2004). However, the processing of subject-verb and adverb-verb anomalies differed in the first stages, since no parsing costs arose in early measures for tense violations. We attribute this difference to the optionality of the adverb and the related, plausible lack of a formal feature checking procedure in this case, which leads the parser to deal with the discourse-related congruency between the verb and the adverb at later stages of processing.

We hypothesized that either optionality or feature-related interpretive properties could independently impact the processing of a morpho-syntactic violation at the verb. Our data on the processing of number, person and tense between the verb and its adjacent controller show that both dimensions matter during processing and determine distinct effects: the optionality of the controller plays a clear role in the early stage (first pass reading), when feature verification is

carried out, while the discourse-relatedness of the features plays a clear role in later stages (go past and total reading time), when morphosyntactic information is mapped onto higher levels of analysis such as the speech act representation.

In general, the reading pattern emerging from the local configuration is in line with theoretical accounts that postulate different structural positions and interpretive properties for each feature (cf. Bianchi 2003, Sigurðsson 2013, Carminati 2005, Mancini et al. 2013). Data also suggest that theoretical accounts that have only considered a distinction at the relational level, proposing a binary opposition between subject-verb agreement and other linguistic dependencies (cf. Chomsky 1995, 2000, 2001; Frazier & Clifton 1996) should be detailed with a richer description of the agreement phenomena. These findings should be thus explicitly considered within parsing models that have not yet formalized a dependency of parsing mechanisms on both relational and interpretive aspects of agreement.

### ***Optionality and anchoring in distal relations***

The interpretation of our data from the distal configuration is less straightforward. Similarly to the local configuration, distal subject-verb violations trigger a rapid mismatch detection effect. Yet, when subject and verb are distally located the difference between person and number emerges even later, as evidenced by the go-past effect in the spillover region. In contrast, the distal location of the adverb with respect to the verb yielded a faster tense mismatch effect compared to the local configuration, showing larger parsing costs for a tense violation from early measures on. In other words, subject-verb and adverb-verb agreement appear to be sensitive to the configuration of the controller-target relation, but in opposite ways: distance

delays the difference in reading times between number and person violations while it fastens the emergence of tense mismatch effects.

This pattern of data can be only partially explained by existing psycholinguistic models that predict a unique mechanism of integration/retrieval during the processing of different agreement relations. In particular, memory-based models of sentence processing would predict that the greater the distance between controller and target, the longer the reading times at the target position because of storage/integration costs (e.g. Gibson 1998, Grodner & Gibson 2005) or because of reactivation difficulty after decay (e.g. Lewis & Vasishth, 2005). Yet, while this explanation can account for the increase in costs for tense anomalies in the distal configuration, it cannot explain the minor disruption generated by person subject-verb violations in late measures when linear distance increases. Overall, it appears that explanations that rely on memory based, non-syntactic properties of parsing inevitably fail to account for the full set of data reported here.

These two findings can be only partially accounted for even under the assumption that subject-verb and adverb-verb agreement are two relations different in nature (primary and non-primary, respectively), as held by a relation-based approach like the Construal model (Frazier & Clifton, 1996). Crucially, this distinction could account for the overall differential pattern evidenced for subject-verb and adverb-verb agreement obtained in the local configuration, but would fall short of an explanation both for the modulation of tense mismatch effects across configuration types, and of the person-number difference that we observe, as this model does not take into consideration any feature-level analysis.

In particular, any model of parsing should be able to predict: i) why the effect of an optional controller, like a temporal adverb, arises earlier and more strongly when this is distally



located with respect to its target; ii) why a person mismatch yields smaller processing disruptions when the obligatory controller (i.e. subject) and its target are distally located.

A better framework for the discussion of these data can be provided by feature-anchoring accounts (Bianchi, 2006; Mancini et al. 2013; Sigurðsson, 2004), which assume independent (and often qualitatively different) anchoring mechanisms for person, number and tense features. Importantly, the data here reported can significantly contribute to widen the scope of these proposals, including a more precise formalization of tense feature processing. These data also underline the impact that linear and structural distance can have in on-line interpretation mechanisms.

Focusing on the interpretive mechanisms licensing features, Mancini and colleagues (2013) have proposed the presence of links – or anchoring relations – between the morpho-syntactic and the discourse representation of the sentence. For a proper feature interpretation, a match must be established between the controller and the target morpho-syntactic values (i.e. 1st, 2nd, or 3rd person, singular or plural number), but also between these and their respective “anchors”, i.e. the semantic-discourse content of feature (cardinality vs. discourse role of speech participants), which can be located respectively within the syntactic structure of the clause (i.e. the inflection layer) or at the more peripheral discourse representation of the clause (i.e. left periphery). Critically, the two features differ not only in the position of their respective anchors, but also in the complexity of the anchoring mechanisms. On the one hand, number interpretation relies on only one anchoring mechanism (i.e. the one linking verb number specification to the subject semantic representation). On the other hand, person interpretation requires that both subject and verb person specifications are anchored to the discourse representation of the sentence, with this multiple anchoring process being motivated by the presence across languages

of grammatical person mismatches between subject and verb (i.e. unagreement patterns, see Mancini et al. 2011, 2013 for a thorough description of this phenomenon and its anchoring procedures). Given the similar discourse-related properties of person and tense, these two features plausibly share equivalent multiple-anchoring mechanisms.

