



Person Features and Lexical Restrictions in Italian Clefts

Cristiano Chesi* and Paolo Canal

Research Center for Neurocognition, Epistemology and Theoretical Syntax (NETS), School of Advanced Studies, Istituto Universitario di Studi Superiori (IUSS), Pavia, Italy

In this paper, we discuss the results of two experiments, one off-line (acceptability judgment) and the other on-line (eye-tracking), targeting Object Cleft (OC) constructions. In both experiments, we used the same materials presenting a manipulation on person features: second person plural pronouns and plural definite determiners alternate in introducing a full NP (“it was [DP₁ the/you [NP bankers]]_i that [DP₂ the/you [NP lawyers]] have avoided _j at the party”) in a language, Italian, with overt person (and number) subject-verb agreement. As results, we first observed that the advantage of the bare pronominal forms reported in previous experiments (Gordon et al., 2001; Warren and Gibson, 2005, a.o.) is lost when the full NP (the “lexical restriction” in Belletti and Rizzi, 2013) is present. Second, an advantage for the mismatch condition, *Art*₁-*Pro*₂, in which the focalized subject is introduced by the determiner and the OC subject by the pronoun, as opposed to the matching *Pro*₁-*Pro*₂ condition, is observed, both off-line (higher acceptability and accuracy in answering comprehension questions after eyetracking) and on-line (e.g., smaller number of regressions from the subject region); third, we found a relevant difference between acceptability and accuracy in comprehension questions: despite similar numerical patterns in both off-line measures, the difference across conditions in accuracy is mostly not significant, while it is significant in acceptability. Moreover, while the matching condition *Pro*₁-*Pro*₂ is perceived as nearly ungrammatical (far below the mean acceptability across-conditions), the accuracy in comprehension is still high (close to 80%). To account for these facts, we compare different formal competence and processing models that predict difficulties in OC constructions: similarity-based (Gordon et al., 2001, a.o.), memory load (Gibson, 1998), and intervention-based (Friedmann et al., 2009) accounts are compared to processing oriented ACT-R-based predictions (Lewis and Vasishth, 2005) and to *top-down Minimalist* derivations (Chesi, 2015). We conclude that most of these approaches fail in making predictions able to reconcile the competence and the performance perspective in a coherent way to the exception of the *top-down* model that is able to predict correctly both the on-line and the off-line main effects obtained.

Keywords: pronominal determiners, *top-down* derivation, complexity, cue-based retrieval, object cleft, intervention, similarity, memory load

OPEN ACCESS

Edited by:

Andrew Nevins,
University College London,
United Kingdom

Reviewed by:

Ellen F. Lau,
University of Maryland, College Park,
United States
Andrea Santi,
University College London,
United Kingdom

*Correspondence:

Cristiano Chesi
cristiano.chesi@iusspavia.it

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 09 December 2018

Accepted: 30 August 2019

Published: 20 September 2019

Citation:

Chesi C and Canal P (2019) Person Features and Lexical Restrictions in Italian Clefts. *Front. Psychol.* 10:2105. doi: 10.3389/fpsyg.2019.02105

INTRODUCTION

A necessary condition for comprehending correctly an Object Relative clause (OR) is to interpret the head of this construction [“the banker” in (1)] as the direct object of the predicate within the relative clause (“praised”):

(1) [The banker]_i [that the barber praised_j] climbed the mountain.

The fact that ORs are generally harder to process than Subject Relative (SR) clauses¹ has been shown systematically using self-paced reading experiments (since King and Just, 1991, a.o.), probe-task paradigms (Wanner and Maratsos, 1978, a.o.), eye movements tracking (Traxler et al., 2002, a.o.), or by monitoring the electrical (Weckerly and Kutas, 1999) or metabolic (Just et al., 1996, a.o.) activity of the brain. Focusing on ORs, Bever (1974) first noticed that their difficulty can be mitigated by varying the type of subject within the relative clause (examples from Gordon et al., 2001²):

- (2) a. [The banker]_i [that you praised _i] climbed
the mountain
b. [The banker]_i [that Ben praised _i] climbed
the mountain

When pronouns are processed in the subject position within the relative clause, as in (2).a, self-paced reading experiments show that the critical verbal regions (“praised” and “climbed”) are read faster than when proper names are present, (2).b; when definite descriptions occupy both the head and the subject position, we obtain the slowest performance on the same critical verbal regions, as in (1)³.

This effect has been extensively studied both from the theoretical/competence perspective (Friedmann et al., 2009; Belletti and Rizzi, 2013, a.o.) and from the psycholinguistic/performance one (Gordon et al., 2004, a.o.), especially in Object Clefts (OCs), when both the subject and the focalized DP can be definite descriptions, proper names or pronouns, (3) (Warren and Gibson, 2005):

- (3) It was [the banker/Pat/you]_i that [the lawyer/Dan/we] avoided _i at the party

In this paper, we present a manipulation of person features [3rd (default) vs. 2nd person] in the paradigm in (3) to investigate the role of person agreement in Italian (an overt subject-verb person agreement language) under the presence of a “lexical restriction” (i.e., the NP introduced by the determiner): second person pronouns will be used as determiners and compared to definite articles under the presence of a plural lexical restriction, as exemplified in (4):

- (4) Sono/siete [gli/voi architetti]_i che [gli/voi ingegneri]
are<sub>3p_PL/are_{2p_PL/} the/you architects that the/you engineers
hanno/avete consultato _i prima di iniziare il lavoro.
have_{3p_PL/have_{2p_PL/} consulted before beginning the work}</sub>

¹In SR, the head is related to the subject position within the relative as in “the banker [that _i praised the barber] ...”.

