Syntax Processing

Syntax Processing

Edited by

Vicenç Torrens

Cambridge
Scholars
Publishing

Syntax Processing

Edited by Vicenç Torrens

This book first published 2021

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-7054-1
ISBN (13): 978-1-5275-7054-2

CONTENTS

INTRODUCTION

VICENÇ TORRENS[1]

Language processing is a crucial part of cognition and studies conducted in this topic are particularly relevant to this field. The present volume deals with research on the processing of a native language, second language learning, bilingualism, typical and impaired syntax processing. The articles presented in this book cover a number of linguistic phenomena, including the following: passives, temporal concord, object pronouns, reflexives, embedded sentences, relative clauses, wh-movement and binding theory.

Syntactic processing has been described by processing models such as the garden path model, constraint-based models, the good-enough theory, serial and parallel processing, modular and interactive theories of language processing, and this volume covers some of these paradigms. The garden path model proposes that a single parse is constructed in a serial manner. When an initial parse of a sentence turns out to be incorrect due to syntactic ambiguity, a reanalysis of the syntactic parse takes place. During reanalysis, some principles take effect, such as late closure or minimal attachment; late closure states that words are attached to the current clause being processed; minimal attachment states that the parser builds the simplest syntactic structure possible in ambiguous sentences. Constraint-based theories are based on statistical learning: the parser takes into account the frequencies and distribution of linguistic structures; speakers apply probabilistic constraints when faced with ambiguous sentences. The good-enough models propose that listeners apply partial and superficial representations and that they do not apply a detailed syntactic analysis to sentences.

Papers in this collection apply various experimental methods, such as eye tracking, reaction times, Event-Related Potentials, picture selection tasks, sentence elicitation, pupillometry and picture matching tasks. The studies included in this book try to cover some of the most representative methods

[1] Corresponding author: Vicenç Torrens, Department of Psychology, National University of Distance Learning, C/ Juan del Rosal, 10, 28040, Madrid; email: vtorrens@psi.uned.es

used in language processing. Eye tracking is a method which measures where a participant looks, giving information about the point of gaze. The eye tracker measures the position of the eyes and their movements. Infrared light is directed to the pupils of the eyes, and the reflections are tracked by a camera. On the other hand, reaction time is a measure of the speed with which a participant reacts to an item: when a participant has to respond to a more difficult sentence, the reaction time is typically longer compared to an easier sentence. Event-related potentials (ERPs) measure the electrical activity in the scalp in reaction to an event. The event is a stimulus like a sound, a word or a sentence, sometimes followed by a question where the participants need to select an answer or remember an item. Pupillometry measures the pupil size and reactivity of participants to sentences. It provides a very accurate and precise data of pupil reactivity to stimulation. Pupillary responses can reflect activation of the brain in response to cognitive tasks: higher pupil dilation is associated with increased processing in the brain. Therefore, it has been found that pupil dilation increases when a participant needs to process a more difficult sentence. Finally, picture matching tasks study the preferential looking of participants, where they are exposed to different pictures or videos side-by-side, and the experimenter records participants' gaze data. Participants are exposed to an auditory or visual stimulus, and only one of the pictures or videos presented afterwards matches the situation depicted by the previous stimulus. The participants that understand the stimulus correctly look at the matching picture or video for a longer period of time. Measures of lexical frequency, familiarity, and imageability of target stimuli are relevant to understand the choice of participants.

The first paper of the book is the chapter "the processing of temporal concord in sentences: an ERP investigation on the role of adverb – verb distance", where **Biondo et al.** focus on the temporal concord between the verb inflection expressing past or future information and a temporal adverb. In particular, the authors try to show how the sentence parser deals with different instances of agreement during online sentence comprehension. Temporal concord entails the coherence in time between a deictic temporal adverb and the main verb of a sentence. Deictic adverbs differ from other adverbs since they need to be anchored to the time of utterance. Biondo et al. studied thirty-nine Italian native speakers who participated in the experiment. Participants' accuracy and reaction times to the grammaticality judgement task were recorded, in addition to the EEG activity collected by 64 electrodes distributed over the scalp. With respect to the results from behavioural data, the authors did not find any significant effect of concord, distance, or concord per distance interaction. With respect to the EEG

activity, in early stages of processing no significant effects were found for distal and adjacent violations, although numerical trends are visible in the grand average ERPs; later in time, a significant effect of concord was instead found for distal and adjacent conditions, matching the properties of the P600 component. The authors conclude that distance is relevant during the processing of temporal concord, and this fact is more important during stages of reanalysis.

The next paper focuses on the syntactic properties in developmental dyslexia: in the paper "Exploring the Syntax-Prosody Interface in Children with Developmental Dyslexia", **Martina Caccia & Maria Luisa Lorusso** study the causes and factors in the origin of reading disorders. These authors suggest that children with SLI and/or dyslexia aged 10–14 years are impaired at disambiguating linguistic structures through prosody. Typically developing (TD) children and children with developmental dyslexia (DD) who learned Italian as a native language took part in the experiment. Each sentence had an ambiguous syntactic structure which was disambiguated through prosody. The results were analyzed in terms of: a) the Response (target, alternative or distracter); b) the Type of linguistic structure; c) the Distance in the syntactic dependencies. Results showed significant main effects of type, distance and grammatical functions assignment. The presence of Dyslexia is thus associated with higher effects of syntactic complexity especially in terms of construction type and the need to restructure syntactic function, which seem not to be efficiently disambiguated by prosody.

Esaulova et al.'s paper assumes that in general, the production of passive forms is considered more effortful than that of active forms, although the use of passives may depend on both the preceding context and whether the referents are animate or patients. In addition, the differences in the outcome of these perceptual priming studies have led to the proposal that the formulation of utterances in the passive voice is dependent on the specific grammatical characteristics of a given language. These authors tested forty-five German-speaking participants in an eye-tracking study where they had to describe scenes depicting an agent and a patient character. In this study they examined whether explicit visual cueing affected the production rate of passive voice utterances in German. In half of the trials, the patient character was explicitly cued. Four lists of stimuli were created with patient Position (on the right or on the left of the scene) and Cueing (cued vs. non-cued patient) as within subjects and within items factors, and patient Animacy (animate vs. inanimate) as a within subjects and between items factor. Esaulova et al. found that explicit visual cueing of patients, as

opposed to implicit cueing, increased the probability of passive utterances produced by German speakers.