Under the assumption that sentence comprehension proceeds in an incremental way, upon encountering any element bearing a discourse-related feature (be it a potential controller or potential target), the parser will access its content and initiate an anchoring process through which the feature will be linked to the discourse representation, with interpretively relevant roles (e.g. speaker, addressee) and speech time being determined. Crucially, accessing deictic information and anchoring it to discourse is likely to be a slow process (see Kreiner, Garrod & Sturt 2013 for a similar assumption related to the analysis of the notional information of collective nouns). Under the hypothesis that anchoring mechanisms operate in a cascaded fashion, it is therefore possible that the triggering of an anchoring operation, for example on the verb, is not contingent on the completion of previous anchoring of the controller, but only on the extraction of discourse-related features from the linguistic input. In this scenario, the linear distance separating controller and target is expected to play a major role in shaping the reading and comprehension correlates: the greater the distance between the two elements, the more the time available to the parser to solidly anchor deictic information to discourse before the same process is initiated at verb position.

As noted in the Introduction, the morphosyntactic expression of person and tense features is anchored to the discourse representation of the sentence, which implies a clause-external link to the left periphery of sentence structure (Bianchi 2003, Sigurðsson 2004, Mancini et al. 2013). When controller and target are adjacent, the parser does not have time to complete and

consolidate the anchoring of the subject/adverb deictic features to this external position before the same process is triggered at verb position. As a consequence, any discourse inconsistency between the subject/adverb's and the verb's deictic content is detected only at later stages. A different scenario arises for number agreement. Because number's anchor is located clause-internally in the semantic representation of the subject, a more rapid anchoring of the semantic information extracted from the controller can be hypothesized. Arguably, during the processing of a mismatching verb, the parser already has access to a definite and solid representation of the subject's cardinality and it can thus easily detect and repair any inconsistency, regardless of the distance that separates the controller from the target.

In other words, whether a feature triggers internal or external anchoring has crucial processing consequences. A mismatch that involves inspection of an externally located anchor, as happens for person and tense violations, is likely to result in greater processing costs compared to internally anchored mismatches in later reading variables. This would explain the difference between person and number in total reading times and the late emergence of tense mismatch effects in the local configuration, but also the similar reading correlates of number violations across configurations.

In a distal controller-target configuration, the linear distance separating the subject/adverb from the verb gives the parser enough time to solidly anchor the controller's deictic information to discourse.

When a person or tense mismatching verb is encountered and anchoring is triggered, the parser can therefore rely on a solid representation of the controller. This would explain why, in the distal configuration, the difference between person and number violations in total reading times disappears. Along similar lines, the increased linear distance between adverb and verb

allows solid anchoring of the adverb's temporal information to discourse, which in turn determines earlier detection of the mismatch and thus the alignment of tense with person and number processing.

## **Conclusion**

The empirical data here presented clearly show that sentence and feature processing cannot be simplistically modelled in a unique and homogeneous mechanism, as already pointed out in Mancini et al. (2014b). Similarly, treating feature processing differences by postulating differences in strength or salience (Carminati, 2005) is also very far from explaining the complexity emerging from the interaction between linear distance and type of feature that our data show.

Concretely, the framework here proposed rests on a cognitive architecture in which bottom-up automatic syntactic processing routines depend on the type of syntactic relation being processed, similarly to that proposed by the Construal model (Frazier & Clifton, 1996) and eADM models (Bornkessel & Schlesewsky, 2006). Yet, at the same time, it capitalizes on the flexible way in which the anchoring of morphosyntactic features to higher-level non-syntactic information (discourse) can actually modify the processing of different syntactic relations.

It could possibly be argued that the scarce literature on the study of feature compatibility checking during sentence comprehension in a wider domain of phenomena makes it difficult to predict which factors determine the speed and efficiency of this anchoring process - most of the empirical studies on agreement refer to subject-verb number agreement, which is indeed the exemplar reference for feature checking also within our approach. In our study, we assume that linear/structural distance has an effect because it gives the system more time to efficiently

conclude the anchoring process, but we cannot exclude that this process may be influenced by other, non-syntactic, contextual factors such as the task, the presence of a larger discourse context or syntactic directional effects (linear order between controller and target), which were not addressed here. For example, optional constituents such as temporal adverbs are not predictable. In other words, if an adverb is presented first in the sentence, the temporal features encoded by verb inflection may be predicted. Conversely, if the verb is presented first in the sentence, a temporal adverb is not necessarily predicted because of its optionality. Moreover, the content of the temporal adverb cannot be fully predicted. Indeed, a past verb such as “played” is compatible with different temporal adverbs such as “yesterday, last year, last week”. Future research will primarily be aimed at testing the robustness of the anchoring mechanism, and at explicitly defining its formal properties.

### **Acknowledgments**

S.M acknowledges funding from the Gipuzkoa Fellowship Program, from grants FFI2016-76432 and SEV-2015-490 (Severo Ochoa Programme for Centres/Units of excellence in R&D awarded by the Spanish Ministry of Industry, Economy and Competitiveness. L.R. acknowledges support from the ERC AG n. 340297 “SynCart”. The authors thank Margaret Dowens for reviewing the manuscript and giving useful suggestions.

### **Disclosure of interest**

The authors report no conflicts of interest.