²In this specific experiment, since “Ben”/“you” items are shorter than “the barber,” the discussed effects could be biased by the relative NP length. However, the same contrast is reported in other experiments [e.g., (Warren and Gibson, 2002), discussed later] where longer names are used, so we assume here that the relevant contrast is genuine.

³Other factors are assumed to induce a facilitation in processing this configuration: for instance animacy of the subject while the object is inanimate (Kidd et al., 2007), but see Belletti and Chesi (2014) for discussion.

This study consists of two new experiments (section Materials and Methods): an acceptability judgment (experiment 1) and an eyetracking study (experiment 2). Comparing off-line (acceptability scores in experiment 1 and accuracy in answering comprehension questions after eyetracking in experiment 2) and on-line evidence (all relevant eyetracking measures in experiment 2) we will evaluate the actual fit of some prominent model discussed in literature (section Predicting Processing Difficulties) aiming at explaining the contrasts revealed in (3): from the analysis (section Discussion) of the results of our experiments (section Results) we conclude that none of the models presented readily predicts the behavioral evidence revealed by this study. We will then argue in favor of the left-right, *top-down* derivational minimalist perspective (Chesi, 2015) where the “complexity” of the non-local dependency is computed using a Feature Retrieval and Encoding Cost (FREC) function: this model better integrates the on-line and off-line results gathered here.

In the first part of this paper, we will introduce the structural properties of the Object Cleft sentences under analysis (section The Properties of the Object Clefts (OCs) Under Analysis). A brief state-of-the-art summary on OC processing effects [section Processing Object Clefts (OCs)] will precede the summary of five major models and their predictions on the contrast previously tested: similarity-based (Gordon et al., 2001, section Similarity-Based Accounts), intervention-based (Friedmann et al., 2009, section Intervention-Based Accounts), and Dependency Locality Theory (DLT, Gibson, 1998, 2000, section *memory-load* Accounts) models will be compared with processing-oriented, memory-usage models that make explicit predictions on reading times and are possibly more transparent in terms of brain mechanisms involved: Lewis and Vasishth (2005) model based on ACT-R (section ACT-R-Based Predictions) and a *top-down*, left-right minimalist derivation based on Chesi (2015) (section *top-down* (Left-Right) Minimalist Derivations). Then we will concentrate on person features on the DP triggering overt verbal agreement, especially focusing on pronouns used as determiners (section Pronouns as Determiners and Agreement): this should clarify the rationale behind the proposed manipulation and the semantic/pragmatic impact of this construction on processing.

THE PROPERTIES OF THE OBJECT CLEFTS (OCs) UNDER ANALYSIS

Object Clefts (OCs) are peculiar focalization constructions in which a direct object is displaced in a prominent left-peripheral position (Rizzi, 1997). Following Belletti’s (2008) analysis, in these structures the copula selects a truncated CP in which the object was moved into the FocP position as shown below:

- (5) ... BE [_{CP} Foree [_{FocP} Object [_{FinP} that [_{TP} Subject ... (Object)]]]]

According to Belletti, OCs, as opposed to Subject Clefts⁴ (SCs), can only convey contrastive/corrective focus (this is the role of FocP), then its realization will be felicitous and perfectly natural only in specific contexts. For instance in correcting/rectifying a statement as below:

- (6) X: Ho sentito che Alberto ha salutato qualcuno prima di partire per le vacanze; ha per caso salutato Beatrice prima di partire? (Dopo il litigio che hanno avuto per colpa di Claudia sarebbe stato un segno distensivo)
I heard that A. said goodbye to someone before leaving for holidays; has he said good bye to B. before leaving?
 (After the fight they had because of C., this would have been a positive sign)
- Y: (no, non-era Beatrice, purtroppo) era CLAUDIA che Alberto ha salutato prima di partire!
 (no, it wasn't B, unfortunately) it was C._{focalized} that A. said goodbye to before leaving!

Notice that a presupposition of existence [p.c. Benincà in (Belletti, 2008), (7).a] and uniqueness, as well as exhaustivity [as in Identificational Focus discussed in E. (Kiss, 1998), (8)], are also implied by the cleft constructions [contra standard focalization, both in root, (7).b, or in embedded contexts, (7).c]:

- (7) a. *(non) è NESSUNO che ho incontrato (non-tutti)
it is (not) NOBODY that I met (not everybody)
- b. NESSUNO ho incontrato (non-tutti)
NOBODY I met (not everybody)
- c. ho detto che NESSUNO assumeranno (non-tutti)
I have said that NOBODY they will hire (not everybody)
- (8) è UNA MELA che ho mangiato (non-una pera o qualcos'altro)
it is AN APPLE that I have eaten (not a pear or anything else)

Despite their peculiarities, these are perfect configurations for testing non-local crossing dependencies in comprehension: from a processing perspective, the distal argument (the focalized DP) must be retained in memory and retrieved, later on, when the verbal predicate is encountered, crucially after the subject has been interpreted as the agent of the predication. Notice that the absence of an appropriate context does not preclude the possibility of correctly processing and interpreting these constructions: in all the experiments that will be mentioned in the next section [section Processing Object Clefts (OCs)], any context introducing OCs was absent.