The goal of **Isobe et al.'s** paper is to investigate whether Japanese-speaking 2-3-year-olds can distinguish short passives from their active counterparts by employing the preferential looking paradigm using an eye-tracker. Previous studies found that children already have the ability to comprehend passives but have difficulty with interpreting the agent role in *ni*-phrases, i.e., '*by*-phrases'. The authors tested sixty children ranging in age from 2;6 to 3;5 (mean = 2;10). They also tested twenty-five children of age 6 and ten adults for comparison. They tested a total of 16 sentences: 8 target sentences, including both active and passive sentences, along with 8 filler sentences. Passive sentences were all short passives, i.e., passives without *-ni* 'by' phrases. The results of the research showed that the 2-3-year-old children look longer at the congruent events for active sentences, whereas they tend to look at the incongruent ones longer for passive sentences; in contrast, they found that the 6-year-old children look longer at congruent scenes regardless of sentence types, a similar pattern compared to adults. Isobe et al. raise the possibility that 2-3-year-old children can notice the passive morpheme but cannot promptly revise the initial interpretation guided by the typical interpretive strategy of taking the first NP as the agent of the verb.

Martohardjono et al. assume that bilingual experience is determined by the relative exposure to the two languages, and that this exposure can vary greatly among speakers. These authors distinguish between heritage speakers, also commonly referred to as second-generation bilinguals; and late bilinguals, also commonly referred to as first-generation bilinguals. Heritage speakers are different in their bilingualism from late bilinguals, who have a more uniform and continuous experience of their first language, are schooled in that language, and acquire the other language only later in life. These authors try to disentangle whether syntactic structures found in the first-learned language that are equivalent to those found in the second-learned language might undergo weakening; also, they compare the judgment and processing between heritage speakers and late bilinguals of a wh- gap structure that differs in grammaticality between Spanish and English: Comp-trace interrogative sentences. This study applies pupillometry as an indicator of cognitive load in real time while participants listen to sentences, and an acceptability judgement task. Forty-five Spanish-English bilingual adults participated in this study and were categorized as either Spanish heritage speakers or adult late bilinguals. Based on these results, these authors conclude that heritage speakers don't show more influence

from English than late bilinguals, nor more weakening of grammatical principles in the first-learned language; however, even as the two groups exhibit similar grammars, their processing strategies show significant differences.

The contribution by **Puebla et al.** consists of research on the presence and timing of interference effects during the interpretation of personal pronouns, which are restricted by Condition B. Condition B states that a pronoun cannot take a c-commanding antecedent from its local domain. In pronoun resolution, there are two variables to take into account: the timing of constraint application and its interaction with other information sources. The research consists of an eye-movement monitoring task designed to investigate the presence and timing of interference effects during L1 and L2 pronoun resolution. Thirty-one native speakers of German and thirty Russian-native learners of L2 German took part in the experiment. These authors compared the L1 and L2 processing of German object pronouns in sentences that contained two c-commanding potential antecedents: a binding-accessible non-local antecedent, and an elaborated but structurally inaccessible local competitor antecedent. They found that L1 participants considered the inaccessible antecedent initially even in the presence of a gender-matching appropriate antecedent; however, the L2 participants only considered the inaccessible antecedent later on during processing, and only in the absence of a gender-matching accessible antecedent. The authors conclude that the cues to anaphor resolution are weighted differently for native and non-native populations.

The study by **Stover et al.** explores the relationship between L1 processing and language dominance in bilingual English-Spanish speakers. The authors treat language dominance as a continuous, relative measure rather than as a categorical or absolute measure. They studied the subject-object asymmetry in relative clause processing with gaze data and behavioral measures. With respect to eye-tracking data, these authors found that mean target fixation proportion is higher for Subject Relative Clauses than Object Relative Clauses in the Relative Clause Region but not in the Matrix Predicate Region. For Subject Relative Clause items, there is a pattern of increased target fixation proportions as Spanish dominance increases; however, Spanish dominance correlates negatively with target fixation proportion for Object Relative Clauses. With respect to behavioral data, Object Relative Clauses had greater response times than Subject Relative Clauses, and increased Spanish dominance also caused slower response times in Object Relative Clauses. Stover et al. found an effect of Object Relative Clauses

that affects the eye movements and response times of highly Spanish-dominant speakers compared to less Spanish-dominant speakers.

Finally, the paper by **Vogelzang et al.** deals with the processing of Principle B Effect in Dutch and German. The main experiment explored pupil dilation information, which was recorded during the presentation of stories. In German, the pupil dilation analysis shows that initially clauses with a reflexive elicit a larger pupil dilation, whereas at the end of the clause, clauses with a pronoun start eliciting a larger pupil dilation with respect to clauses with a reflexive. In contrast, the Dutch data shows that reflexives were not more effortful to process than pronouns at any point during the sentence, whereas pronouns were statistically more effortful to process than reflexives in the middle and at the end of the clause. Their results support the hypothesis that German adults in comparison to Dutch adults show less increased processing effort when resolving a pronoun compared to a reflexive. The authors argue that adults' processing effects in Dutch could be due to the Delay of Principle B Effect, and that differences in the pronominal system with respect to the use of pronouns and reflexives explain the acquisition difficulties found in children. It might be the case that if less relative processing effort is required for adults to resolve pronouns in German than in Dutch, cognitive resources could influence pronoun processing in these languages.

In conclusion, the present volume consists of a set of studies on syntax processing that were presented at the Experimental Psycholinguistics Conference in Palma de Mallorca (EPC) in June 2019. I would like to thank the plenary speakers of the Conference (Willem Mak, Esther Ruigendijk and Juan Uriagereka) and the members of the Scientific Committee (Sergi Balari, Lluís Barceló-Coblijn, Joe Barcroft, Antonio Benítez-Burraco, Denisa Bordag, Armanda Costa, Antoni Gomila, Pedro Guijarro, Aritz Irurtzun, Christer Johansson, Loes Koring, Evelina Leivada, Paulina Łęska, Paolo Lorusso, Manuela Pinto, Ankelien Schippers, and Maria del Mar Vanrell). I'd like to thank all the anonymous reviewers of this volume and also the contribution of the coordinators of this book series. Finally, I am also very grateful to Universidad Nacional de Educación a Distancia for their support on the organization of this Conference.