## References

- Alexiadou, A. (1997). *Adverb placement: A case study in antisymmetric syntax* (Vol. 18). John Benjamins Publishing.
- Alexiadou, A. (2013). Adverbial and adjectival modification. *The Cambridge handbook of generative syntax*, 458-484.
- Altmann, G. T., van Nice, K. Y., Garnham, A., & Henstra, J. A. (1998). Late closure in context. *Journal of Memory and Language*, 38(4), 459-484.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4), 390-412.
- Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, 17(1), 137-153.
- Baggio, G. (2008). Processing temporal constraints: An ERP study. *Language Learning*, 58(s1), 35-55.
- Benincà, P., & Poletto, C. (2004). Topic, focus, and V2. *The Structure of CP and IP*, 52-75.
- Bianchi, V. (2003). On finiteness as logophoric anchoring. *Temps et point de vue/Tense and point of view*, 213-246.
- Bianchi, V. (2006). On the syntax of personal arguments. *Lingua*, 116(12), 2023-2067.
- Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological review*, 113(4), 787.
- Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). The role of prominence information in the real time comprehension of transitive constructions: A crosslinguistic

approach. *Language and Linguistics Compass*, 3, 19–58.

Carminati, M. N. (2005). Processing reflexes of the Feature Hierarchy (Person> Number> Gender) and implications for linguistic theory. *Lingua*, 115(3), 259-285.

Chafe, W. (1984, October). How people use adverbial clauses. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 10), 437-449.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT press.

Chomsky, N. (2000). Minimalist inquiries: The framework. *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. by Roger Martin, David Michaels, and Juan Uriagereka, 89–155.

Cinque, G. (1999). *Adverbs and functional heads: A cross-linguistic perspective*. Oxford University Press.

Corbett, G. G. (2003). Agreement: terms and boundaries. In *The Role of Agreement in Natural Language*. Proceedings of the 2001 Texas Linguistic Society Conference, Austin, Texas, 109-122.

Deutsch, A. (1998). Subject-predicate agreement in Hebrew: Interrelations with semantic processes. *Language and cognitive processes*, 13(5), 575-597.

Deutsch, A., & Bentin, S. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew: Evidence from ERPs and eye movements. *Journal of Memory and Language*, 45(2), 200-224.

De Vincenzi, M., Rizzi, L., Portolan, D., Di Matteo, R., Spitoni, G., & Di Russo, F. (unpublished). Mapping the language: A reading time and topographic ERP study on tense, agreement, and Aux-V violations. Ms., University of Chieti.

- Dickey, M. W. (2001). *The processing of tense: Psycholinguistic studies on the interpretation of tense and temporal relations* (Vol. 28). Springer Science & Business Media.
- Enç, M. (1987). Anchoring conditions for tense. *Linguistic inquiry*, 633-657.
- Fonteneau, E., Frauenfelder, U. H., & Rizzi, L. (1998). On the contribution of ERPs to the study of language comprehension. *Bulletin suisse de linguistique appliquée*, 68, 111-124.
- Frazier, L., & Clifton, C. (1996). *Construal*. Mit Press.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1), 1-76.
- Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. *Universals of language*, 2, 73-113.
- Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive science*, 29(2), 261-290.
- Hagoort, P. (2003). How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *Neuroimage*, 20, S18-S29.
- Hagoort, P. (2013). MUC (memory, unification, control) and beyond. *Frontiers in Psychology*, 4, 416.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language*, 59(4), 434-446.
- Kreiner, H., Garrod, S., & Sturt, P. (2013). Number agreement in sentence comprehension: The relationship between grammatical and conceptual factors. *Language and Cognitive Processes*, 28(6), 829-874.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmerTest'. R package version, 2-0.



Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive science*, 29(3), 375-419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in cognitive sciences*, 10(10), 447-454.

Mancini, S., Molinaro, N., & Carreiras, M. (2013). Anchoring agreement in comprehension. *Language and Linguistics Compass*, 7(1), 1-21.

Mancini, S., Molinaro, N., Davidson, D. J., Avilés, A., & Carreiras, M. (2014b). Person and the syntax–discourse interface: An eye-tracking study of agreement. *Journal of Memory and Language*, 76, 141-157.

Mancini, S., Molinaro, N., Rizzi, L., & Carreiras, M. (2011). A person is not a number: Discourse involvement in subject–verb agreement computation. *Brain research*, 1410, 64-76.

Mancini, S., Postiglione, F., Laudanna, A., & Rizzi, L. (2014a). On the person-number distinction: Subject-verb agreement processing in Italian. *Lingua*, 146, 28-38.

Mancini, S., Quiñones, I., Molinaro, N., Hernandez-Cabrera, J. A., & Carreiras, M. (2017). Disentangling meaning in the brain: Left temporal involvement in agreement processing. *Cortex*, 86, 140-155.

Molinaro, N., Vespignani, F., & Job, R. (2008). A deeper reanalysis of a superficial feature: An ERP study on agreement violations. *Brain Research*, 1228, 161-176.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and language*, 41(3), 427-456.

Rispens, J., & de Amesti, V. S. (2016). What makes syntactic processing of subject–verb agreement complex? The effects of distance and additional agreement. *Complexity in human languages: A multifaceted approach*, 30, 2588.

Qiu, Y., & Zhou, X. (2012). Processing temporal agreement in a tenseless language: An ERP study of Mandarin Chinese. *Brain research, 1446*, 91-108.

Sagarra, N. (2008). Working memory and L2 processing of redundant grammatical forms. *Understanding second language process*, 133-147.

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, & Psychophysics, 74*(1), 5-35.

