Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

Tanulmány

Ameni Mejri

L2 online anaphora processing by Tunisian Arabic speakers

Abstract

This second language study explores anaphora resolution by Tunisian Arabic learners of ESL. The aim of the study was to investigate if there are any intrusion effects and if second language proficiency has a direct effect on the time and accuracy of anaphora resolution. Therefore, an online anaphora interpretation task in different experimental conditions was performed. These conditions include manipulating the gender information of the antecedent or an NP distractor, where the anaphora might be temporarily distanced from the antecedent. The study results confirmed the intrusion effects of the multiple NPs in the test sentences. The Tunisian participants were distracted by the competing NP before ultimately selecting the right antecedent. Constructions where the inaccessible antecedent mismatched the reflexive did not constitute a problematic case of anaphora resolution. However, when the inaccessible antecedent gender matched the reflexive, the accuracy rates diminished. The proficiency of the participants showed a significant correlation with their accuracy of anaphora resolution. In addition, there was a negative correlation between the proficiency of the participants and the time of response. Based on these results, it is postulated that anaphora interpretation in the second language setting can be characterized by the intrusion effect, unlike the L1 context. The proficiency results are taken as a confirmation that L2 competence can directly affect the accuracy and the time needed to interpret anaphora by English second language speakers. Accordingly, teachers of ESL may consider the integration of pragmatic gender-related information in their syntactic curricula and train learners to handle gender cues in syntactically complex structures. Addressing the pragmatics-syntax interface in ESL curricula can be equally beneficial to intermediate and advanced learners.
Keywords: Anaphora, Binding Constraints, Intrusion Effect, Second Language Acquisition, Tunisian Arabic

1 Introduction

In second language research, the utilization of experimental conditions, along with the consideration of cultural influences, investigation into L2 proficiency, and assessment of cross-linguistic applicability, collectively enhance current grasp of anaphora resolution. Within this framework, delving into the interpretation of reflexive pronouns among non-native speakers provides a comprehensive outlook on the complex processes that form the foundation of language comprehension.

Research on reflexive pronoun interpretation has focused on the role of the binding theory in constraining the potential antecedent in different experimental conditions. In particular, various L1 studies have examined two antithetical views of anaphora resolution: the structure-

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

based approach versus cue-based antecedent retrieval (Philip et al. 2011). In the former, identifying the permissible antecedent for a referring expression entirely depends on structural information relevant to syntactic hierarchies. Strictly pairwise comparisons between the two NPs in question are run at a time. This approach predicts that stimuli with or without competing antecedents (e.g., *the pilot$_i$ injured himself$_i$,* vs. *the pilot$_i$ who scared Mary$_k$ injured himself$_{i/*k}$*) do not pose a processing demand. Such a claim has been experimentally supported by different studies that deployed a plethora of research methods, akin to eye-tracking (Sturt 2003; Dillon et al. 2003), cross-modal priming (Nicol & Swinney 1989), ERP technique (Xiang et al. 2009). The structure-based approach encounters a formidable challenge from the cue-based model, which posits that a collection of content cues interferes in the identification of the correct antecedent. Processing guided by cues beyond the binding structure involves the consideration of all potential candidates within a given stimulus, irrespective of their grammatical accessibility or gender compatibility. Consequently, this primary candidate set may encompass both ungrammatical and gender-mismatched noun phrases, alongside the admissible ones. While this processing type exhibits notable accuracy, it comes at the cost of increased time required to process all the noun phrase candidates. In simpler terms, the deliberation of competing antecedents with inaccurate phi-features (sometimes intentionally manipulated in select experiments) culminates in extended reading and processing durations, attributable to the reliance on multiple cues and the activation of improper candidates (Philips et al. 201; Lewis et al. 2006). Straub and Badecker (2002) corroborated this approach through a self-paced reading experiment, revealing the interference effect caused by distractor noun phrases in anaphora resolution.

Continuing the same research trajectory, the exploration extends into the realm of second language acquisition. However, consensus remains elusive regarding the timing and status of each processing model. Studies conducted in second language (L2) acquisition, such as those by Felser and Cunnings (2012), Bertendshaw (2009), and Felser et al. (2009), have undertaken the investigation of whether L2 learners possess an inclination towards structural information or instead lean on non-structural cues for the establishment of anaphoric connections. These inquiries aimed to delineate the role binding-related information plays in locating the apt antecedent. Among the findings, studies involving German L1 speakers (Felser & Cunnings 2012) and Japanese L1 speakers (Felser et al. 2009) underscored the consideration of discourse-driven, albeit binding-unavailable, antecedents in the cognitive process. While these results spotlight the interfering impact of distractors, both L2 participants and the native control group displayed a consistent adherence to the principled understanding of antecedent identification.

Building upon this foundation, the present study endeavors to extend the trajectory into the L2 context. It seeks to discern whether Tunisian Arabic speakers' anaphor resolution is informed by structural cues or by non-structural cues. In pursuit of this objective, the manipulation of gender within both legitimate and distractor noun phrases follows the method established in Sturt (2003). Additionally, a set of appropriate and stereotypical noun phrases is introduced to further challenge participants. Essentially, the study adopts experimental conditions from Sturt and Cunnings (2014), which explore the alignment and misalignment of possible antecedents with the reference expression in terms of gender features. This strategic manipulation aims to unravel whether such information introduces confusion among participants. The linguistic stimuli and research conditions draw upon diverse studies conducted in first language (L1) processing. Concretely, the test sentences amalgamate resources from Dillon et al. (2013), Sturt and Cunnings (2014), and Sturt (2003). The goal of this endeavor is

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

to ascertain the validity of L1 claims when transposed to the L2 environment. A lens sharpened by the L2 proficiency of the Tunisian speakers is applied to assess this research objective. As a result, the following inquiries come to the forefront:

A. Does L2 proficiency of Tunisian ESL learners affect the accuracy of antecedent identification?
B. Do Tunisian ESL learners use the binding information to identify reflexive pronoun antecedents in gender-manipulated sentences?
C. Is the time needed for antecedent identification affected by gender-manipulation?

The structure of this paper is divided as follows: section 2 treats the relevant theoretical and linguistic background of the L2 participants, section 3 is devoted to surveying relevant studies in L1 and L2 contexts that treat similar research foci, and finally sections 4 as well as 5 report the study methodology, results and discussion, respectively.

2 Background

2.1 Binding

The binding theory establishes the binding conditions governing the interpretation and distribution of noun phrases. Concerning anaphors, their distribution typically adheres to the guidelines set forth in Principle A, while Principle B outlines the distribution of non-reflexive pronouns.

1. Binding Principles
   Principle A: An anaphor is bound in a local domain.
   Principle B: A pronoun is free in a local domain (Hageman 1994).

Principle A stipulates that anaphors must be bound within a clause-level scope to a nearby binder. In contrast, Principle B suggests that pronouns can be utilized deictically within discourse; they are not bound at the clause level. Instead, they gain interpretation based on their co-reference with a prominent NP/entity within the discourse. The distinct syntactic contexts that host pronouns and anaphors, alongside their respective distribution patterns, ensure their complementary distribution. Reflexive pronouns find their binding where non-reflexives remain unbound.

In the following example (2), the anaphor *herself* is dependent and syntactically bound by the antecedent *Mary*.

2. Mary_{i} hurt $\text{herself}_{i/*j}$
3. Mary_{i} hurt $\text{her}_{*i/j}$

In (3), the pronoun, *her*, cannot be bound by the antecedent, *Mary*, given its free distribution. Examples (2) and (3) show the complementary distribution of reflexives and non-reflexives.

Following condition A in the binding theory, it is correctly predicted that the licit antecedent in (4) is matrix subject.

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

4. a. The pilot$_i$ injured himself$_i$
   b. The pilot$_i$ who scared Mary injured himself$_i$ (King et al. 2012: 67)
   c. The pilot$_i$ who scared Mary$_j$ injured herself$_{j*}$

The pilot in these stimuli is the local and c-commanding NP that can syntactically bind the referring expressions. Placing an embedded phrase in (4a) temporarily separates the binder and the bindee, yet the binding relation stays intact. Both constituents in question satisfy the locality and c-command requirements of condition A. In other words, *the pilot* is still higher in position to c-command the dependent expression.

In line with the binding theory, in (4b), the anaphor, *herself,* cannot corefer with the distractor NP, *Mary*. The syntactic position of *Mary* in the sentence does not satisfy the c-command requirement of being higher in position, and being under the same branching node [IP] that dominates the referring expression and *Mary*. This latter is not available as a licit antecedent in this configuration.