THE PROCESSING OF TEMPORAL CONCORD IN SENTENCES: AN ERP INVESTIGATION INTO THE ROLE OF ADVERB – VERB DISTANCE

NICOLETTA BIONDO[1], EMMA BERGAMINI & FRANCESCO VESPIGNANI

Abstract

The processing of temporal concord anomalies (e.g., Last week/*Tomorrow I bought a car) has been rather understudied compared to the processing of other agreement relations. Moreover, previous event-related potential (ERP) studies investigating temporal concord violations reported quite heterogeneous findings. One crucial aspect that was not kept constant across studies was the linear distance between the verb and the adverb. After providing a review of the previous ERP literature investigating agreement relations in general and temporal concord in particular, we present and discuss new ERP data on the processing of temporal concord in Italian. In particular, we investigated whether adverb-verb linear distance affects the processing of temporal anomalies. Our results show that adverb-verb distance affects the amplitude of late components (P600, SFN) elicited by temporal concord violations in Italian. We conclude that distance plays a role during the processing of temporal concord, in particular during stages of reanalysis.

1. Introduction

One of the most interesting properties of human language is its redundancy. The speakers of a language can express the same formal

[1] Corresponding author: Nicoletta Biondo, Department of social, political and cognitive sciences, University of Siena, Via Roma 56 (Room 303), 53100 Siena, Italy, email: nicoletta.biondo@unisi.it

property on several words within the same sentence. For example, in a sentence like (1) in Italian, the reader comprehends that the entity performing the action of playing in a football team (the girls) refers to a plurality of individuals. We extract this information thanks to the plural number feature, which is expressed by both the determiner *le* and the noun inflection *ragazz-e* as well as by the verb inflection *gioc-ano*. Understanding the relation among constituents is pivotal to efficiently process sentences. Indeed, if we change the feature of the noun *ragazze* or the feature of the verb *giocano* as in (2) and (3) respectively, the sentences become ungrammatical.

(1) *Le ragazze che abbiamo visto all'aeroporto giocano in una famosa squadra di calcio.*

(The$_{PL}$ girls$_{PL}$ that (we$_{PL}$) have$_{PL}$ seen at the airport play$_{PL}$ in a famous football team)

(2) **Le ragazza che abbiamo visto all'aeroporto giocano in una famosa squadra di calcio.*

(The$_{PL}$ girl$_{SG}$ that (we$_{PL}$) have$_{PL}$ seen at the airport play$_{PL}$ in a famous football team)

(3) **Le ragazze che abbiamo visto all'aeroporto gioca in una famosa squadra di calcio.*

(The$_{PL}$ girls$_{PL}$ that (we$_{PL}$) have$_{PL}$ seen at the airport plays$_{SG}$ in a famous football team)

Linguists and psycholinguists have been using the term "agreement" or "concord" to define the consistent covariance of features between two or more words (Corbett, 2003), which is necessary in order to have grammatical and comprehensible sentences. One widely-studied instance of agreement is the one between the verb and the subject, or between the determiner and the noun. Another well-studied phenomenon is the feature covariance between a pronoun and its antecedent in sentences such as *'The lady is brushing herself'*.

In this chapter we will focus on a relatively understudied instance of concord, namely the temporal concord between the verb inflection expressing past or future information and a temporal adverb such as *last week,* in sentences as in (4).

(4) *Domani le ragazze giocheranno/*giocarono una partita.*

(Tomorrow$_{FUT}$ the girls will play$_{FUT}$/*played$_{PST}$ a match)

This investigation aims at widening the study of agreement phenomena and enriching the current debate on how the sentence parser deals with different instances of agreement during online sentence comprehension.
In the next sections, first we summarize the main findings on the electrophysiological correlates of agreement in sentences, then we provide a theoretical description and a review of previous experimental studies investigating temporal concord and, finally, we present new electrophysiological data on the effect of adverb-verb distance during the processing of a temporal violation within a sentence in Italian. The manipulation of linear distance can be extremely useful for the study of agreement phenomena. In the specific case of temporal concord, linear distance can affect the way the temporal information provided by an adverb is tracked before encountering a tensed verb, the way the information encoded in the verb is integrated with the information of the adverb, and the way an adverb-verb temporal inconsistency is dealt with. Our results show that distance plays a role during the processing of temporal violations, in particular during stages of reanalysis.

1.1. Event-related potential studies (ERPs) of agreement relations

ERPs is a derived measure of the electroencephalogram (EEG) which consists in the measurement of the electrical activity of the brain through electrodes placed on the scalp. One of the advantages of electrophysiological data (over many types of behavioral data) is that they reflect the neural activity continuously, without time delay. In particular, ERPs are computed by extracting and averaging the portions of EEG signal time-locked to a relevant stimulus (e.g., the verb of a sentence). A large number of trials per condition is required in order to gather a good signal to noise ratio. Psycholinguists often adopt the violation paradigm: the electrophysiological activity generated by a word containing a grammatical error (e.g., the verb in the sentence '*That girls dances'*) is generally compared with the activity generated by the same word in a sentence without errors (e.g., '*That girl dances'*). From the very first papers investigating the ERP generated by anomalous sentences (e.g., Kutas & Hillyard, 1980), researchers have been trying to characterize the ERP deflections specific to different linguistic manipulations in terms of *components*, which are defined by a functional interpretation.

For example, past ERP literature on agreement violations report that (adjective-noun, subject-verb, determiner-noun) agreement errors typically trigger a biphasic pattern, namely an early negative deflection in the 300-500ms interval followed by a positive deflection in the 500-900ms time window after the presentation of the violating word (see Molinaro, Barber & Carreiras, 2011 for an overview).