Sybesma, R. (2007). Whether we tense-agree overtly or not. *Linguistic Inquiry, 38*(3), 580-587.

Sigurðsson, H. Á. (2004). The syntax of Person, Tense, and speech features. *Rivista di Linguistica-Italian Journal of Linguistics, 16*(1), 219-251.

Sigurðsson, H. Á. (2016). The split T analysis. I: Kristin Melum Eide (red.): Finiteness Matters. *On finiteness-related phenomena in natural languages. Amsterdam. S*, 79-92.

Silva-Pereyra, J. F., & Carreiras, M. (2007). An ERP study of agreement features in Spanish. *Brain Research, 1185*, 201-211.

Steinhauer, K., & Ullman, M. T. (2002, October). Consecutive ERP effects of morpho-phonology and morpho-syntax. *Brain and Language, 83*(1), 62-65.

Stepanov, A. (2001). Late adjunction and minimalist phrase structure. *Syntax, 4*, 94-125.

Van Gompel, R. P., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language, 52*(2), 284-307.

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The quarterly journal of experimental psychology, 47*(3), 631-650.

Zawiszewski, A., Santesteban, M., & Laka, I. (2016). Phi-features reloaded: An ERP study on person and number agreement processing. *Applied Psycholinguistics* 37, 601-626.

- 
- <sup>i</sup> Note that the concept of obligatoriness needs to be interpreted within the theoretical framework of the EEP proposed by Chomsky (1981). In other words, the subject is considered obligatory, even when it is omitted in the sentence. When the subject is omitted, which is possible in languages such as Spanish (e.g. (Yo) fui a un concierto), an empty category called *little pro* is assumed to occupy the subject position and to fulfil the EPP requirements.
- <sup>ii</sup> There is no clear and shared opinion about the use of the term ‘agreement’ and ‘concord’ that have been adopted to identify several syntactic phenomena (Corbett, 2003). We will use the term ‘agreement’, in line with previous empirical literature adopting terms such as ‘Tense agreement’ (cf. Sybesma, 2007; Sagarra, 2008) or ‘temporal agreement’ (Qiu & Zhou, 2012; Baggio, 2008).
- <sup>iii</sup> In Romance languages such as Spanish, the preverbal position of deictic temporal adverbs is natural and does not require any specific prosodic contours. If some difference has to be assumed, one may say that the sentence-initial position is, in general, more prominent at a discourse level. With reference to temporal adverbs, some theories assume that the adverbial sentence-initial position can be favoured in that it allows ‘setting the scene’ (Benincà & Poletto, 2004) with no need of a previous context (cf. Chafe, 1984; Dickey, 2001) and may be more appropriate in de-contextualized sentences.
- <sup>iv</sup> The realization of adverbs in the sentence structure is still a matter of debate in theoretical linguistics (Alexiadou, 2013 for an overview), and theoretical accounts have suggested both that adverbs are considered syntactically (e.g. Cinque, 1999) or post-syntactically (e.g. Stepanov, 2001). No discussion will be provided in favour of a specific theoretical account, since the findings here reported are compatible with both a syntactic and a post-syntactic analysis of adverbs.

---

<sup>v</sup> Separate analyses run using plural control sentences in subject-verb manipulations did not evidence differences in the results in any dependent variable analysed.

<sup>vi</sup> In this measure, number violations led to smaller reading times on the post-target region compared both to the control condition and to the person violations. A possible interpretation of this effect is that number violations (in the local configuration) are repaired parafoveally, i.e. before entering the post-target area (the so-called preview benefit effect e.g. Schotte, Angele & Rayner 2012), which would be in line with the assumption that number violations are easier to repair than person violations (see also findings from a self-paced reading experiment in Italian reported by Mancini et al. 2014).