Processing Object Clefts (OCs)

The performance contrasts elicited by OCs suggest that the nature of both DPs present in the construction plays a major role (Gordon et al., 2001, 2004; Warren and Gibson, 2005): a definite

⁴Belletti (2008) assumes that, in Subject Clefts, the Subject raises to CP to satisfy an EPP feature (hence the CP becomes a small clause in the sense of Stowell, 1983); this would create an intervention context in case the object or the indirect object would move across the subject position. As a consequence, only the subject can realize the new information focus, using the *vP* periphery of the matrix copula while the focalized (indirect) object moves to the CP peripheral FocP, where no EPP is present (exactly as a *wh*-item). In this position, it can only express contrastive/corrective (and not simply new information) focus.

DP (D), a proper name (N), or a pronoun (P) occupying the two relevant positions produce different effects according to their relative distribution. The full prototypical paradigm (Warren and Gibson, 2005), introduced in (3) and expanded below in (9) for convenience, is used to illustrate these contrasts.

- (9) *object_{focalized} subject verb spill-over condition*
- a. It was **the banker** that **the lawyer** **avoided** _ at the party [D_1-D_2]
 a'. It was **the banker** that **Dan** **avoided** _ at the party [D_1-N_2]
 a". It was **the banker** that **we** **avoided** _ at the party [D_1-P_2]
- b. It was **Patricia** that **the lawyer** **avoided** _ at the party [N_1-D_2]
 b'. It was **Patricia** that **Dan** **avoided** _ at the party [N_1-N_2]
 b". It was **Patricia** that **we** **avoided** _ at the party [N_1-P_2]
- c. It was **you** that **the lawyer** **avoided** _ at the party [P_1-D_2]
 c'. It was **you** that **Dan** **avoided** _ at the party [P_1-N_2]
 c". It was **you** that **we** **avoided** _ at the party [P_1-P_2]

Warren and Gibson (2005) evidence [average reading times reported in (10)⁵], based on the paradigm in (9), shows that the greatest slowdown in self-paced reading at the critical verbal segment (*avoided*) is associated to the D_2 condition [i.e., when the subject of the cleft is a definite description, (9).a,b,c]. This correlates with the lowest accuracy rate in comprehension questions. Similar (non-significantly different) reading times are revealed for the N_1-N_2 matching condition (9).b', while the P_1-P_2 matching condition (always presenting a person mismatch), as well as the other conditions in which *P* is the subject of the cleft, produce the fastest reading times of the critical verbal region.

(10) Condition	Average RT (SE) ms
D_1-D_2	365 (19)
D_1-N_2	319 (12)
D_1-P_2	306 (14)
N_1-D_2	348 (18)
N_1-N_2	347 (21)
N_1-P_2	291 (14)
P_1-D_2	348 (18)
P_1-N_2	311 (15)
P_1-P_2	291 (13)

The authors reported a reliable effect of subject type with P_2 conditions averaging 30 ms faster than N_2 conditions, which is 28 ms on average faster than D_2 conditions. A marginally significant interaction (mainly driven by the slowest D_1-D_2 and N_1-N_2 conditions) indicates that the matching conditions, overall, are significantly slower than the mismatch conditions.

Also in comprehension, Warren and Gibson reported a main effect on the subject type (D_2 condition is harder than N_2 , which is harder than P_2) and an interaction between focalized object type and subject type, with all matching conditions inducing lower accurate results.

These results confirmed and expanded other results discussed in literature (e.g., Gordon et al., 2001).

⁵Thanks to Tessa Warren for sharing the original data. Differences in gray shades indicate statistically significant differences.

Predicting Processing Difficulties

Various models have been proposed to account for the performance data presented in the previous section. Here five models will be discussed, all considering as key factors: (i) the nature of the DPs involved in the dependency (*memory-load* account, Warren and Gibson, 2005, section *memory-load* Accounts), (ii) the similarity between the two DPs (*similarity-based*, Gordon et al., 2001 and section *Similarity-Based Accounts* *intervention-based* Friedmann et al., 2009, section *Intervention-Based Accounts*, accounts), (iii) the distance/activation of the focalized object with respect to the predicate (Lewis and Vasishth, 2005, section *ACT-R-Based Predictions*), or a combination of these factors (*top-down Minimalist* account, Chesi, 2015 section *top-down* (Left-Right) *Minimalist Derivations*). The predictions these models differ substantially both in terms of general complexity factor (DP types vs. matching/mismatching conditions), relevant features inducing difficulty and the exact moments in which such difficulty can be revealed (encoding at DP vs. retrieval at VP; on-line as slow down at specific regions vs. off-line as comprehension accuracy).