The presence, amplitude and topographic distribution of the early negativity was found to vary across studies, languages and manipulations, and its functional interpretation has been largely debated (e.g., Tanner & Van Hell, 2014; Molinaro, Barber, Caffarra & Carreiras, 2015; Tanner, 2015; Caffarra, Mendoza & Davidson, 2019). In some studies, there was no evidence of a negativity (e.g., Nevins, Dillon, Malhotra & Phillips, 2007; Frenck-Mestre, Osterhout, McLaughlin & Foucart, 2008; Alemán Bañón & Rothman, 2019). In other studies, the negativity was either distributed in the left-anterior part of the scalp or it was broadly distributed, spanning over central-posterior areas. When the negativity is anterior and left-lateralized, it is labeled as Left-Anterior Negativity (LAN). Researchers have argued that the LAN reflects automatic morpho-syntactic processing (e.g., Friederici, 1995; Friederici, Hahne & Mecklinger, 1996; De Vincenzi, Job, Di Matteo, Angrilli, Penolazzi, Ciccarelli & Vespignani, 2003; Mancini, Molinaro, Rizzi & Carreiras, 2011; Molinaro et al., 2011) or an increase in the use of working memory resources caused, for example, by the processing of long-distance dependencies (e.g., Kluender & Kutas, 1993; King & Kutas, 1995; Fiebach, Schlesewsky & Friederici, 2002; see also Martín-Loeches, Muñoz, Casado, Melcon & Fernández-Frías, 2005). Moreover, it has been argued that the LAN is more likely to appear more consistently in local within-phrase agreement such as in the determiner-noun relation than in across-phrase agreement such as the one between the subject noun phrase and the verb (Molinaro et al., 2011). When the negativity is more broadly distributed and its peak arises at 400ms after the stimulus onset, it is labeled as N400, a component that is also elicited by semantic integration problems (Kutas & Hillyard, 1980, 1984; Berkum, Haagort & Brown, 1999; see Lau, Phillips & Poeppel, 2008 for a review), and discourse-related processing (e.g., Mancini et al., 2011; see also Nieuwland & Van Berkum, 2006).

The following positive-going deflection was found more consistently across studies investigating syntactic anomalies. It typically emerges in the central-posterior areas of the scalp and it is often called P600, although sometimes in the past it was dubbed as Syntactic Positive Shift (SPS). The

P600 has been traditionally interpreted as an index of syntactic difficulty, in terms of repair, reanalysis and integration. Indeed, a P600 was found for morpho-syntactic violations (e.g., Hagoort, Brown & Groothusen, 1993; Osterhout & Mobley, 1995; Friederici et al., 1996), for garden-path sentences (e.g., Osterhout, Holcomb & Swinney, 1994; Kaan & Swab, 2003; Gouvea, Phillips, Kazanina & Poeppel, 2010), and for the processing of (grammatical) long-distance dependencies (e.g., Kaan, Harris, Gibson & Holcomb, 2000; Phillips, Kazanina, Abada, 2005). Molinaro et al. (2011), among others, proposed that the P600 could reflect two different processing stages. The early stage of the P600 (500-750ms) is more broadly distributed and it is related to the difficulty in the integration of the mismatching constituent with the previous portion of the sentence, while the later stage of the P600 (750-1000ms) is more posteriorly distributed and it represents reanalysis or sentence processing repair mechanisms (see also Kasparian, Vespignani & Steinhauer 2017 for a three-stages proposal). It should be noted, however, that the syntactic specificity of the P600 has been called into question. Indeed, a P600 emerged for non-syntactic violations, such as semantic, animacy, and thematic role violations (e.g., Kuperberg, Sitnikova, Caplan & Holcomb, 2003; Kim & Osterhout, 2005; Kuperberg, Caplan, Sitnikova, Eddy, & Holcomb, 2006; Van Petten & Luka, 2012 for a review) and for musical syntactic violations (e.g. Patel, Gibson, Ratner, Besson, & Holcomb, 1998; Patel, 2003). For this reason, several accounts have proposed that the P600 reflects an index of general reanalysis[2].

Many studies investigating the processing of agreement with the violation paradigm also reported a sustained negativity in the violation condition compared to the control condition (e.g. De Vincenzi et al., 2003; Hagoort et al., 1993; Osterhout & Mobley, 1995; Molinaro, Vespignani & Job, 2008). This negativity starts arising 300ms after the onset of the last word of the sentence, or as soon as the P600 response to the ungrammaticality ends, in central and posterior areas of the scalp. The functional interpretation of this negativity is still debated (see Stowe, Kaan, Sabourin & Taylor, 2018 for a recent review). One possible interpretation of this

[2] For example, the P600 has been related to a continued combinatorial analysis of mismatching morpho-syntactic and semantic-thematic constraints (e.g., Kuperberg, 2007), to mechanisms of well-formedness checking following a failed mapping between animacy/thematic roles and plausibility (e.g., Bornkessel-Schlesewsky & Schlesewsky, 2008), to a conflict-monitoring mechanism triggered by the encountered input that does not match top-down expectations (e.g., Van De Meerendonk, Kolk, Vissers, & Chwilla, 2010).

late negativity is that it reflects the tension between the impossibility of fully analyzing and storing the sentence (because of the mismatch) and the impossibility of discarding the sentence completely (because new upcoming input could help solving the incongruities). Given that this effect often arises in presence of a metalinguistic task (i.e., grammaticality judgement task), it has also been proposed that this negativity arises when some information needs to be maintained to successfully perform the decision task.

1.2 Widening the study of agreement relations: the temporal concord

In the generative framework, the covariance of features between the subject and verb is guaranteed by a unique operation in which all features are transmitted/copied from the subject to the verb (e.g., Chomsky 1995, 2000). Similarly, in psycholinguistics, a unique mechanism for the processing of subject-verb agreement has been predicted in many mainstream models of sentence parsing (e.g., Friederici, 2002, 2011; Bornkessel & Schlesewsky, 2006; Hagoort, 2003, 2013). However, recent experimental evidence showed that the mechanism underlying agreement processing may not be unique, neither during the processing of different features (e.g., number, person) within the same relation (e.g., subject-verb), nor during the processing of the same feature (e.g., number) across different relations (e.g., determiner-noun, subject-verb, noun-reflexive). Evidence for a differentiation in the mechanisms underlying the processing of different agreement features comes from experimental studies showing longer reading time (e.g., Mancini, Postiglione, Laudanna & Rizzi, 2014; Biondo, Vespignani, Rizzi & Mancini, 2018) and different ERP responses (e.g., Mancini et al., 2011; Zawiszewski, Santesteban & Laka, 2015; Mancini, 2018) during the processing of person violations compared to number violations during in sentences. For example, Mancini et al. (2011) found the classic LAN-P600 pattern in response to number violations and an N400-like negativity followed by a P600 for person violations, compared to the control condition. Evidence for a differentiation in the processing of different concord relations comes from studies (e.g., Sturt, 2003; Phillips, Wagers & Lau, 2011; Dillon, Mishler, Sloggett & Phillips, 2013; Jäger, Engelmann & Vasishth, 2017; but see Jäger, Mertzen, Van Dyke & Vasishth, 2020) suggesting that the retrieval mechanisms implied in the resolution of subject-verb agreement and anaphora (e.g., '*John likes himself*') could differ qualitatively, being the retrieved features equal (e.g., number). In particular, anaphora resolution