**Figures**

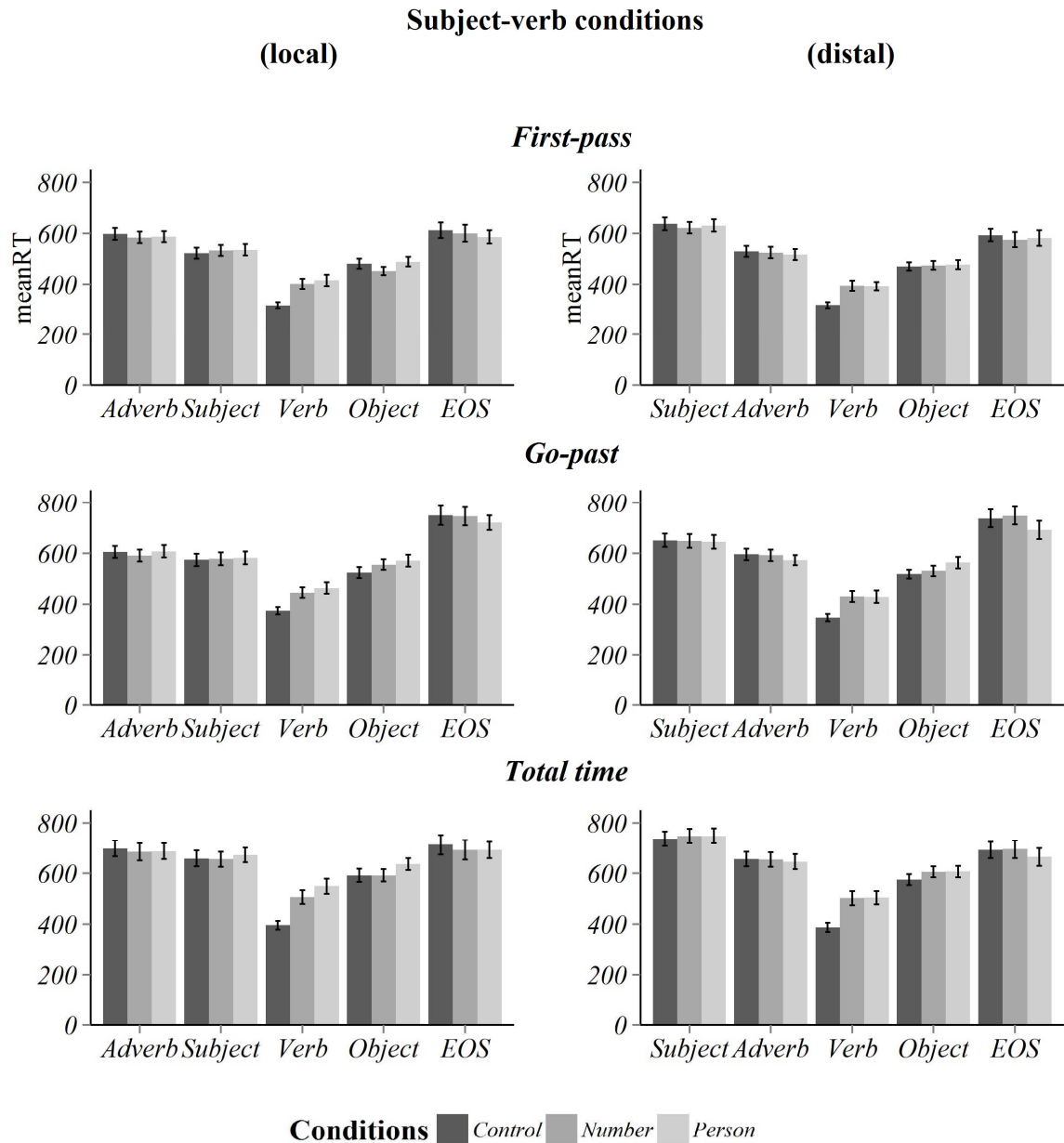
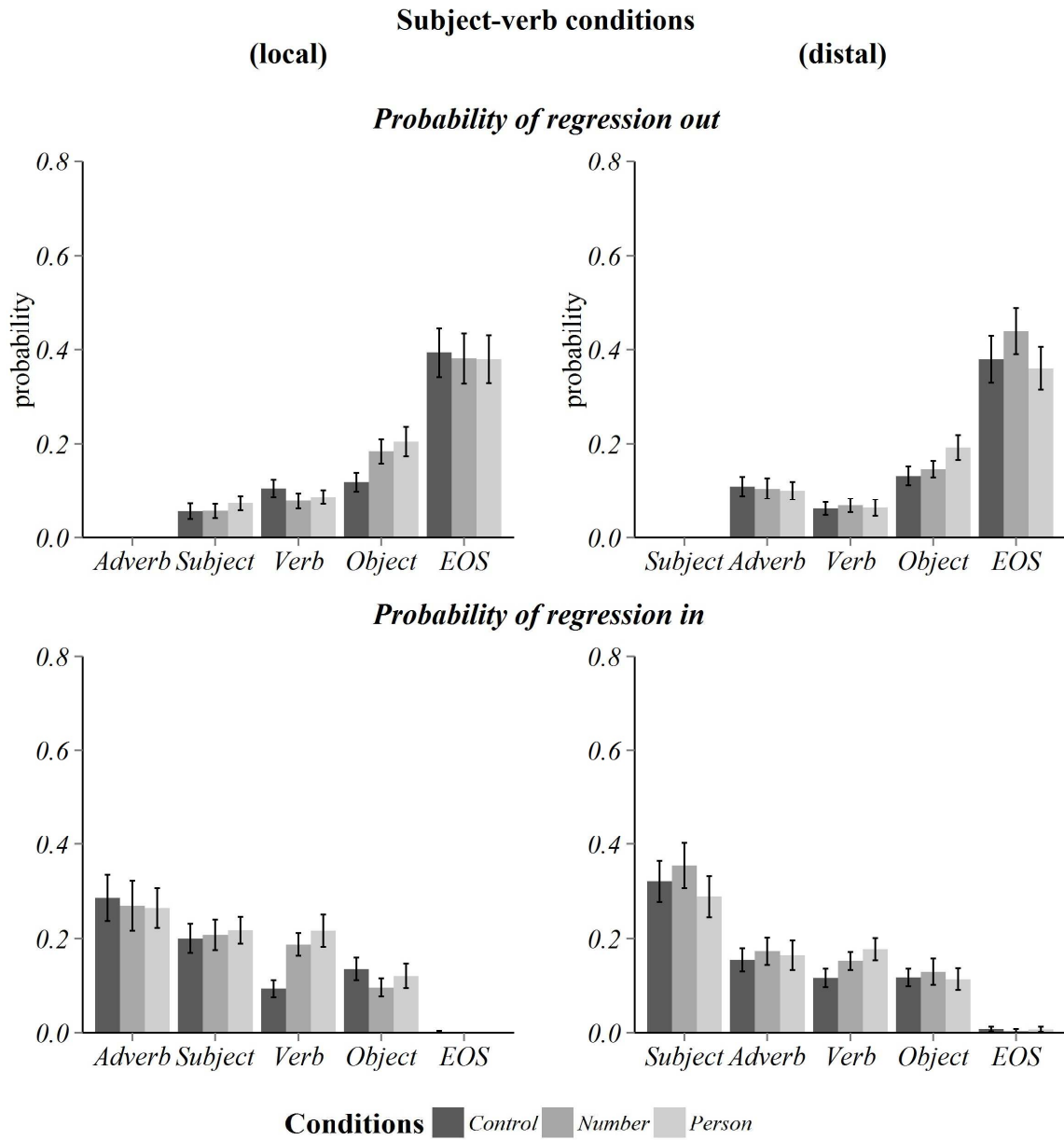