Moreover, Sentence (4c) poses a processing load given that the gender features of the anaphor do not match the permissible antecedent, *the pilot*, yet they match the impermissible NP, *Mary*. In online experiments of anaphora resolution that rely on cue-based processing, it is expected that (4c) would require higher processing times due to evaluating the binding possibilities of all involved NPs, and due to the gender-mismatch between the NPs in question. Considering *Mary* as a potential binder by referring to its gender-compatibility would be costly and end up by ruling out this possibility. However, relying on structure-based processing, (4c) might be demanding only because of the mismatch between *the pilot* and *herself*. From this perspective, *Mary* would not be part of the candidate list of antecedents. The antecedence relationship will be evaluated based on a pairwise fashion, and it would be ruled out due to the noted mismatch. In that, the embedded clause engenders a processing load by supplying an NP distractor that is evaluated for binding possibilities. In experimental studies on anaphora resolution, considering the NP, *Mary,* as a potential binder based on its gender features is an indication of an intrusion effect, as will be presented in the subsequent section.

2.2 Intrusion effects

In antecedent-based binding accounts, the establishment of an anaphoric dependency relies on binding constraints. This concurs with the predictions of the canonical binding theory. According to these predictions, it is exclusively the binding information that delineates all possible antecedents. Any other information involved in this process is deferred, as binding functions as an initial filter to define legitimate antecedents. Sturt (2003) referred to this role of the binding theory as "binding-as-initial-filter," operative from the earliest stages of processing. Any competing, albeit unsuitable, antecedent is excluded from computation and disregarded in subsequent antecedent identification.

Nevertheless, the role of the binding theory isn't universally perceived as an initial filter. In the cue-based retrieval model, non-structural cues are harnessed in the task of antecedent identification. These cues encompass factors such as discourse prominence, gender features, grammatical position, or subjecthood. Instead of engaging in dual computations for the relevant constituents, gender agreement, for instance, might be employed to assess all candidate noun phrases for an anaphoric dependency (Dillon et al. 2013).

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

5. The pilot$_i$ who scared John$_k$ injured himself$_i$ (King et al. 2012: 67)

In sentence (5), the NPs, *the pilot* and *John*, are both considered members of the early set of candidates. Later in the processing, one of these NPs will have to be ruled out. Membership in this primary set is based on all the gender-matching NPs, which is taken as an indicator of the intrusion effect where information, other than binding, is informing the process.

The cue-based model operates by activating all potential candidates within the given stimuli to discern the appropriate antecedent. In contrast, the structure-based model confines its computation to a pairwise analysis involving only the relevant noun phrases (Philips et al. 2011). The former approach introduces the concept of the intrusion effect, suggesting that binding doesn't serve as an initial filter for establishing anaphoric relationships. Within cue-oriented processing, the timing, role, and potency of binding information remain less precisely defined. Instead, a suite of non-structural cues takes center stage, yielding heightened accuracy albeit accompanied by prolonged reading and processing times.

Through this study, the aim is to probe whether Tunisian Arabic speakers demonstrate intrusion effects in their performance, triggered by their reliance on non-structural cues for anaphora resolution. By tracking the temporal sequence, the aspiration is to unravel how L2 Tunisian speakers evaluate and ultimately determine antecedents.

2.3 Reflexive pronouns in Tunisian Arabic

The English and Tunisian Arabic (TA) pronominal systems are generally different from each other. However, both language systems have a common ground when it comes to the structure and distribution of reflexive pronouns. The reflexive pronoun in TA is composed of the invariant noun "ruh" that stands for "self", to which a pronominal suffix is attached. Thus, the equivalents of the English "himself" and "herself" are "ruhu" and "ruha", respectively. The following table summarizes the TA pronouns and their English equivalents.

TA reflexives	**Features**	**English counterpart**
Ruh-i	1st singular	myself
Ruh-na	1st plural	ourselves
Ruh-ik	2nd singular	yourself
Ruh-kum	2nd plural	yourselves
Ruh-u	3rd masculine singular	himself
Ruh-a	3rd feminine singular	herself
Ruh-hum	3rd plural	themselves

Table 1. Reflexive pronouns in TA

The binding similarities between these two languages are exemplified in the binding of the reflexive to a local antecedent in the strict sense. For instance, in (6) the referent, *ruhu*, is the verbal coargument of the antecedent, *Jad*, representing in this way a typical reflexive pronoun distribution.

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

6. Jad$_i$ jrah ruhu$_i$
   Jad hurt.PST.3SM himself
   Jad hurt himself

Akin to English, TA reflexives fall within the same category of having verbal coargument reflexives that strictly abide by Principle A in their distribution. This is congruent with the complementary pronominal distribution where the reflexive is grammatical and the pronoun is ungrammatical.

The test sentences used in this study represent a similar context for the distribution of anaphors in English and Tunisian Arabic. The following examples (7) and (8) as well as (9) and (10) are the English stimuli and their TA equivalents, respectively. This distribution instances of the reflexive in both languages is identical.

7. Jonathan was walking through the military barracks. He heard that the soldier$_i$ had positioned himself$_i$ in the middle of the mess hall. (Cunnings & Sturt 2014: 59).
8. Jonathan kan yemshi fi thakanat al-jaysh. Sma'a annou al-jundi$_k$
   Jonathan walk.PST.PRGRV through the military barracks. Heard.PST that the soldier
   waqaf ruhu$_k$ fi el-qa'at al-mekla
   position.PST himself in the middle of the mess hall
9. The new executive who oversaw the middle managers doubted himself on most major decisions. (Dillon et al. 2013: 89).
10. Al-mudir al-jdid illi ken yshrif a'al mudara' ken yshok
    The new executie who.RELATIV oversee.PSTthe middle managers doubt.PST
    fi ruhu fi mo'dham al-qarart al-ra'isa
    himself on most decisions major

The binding possibilities in the above-mentioned data represent the commonalities in the distribution of anaphors in English and TA, to survey the major reflexive pronoun distributions in the potential participants' L1. Given that L1 anaphors are guided by the binding constraints, it is expected that in the L2 context the participants will not be faced with a distributional pattern with which they are not familiar. Their performance in the L2 context, with stimuli that might activate an intrusion effect, will not be treated as an instance of L1 transfer. There, the time course and the interpretation patterns of data that might involve more than one candidate like sentence (5) (*the pilot$_i$ who scared John$_k$ introduced himself$_i$ to the passengers* (King et al. 2012: 67)) will be used to examine if the Tunisian participants would rely on gender features to identify the antecedent. The adapted linguistic stimuli used in this piece of research do not pose an entirely novel mechanism in language processing that might place any cognitive demands on the L2 participants**.**

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

3 Related Work

3.1 Antecedent retrieval in L1 studies

Psycholinguistic exploration of anaphora resolution delves into the factors influencing antecedent identification within a stimulus-response framework. When investigating the early studies addressing antecedent assignment, the review conducted by Nicol and Swiney (1989) centered on the examination of syntactic constraints' role in anaphora resolution. In essence, Nicol and Swiney aimed to determine whether additional elements, such as grammatical function, agreement, or gender features, were employed alongside or prior to the syntactic rule in the task of sentence comprehension. Their focal point also encompassed the analysis of how the initial set of antecedent candidates is formulated. Notably, their emphasis predominantly rested on online studies, chosen for their precise portrayal of unconscious sentence processing, as exemplified by Nicol (1988).

This study reviewed and included stimuli such as *the boxer told the skier that the doctor for the team would blame himself for the recent injury* (Nicol 1988: 13). The participants engaged in a cross-modal priming study, wherein they encountered a visual representation while simultaneously receiving auditory input, followed by a lexical decision task. The findings illuminated that the binding constraints served to constrain the set of candidates forming the initial roster of possible antecedents. Notably, the licit antecedent within the sentence exhibited significant priming effects, in contrast to the competing NP alternatives. In this investigation, Nicol (1988) reaffirmed that binding constraints acted as an early filter, steering the initial creation of the candidate set and ultimately guiding the selection of the appropriate contender, while disregarding other syntactic information. This research finding is further substantiated by a collection of studies (Cunnings & Sturt 2014; Dillon et al. 2013; Sturt 2003) while being challenged by other research emphasizing the pertinence of a cue-based model for antecedent retrieval and identification (Patil 2016).