Similarity-Based Accounts

Gordon et al. (2001) explicitly focus on working memory demands in their studies using self-paced reading paradigms. Their proposal is based on the idea that having two DPs “of the same kind” stored in memory makes the OR/OC more complex than SR/SC. This is sufficient to model memory interference during encoding, storage, and retrieval (Crowder, 1976). When similarity between DPs is calculated considering noun type (proper vs. common), gender, number, animacy, case, and person, this theory is sufficient to predict asymmetries for most of the contrasts presented in section *Processing Object Clefts* (OCs): D_1 - D_2 and N_1 - N_2 matching condition are expected to be the hardest configurations, while the P_1 - P_2 matching configuration might result slightly easier than the other matching conditions because of person features mismatch. In all other mismatch cases, this approach predicts (both on-line and off-line) lighter effects because of the difference in type/features without being able to distinguish between D_1 - N_2 [easier, according to (10)] and N_1 - D_2 [harder in (10)] conditions or between P_1 - D_2 [harder in (10)] and D_1 - P_2 [easier in (10)]. This is also expected under the assumption that all features equally contribute to memory confusion.

Memory-Load Accounts

Memory-load accounts (Gibson, 1998; Warren and Gibson, 2002, 2005, a.o.) explain most of the contrasts presented in section *Processing Object Clefts* (OCs) by postulating an “integration cost” (Gibson, 1998, *Syntactic Prediction Locality Theory*, SPLT) proportional to new discourse referents⁶: since pronouns do not introduce new discourse referents and names are referentially lighter than definite descriptions (Warren and Gibson, 2005), *memory-load* accounts predict faster reading

⁶Assuming an accessibility hierarchy (Ariel, 1990), the model postulates a discourse referent cost following this referentiality scale: (less accessible) definite description > proper names > referential pronoun (*he/she/him/her*) > deictic pronoun (*I/we* and *you*) (more accessible). The structural integration cost is proportional to the referent cost.

times at the cleft verbal region when the subject is a pronoun and slightly longer reading times when it is a proper name. On the other hand, this account incorrectly predicts faster reading times for the N_1 - N_2 matching condition (“it was Patricia that Dan *avoided* at the party”) than for the D_1 - D_2 condition (“it was the lawyer that the businessman *avoided* at the party”), even though no significant differences emerged from this contrast.

Intervention-Based Accounts

The *intervention-based* accounts (Friedmann et al., 2009; Belletti and Rizzi, 2013, a.o.) can explain the symmetry revealed in the D_1 - D_2 and N_1 - N_2 matching conditions in terms of similarity of the critical intervening features: Friedmann et al. (2009), building on Rizzi (1990) locality constraint, assume that whenever features are shared between a filler, X [e.g., “the banker” in (1)] and a structural intervener, Z [e.g., “the barber” in (1)], the relation between X and the related selected gap, Y , gets disrupted in a way that is proportional to the kind (and number) of features involved. Assuming that lexical restriction, rather than referentiality (c.f. section *memory-load* Accounts), is computed and that features expressing such lexical restriction in definite descriptions, proper names and pronouns are distinct (they assume N for common nouns, N_{prop} for proper names, and a null N for pronouns), the *intervention-based* account predicts exactly that the matching conditions in which common nouns and proper names are present are comparable, while pronouns induce easier processing since N is absent. A crucial assumption here is that only features triggering movement should cause intervention (Friedmann et al., 2009:83). In this respect, the lexical restriction should not play a significant role, since this “feature” is buried within the DP and does not seem to trigger movement. However, (Belletti and Rizzi, 2013) explicitly consider the lexical restriction as a movement trigger⁷, hence rescuing the idea that its presence has an impact in terms of intervention. This model does not predict differences when pronouns are in the focalized object or in the subject position, neither it makes explicit predictions in the D_1 - N_2 and N_1 - D_2 cases: being N and N_{prop} distinct, either we assume that they play a role in triggering movement, hence expecting milder effects than in the matching conditions, or we assume that N and N_{prop} are not involved in movement hence these cases should be comparable with respect to the other matching cases. Since verb region in the D_1 - N_2 is significantly faster than in the N_1 - D_2 condition according to (10), neither assumptions lead to the correct prediction.

⁷The relevant opposition they propose is the one discussed in Munaro (1999):

- (i) a. Con **che** **tosat** à-tu parlà?
with which boy did you speak?
- b. Avé-o parlà de **chi**?
have you spoken of whom?

Notice that the difference between the *wh*-items involved in this construction, *per se*, could be responsible for the different landing sites; the relevant comparison should involve a minimal pair of the kind “that” vs. “that boy” for which distinct positions are targeted according to the presence of the lexical restriction.