Figure 1a. Bar plots of mean reading times in milliseconds (and standard errors) for subject-verb agreement in eye-tracking latency measures. Mean reading times were divided into five regions: the adverb phrase, the subject phrase, the verb (target), the in/direct object phrase, the end of sentence (eos) containing the remaining phrases.



*Figure 1b.* Bar plots of mean probabilities of regression (and standard errors) for subject-verb agreement conditions in five different regions: the adverb phrase, the subject phrase, the verb (target), the in/direct object phrase, the end of sentence (eos) containing the remaining phrases.

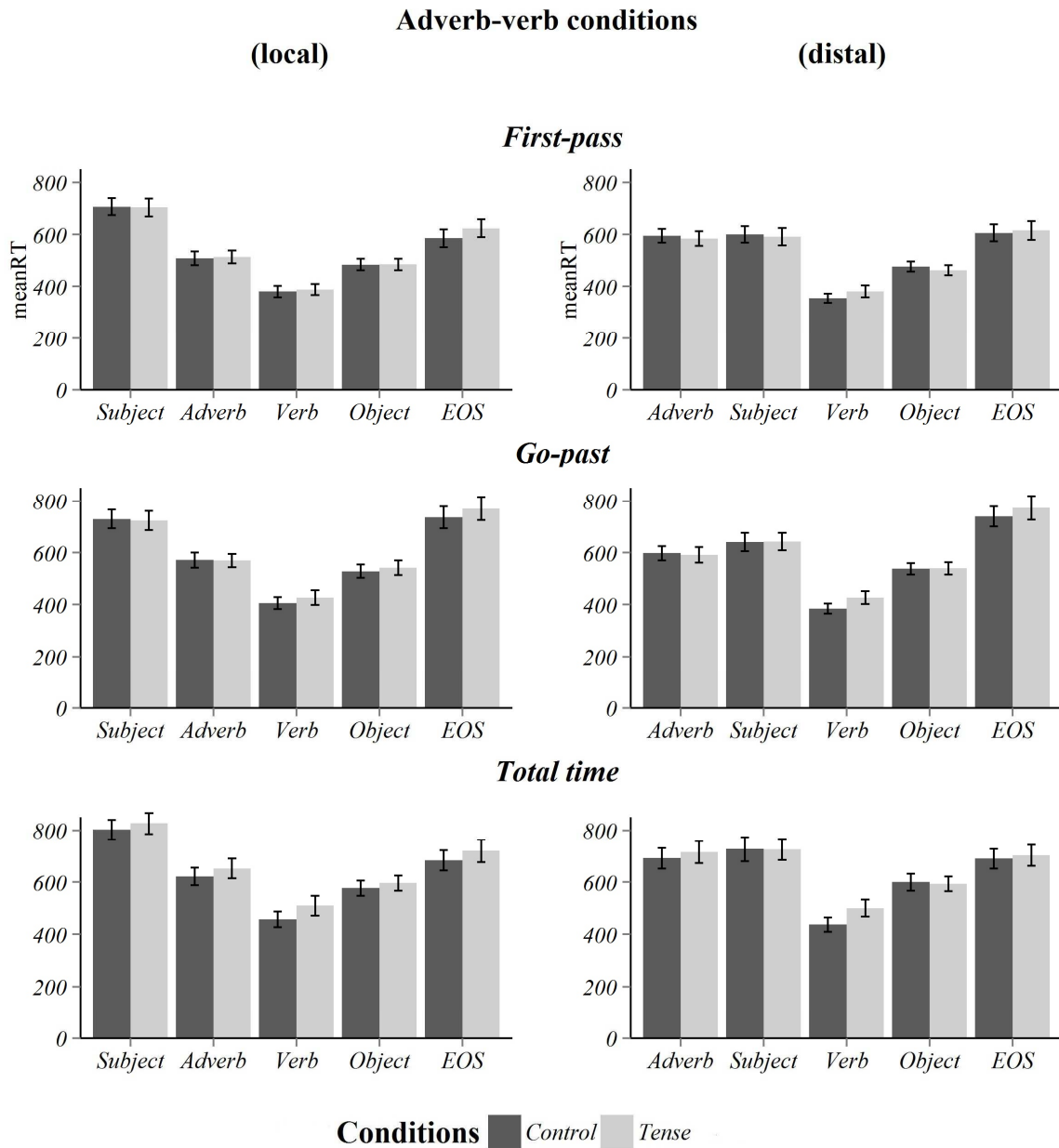
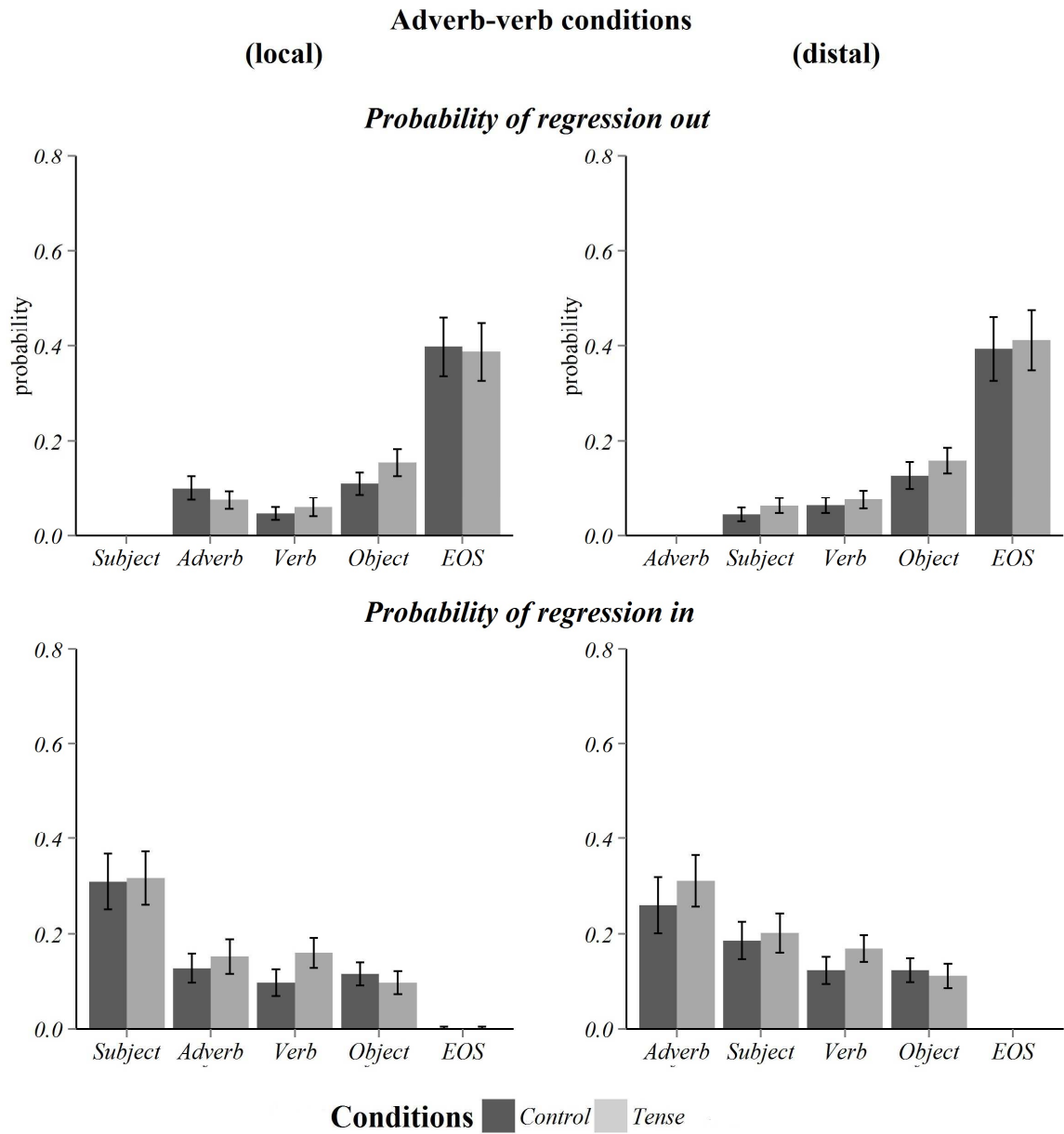