Raising inquiries into the robustness of principled grammatical knowledge in molding the candidate set, a series of studies sought to either validate or challenge this discovery through a diverse range of methodologies. Badecker and Straub (2002) embarked on a self-paced word-per-word reading task encompassing six experiments to dissect the interplay and synergy among grammatical binding relations, agreement features, and discourse focus status during online pronoun processing (with a specific emphasis on anaphors). This inquiry follows a lineage of investigations (Nicol 1988; Harris et al. 2000) that postulated the exclusive role of syntactic constraints in shaping the candidate set, aiming to ascertain whether these constraints exclusively serve as the filtration mechanism to eliminate illicit candidates. Their experimental conditions included:

11. *Multiple match: Jane thought that that Bill$_k$'s brother$_i$ owed himself$_i$ another opportunity to solve the problem*
    Accessible match: Jane thought that that Beth$_k$'s brother$_i$ owed himself$_i$ another opportunity to solve the problem (Badecker & Straub 2002: 761)

This choice of the stimuli was based on the researchers' findings that the grammatical function of an inaccessible antecedent in a subject position would lead to slowing the processing time. To redress that, they chose to manipulate the salience of the NP that can be a candidate

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

antecedent. The genitive NP in this stimuli is not prominent in the discourse and the stimuli comprised:

12. *Multiple match: It appeared to John$_i$ that Bill$_i$ owed himself$_i$ another opportunity to solve the problem*
    Accessible match: It appeared to Jane$_k$ that Bill$_i$ owed himself$_i$ another opportunity solve the problem (Badecker & Straub 2002: 762)

John and *Jane* are non-subject candidates that might compete with the accessible/local subject, *Bill*. The expectation would be a higher reading time in the case of the inaccessible match (when the NP, *Jane,* is considered as a potential antecedent), hence the possibility of processing an illicit member in the initial list.

The reported results indicate an antecedent identification accuracy level exceeding 80%, alongside the absence of discernible reading time disparities among various experimental conditions and in the reflexive regions. Badecker and Straub concluded that the absence of effects associated with gender cues and position suggests that antecedents find definition through their focus positions. Subjects emerge as the most salient elements within a grammatical construction, accounting for their role in both subject position and initial membership within the candidate set during anaphoric processing. The immediate impact of subjecthood in parsing underscores the significance of local discourse prominence in the initial formulation of a list of candidates. Proposeing an alternative perspective on the roles assumed by structural binding relations and discourse prominence—specifically, salience via subjecthood or coargumenthood—in incremental resolution, Badecker and Straub deviate from the exclusive emphasis on binding as an initial filter in anaphoric resolution, as advanced by Nicol (1988) and Nicol and Swinney (1989), in favor of a more comprehensive model they designated the "interactive-parallel-constraint model." There, the inception of the candidate set is shaped by local discourse prominence and gender, followed by the rapid operation of binding constraints to select the appropriate antecedent. This interactive model supplants the notion of binding as the primary filtering mechanism. Notably, based on the priming methodology detailed in Nicol and Swinney (1989), there isn't an accurate representation of the initial processing involved in formulating the candidate set. The binding constraints do not eliminate inaccessible antecedents from the early set, thereby accounting for the parallel functioning of salient grammatical entities (subjects) and gender features to restrict the list to feature-compatible noun phrases.

Badecker and Straub (2002) posited the interactive-parallel-constraint model to account for the limitations of Nicol and Swinney's (1989) study. Yet the account they forwarded lacks precision, given the type of online measurement technique used to investigate the role of binding theory in antecedent retrieval. To address some gaps in previous research, Sturt (2003) took up the same research mantle on the role of binding information in the same context. Sturt designed two eye-movement experiments and twisted the experimental conditions to use stimuli that satisfied accessible/inaccessible match, accessible/inaccessible mismatch, accessible mismatch with inaccessible match, and accessible with inaccessible mismatch conditions. The used test sentences (*Jonathan/Jennifer was pretty worried at the City Hospital. He remembered that the surgeon had pricked himself/herself with a used syringe needle. There should be an investigation soon* (Sturt 2003: 546)) included the characters *Jonathan* and *Jennifer*, who were presented as pictures then referred to as *he* or *she*, in order to put them in a position of discourse

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

focus. Sturt (2003) hypothesized that a processing difficulty would arise when the gender of the stereotypical NP, *the surgeon*, and the dependent expression do not match (*Jonathan$_k$ was pretty worried at the City Hospital. He remembered that the surgeon$_i$ had pricked herself$_{i*/k*}$ with a used syringe needle. There should be an investigation soon* (Sturt 2003: 546)). It was also expected that by placing an inaccessible antecedent in a position of discourse prominence, an effect of "binding-inaccessible" would be created, for instance for the NP, *Jonathan*. Binding as an initial filter will operate from the onset of the retrieval and no gender effect would be noted, in the opposite case of the late filter, gender effects would be detected. Should the binding constraint function as defeasible filter, from the beginning of the processing accessible antecedent, intrusion effects will be noted.

The second experiment followed the same experimental conditions, yet addressed the strict linear positioning of the accessible antecedent followed by the reflexive and aimed at excluding the strict linear explanation. Sturt wanted also to examine if distancing the referent and the antecedent would also create a particular intrusion effect. For this reason, a relative clause between the two entities was introduced (*Jonathan/Jennifer was pretty worried at the City Hospital. The surgeon who treated Jonathan/Jennifer had pricked himself/herself with a used syringe needle. There should be an investigation soon* (Sturt 2003: 546)). While the position of the two NPs in question was reversed, their structural binding relation was intact.

In reporting the study findings, Sturt indicated that in the first experiment, the first-fixation and the first-pass reading times were exceedingly short. This was emblematic of the early application of the binding constraints in sorting out the potential antecedent. However, in later processing when the antecedent was not feature-compatible, it was noted that the second-pass times were longer, coupled with regression times to the pre-final region. In other words, when the readers were answered sentences that combined the mismatching and inaccessible antecedent, *Jennifer*, the accessible antecedent, *the surgeon*, and the referent, *himself*, a late processing effect was remarkable, leading to the conclusion that even though the binding constraint was applied at an early stage, the inaccessible antecedents still had a say in the retrieval. In the subsequent eye-movement experiment, Sturt reported no significant effects of the linear order of the reflexive and its antecedent. It was confirmed that the syntactic architecture of the sentence and the interference of another clause between the antecedent and the dependent expression did not alter the course of processing.

Sturt (2003) responded to these findings by confirming the role of gender-related information in anaphoric processing. However, he clarified that this effect did not imply a delayed binding filter. Instead, the binding constraint was operational at an early processing stage, surpassing the impact of these cues, which primarily manifested during later processing phases or recovery. Sturt's results effectively dispelled the notion of a late filter model, where binding remains inactive in the early phases. In fact, Sturt (2003) advocated for the concept of a defeasible filter model, delineating its dual stages: the initial "bonding" phase for establishing anaphoric connections, followed by the subsequent "resolution" phase to evaluate the created dependency in terms of feature compatibility.

The debate surrounding the role of binding information in shaping the set of potential candidate antecedents extended beyond determining the timing of its application. Certain studies, such as those by Dillon et al. (2013) and Patil et al. (2016), addressed this quandary from the perspective of memory-based cueing systems. Their objective was to assess the contribution of various cues, including gender cues. Their approach was grounded in categorizing Sturt's (2003) findings, which collectively suggested that establishing reference

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

necessitated syntactic knowledge combined with gender specifications. This is where a convergence of ideas takes place. On the one hand, Dillon et al. (2013) endorsed Sturt's initial claims that binding is initially operative in antecedent identification, and it is not coupled with any information external to binding. Their confirmation is obtained from combining online and offline methodologies and manipulating the stimuli used. In fact, they manipulated the syntactic environment hosting the NP distractor, and hypothesized that the intrusion effect would result in faster reading times for ungrammatical entries with a matching intruder and a more complicated processing for grammatical stimuli without a matching intruder or NP (*The new executive$_i$ who oversaw the middle manager$_k$(s) apparently doubted himself$_i$/themselves on most major decisions* (Dillon et al. 2013: 89). On the other hand, Patil et al. (2016) found that binding cues and other structural cues were deployed based on the high errors and long-time retrievals for the mismatch condition. These findings were obtained from an eye-movement experiment using (adapted) stimuli from Xiang et al. (2009) (*the tough soldier$_i$ that Fred$_k$ treated in the military hospital introduced himself$_i$ to all the nurses* (p. 44)).

The observed disparities regarding the function of binding information in retrieving antecedents and in constraint-based systems stem from an inaccurate portrayal of the human parser. According to Patil et al. (2016), the human parser does not solely rely on syntactic cues; instead, a parallel cue system might also be operational, with binding configurations as an illustrative example.

In essence, ongoing research exploring the impact of structural and non-structural cues on anaphora resolution has given rise to two distinct research perspectives. The first perspective posits that structural information is crucial in forming the initial candidate set. This processing approach is often considered representative of how native speakers (L1) process language. This viewpoint is supported by studies that focus solely on anaphora configurations (Sturt, 2003), as well as those that compare different language structures, such as subject-verb agreement and anaphora (Dillon et al. 2011).