ACT-R-Based Predictions

Lewis and Vasishth (2005) present an explicit moment-by-moment model of parsing⁸ based on independently motivated working memory principles. Their predictions, both on similarity effects (c.f. section Similarity-Based Accounts) and probability to retrieve the correct, accessible, syntactic chunk over time (c.f. section *memory-load* Accounts), follow from their assumptions based on an implementation of some components of the Adaptive Control of Thought-Rational (ACT-R) architecture (Anderson and Matessa, 1997; Anderson, 2005) within the sentence comprehension perspective. By focusing on working-memory retrieval, this model is able to estimate precisely the integration cost of a non-local constituent relying on its distance from its re-attachment point: the structural chunks are stored in memory and their activation (i.e., a purely numerical value) fades over time; stored chunks receive an activation boost whenever re-accessed. The longer is the time passed after the last re-activation, the longer it will take to retrieve the correct chunk. This plainly explains the difference between retrieving the subject in a SC, which is relatively fast, or the focalized object in OCs, which is relatively slow due to the time spent in attaching the intervening subject. This model can be used to simulate memory decay and difficulty in re-accessing specific constituents. Moreover, since their attempt is to explicitly describe processes and memory structures giving rise to specific linguistic configuration by providing a psychologically motivated theory of processing, this approach has a high explanatory potential. However, unless specific cues pre-activate the object (e.g., agreement as in clitic doubling constructions), this model can hardly predict the relevant asymmetries in the paradigm discussed in (9): we could assume retrieval and attachment of 1st and 2nd person pronouns to be slightly faster than default 3rd person DPs because of their higher saliency in the context; in this sense when P is in the focalized position (P_1) or in the subject position (P_2), this might somehow reduce the general cost paid for retrieving a distal argument, but any prediction in all other cases requires extra assumptions.

Top-Down (Left-Right) Minimalist Derivations

An alternative way to look at processing, without assuming any specific parsing algorithm or declarative grammatical rule format (as in Lewis and Vasishth, 2005), while maintaining an incremental left-right perspective, is presented in Chesi (2015, 2017): in the *Minimalist top-down* derivation proposed, *merge* is the sole structure building operation and it operates by attaching new (incoming) items always to the right of the phrase structure built so far (c.f. Phillips', 1996, *merge right*). The integration of new items is guided by the expectations triggered by the select feature(s) lexically encoded in the items already merged: for instance, a lexical entry like $[v \text{ run } =D]$ ⁹ indicates that the verb ("V") "run" selects a determiner ("=D"), namely that a DP is

⁸Context-Free Grammar rules (Chomsky, 1965) and a Left Corner Parsing algorithm (Aho and Ullman, 1972) is assumed as well as various HPSG details (Pollard and Sag, 1994). Refer to **Supplementary Materials** section ACT-R-Based Model for technical details.

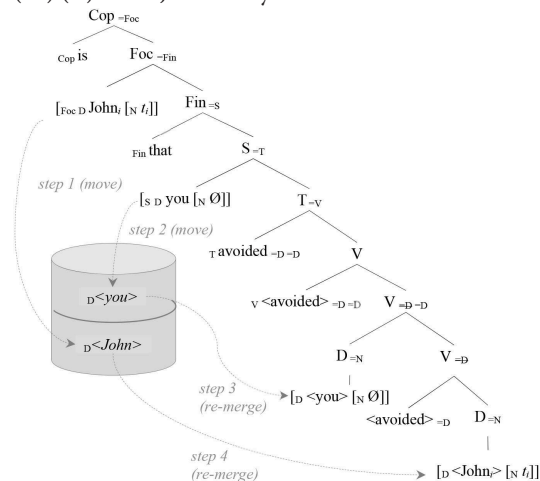
⁹We adapt here Stabler's (1996) Minimalist Grammar formalism: features associated to a lexical item are enclosed under squared brackets; categorial features are capital letters to the left of the lexical entry (e.g., "V" in $[v \text{ run } \dots]$), select

expected next; expectations are always projected after the lexical item is processed. If a category $[X]$ is expected (as result of the expansion of " $=X$ " select feature) either a new $[X \text{ item}]$ or $[X \ Y \ \text{item}]$ can be merged next; in the second case, $[Y \ <\text{item}>]$ must be stored in memory since the "Y" categorial feature was unexpected (i.e., unselected) and must be remerged in the structure as soon as a lexical item with a " $=Y$ " select feature is processed. This is how a non-local, filler-gap dependency (movement) is implemented. In (11), we exemplify the OC derivation using the grammatical knowledge (*Lex*) for the relevant paradigm in (9)¹⁰:

$$(11) \text{ Lex} = \left\{ \begin{array}{l} \text{Categories and default selections} = \{ \\ [\text{Cop} = \text{Foc}] [\text{Foc} = \text{Fin}], [\text{Fin} = \text{S}], [\text{S} = \text{T}], [\text{T} = \text{V}], \\ [\text{D} = \text{N}], [\text{N}], [\text{V}] \\ \} \\ \text{Lexical items} = \{ \\ [(\text{S}) (\text{Foc}) \text{D} \text{John}_i [\text{N} \ t_i]], [(\text{S}) (\text{Foc}) \text{D} \text{you} [\text{N} \ \emptyset]], \\ [(\text{S}) (\text{Foc}) \text{D} \text{the}], [\text{N} \ \text{banker}], \\ [\text{Fin} \ \text{that}], [\text{T} \ \text{v} \ \text{avoided} =\text{D} [\text{D}]], [\text{Cop} \ \text{is}] \\ \} \end{array} \right.$$

The relevant part of the derivation (equivalent both in generation and in parsing) can be schematized as follows:

(12) (It) is the John that you avoided ...