seems to be less prone to interference (i.e., in the retrieval of a syntactically illicit but feature matching antecedent) compared to subject-verb agreement, which shows clear interference effects from illicit antecedents.

This piece of evidence suggests that a more in-depth investigation of other concord phenomena is needed and pivotal for a more accurate formalization of the parsing mechanisms at play during sentence processing. To this end, we attempt to offer new insights about the computation of the adverb-verb temporal concord.

Past literature has been using terms such as "Tense agreement" (e,g, Sybesma, 2007; Sagarra & Han, 2008) or "temporal agreement" (Qiu & Zhou, 2012; Baggio, 2008) to label the adverb-verb relation, in line with a "broad" interpretation of the term agreement, which identifies the covariance of a semantic or formal property between two elements in a sentence. However, the term agreement has also been used in a "narrower" sense, so as to refer to the feature-checking mechanism specifically formalized to describe the feature sharing between the subject and the verb. In order to avoid terminological ambiguities, we will adopt the term "temporal concord" to refer to the adverb-verb relation. We think that this terminological choice is more appropriate. First, the term "concord" has been used to identify the adjective-noun relation (Chomsky, 2001), that is a modifier-phrase relation, such as the one between the adverb and the verb. Second, this terminological choice would also account for another property that makes adverbs and adjectives similar, namely their optionality (compared to the obligatoriness of elements such as the subject DP in the subject-verb relation).

Temporal concord entails the coherence in time between a deictic temporal adverb as *yesterday, next year* and the main verb of a sentence. Deictic adverbs differ from other adverbs such as clock-calendar adverbs (e.g., *on Monday, at 10 PM*), since they need to be "anchored" to the time of utterance (Smith, 1978; 1981). *Yesterday* is always interpreted as the 24-hour time interval preceding the time of utterance. As argued by Alexiadou (1997, 2000), deictic adverbs are marked [±PAST], and for this reason they can lead to ungrammaticality ('*Yesterday Maurizio went/*will go home*'), while clock-calendar adverbs cannot ('*At 10 PM Maurizio went/will go home*').

From a generative perspective, the features expressed by verb inflection, namely tense and phi-features (i.e., number, person and gender), are all

encoded in a unique projection[3] (TP). The co-existence of different features is easy to be seen in morphologically rich languages (e.g., Italian) where these features are all morphologically realized on the verb (e.g., *gioc-her*$_{Tense}$*-anno*$_{Number/Person}$), and where the verb can establish a concord relation with other phrases, be it a subject DP in subject-verb agreement or an adverbial phrase in the temporal concord.

One licit question at this point is whether the same mechanisms are at play when the verb is encountered, and subject-verb and adverb-verb relations need to be processed. These relations differ in two salient aspects at least. First, subject agreement with the inflected verb is an obligatory dependency while temporal concord is not. Second, the morphosyntactic features we are considering have an interpretive counterpart that intrinsically differ in nature.

As for the first, more formal aspect, both the subject position (EPP requirement of Government & Binding approaches, e.g., Chomsky, 1981) and the inflectional node are obligatory components of the clausal structure, and there is an obligatory requirement of match in person and number features when expressed in both constituents (in languages like Italian). In minimalist approaches, this requirement is expressed as an obligatory feature checking procedure that must take place both from the vantage point of the inflectional node, which must have its features valued by the subject, and of the subject DP, which must have its case licensed: the subject acquires nominative case through this local relation with the inflectional node[4]. Adverb concord with the temporal inflection on the verb is very different in this respect: tense is an obligatory position in the functional structure of (finite) clauses, but the adverb is an adjunct, an optional position[5]. Adverbial DPs such as *last week, next week* bear temporal features for past and future (Alexiadou, 1997; Enç, 1987), but the

[3]It should be noted that within the generative framework there are also accounts proposing and showing evidence for the existence of two distinct structural projections, AgrP and TenseP (Pollock, 1989 and subsequent cartographic work, e.g., Belletti, 1990; Shlonsky, 2010; Rizzi & Cinque, 2016).

[4]Indeed, in infinitives where the inflectional node is not active, the subject typically cannot be nominative.

[5]For many theories (e.g., Chomsky, 1986, 1995; Sportiche, 1988 among others), adverbs behave as satellites attached to the maximal projection of the phrase they modify, since they are not obligatory in a sentence. However, it is worth mentioning that there are other theories proposing that adverbs occupy the specifier position of hierarchically organized functional projections (e.g., Kayne, 1994; Alexiadou, 1997, 2000; Cinque, 1999, 2004).

temporal inflection on the verb does not need the presence of the adverbial, it has an inherent temporal value (in the minimalist terminology in Chomsky 1995, 2000 it carries an "interpretable feature"). Reciprocally, these adverbs do not need a structural case to be checked, they have an inherent case[6] that does not need to be structurally licensed, differently from structural cases like nominative and accusative. In other words, differently from subject-verb agreement, no formal feature checking is assumed to take place for temporal concord.

The second salient difference between temporal concord and subject-verb number agreement is that tense and number have clearly different interpretive counterparts that impact different aspects of meaning construction. Number features relate to the cardinality of the set referred to by the subject argument (one entity, more entities). Tense features express the location of the event described by the verb on the temporal axis with respect to the time of utterance (past, present, future). According to the anchoring approach (e.g., Bianchi 2003, 2006; Sigurðsson 2004, 2016), we can say that the features are "anchored" to a distinct interpretive specification. Number is anchored to the cardinality of the entity referred to, a property expressed by the DP itself. Tense, on the other hand, must connect to the time of the speech (now), a specification that is expressed in the complementizer system (the left periphery of the sentence). For this reason, we can say that number features are "internally anchored" while tense features are "externally anchored".