On the flip side, the constraint-guided model, which characterizes the second perspective, doesn't offer a consistent processing pattern due to its limitations in handling various language elements during computations. This model relies on multiple information sources, which can strain memory, and subsequently results in longer processing times (Philips et al. 2011). While consensus remains elusive within these two perspectives, the broader research consensus favors the structure-based model. On the other hand, the cue-oriented model is commonly associated with characterizing how second language (L2) learners process language. Due to their limited familiarity with L2 syntax, L2 learners often depend on lexical and structural cues to navigate anaphora resolution (Felser 2016). Given that L2 processing tends to take more time, the cue-based model seems to be more suitable for accommodating the various complexities of L2 processing. It is, thus, worth noting that examining this issue within the context of L1/L2 learners, is still in its nascent stages.

3.2 Antecedent retrieval in L2 studies

Numerous studies have explored anaphora processing in second language research, comparing it with native language context. In L1 contexts, condition A of binding is seen as the initial filter guiding the initial candidate set (Dillon et al. 2013; Sturt 2003). However, non-native reflexive processing doesn't rely on this early filter (Felser 2016).

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

In a study involving 32 Japanese speakers and 37 English controls, Bertanshaw (2009) investigated Japanese speakers' interpretation of English reflexive pronouns. Two experiments were conducted - an online eye-movement test and an offline test - using gender manipulation conditions. In offline performance, learners identified local antecedents but with imperfect interpretation. The online test showed Japanese speakers impacted by manipulations, exhibiting higher fixation and regression for inaccessible antecedents compared to native speakers. Proficient speakers demonstrated faster processing, suggesting binding constraint acquisition. However, L2 learners' anaphora processing differs from native speakers, influenced by gender manipulations and discourse prominence, resulting in intrusion effects and extended reading times, conferring a divergence in anaphora processing between the studied groups.

The same foci were also the subject of investigation in another study conducted by Felser, Sato and Bertenshaw (2009) on Japanese L1 speakers, whose language can allow for the long-distance binding of the anaphor *zibun*, the question was to examine if these speakers would opt for long-distance binding with the English reflexives that allow only for the local argument binding. Their paradigm was gender mismatch which targeted only the inaccessible antecedents based on the hypothesis that binding as an initial filter would eliminate this set of antecedents in this first computation. By endorsing the eye-movement monitoring methodology while the participants read sentences contextualized in some discourse, the researchers found that the Japanese speakers fixated on the gender-matching yet inappropriate antecedent, unlike the native control group. In other words, the informants were distracted by the antecedents in the following test data:

13. *Jane/John$_k$ noticed that Richard~~$_i$~~ had cut himself$_i$ with a very sharp knife.*
14. *It was clear to Jane/John$_k$ noticed that Richard$_i$ had cut himself$_i$ with a very sharp knife*. (Bertenshaw 2009: 126).

The distraction resulted from cues such as the phi-features of the reflexive that matched those of the inappropriate antecedent. The other finding forwarded in this study was this distraction marked only the antecedents in the first sentence type where they are in a matrix position (13) unlike the sentence type (14) where they are in a less prominent position. Such a fact suggests, therefore, that antecedents in a non-subject position were not a source of a remarkable distraction in the processing, an effect which was also documented by Badecker and Straub in L1 context where subjecthood or discourse prominence was central in early antecedent processing. It is worthwhile to note also this distraction occurred only in the first reading/processing phase. While this study showed the effect of the gender-mismatch paradigm and subjecthood effect noted in L1 studies, the used methodology could not provide a categorical explanation in case the Japanese speakers transferred the L1 long-distance binding property or they had a tendency to compute the syntactically prominent antecedents regardless of being accessible or inaccessible.

Given the limitations of the previous experiment and the raised questions that the data generated, it was necessary to conduct another study to address those inquiries with more stratified test data and design. Felser and Cunnings (2012), thus, employed the same eye-movement technique on German participants who had advanced proficiency in English. The gender manipulation they created resulted in 4 sets such as double match, accessible match/ inaccessible match, accessible mismatch/ inaccessible match, and double mismatch.

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

The effects of this layered gender-mismatch manipulation revealed that the German speakers fixated on the inaccessible antecedents and seemed to recognize the appropriate antecedent only after reading the entire test sentence. It is worth noting that in German, that 3rd person reflexive pronouns do not display any gender feature differences. The remarkable processing pattern was that the German participants interpreted the reflexive as referring to the matrix subject (*John* or *Jane*), unlike their native counterparts, who showed no signs of the gender-mismatch effects of inappropriate antecedents. The researchers accentuated that while non-native speakers relied on discourse-related information to establish anaphoric links, the native speakers relied on their principled syntactic knowledge. Such a fact was also indicative that L2ers were not guided by their L1 syntactic knowledge but rather by discourse. The results of both the L2ers and L1 speakers indicate that intrusion effects are only part of L2 processing. It is worthwhile to note that binding as an initial filter is confirmed in L1 processing as advocated by Sturt (2003) and Dillon et al. (2013), among others, while in L2 anaphora processing, the interactive-parallel-constraint model advocated by Badecker and Straub (2002) seems to be more applicable than in the L1 context. Such a claim needs further examination on a multitude of populations and using various methodologies.

The obtained results in L2 research seem inconclusive on the mechanisms involved in anaphor processing in L2. This is traced back to a set of factors that include the methodology used (eye-movement versus self-paced reading in the case of L1 and L2), the type of the constructions tested (reflexives in object positions following the verb (in the strict linear order)), and the populations involved (German or Japanese among other languages, but not Arabic so far). For this reason, the current study aims at testing the role of the binding constraint in L2 anaphor processing via combining test sentences that include experimental conditions of twisting the compatibility features of the accessible and inaccessible matches and different construction types where the reflexive is in a strict adjacent position with the antecedent or is temporarily distanced from the antecedent (non-adjacent). According to King et al. (2012), this type of test construction can be more informative of the role of the binding constraints in establishing immediate links in case of an NP distractor is or is not present, hence more informative in reporting any intrusion effects. By examining these effects in L2 anaphor resolution, this study seeks to draw on the findings of the L1 setting where no intrusion was detected, although sometimes predicted, to forward the hypothesis that in the L2 context relying on cues other than the syntactic constraints will result in longer reading times and will be indicative of a limited L2 proficiency. Proficient participants would rely on their L2 syntactic knowledge to establish anaphoric links whether or not an NP distractor is present.

4 Methodology

4.1 Hypotheses

The present study explores antecedent identification by Tunisian ESL learners in different syntactic contexts by relying on stimuli used in the L1 context (Dillon et al. 2013; Sturt 2003). Given the difference between both settings, the current L2 study focuses on L2 proficiency in the accuracy of identifying antecedents. On the other hand, this study aims at measuring the time needed to identify the antecedent of the reflexive. To achieve these aims, the test sentences are based on the following experimental conditions in the presence of a competing antecedent:

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

a. Accessible antecedent match, inaccessible antecedent mismatch
b. Accessible antecedent match, inaccessible antecedent match
c. Inaccessible antecedent mismatch, accessible antecedent mismatch

Accessible in the present study refers to the structurally appropriate antecedent according to the syntactic binding conditions (i.e., the closest antecedent c-commanding the reflexive), whereas inaccessible refers to structural inappropriateness according to the same binding conditions. For instance, *John/Jane in (13)* are inaccessible antecedents, whereas Richard is the accessible antecedent. Whereas match denotes number, case and gender agreement between the reflexive and the antecedent, the present study focuses primarily on gender match and mismatch, which arise from the association of the proper noun with one gender or the gender stereotypes associated with NPs such as *footballer*.

The following hypotheses will be evaluated based on the obtained data:

H1: self-reported L2 proficiency of the Tunisian Arabic participants correlates with the accuracy of anaphora interpretation in this experiment.
H2: self-reported L2 proficiency of the Tunisian Arabic participants negatively correlates with the time needed to process anaphors in this experiment.
H3: Gender manipulation will be associated with a shorter reading/processing time for conditions (a) and (b) and a longer reading/processing time for condition (c).
H4: The use of stereotypical NPs will correspond to longer processing time and less accurate responses than the use of proper nouns in NP.