This tree diagram summarizes the history of the derivation, which implements the OC analysis presented in (5) and it is transparent with respect to processing (both parsing and generation).

As we see after step 1 and step 2, both arguments must be stored in memory because of their unexpected (i.e., unselected) "argumental" (D) feature right after they get first merged into the

features are ordered after a lexical entry and are prefixed by the equal sign (e.g., " $=D$ " in $[\dots \text{run } =D]$).

¹⁰S indicates the subject (topic-related/EPP) functional position (see Bianchi and Chesi, 2014 for a discussion of the subject-related positions in this framework). Round brackets indicate optional features, i.e., (S) and (Foc) can be associated to D or not. The structure of the various DPs implements Longobardi's (1994) raising analysis. Retrieval from *M*(emory) will always preempts lexical insertion; *M* is a last-in-first-out memory. Angled brackets indicate unpronounced copies of an item. See the **Supplementary Materials** section Top-Down Minimalist model for technical details and the full derivation.

structure. Both arguments are retrieved after the verb is merged and its selection requirements are expressed (steps 3 and 4). To predict processing difficulties at retrieval, we associate a cost to the memory buffer access: this cost grows exponentially with respect to the number of items stored (m), linearly with respect to the number of new features to be retrieved from memory (nF), and it is mitigated (linearly, again) by the number of distinct cued features (dF) by x (the region where retrieval is requested, in this case the verbal predicate). This is the core of the Feature Retrieval Cost (FRC)¹¹ function:

$$(13) \text{ FRC}(x) = \prod_{i=1}^n \frac{(1+nF_i)^{m_i}}{(1+dF_i)}$$

Using the lexicon in (11), we expect to retrieve definite descriptions like [_D *the* [_N *lawyer*]] (namely a D and a N category), proper names like *John* = [_D *John*_{*i*} [_N *t_i*]] (i.e., a contextually salient D index and a coindexed N) and pronouns like *you* = [_D *you* [_N \emptyset]] (i.e., just a contextually salient D index; for more details on these DP analyses, see section Pronouns as Determiners and Agreement). With these feature structures, we obtain the following FRCs at the verb segment for each condition in (9). These can be easily compared with the average reading times in (10):

(14) Condition	Average RT (SE) ms	log(FRC) ¹²
D ₁ -D ₂	365 (19)	1,43
D ₁ -N ₂	319 (12)	1,08
D ₁ -P ₂	306 (14)	0,78
N ₁ -D ₂	348 (18)	1,26
N ₁ -N ₂	347 (21)	1,26
N ₁ -P ₂	291 (14)	0,6
P ₁ -D ₂	348 (18)	1,26
P ₁ -N ₂	311 (15)	0,9
P ₁ -P ₂	291 (13)	0,6

Pearson correlation between average reading times and log(FRC) is extremely significant: $r(7) = 0.98$, $p < 0.001$.

Notice that the FRC plainly subsumes (and integrates) Friedmann et al. (2009) account. Moreover, it precisely

¹¹The intent of this formula (compared to more complex and complete ones, Van Dyke, 2007; Chesi, 2017) is to highlight the factors that are necessary and sufficient to account for the minimal variations in the paradigm under discussion.

¹²Logarithmic function (i.e., log(FRC)) is provided for making this cost directly usable in on-line predictions at a comparable scale with respect to other metrics [e.g., (Gibson, 1998) or (Lewis and Vasishth, 2005)]. This is how the FRC is calculated, condition by condition (see **Supplementary Materials** section Top-Down Minimalist model for more details):

D₁-D₂ = 9·3 = 27: 9 for retrieving D₂, since $nF = 2$ (D and N count as one), $m = 2$ because two DPs are in memory at this time, and $dF = 0$ because no feature is cued by the verb distinguishing one DP from the other; 3 for retrieving D₁, since $nF = 2$ (D and N are new), $m = 1$ and $dF = 0$;

D₁-N₂ = 4·3 = 12: 4 for retrieving N₂ ($nF = 1$, that is N, since D is contextually salient as we will see in section 0, $m = 2$, $dF = 0$), 2 for retrieving D₁ ($nF = 2$, $m = 1$, $nF = 0$);

D₁-P₂ = 2·3 = 6: 2 for retrieving P₂ ($nF = 1$ even if deictic pronouns are contextually salient, the correct person must be retrieved, $m = 2$, $dF = 1$ since a distinct case on pronouns is cued by the verb), 3 for retrieving D₁ ($nF = 2$, $m = 1$, $nF = 0$);

N₁-D₂ = 9·2 = 18: 9 for retrieving D₂ ($nF = 2$, since D and N are new, $m = 2$, $dF = 0$), 2 for retrieving N₁ ($nF = 1$, $m = 1$, $nF = 0$);

characterizes the triggers of the filler-gap dependency similarly to the cue-based memory retrieval approach (Van Dyke and McElree, 2006): we expect confusion (higher FRC) when the cued characteristic features are non-unique at retrieval. This also explains the cross-linguistic variation revealed, for instance, in Hebrew vs. Italian with respect to gender vs. number (Belletti et al., 2012): since in Hebrew the verb agrees with its subject also in gender, in Hebrew, but not in Italian, gender mismatch facilitates ORs processing. Under this perspective, this is because the verb uses such cues to retrieve the relevant argument from memory (hence, in FRC terms, gender mismatch increases dF in Hebrew since cued by the verb), while just number mismatch, but not gender mismatch, helps in Italian for the same reason (dF increases when cued number is in a mismatch configuration).