Figure 2a. Bar plots of mean reading times in milliseconds (and standard errors) for adverb-verb temporal concord conditions in each eye-tracking latency measure. Mean reading times were divided into 5 regions: the adverb phrase, the subject phrase, the verb (target), the in/direct object phrase, the end of sentence (eos) containing the remaining phrases.



*Figure 2b.* Bar plots of mean probabilities of regression (and standard errors) for adverb-verb temporal concord conditions in five different regions: the adverb phrase, the subject phrase, the verb (target), the in/direct object phrase, the end of sentence (eos) containing the remaining phrases.



## Tables

*Table 1.* Summary of the predicted effects for the mismatch conditions (compared to the control conditions). The table summarize the factors under investigation, as well as the relative eye-tracking measures which are predicted to be affected by a mismatch effect during processing of subject-verb number/person and adverb-verb tense violations on the verb.

<b>Factors</b>		<b>Affected eye-tracking measures</b>
Obligatoriness	+ obligatory (S-V)	first-pass, go-past, total
	- obligatory (A-V)	go-past, total
Anchoring	+ deictic (P, T)	total (larger RT compared to Number violations)
	- deictic (N)	total
Distance	+ distal (N, P, T)	first-pass, go-past, total
	- distal (N, P, T)	first-pass, go-past, total (larger RT compared to the distal conditions)

*Note.* SV= subject-verb violation, AV= adverb-verb violation, N= number violation, P= person violation, T= tense violation; RT= reading times.