4.2 Instrument

4.2.1 The survey

In order to examine the hypotheses mentioned above, the current study is based on an online survey that measures the time-course of reflexive pronoun interpretation by Tunisian second language learners. Time measurement covers the entire reading time for every test sentence. For this reason, time information will be interpreted by reference to anaphora processing. Furthermore, this survey aims at using and adapting some of the stimuli used in L1 research in second language research. To this end, three experimental conditions are created in order to evaluate the intrusion effect, if any, for L2 speakers. The following are the conditions of this study and their respective stimuli:

15. *Experimental conditions*
   a. Accessible antecedent match, inaccessible antecedent mismatch
      The pilot$_i$ who scared Mary$_k$ had injured himself$_i$ quite badly during the journey.
      Victoria$_i$ worked in the fire station in town. She knew that Bill$_i$ had injured himself$_i$ by the table in the room.
   b. Accessible antecedent match, inaccessible antecedent match
      Jonathan$_k$ was walking through the military barracks. He heard that the soldier$_i$ had positioned himself$_i$ in the middle of the mess hall.
   c. Inaccessible antecedent match, accessible antecedent mismatch

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

Richard$_k$ remembered the tour of the national stadium. He recalled that Jack$_i$ had prepared herself$_j$ in the brand new locker room.

Starting with (15a), the used stimuli include a stereotypical gender NP (*the pilot*) following Dillon et al. (2013), and a proper noun antecedent, in order to provide both alternatives for the informants. The aim of using a stereotypical NP is to further increase the confusion for the L2ers and see if such a particular manipulation can affect the course of their interpretations. As indicated by Sturt (2003), when the stereotypical gender NP does not match the binding information of the sentences, processing might be delayed. Measuring such an effect is also part of the current study.

In (15a), the only accessible antecedents for *himself* are *the pilot* and *Bill*, respectively. In each test item, the reflexive matches the stereotypical NP and the proper noun antecedent, and the NP distractor is both gender-incompatible and binding-inaccessible. The opposite scenario is evoked in (15b), where the binding-inaccessible antecedent, *Jonathan*, matches the phi-features of the reflexive pronoun. With (15c), the situation is even more complicated when there is no available antecedent for the reflexive. The accessible NP, *footballer*, does not match the gender traits of the reflexive and the matching NP, *Richard*, is not available for binding in this construction.

4.2.2 Procedure

We used Psytoolkit (Stoet 2012; 2017), a software package for programming psychological studies, for experimenting with the participants. Psytoolkit measures the responses and the time course needed to reach every response in the test. A link was generated to be distributed online to the potential comprehenders. The initial screen of the link presented the research focus and provided contact details for the informants should they need any further information. Following that, they had to answer a total of 24 test items that included 7 test sentences and 7 fillers, followed by 7 biographical items presented at the end of the survey. The test sentences were associated with a (highlighted) question and four alternatives. They were presented as follows:

16. The pilot who scared Mary had injured himself quite badly during the journey.
    Himself refers to?
    – Mary
    – Either the pilot or Mary
    – Neither the pilot not Mary
    – The pilot

The questionnaire contained an equal number of filler sentences to distract the participants from the main focus on reflexive pronouns. The same structure of the target test sentences was used in the fillers: a stimulus coupled with a question and four alternatives. The choice of vocabulary in the questionnaire aimed at accommodating different proficiency levels and avoiding any meaning-related confusion or delays.

In order to collect responses, the web-based questionnaire was sent by email to teachers and colleagues who shared it with their students. The potential participants were informed that they needed to access the link on their own devices. They were also informed that the time was being measured while they filled in the survey, so that they would be encouraged not drop in the

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

middle of the questions. Data collection remained active for five days in order to obtain more answers. Given that this is a pilot study, the initial number of responses, constituting 24 complete online surveys, was sufficient to run the analysis.

4.3 Participants

Forty-five adult Tunisian participants aged between 18 and 35 years old were recruited. They were learning English as a second language at the time of the study. Participants learned Arabic as their first language and started learning English at least three years ago, indicated they have no experience living in an English-speaking country for more than six months, and were enrolled in any English language courses.

Twenty-six of the participants were females, while 19 were males. Their age distribution is shown in Table 2.

Age category	Number of participants	Percentage
18-20 years old	6	13.3%
21-23 years old	10	22.22%
24-26 years old	13	28.88%
27-29 years old	7	15.55%
30+ years old	9	20%

Table 2. Participants' age distribution

As far as their educational level is concerned, 22 participants are BA level (48.88%), while 20 are MA level (44.44%) and 3 are PhD level students (6.66%). The participants belong to different majors and are not English linguistics practitioners, they are all naïve to linguistics.

Most of the students who took part in this study belonged to study programs that are based on French as a language of instruction in Tunisia, or English as a language of instruction in Hungary. When self-reporting their English level proficiency, they indicated some varied values, associated with some standardized assessment scores (TOEFL or IELTS), if any (Table 3).

Proficiency levels	Number	Standardized assessment score	Number	Experience in an English-speaking country	Number
Beginner level	1	0-5 IELTS/ 0-45 TOEFL	3	0-3 years	5
Elementary level	4	5.5-7.5 IELTS/ 46-109 TOEFL	6	4-6 years	2
Intermediate level	15	8-0 IELTS/ 110-120 TOEFL	1	7-10 years	1
Upper-intermediate level	14	No IELTS or TOEFL scores	35	10 years and more	2
Advanced level	11			No experience	35

Table 3. Participants' English language proficiency

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

The biographical data obtained in this research indicates that the participants' proficiency ranges mainly between the intermediate, upper-intermediate and advanced levels than the beginner level. Such rates are expected to affect the accuracy and the time-course of the responses. The limited access to English-speaking country experience can also account for the limited proficiency. The participants are mainly exposed to English in the formal context of education. It is important to note that self-reported proficiency levels are not entirely based on accurate assessment scores. They are primarily based on students' impressions of their English language level, and they are taken as approximations of their authentic L2 competence. In fact, their performance on the interpretation of the reflexive pronouns is going to be taken as a representation of their L2 level.

5 Results and Discussion

5.1 Results

The findings demonstrated anaphora resolution accuracy ranging from 95.55% to 60%. Since this L2 study lacks a native control group, these accuracy levels will be benchmarked against an ideal theoretical native performance of 100%. Table 4 provides a summary of the results, indicating the lowest accuracy levels for critical sentences, particularly under more complex experimental conditions.

Experimental conditions	**Test sentences**	**Antecedent**	**Number of correct responses**
Accessible antecedent match, inaccessible antecedent mismatch	1. The pilot who scared Mary had injured himself quite badly during the journey	The pilot	43
	2. The new executive who oversaw the middle managers doubted himself on most major decisions.	The new executive	39
	3. Miranda was confused by the size of the depot. She knew that Jenny had lost herself near the back of the store.	Jenny	42
	4. Victoria worked in the fire station in town. She knew that Bill had injured himself by the table in the room	Bill	41
Accessible antecedent match, inaccessible antecedent match	5. Jonathan was walking through the military barracks. He heard that the soldier had positioned himself in the middle of the mess hall	The soldier	37
	6. The pilot who scared John introduced himself to the passengers.	The pilot	38
	7. Joana worked out after a long day at the office. She learnt that her colleague Marry trained herself in the same gym as well.	Marry	39

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

	8. Anthony watched the operation in the hospital. He remembered How Bernard taught himself the necessary tricks to survive all the stress.	Bernard	41
Inaccessible antecedent mismatch, accessible antecedent mismatch	7. Richard remembered the tour of the national stadium. He recalled that the Jack had prepared herself in the brand new locker room.	Jack	41
	10. Harry enjoyed walking in the local sports ground. He was glad that Tony taught herself how to paint the lines on the pitch.	Tony	32
	11. Alice found the surgery very busy that morning. She was surprised that Anna criticized himself for being late for the appointment.	Anna	24
	12. Peter knows some strange stories of the seaside hotel. The receptionist who has been working there told him that John killed herself early May.	John	27

Table 4. Summary of the performance results

Analyzing this data reveals that instances in which the inaccessible antecedent did not align with the reflexive, posed no anaphora resolution challenges. However, when the gender of the inaccessible antecedent matched the reflexive, accuracy rates dropped.

Shifting focus to the experiment's time course, the provided values correspond to the performance rates. Timed data illustrates that accuracy levels increase as the time required to identify the antecedent decreases. Responses with lower accuracy tend to exhibit lengthier reading times, as seen in sentences (1), (3), and (4). However, certain critical sentences such as (7) and (8) indicate that while processing time was longer, accuracy remained intact. This contrasts with sentences (10), (11), and (12), where accuracy suffered. Notably, the 45 participants consistently showed either faster or slower processing with all study stimuli. The most time-intensive sentences, taking longer average reading and processing times, were (5), (7), (11), and (12). These sentences included stereotypical NPs that matched the reflexives in terms of gender traits but lacked binding compatibility. Table 5 displays an overview of average time per sentence in seconds alongside accuracy levels.