Do these different properties (non-/optionality, internal/external anchoring) play any role during online sentence comprehension? According to some accounts, they do. The Construal model by Frazier & Clifton (1996), for example, predicts different parsing routines at play during the processing of primary and non-primary relations. Primary relations entail obligatory phrases such as verbs and verb's arguments. These phrases and relations are assumed to be processed through the attachment mechanisms described by classical syntax-first models (Frazier, 1987). Non-primary relations entail optional phrases such as adjuncts. The processing of these phrases and relations follows the construal principle: an upcoming non-primary phrase is associated to the current thematic domain (i.e., the last theta assigner) and it is interpreted by using both structural and non-structural information. It follows that the processing of primary relations should be more "resistant" and automatic while the processing of non-primary relations can be influenced by extra-syntactic (e.g., semantic,

[6] As we see in languages with overt morphological case.

pragmatic) factors. Moreover, other accounts predict different mechanisms at play based on the (discourse-related) interpretive properties of different features. Following the anchoring approach proposed in linguistic theory, Mancini, Molinaro & Carreiras (2013) theorized and showed in several experimental studies that the anchoring properties of different features (i.e., number, person) can affect sentence processing. For example, in an ERP study Mancini et al. (2011) compared the processing of number and person subject-verb violations. While number violations gave rise to the classic LAN-P600 response compared to the control condition, person violations elicited an N400-like negativity followed by a P600. Moreover, the P600 was larger in amplitude and more frontally distributed for person than for number violations. The N400-like negativity for person violations was interpreted as an index of failure in the establishment of the interpretive relations among constituents (i.e., is the subject the speaker, the addressee or neither of the two, in the speech act?), which goes beyond the level of morpho-syntax (for which a LAN effect is expected). The frontally distributed P600 effect (compared to the more posterior topography of the P600 generated by number violations) was also linked to discourse-related integration difficulties, namely to the impossibility of integrating the (mismatching) participants' role expressed by the subject and the verb in the same discourse representation. More recently, in one eye-tracking study Biondo et al. (2018) compared the processing of number, person and tense violations in the same experimental paradigm. Number and person violations led to similar parsing costs in early stages of processing while the parsing costs for tense violations appeared only in later stages. In other words, the "primary" subject-verb violations gave rise to earlier and stronger parsing costs than the "non-primary" temporal violations. Moreover, person anomalies caused larger parsing costs than number anomalies in the same stage in which tense anomalies started showing a cost. In other words, the violation of person and tense, the two features that need a reference to discourse to be interpreted (speech participants, speech time), affected the same (late) stages of processing. These findings suggest that online sentence comprehension is differently affected by both the obligatoriness of the relation and by the discourse-related properties of the features under computation.

Despite the theoretical and experimental evidence suggesting that a deeper investigation of other concord phenomena is needed, the temporal concord has been rather understudied. One possible reason can be related to the

optionality of the temporal adverbs[7]. However, the processing and attachment preferences of optional phrases/adjuncts such as prepositional phrases or relative clauses has been quite a debated topic in psycholinguistics (Cuetos & Mitchel, 1988; De Vincenzi & Job, 1993; Carreiras & Clifton, 1993; Carreiras, Salillas & Barber, 2004). More importantly, the temporal concord has played a pivotal role in other research fields. For example, several studies investigating language impairment in aphasia reported that agrammatic patients show marked difficulties during the comprehension and production of temporal concord while subject-verb agreement is less impaired in the same patients. The source of this linguistic impairment has been related to a difficulty in the representation of the syntactic structure (TenseP) encoding temporal information (e.g., Friedmann & Grodzinsky, 1997), or to a difficulty in the representation of tense features encoded in the sentence structure (e.g., Nanousi, Masterson, Druks & Atkinson, 2006; Clahsen & Ali, 2009). Temporal concord and subject-verb agreement also seem to be acquired in different stages both in child first language acquisition (Weist, 2014; Belletti & Guasti, 2015) and in adult second language acquisition (Biondo & Mancini, 2019). We thus believe that a deeper investigation of the temporal concord processing, in unimpaired sentence comprehension is necessary to have a wider look at concord and agreement processing and to implement current mainstream formalizations of sentence parsing routines (e.g., Friederici, 2011; Haagort 2013; Bornkessel & Schlesewsky 2006). A richer knowledge of the mechanisms involved in "typical" sentence processing can help us to understand which mechanisms develop earlier/later during (first and second) language acquisition, and which routines are impaired in presence of a language disorder and why.

1.3 ERP literature on the processing of temporal concord violations

Compared to the extended ERP literature on the processing of subject-verb agreement, there are few ERP studies that investigated the processing of temporal concord anomalies in adult (unimpaired) sentence reading. Table 1 summarizes the studies where adult native speakers of a language (without language impairments) were asked to read sentences containing either a correct or a mismatching temporal relation between a deictic temporal adverb and a verb expressing a temporal value (either through tense or through temporal particles). Despite their intrinsic differences

[7] Even when temporal adverbs are expressed in a sentence, a feature consistency is not always required, e.g., for calendar-clock adverbs.

(e.g., type of language, experimental material tested), these studies showed that temporal violations are detected quite early in sentence processing (300-500ms after the stimulus onset). In almost all the studies, the mismatching verb compared to its correct counterpart elicited an early negativity. The topography of this negativity, however, spans from a left anterior to a more central/posterior distribution. Conversely, a P600 component was reported consistently, in all the studies. Finally, some of the studies reported a widely distributed sustained negativity arising between 300ms and 800ms after the onset of the final word of the sentence when a temporal violation was encoded on the verb of the sentence.

One licit question is if/how the electrophysiological correlates of the temporal concord differ from the correlates of other agreement phenomena, and where the source of heterogeneity resides, especially in the early time window. In the following paragraphs, we first look at pure methodological variability[8] (e.g., in the acquisition/analysis of the EEG data) which can explain the presence/absence/topographic distribution of the components involved in the processing of temporal violations. Then, we consider and test one of the linguistic factors that could have affected the outcome of previous studies, namely adverb-verb distance.