Table 2. Sample of the material (Adverb-verb temporal concord: *control1, tense*; subject-verb agreement: *control2, number, person*) in the two configurations (adjacent controller: *local*; distal controller: *distal*)

	Local	Distal
<b>CONTROL1</b>	Los viajeros cansados   mañana a mediodía   <b>regresarán</b>   a casa   con mucho equipaje. <i>(The tired travelers tomorrow at noon<sub>(FUT)</sub> will go<sub>(FUT)</sub> back home with a lot of bags)</i>	Mañana a mediodía   los viajeros cansados   <b>regresarán</b>   a casa   con mucho equipaje. <i>(Tomorrow at noon<sub>(FUT)</sub> the tired travelers will go<sub>(FUT)</sub> back home with a lot of bags)</i>
<b>TENSE</b>	Los viajeros cansados   mañana a mediodía   <b>regresaron</b>   a casa   con mucho equipaje. <i>(The tired travelers tomorrow at noon<sub>(FUT)</sub> went<sub>(PST)</sub> back home with a lot of bags)</i>	Mañana a mediodía   los viajeros cansados   <b>regresaron</b>   a casa   con mucho equipaje. <i>(Tomorrow at noon<sub>(FUT)</sub> the tired travelers went<sub>(PST)</sub> back home with a lot of bags)</i>
<b>CONTROL2</b>	Mañana a mediodía   el viajero cansado   <b>regresará</b>   a casa   con mucho equipaje. <i>(Tomorrow at noon the tired traveler<sub>(3SG)</sub> will go<sub>(3SG)</sub> back home with a lot of bags)</i>	El viajero cansado   mañana a mediodía   <b>regresará</b>   a casa   con mucho equipaje. <i>(The tired traveler<sub>(3SG)</sub> tomorrow at noon will go<sub>(3SG)</sub> back home with a lot of bags)</i>
<b>NUMBER</b>	Mañana a mediodía   el viajero cansado   <b>regresarán</b>   a casa   con mucho equipaje. <i>(Tomorrow at noon the tired traveler<sub>(SG)</sub> will go<sub>(PL)</sub> back home with a lot of bags)</i>	El viajero cansado   mañana a mediodía   <b>regresarán</b>   a casa   con mucho equipaje. <i>(The tired traveler<sub>(SG)</sub> tomorrow at noon will go<sub>(PL)</sub> back home with a lot of bags)</i>
<b>PERSON</b>	Mañana a mediodía   el viajero cansado   <b>regresarás</b>   a casa   con mucho equipaje. <i>(Tomorrow at noon the tired traveler<sub>(3RD)</sub> will go<sub>(2ND)</sub> back home with a lot of bags)</i>	El viajero cansado   mañana a mediodía   <b>regresarás</b>   a casa   con mucho equipaje. <i>(The tired traveler<sub>(3RD)</sub> tomorrow at noon will go<sub>(2ND)</sub> back home with a lot of bags)</i>

*Table 3.* Mean score and standard errors of the naturalness judgment task.

	<b>Local</b>	<b>Distal</b>
<b>CONTROL1</b>	5.7 (0.3)	6.1 (0.2)
<b>TENSE</b>	2.4 (0.4)	2.2 (0.4)
<b>CONTROL2</b>	6.2 (0.2)	5.7 (0.3)
<b>NUMBER</b>	1.8 (0.4)	1.8 (0.4)
<b>PERSON</b>	1.9 (0.4)	1.7 (0.4)

*Table 4.* Grammaticality effects and interactions at the target (verb) region and post-target region for subject-verb (number/person) and adverb-verb (tense) violations. Stars represent significant differences between the violation and the correct condition in each measure and configuration

Configuration	Relation	Target region					Post-target region				
		FP	GP	TT	PO	PI	FP	GP	TT	PO	PI
Local	Adverb-V			*		*					*
	Subject-V	*	*	*		*	*	*		*	
Distal	Adverb-V	*	*	*		*					
	Subject-V	*	*	*		*		*	*	*	
<i>configurationxrelationxgrammaticality</i>		*					*				

*Note.* FP= first-pass, GP= go-past, TT= total time, PO= probability of regression out, PI= probability of regression in

*Table 5.* Grammaticality effects and interactions at the target (verb) region and post-target region for subject-verb number and person violations. Stars represent significant differences between the violation and the correct condition and between the two violation conditions in each measure and configuration

Configuration	Condition	Target region					Post-target region				
		FP	GP	TT	PO	PI	FP	GP	TT	PO	PI
Local	Number	*		*	*	*				*	*
	Person	*	*	*		*		*	*	*	
	Number<Person			*			*	*	*	*	
Distal	Number	*	*	*							
	Person	*	*	*				*		*	
	Number<Person							*			
<i>configurationxcondition</i>							*		*		

*Note.* FP= first-pass, GP= go-past, TT= total time, PO= probability of regression out, PI= probability of regression in

*Table 6.* Grammaticality effects and interactions at the target (verb) region and post-target region for tense and person violations. Stars represent significant differences between the violation and the correct condition in each measure and configuration

Configuration	Relation	Target region					Post-target region				
		FP	GP	TT	PO	PI	FP	GP	TT	PO	PI
Local	Tense			*		*					*
	Person	*	*	*		*		*	*	*	
Distal	Tense	*	*	*		*					
	Person	*	*	*		*		*		*	
<i>configurationxrelationxgrammaticality</i>		*		*							

*Note.* FP= first-pass, GP= go-past, TT= total time, PO= probability of regression out, PI= probability of regression in