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

Experimental conditions	**Stimuli**	**Average in seconds**	**Accuracy in percentage %**
Accessible antecedent match, inaccessible antecedent mismatch	1. The pilot who scared Mary had injured himself quite badly during the journey.	17.53	95.55
	2. The new executive who oversaw the middle managers doubted himself on most major decisions.	22.82	86.66
	3. Miranda was confused by the size of the depot. She knew that Jenny had lost herself near the back of the store.	19.70	93.33
	4. Victoria worked in the fire station in town. She knew that Bill had injured himself by the table in the room.	17.06	91.11
Accessible antecedent match, inaccessible antecedent match	5. Jonathan was walking through the military barracks. He heard that the soldier had positioned himself in the middle of the mess hall.	29.08	82.22
	6. The pilot who scared John introduced himself to the passengers.	25.28	84.44
	7. Joana worked out after a long day at the office. She learnt that her colleague Marry trained herself in the same gym as well.	13.29	86.66
	8. Anthony watched the operation in the hospital. He remembered How Bernard taught himself the necessary tricks to survive all the stress.	42.33	91.11
Inaccessible antecedent mismatch, accessible antecedent mismatch	7. Richard remembered the tour of the national stadium. He recalled that the footballer had prepared herself in the brand new locker room.	34.60	91.11
	10. Harry enjoyed walking in the local sports ground. He was glad that Tony taught herself how to paint the lines on the pitch.	23.11	71.11

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

	11. Alice found the surgery very busy that morning. She was surprised that Anna criticized himself for being late for the appointment.	30.01	53.33
	12. Peter knows some strange stories of the seaside hotel. The receptionist who has been working there told him that John killed herself early May.	36.46	60

Table 5. Average in seconds and accuracy per sentence

The study aimed to examine a set of hypotheses and answer particular research questions. By referring to the first raised concern, this piece of datum indicates that L2 proficiency plays a central role in determining the accuracy levels of the study informants. To validate H1, the correlation value between proficiency and accuracy levels was computed. L2 proficiency is based on a 5-point scale: the beginner level (20%), elementary level (40%), intermediate level (60%), upper-intermediate level (80) and advanced level (100%).

The study examined the connection between participants' accuracy in interpreting reflexives and their L2 proficiency. The first hypothesis proposed a positive correlation between accuracy and L2 proficiency, which was supported by a moderate correlation coefficient of 0.33 ($p < 0.05$). This indicates that as L2 proficiency improves, accuracy in interpreting reflexives also improves.

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

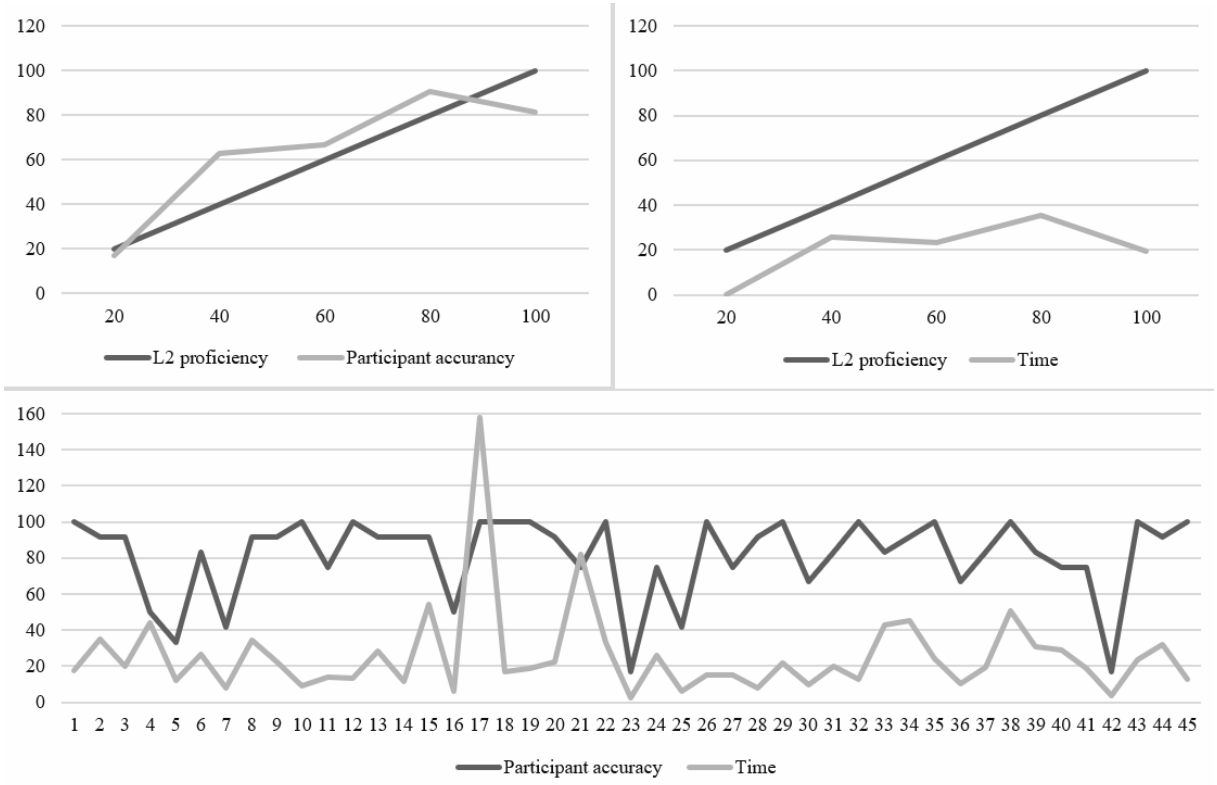

Figure 1. the effect of L2 proficiency on time and anaphora resolution

Turning to H2, it is predicted that L2 competence will negatively correlate with the time needed to process anaphors. As a confirmation, the findings revealed a statistically significant but practically weak negative correlation, with a correlation coefficient of -0.02 ($p < 0.05$). This suggests that while participants with higher L2 competence may exhibit slightly shorter response times, the effect is not substantial. To elucidate, as the L2 competence increased, the time needed to resolve anaphoric dependencies diminished. This suggests that language skills play a role in understanding the complexity of reflexive language.

It is worth mentioning that 11 answers out of 540 were discarded from the statistical analysis for their lack of accuracy was coupled with considerably short response time.

Regarding H3, it is predicted that an intrusion effect would be detected mainly in experimental conditions (b) and (c), unlike (a). To examine this effect, the accuracy per sentence and the time measurements per sentence were compared in order to filter out the demanding sentences and conditions. Table (6), which replicates the same values from Table 4, contrasts these 2 values to show the average time and the accuracy level per sentence.

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

Test sentences	1	2	3	4	5	6	7	8	9	10	11	12
Accuracy per sentence	95.55	86.66	93.33	91.11	82.22	86.66	86.66	91.11	91.11	71.11	53.33	60
Average time per sentence	17.53	22.82	19.70	17.06	29.08	25.28	13.29	42.33	34.60	23.11	30.01	36.46

Table 6. Time measurements vs. accuracy

As highlighted in table (6) indicated that the intrusion effect, triggered by the NP distractor and its gender manipulation, is associated with higher reading times and a compromised accuracy level. Given that mainly sentences (5), (6), (10), (11), (12) are remarkably demanding, H3 is confirmed, in that, conditions (b) and (c) are more problematic for the comprehenders than condition (a). This performance rate suggested that gender manipulation significantly impacted the accuracy and time required for antecedent identification. Therefore, based on all the obtained results, all three hypotheses were supported.

These findings suggest that the intrusion effect is mainly detected with low proficiency participants translated in the longer reading times and inaccurate responses. When the intrusion was noted among high-proficiency participants, higher reading times were detected yet the accuracy was still high. Such a result suggests L2 proficiency is a parameter that either enhances or jeopardizes accuracy levels.

The obtained time-course data prove that the binding information was not readily called upon when a competing antecedent was part of the processing. In fact, processing became more problematic for the informants when both candidate NPs matched the reflexive on the one hand, and when the matching antecedent did not grammatically satisfy the binding requirements. The time oscillation supports the claim that an intrusion effect was detected among L2 speakers of English (Bertenshaw et al. 2009), as opposed to L1 speakers (Dillon et al. 2013).

As for the last hypothesis, there was a remarkable pattern that the proper NPs, such as *Bill* or *Jenny*, were slightly clearer than the gender stereotypical NPs, *the pilot*. This was reflected in the longer processing time of sentences such as 2, 5, 7 and 12 where the antecedent was not a proper noun. However, this pattern did not significantly correlate with the accuracy of the responses. Therefore, the fourth hypothesis was not statistically supported.

Notwithstanding the reached validation for the study hypotheses, it is important to note that the reposted proficiency levels are based on the informants' self-evaluation. Although these values seem to be proportional with their performance rates and the time needed to process anaphora, it is central to indicate that they are an approximation of their authentic L2 competence.