Table 1 (next page). Linguistic details, tasks and results of the ERP studies investigating the processing of temporal concord violations. The ERP components refer to the onset of the target words, which are underlined. LPN stands for left posterior negativity while RAN stands for right anterior negativity. The studies marked with an asterisk (see * in the Authors column) also reported a sustained SFN (Sentence Final Negativity) for tense violations.

[8] Some studies (E1, E2) lacked methodological details so they could not be considered.

Code	Authors	Language	Experimental sentences (translated maintaining the same word order)	Task	0	300	500	750	1000 ms
E1	Fonteneau et al., 1998	French	Tomorrow the student will read/was reading the book.	Acceptability		Post. negativity & Front. positivity			
E2	Steinhauer & Ullman, 2002	English	Yesterday, I sailed/sail Diane's boat to Boston. Yesterday, we ate/eat Peter's cake in the kitchen.	Acceptability		LAN		P600	
E3	Newman et al., 2007	English	Yesterday I frowned/frown at Billy.	Acceptability		LAN		P600	
			Yesterday I ground/grind up coffee.			LPN		P600	
E4	Baggio, 2008*	Dutch	Last Sunday painted/paints Vincent the window-frames of his country house.	Passive reading		LAN		P600	
E5	Qiu & Zhou, 2012*	Mandarin Chinese	Next month/*Last month United Nations *jiangyao* dispatch a special investigation team. Last month/*Next month United Nations *cengjing* dispatch a special investigation team.	Acceptability		N400	P600		
E6	Dillon et al., 2012	Hindi	Although last night that traveler stone upon fall-*aa*/*fall-*e*-*gaa*, but to him injures not happen.	Acceptability		Posterior negativity (N400-like)		P600	
E7	Dragoy et al., 2012*	Dutch	The waiter who now/*just before the pepper grinds gets no tip.	Comprehension		P300		P600	
			The waiter who now the pepper grinds/*ground gets no tip.			-		-	
E8	Bos et al., 2013	Dutch	The granpa who a moment ago the coffee has ground/*will grind looks after his visitors.	Comprehension		P300		fP600	pP600
E9	De Vincenzi et al., unpublished*	Italian	The secretary long time ago called/*will call for a meeting.	Comprehension		RAN			P600

1.3.1 Methodological review

Let's first focus on the early component elicited in the 300-500ms time window. Some studies reported a left-lateralized negativity (E2, E3, E4), other studies reported a more distributed negativity (E1, E5, E6), one study reported a right-lateralized negativity (E9) and, finally, other studies reported a posterior positivity (E7, E8). It should be noted that the reference choice can be crucial for the detection of lateralized components (such as LAN). The activity detected by the reference (and the surrounding electrodes) is subtracted from the activity of the other electrodes. Consequently, if the reference is located on the same (e.g., left) side of the component under investigation, the amplitude of that effect could be reduced (Molinaro et al., 2011; 2015; but see Tanner, 2015). This could be the case of E9, where all scalp channels were referenced to the left mastoid and a right-lateralized negativity was found for temporal violations. The presence of a more broadly distributed negativity in E5, E6, on the other hand, should not be due to reference issues given that linked mastoids[9] were used as reference. Some of these studies (E7, E8) also reported a very different component for the items containing a temporal violation, namely a sustained positive activity that started around 300ms after the stimulus onset. Crucially, both studies reported that less than/only a third of the sentences presented in each list contained violations (Dragoy et al., 2012:313; Bos et al., 2013:290). As pointed out by Molinaro and colleagues (2011), the P600 may be sensitive to the proportion of violations in the experimental set. Coulson, King & Kutas (1998) indeed showed that the rarest stimuli (i.e., the stimuli with the smallest proportion in the whole set) elicit a larger P3b, a positive going component with central-parietal maximum that can start around 300ms after the stimulus onset. It can thus be possible, that the early latency of the P600 in E7 and E8 is related to the rarity/small proportion of sentences containing a violation.

Differently from the early components, the P600 arising in the 600-900ms time window showed up quite clearly and constantly in all the studies on temporal concord violations. Whether this component is modulated by linguistic factors cannot be assessed. In the current study, we will be able to test whether the P600 is affected by the distance of the two constituents.

[9] In particular, online linked mastoids were used as reference in E6 while offline linked mastoids were used in E3, E4, E5, E7, E8.

Finally, the presence of a broadly distributed sentence-final negativity (SFN) in response to temporal violations was partially reported. Some studies did report an SFN in response to temporal violations (E4, E5, E7, E9), while one study did not find any final negativity (E8). The remaining studies did not report any analysis of the ERPs triggered by the last word of the sentence (E1, E2, E3, E6). In other words, we cannot assess whether in these studies no SFN was present, or whether the SFN was present and not reported. One observation that should be considered is that the presence of the SFN does not seem to be related to a specific task, since it was found in passive reading (E4), in comprehension (E7, E9), and in acceptability (E5) tasks.

To sum up, the methodological review of the previous studies suggests that the presence of a right-lateralized early negativity or of an early positivity in response of a temporal violation may be related to specific methodological choices rather than to the processing of temporal concord per se. One issue that is still unsolved, however, is the left-anterior (LAN-like) or more central (N400-like) topographic distribution of the early negativity found in previous studies. Although we can exclude technical factors, there are still several linguistic factors that could account for this variability. The role of one of these factors, namely the distance between the adverb and the verb, is addressed in the current study. Moreover, the current study also allows to test/replicate the presence of the P600 and SFN in response to temporal violations, as well as their hypothetical modulation as a function of adverb-verb distance.

1.3.2 Does distance play any role in the processing of temporal concord violations?

A deeper look at the experimental material which was adopted in previous studies shows that the configuration between the adverb and the verb was not kept constant, both linearly and structurally. In some studies, the temporal adverb and the verb were adjacently located (E4, E9), while in others the two elements were separated by a pronoun (E2, E3), by a lexical subject (E5), by a lexical object (E7, E8), or by several phrases (E6). Some representative examples are reported in (5 - 7).