5.1 Discussion

The primary focus of this research was to investigate if anaphora resolution in the L2 context was affected by any intrusion effect by investing an L1 stimulus that was adapted in this study. On the one hand, the aim was to see if manipulating the gender of an NP distractor would obstruct the L2 participants from strictly relying on the structural information of binding available in every stimulus. On the other hand, examining the effect of L2 proficiency was

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

central to this quest. Answers to these questions were sought by developing a set of experimental conditions and a time-based measurement. Studies such as Dillon et al. (2013), Sturt (2003) and Cunnings and Sturt (2014) were used in order to replicate some of their L1 linguistic data in the current research. As opposed to the L1 findings of Sturt (2003), the results generated from this online experiment confirmed the intrusion effects through participants' accuracy and time needed to identify the relevant antecedent. It was also confirmed that L2 proficiency significantly impacted the accuracy and time rates. It was observed that the intrusion effects were mainly exhibited in instances where the participants' accuracy shrunk and the time-course augmented. These findings are discussed against the L1 background concerning the role of binding information in second language acquisition. Implications retrieved from this study will also be discussed against the L2 research studies.

It should be noted that gender indexicality in proper nouns may not always be as clear to L2 as to L1 speakers. Whereas Mary and John, for instance, are generally associated with female and male referents, Tony can be used as a short form for Anthony and Antonia. Therefore, despite the relatively frequent use of some proper nouns, such as Tony, with masculine names, they still can be used with female referents. The same is applicable to gender-stereotyped NPs. Therefore, future research may elicit information about the participants' perception of the gender of the NPs used in the stimuli.

Studies on L1 anaphora resolution concurred on negating intrusion effects while processing anaphora (Dillon et al. 2013; Xiang et al. 2009). In that, binding is the strongest cue when compared to other cues, such as the agreement information. For native speakers, binding-as-initial-filter functions from the very initial stages of anaphora processing and is the ultimate filter in setting the preliminary candidate set. However, Cunnings and Sturt (2014) clearly indicated that confirming that gender cues were not at all used in the processing would not be a viable claim. Their findings support the view that binding is a solid filter but do not utterly nullify gender cues were not excluded from the processing in constructions where strict syntactic binding applies. Other cues seem to be part of the processing, yet not as apparent and prevalent as binding. With these claims in the background, it is plausible to contend that the observed intrusion effects in the current performance are not only characteristic of L2 anaphora resolution. Some L1 studies showed that reliance on different sorts of cues, besides binding, is a processing route for informants.

In the present L2 speakers did not seem to apply binding as an initial filter to choose the antecedent of the reflexive. They were more inclined to search for agreement information regardless of the antecedent being accessible from a structural binding perspective or not. This can be attributed to the influence of the native language of the participants. Arabic and the Tunisian variety of Arabic follow a free word order in which the antecedent of the reflexive can be relatively mobile and structurally covert. Moreover, gender agreement is fundamental to the Arabic language and is more salient than it is in English. Therefore, the Tunisian participants may have applied the rules of their L1 to their interpretation of the L2 sentences, given that they were linguistically naive.

Intrusion effects were not detected in different experimental conditions where the NP distractor sometimes matched the reflexive and was binding-incompatible. However, the same claim cannot be generated to constructions where the inappropriate antecedent is close to the reflexive. In fact, in this study, all the distractor NPs were not distanced from the reflexive, unlike the legit antecedent that was only sometimes temporarily separated from the reflexive. Yet, both situations seemed to be relatively problematic for the participants. When the

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

intervening NP agreed with the reflexive, the syntactic distance mattered for antecedent identification. Distance from the antecedent was demanding in terms of processing, and gender manipulation complicated the entire task of anaphora resolution.

Competing antecedent retrieval models in L1 have taken two directions, admitting (Dillon et al. 2013) and nullifying (Patil et al. 2016) the role of the intrusion effect. While these models might not be explicitly applicable in L2 anaphora resolution, they can be informative about the role of the syntactic information in both L1 and L2 settings. The difference in processing might further be clarified when retrieval models are equally tested and adapted in the L2 setting. For instance, Dillon et al.'s (2013) view and memory model completely denied any intrusion effect. They based their model on the primacy of the binding syntax, which seems to be unrealistic in terms of the active parser. Patil et al. (2016) and Cunnings and Sturt (2014) considered that excluding any other cues besides binding might not be accurate. Manifest gaps between L1 and L2 anaphora resolution are resultant from models that do not acknowledge the relevance of gender cues. Consequently, the difference of L1 and L2 acquisition in general and anaphora interpretation, in particular, will not be attributed to the relevant factors.

In the L1 context, Badecker and Straub's (2002) conclusions on the integration and effect of different sources of information seem to be more accurate both in L1 and L2. Instead of relying on principled grammatical information, they argued that information related to morphosyntactic cues, discourse-based cues and subjecthood might all be invested in reference assignment. According to their study, the effect these cues exert cannot be neglected. If L2ers in this study reached a considerable level of accuracy, and relied on information outside the rigid binding view, this performance cannot be always reduced to proficiency levels only. It can also be a processing pattern that is tested and confirmed to be part of the L1 context.

Turning to L2 studies, performance patterns have always been reported as differing from L1. In that, Clahson and Felser (2017) reported that L2ers rely on lexical or discourse-related information instead of strict reliance on binding information. In the current study, this claim is confirmed as far as the strict recourse to the syntactic information is concerned. It was observed that L2ers' syntactic knowledge was not the sole information source, particularly in some experimental conditions. It needed processing or an evaluation backup from other morpho-syntactic cues similar to gender. Across multiple L2 anaphora resolution studies, it was reported that a set of cues always constituted a processing repertoire. While this is coupled with oscillating reading times and performance, it is legitimate to consider it, as integral to L2 anaphora. The current study resonates with a multitude of findings revealing the distraction effect of gender manipulation on L2 participants (Felsher & Cunning 2012; Bertenshaw 2009; Felser et al. 2009).

In conditions (b) and (c), the gender-matching, yet grammatically unavailable for binding, antecedent exerted a potent role in evaluating the set of candidates. This pattern might indicate that the initial set comprised every agreeing NP with the anaphoric element. This was justified by the longer time rates that were needed to evaluate all the candidates and discard the binding-illicit NP. Another plausible view might be the impact of the inaccessible NPs in slowing down antecedent identification without being part of the candidate set. Both views are tenable insofar as they confirm that binding is not an initial filter in L2 anaphora resolution. Based on the visible effect of gender information in this study, the results might be approached as a representation of a dual-interpretation stage, generalizing Sturt's (2003) contention. The initial stage is the *bonding*: an anaphoric link is established between the dependent expression and all the possible NPs. The following stage is *resolution*: an evaluation is run to filter the proper

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

antecedent. This two-phase interpretation view allows for the possibility of developing a candidate set with all the available NPs, even if they are not appropriate binders. Explaining the role of the NP distractor in the processing becomes clear, notably when it is not only reduced to a matter of L2 proficiency.

English language proficiency intensively affects the course of anaphora resolution. Bertenshaw (2009) noted that proficient L2ers recorded shorter processing times when performance was compared to that of their native counterparts. In this study as well, it was observed that L2 proficiency impacted the accuracy level of anaphor interpretation. High-proficiency informants had ceiling performance and relatively limited processing times. Such a finding underscores the role of L2 competence in evaluating how instruction functions in the L2 context. In this study, it was noted that higher processing times did not undermine the accuracy rates on some occasions, which is attributed to high performers in this regard. It is worthwhile to note that relying on L2 proficiency as an evaluation metric makes it possible to test and refine Sturt's contention of the 2-phase *bonding-resolution* processing route for L1 and L2 speakers. Given that this processing pattern seems more representative, particularly in an L2 context, it is central to investigate its viability in disparate linguistic settings and populations.

6 Conclusion

This study aimed at investigating intrusion effects in anaphora resolution among Tunisian Arabic learners, primarily exposed to English within formal educational settings, the outcomes of this research reiterated the observed impact of NP distractors on identifying suitable antecedents. We established a clear correlation between L2 proficiency and informant performance, evident through moderately constrained accuracy levels and extended reading times in cases of contentious stimuli. Our investigation not only verified the presence of intrusion effects via gender cues but also emphasized the influential role of NP distractors. Drawing on the dual-interpretation perspective of bonding and resolution, heightening that each NP within a sentence forms part of an initial candidate set, we posit that anaphoric binding does not rely on primary grammatical cues among L2ers. Thus, Sturt's approach (2003) should hold greater validity than the Binding-as-initial-filter convention. Therefore, we encourage ESL educators to incorporate pragmatic gender-related insights into their syntactic curricula. Equipping learners with the ability to navigate gender cues within complex syntactic structures can benefit intermediate and advanced learners. Furthermore, ESL instructors should exhibit sensitivity towards learners' L1 backgrounds and the applicability of L2 rules within that context. L1 and L2 processing are intertwined, making a contrastive analysis of syntactic rules valuable for L2 learners. This approach can empower them to identify the origins of their errors and adopt more effective strategies for rectification.