(5) From Baggio (2008):

Afgelopen zondag$_{PST}$ lakte$_{PST}$/lakt$_{PRS}$ Vincent de kozijnen van zijn landhuis. (Dutch)

('Last Sunday painted/*paints Vincent the window frames of his country house')

(6) From Fonteneau et al. (1998):

Demain$_{FUT}$ l'étudiant lira$_{FUT}$/*lisait$_{PST}$ le livre. (French)

('Tomorrow the student will read/*read the book')

(7) From Dillon et al. (2012):

Haalaanki *pichle shaam$_{PST}$* vo raahgiir patthar ke-uupar gir-aa$_{PERF}$/*gir-e-gaa$_{AGR\text{-}FUT}$, lekin use choT nahiin aa-yii. (Hindi)

('Although last night that traveler stone upon fell/*fall, but to him injures not happen')

Interestingly, the studies where the adverb and the verb were adjacent, as in (5), reported a left anterior negativity, while the studies where the adverb and the mismatching verb were divided by one or more lexical constituents, as in (6) and (7), triggered a more central/posterior negativity. It thus seems that distance may play a role in the detection of the temporal violation. No ERP study has ever explicitly tested the effect of distance on temporal violations, but there is some behavioral evidence.

In an eye-tracking study, Biondo et al. (2018) manipulated both the grammaticality of the tensed verb and the distance between the verb and the adverb. Crucially, the authors found that the distance of the two constituents affected the processing of the concord violation. When the adverb and the verb were adjacent the effects of temporal mismatch showed up in late measures, while when the adverb and the verb were distally located the mismatch effect was found from early measures on. In other words, the larger the distance between the adverb and the verb, the earlier the detection/processing of the temporal violations. The authors interpreted these findings within a predictive framework. If phrases encoding a complex lexical content, such as nouns, require time to be semantically interpreted (e.g., Frazier & Clifton, 1998; Kreiner, Garrod, & Sturt, 2013; see also Chow, Lau, Wang & Phillips, 2018), some time may also be needed to anchor the temporal specification of the adverb to discourse (i.e., Speech Time in the left periphery of the sentence). Under the assumption that sentence comprehension proceeds in an incremental way and that anchoring requires time to be completed, one may expect that when the verb immediately follows the adverb, the temporal specification

retrieved from the lexical representation of a deictic temporal adverbial phrase could not be immediately available to syntactic processing. The system does not have enough time to correctly and fully activate a temporal representation provided by the adverb, leading to a more delayed detection of the violation. On the contrary, if some words intervene between the already parsed adverb and the verb, it can be more likely that there is more time to solidly anchor deictic information to discourse, allowing an earlier detection of the temporal mismatch. These findings suggest that distance may play a role in the unfolding of some mechanisms such as the extraction of discourse-related properties from the linguistic input, thus making the study proposed here specifically relevant to better understand the processing of temporal concord. Moreover, ERPs are particularly sensitive to qualitative changes in the processing of a relation, thus helping to qualify the nature of these extra costs found behaviorally.

2. The current study

The main aim of this study was to investigate how the parser deals with temporal concord violations, and whether the processing of temporal concord is affected by the distance between the two relevant constituents, the adverb and the verb. All previous ERP studies on temporal concord violations reported an early negativity, so we expect both adjacent and distal temporal violations to be detected early in processing (300-500ms post-verb). Moreover, the ERP literature on temporal concord, together with recent behavioral studies (e.g., Biondo et al. 2018), suggested that the distance between the adverb and the verb should affect early stages of processing, probably in terms of earliness and/or easiness of detection of the violation. We thus expected distance to affect the early ERP components. If the anchoring to discourse requires time (Biondo et al., 2018) and discourse-related processing can affect early ERP components (Mancini et al., 2011) we should expect an N400-like component for distal violations, where more time is given to the anchoring process to be completed, compared to the adjacent violations where a LAN effect could be found (i.e., a pure morphological mismatch detection). This pattern would also be in line with previous ERP findings reported in Table 1, showing LAN effects for adjacent temporal mismatches and N400-like effects for distal temporal mismatches.

Whether the adverb-verb distance also affects later stages of processing is hard establish because all previous studies reported a P600 effect, and possible amplitude and topographic differences were not directly

documented, and thus hard to be quantified across studies. However, a modulation of the P600 amplitude as a function of the adverb-verb distance cannot be excluded, especially in light of the previous studies investigating the effect of distance during subject-verb agreement processing. Kaan (2002) tested whether intervening material between the subject and the verb affects feature tracking, integration, or the revision processes related to the detection of a subject-verb mismatch. The ERP components triggered by the mismatch (broad negativity followed by a P600 and a sustained end-of-sentence negativity) were not affected by distance, neither in early nor in late stages of processing. The effect of distance was only visible in the behavioral response: the readers were less accurate in judging the ungrammaticality of the sentences where the subject and the verb were distally located (compared to the sentences with an adjacent mismatch). Shen, Staub & Sanders (2013) also tested the effect of distance in an ERP paradigm. The ungrammatical sentences where the subject and the verb were adjacent (compared to their correct counterpart) showed an early anterior negativity followed by a P600. The ungrammatical sentences in the distal condition (compared to their control) triggered an early posterior negativity that the authors interpreted as an instance of N400. It should be noted, however, that the authors opted for a more naturalistic task in this study, namely a text describing a story in which semantic and pragmatic factors clearly played a stronger role and possibly induced a larger involvement of N400-related processes. It is thus hard to make a comparison with previous studies testing the same conditions in isolated sentences. More recently, Rispens & Amesti (2017) manipulated both the number (0, 1, 2) and the type of constituents (adverb, prepositional phrase containing agreement features) located between a mis/matching subject-verb relation in isolated sentences. They replicated the behavioral finding of Kaan (2002), since participants were less accurate in judging the distal violations than the adjacent ones. In the early time window (300-500ms) a posterior negativity was found for all the violations, independently from the number and type of interveners (the type of constituent did not affect any behavioral response or ERP component). Crucially, an effect of distance was found in the 500-1000ms time window: the distal condition elicited a larger P600 than the adjacent condition. The authors interpreted these data as evidence that two distally constituents are harder to integrate and more complex to process as compared to two adjacent constituents[10]. In sum, based on previous studies

[10] This interpretation, however, is rather speculative since the authors did not find any significant grammaticality x distance interaction, but only two main effects of grammaticality and distance (Rispens & Amesti, 2017: 169)