References

Badecker, W., & Straub, K. (2002): The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory and Cognition* 28, 748–769. https://doi.org/10.1037/0278-7393.28.4.748

Bertenshaw, N. (2009): *The application of binding constraints by Japanese L2 learners of English*. Doctoral dissertation. University of Essex.

Clackson, K., Felser, C. & Clahsen, H. (2011): Children's processing of reflexives and pronouns in English: Evidence from eye-movements during listening. *Journal of Memory and Language* 65, 128–144. https://doi.org/10.1016/j.jml.2011.04.007

Cunnings, I. & Sturt, P. (2014): Coargumenthood and the processing of reflexives. *J.Mem.Lang.* 75, 117–139. https://doi.org/10.1016/j.jml.2014.05.006

Dillon, B., Mishler, A., Sloggett, S. & Phillips, C. (2013): Contrasting intrusion profiles for agreement and anaphora: Experimental and modelling evidence. *Journal of Memory and Language* 69, 85–103. https://doi.org/10.1016/j.jml.2013.04.003

Felser, C., & Cunnings, I. (2012): Processing reflexives in a second language: The timing of structural and discourse-level constraints. *Applied Psycholinguistics* 33(3), 571–603. https://doi.org/10.1017/S0142716411000488

Felser, C. (2016): Binding and coreference in non-native language processing. *Empirical Perspectives on Anaphora Resolution* 563, 241. https://doi.org/10.1515/9783110464108-012

Felser, C., Sato, M. & Bertenshaw, N. (2009): The on-line application of Binding Principle A in English as a second language. *Bilingualism: Language and Cognition* 12(4), 485–502. https://doi.org/10.1017/S1366728909990228

King, J., Andrews, C. & Wagers, M. (2012): *Do reflexives always find a good antecedent for themselves? Poster at the 25th annual meeting of the CUNY conference on human sentence processing*. New York, NY, March 14–16.

Lewis, R., Vasishth, S. & Van Dyke, J. (2006): Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science* 10, 447–454.

Marinis, T. (2018): Cross-modal priming in bilingual sentence processing. *Bilingualism: Language and Cognition* 21(3), 456–461. https://doi.org/10.1017/S1366728917000761

Nicol, J. L. (1988): *Coreference processing during sentence comprehension*. Doctoral dissertation. Massachusetts Institute of Technology.

Nicol, J. & Swinney, D. (1989): The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research* 18, 5–19. https://doi.org/10.1007/BF01069043

Patil, U., Vasishth, S. & Lewis, R. L. (2016): Retrieval interference in syntactic processing: The case of reflexive binding in English. *Frontiers in Psychology* 7, 329. https://doi.org/10.3389/fpsyg.2016.00329

Phillips, C., Wagers, M. & Lau, E. F. (2011): Grammatical illusions and selective fallibility in real-time language comprehension. In: Runner, J. (ed.): *Experiments at the interference, syntax and semantics*. Vol. 37. Bingley, UK: Emerald Publications. https://doi.org/10.1163/9781780523750_006

Stoet, G. (2010): PsyToolkit – A software package for programming psychological experiments using Linux. *Behavior Research Methods* 42(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017): PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology* 44(1), 24–31. https://doi.org/10.1177/0098628316677643

Sturt, P. (2003): The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language* 48, 542–562. https://doi.org/10.1016/S0749-596X(02)00536-3

Xiang, M., Dillon, B. & Phillips, C. (2009): Illusory licensing across dependency types: ERP evidence. *Brain and Language* 108, 40–55. https://doi.org/10.1016/j.bandl.2008.10.002

Ameni Mejri
University of Debrecen, Doctoral School of Linguistics
University of Debrecen, Institute of English and American Studies
Pf. 400
H-4002 Debrecen
amany.mejri@gmail.com

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

Appendix A: Experimental materials

All the test sentences are represented based on the study conditions:

A. Accessible antecedent match, inaccessible antecedent mismatch
B. Accessible antecedent match, inaccessible antecedent match
C. Inaccessible antecedent match, accessible antecedent mismatch

Every test entry is provided with a set of alternatives presenting the potential antecedent, inaccessible antecedent, and 2 possibilities of considering either both antecedent or neither of them. Below are the test sentences in the same order as they appear in the survey (separated by the fillers).

1. The pilot who scared Mary had injured himself quite badly during the journey
2. Jonathan was walking through the military barracks. He heard that the soldier had positioned himself in the middle of the mess hall.
3. Miranda was confused by the size of the depot. She knew that Jenny had lost herself near the back of the store
4. Victoria worked in the fire station in town. She knew that Bill had injured himself by the table in the room.
5. Richard remembered the tour of the national stadium. He recalled that the footballer had prepared herself in the brand new locker room.
6. The new executive who oversaw the middle managers doubted himself on most major decisions.
7. The pilot who scared John introduced himself to the passengers.
8. Joana worked out after a long day at the office. She learnt that her colleague Marry trained herself in the same gym as well.
9. Anthony watched the operation in the hospital. He remembered How Bernard taught himself the necessary tricks to survive all the stress.
10. Harry enjoyed walking in the local sports ground. He was glad that Tony taught herself how to paint the lines on the pitch.
11. Alice found the surgery very busy that morning. She was surprised that Anna criticized himself for being late for the appointment.
12. Peter knows some strange stories of the seaside hotel. The receptionist who has been working there told him that John killed herself early May.

The sentences used as fillers in this survey are listed below in the same order as they appear in the survey, among the target test sentences:

1. The butcher heard that every boy in the street loudly said that he could not lift the heavy box.
2. The secretary is the only employee who understood the stamp system of the company. Even Andrew was not able to keep up with it. He had to ask for her help all the time.

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

3. A fight had broken out at the celebration dinner. Bill in the hotel noticed quickly that his friend was in danger. He should move away from the table.
4. Anna was scared of the scandal. She wanted to protect her sister Mary, but always in vain. The two girls had to face it sooner or later.
5. Had Peter known what was going to happen to Maria, he would never have allowed her to travel alone.
6. Edith should decide if she will brush her teeth or paint her nails before going out with Nelly. She still feels lazy to go out though.
7. The bus driver was explaining the history of the city to the teenage girls group. He enjoyed the ride as they stood beside him taking picture of all the monuments he described.
8. The king trusted that every prince at the royal celebration clearly understood that he should not join in the dance.
9. Writing a report is harder than what Sally thought. She needed the help of a freelancer to be able to finish the document before the deadline.
10. After a long introspection, Steven found that his art was not evolving. He was misled by the comments of people around him and forgot his real inspiration.
11. Jack had the tendency to lend his books to anyone who would ask for them. Lately Fred warned him that his collection started to shrink.
12. Peter was visiting the dairy farm of his relatives. He watched Mr. Jones while he made some special cheese for them from a fresh goat’s milk.

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

Appendix B: Participants' accuracy and L2 proficiency levels

Participants	Accuracy %	Proficiency level	Average time in seconds
1	100	5	17.33
2	91.66	3	35.32
3	91.66	4	20.19
4	50	5	43.84
5	33.33	5	12.25
6	83.33	3	26.33
7	41.66	2	7.89
8	91.66	4	34.27
9	91.66	4	22.19
10	100	4	8.96
11	75	5	14.05
12	100	4	13.14
13	91.66	3	28.47
14	91.66	5	11.79
15	91.66	4	54.03
16	50	3	6.23
17	100	4	157.72
18	100	3	17.04
19	100	5	18.92
20	91.66	4	22.16
21	75	2	81.75
22	100	5	33.34
23	16.66	1	2.67
24	75	3	26.24
25	41.66	2	6.04
26	100	3	15.44
26	75	5	15.3
28	91.66	2	7.75
29	100	3	22.02
30	66.66	5	9.79
31	83.33	4	19.98
32	100	5	12.52

Ameni Mejri:
L2 online anaphora processing by Tunisian Arabic speakers
Argumentum 19 (2023), 196–225
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2023/10

33	83.33	3	42.56
34	91.66	4	45.46
35	100	5	24.01
36	66.66	4	10.2
37	83.33	3	19.13
38	100	3	50.78
39	83.33	4	30.51
40	75	3	28.76
41	75	3	18.74
42	16.66	3	3.79
43	100	4	23.56
44	91.66	4	3214
45	100	3	12.71