In addition to FRC, an encoding cost must be considered whenever an element is merged into the structure (similarly to Gibson's 1998 new discourse referent cost): this is the Feature Encoding Cost (FEC), a numerical value associated to each new item merged that is proportional to the number of new relevant features integrated in the structure:

$$(15) \text{ FEC}(x) = \sum_{i=1}^n eF_i$$

eF is the cost of each new relevant feature to be encoded at x . For simplicity $eF = 1$ for a new categorial feature introduced (e.g., 1 for D and 1 for N), 2 for a duplication of the same lexical category still requiring a structural integration as selected argument (i.e., 2 for the second N both in D₁-D₂ and N₁-N₂), 0 otherwise. In the paradigm (9) the FEC predictions are the following ones:

(16)	<i>object</i> _{focalized}	<i>subject</i>	<i>verb</i>	<i>spill-over</i>	<i>condition</i>
a.	It was the banker	that the lawyer	avoided _	at the party	[D ₁ -D ₂]
	1 2	1 3	2	3	
a'.	It was the banker	that Dan	avoided _	at the party	[D ₁ -N ₂]
	1 2	1 1	2	3	
a".	It was the banker	that we	avoided _	at the party	[D ₁ -P ₂]
	1 2	1 0	2	3	
b.	It was Patricia	that the lawyer	avoided _	at the party	[N ₁ -D ₂]
	1 1	1 2	2	3	
b'.	It was Patricia	that Dan	avoided _	at the party	[N ₁ -N ₂]
	1 1	1 2	2	3	
b".	It was Patricia	that we	avoided _	at the party	[N ₁ -P ₂]
	1 1	1 0	2	3	
c.	It was you	that the lawyer	avoided _	at the party	[P ₁ -D ₂]
	1 0	1 2	2	3	
c'.	It was you	that Dan	avoided _	at the party	[P ₁ -N ₂]
	1 0	1 1	2	3	
c".	It was you	that we	avoided _	at the party	[P ₁ -P ₂]
	1 0	1 0	2	3	

N₁-N₂ = 9·2 = 18: 9 for retrieving N₂ ($nF = 2$ even though D should be contextually salient, being two proper names presents, the same D, i.e., a co-referential index, cannot be sufficient to distinguish them, then an extra cost must be paid here as in the D-D condition, $m = 2$, and $dF = 0$), 2 for retrieving N₁ ($nF = 1$, just N is new since the determiner is now contextually salient and unique, $m = 1$ and $dF = 0$);

N₁-P₂ = 2·2 = 4: 2 for retrieving P₂ ($nF = 1$, $m = 2$, $dF = 1$); 2 for retrieving N₁ ($nF = 1$, $m = 1$, $dF = 0$);

P₁-D₂ = 9·2 = 18: 9 for retrieving D₂ ($nF = 2$, $m = 2$, $dF = 0$); 2 for retrieving P₁ ($nF = 1$, $m = 1$, $dF = 0$);

P₁-N₂ = 4·2 = 8: 4 for retrieving N₂ ($nF = 1$, $m = 2$, $dF = 1$); 2 for retrieving P₁ ($nF = 1$, $m = 1$, $dF = 0$);

P₁-P₂ = 2·2 = 4: 2 for retrieving P₂ ($nF = 1$, $m = 2$, $dF = 1$); 2 for retrieving P₁ ($nF = 1$, $m = 1$, $dF = 0$);

More precisely, in accordance with the structural assumptions expressed in (11), definite descriptions generally require the encoding of two critical new features (a determiner and a nominal restriction), proper names one (the contextually salient determiner is “free” and the proper name nominal restriction costs 1), while deictic pronouns have no encoding cost since contextually already present in the context (hence already pre-activated).

The absence of cost and the extra cost associated, respectively, to an already introduced feature and to the duplication of a category is coherent with a conception of memory as a pattern associator: if the pattern p_f , encoding feature f , has been just activated, re-activating it should have a minor cost (priming effect), while forcing a differentiation in a fully-overlapping pattern should induce the recruitment of extra memory units, hence an extra cost.

Summary

Summarizing, **Table 1** reports the predictions made by the models just described and the average reading times revealed at the verb segment in (9) [data from (10)]:

Theories based on the referentiality hierarchy (Ariel, 1990; Gibson, 1998; Warren and Gibson, 2005, a.o.; *memory-load* prediction in the table) fail to predict that also N_1 - N_2 matching condition induces a low performance comparable to the D_1 - D_2 matching condition. Similarity-based accounts (Gordon et al., 2004, a.o.) capture this fact, but fail in distinguishing any order permutation in mismatching conditions (e.g., D_1 - P_2 vs. P_1 - D_2); Intervention-based accounts (e.g., Belletti and Rizzi, 2013, a.o.) correctly predict harder times with both D_1 - D_2 and N_1 - N_2 matching condition, also expecting better performances with pro-intervening conditions (i.e., D_1 - P_2 , N_1 - P_2 , P_1 - P_2), but fail in predicting any distinction among other conditions (e.g., P_1 - D_2 vs. D_1 - P_2 or D_1 - N_2 vs. D_1 - N_2). Notice moreover that the processing costs at the verb segment do not follow from this perspective (as in any other bottom-to-top, movement-based approach). The ACT-R-based models, only relying on distance and pre-activation of the relevant argument, can predict easier retrieval only when P is present either at the subject or at the focalized object position. Also in these cases, P_1 - D_2 (and more marginally P_1 - N_2) would be predicted to be easier than it actually is. In the end, the *top-down* model (Chesi, 2015) correctly predicts more efforts in processing the D_1 - D_2 and N_1 - N_2 matching conditions (with D_1 - D_2 being the hardest configuration), medium difficulty when D_2 and N_2 are integrated in the subject position (different encoding costs) and lighter effects when P_2 is present because of case (nominative “we” vs. “us” morphology in English, while 1st/2nd vs. 3rd person

asymmetry can not be used as a cue because of past tense of the cleft predicate).

In conclusion, *memory-load* and *top-down* predictions present the highest level of correlation with respect to the revealed average reading times, but they make quite different predictions: in *memory-load* theories interference is irrelevant, while the *top-down* prediction crucially relies on the fact that retrieval is at issue, especially when features overlapping among items to be re-merged occurs. For the *top-down* model, also an encoding cost is considered, but the prediction is that the items already present (salient) in the discourse environment and, more generally, those features already merged in the structure, pay a minor cost at encoding (providing a precise characterization of the “more accessible” referents in *memory-load* accounts), with the exception of the re-introduction of a categorial feature (N in this case) and, possibly, its saliency specification (D), whenever they must be kept distinct in memory.

Pronouns as Determiners and Agreement

One way to dig further into the predictions emerging from these different assumptions is to keep all peculiar factors of OCs constant while investigating the specific contribution of single cued features using an overt subject-verb agreement language: focusing on person features, 1st and 2nd person, unlike 3rd person, are anchored to the speech event, being always present in a speech act (and in a left-peripheral structural dedicated position, Bianchi, 2003, 2006; Sigurdhsson, 2004). Because of their saliency (and dedicated structural position), we might expect 1st and 2nd person features to facilitate the integration of an argument better than default 3rd person (a non-person, in Sigurdhsson, 2004 terms). This could have an impact in terms of encoding: “highly accessible” deictic pronouns are lighter both for the *memory-load* (higher position in the accessibility hierarchy, section *memory-load* Accounts) and for the *top-down* models (being already present in phrase structure, they do not pay an extra FEC, section *top-down* (Left-Right) Minimalist Derivations).

On the other hand, at retrieval, different hypotheses can be formulated: considering person mismatch as a general facilitation (default hypothesis, H1), both the *similarity-based* (under any condition) and the *top-down* model (only when the relevant person mismatch is cued by the verbal agreement morphology) predict a facilitation¹³; an alternative hypothesis (H2), considering the salience of 1st and 2nd person features,

¹³Intervention-based approach would predict a specific facilitation for the 1st/2nd vs. 3rd person mismatching condition only under the further assumption that these features are relevant as movement triggers.

TABLE 1 | Summary of the predictions for the paradigm in 10 (data from Warren and Gibson, 2005).

Condition	D_1 - D_2	D_1 - N_2	D_1 - P_2	N_1 - D_2	N_1 - N_2	N_1 - P_2	P_1 - D_2	P_1 - N_2	P_1 - P_2
Read. time (SE) ms	365 (19)	319 (12)	306 (14)	348 (18)	347 (21)	291 (14)	348 (18)	311 (15)	291 (13)
Memory-load prediction	Hard	Medium	Easy	Hard	Medium	Easy	Hard	Medium	Easy
Similarity-based prediction	Hard	Medium	Easy	Medium	Hard	Easy	Easy	Easy	Medium
Intervention-based prediction	Hard	?	Easy	?	Hard	Easy	Easy	Easy	Easy
ACT-R-based prediction	Hard	Hard	Medium	Hard	Hard	Medium	Medium	Medium	Easy
Top-down prediction	Hardest	Medium	Medium-Easy	Hard	Hard	Easy	Hard	Medium	Easy

should predict a facilitation only for 1st and 2nd person and not for 3rd person. Under this second hypothesis, the only model making different predictions, based on the arguments involved, is the *top-down* model: only the subject and only when verb agreement is overt, 1st/2nd vs. 3rd person mismatch should produce a facilitation [this is because only 1st/2nd person feature mismatch would be considered as a *dF* facilitation for the FRC in (13)]. A neurophysiological evidence supporting the idea that (1st/2nd vs. 3rd) person features are peculiar in terms of subject-verb agreement (vs. number) is discussed in Mancini et al. (2011).

Notice that the introduction of a lexical restriction after the pronoun (e.g., “[D you [N bankers]]”) would remove any advantage of the bare pronominal condition according to the Intervention-based model (section Intervention-Based Accounts) and, for different reasons (increased number of features to be retrieved and compared), for the *top-down Minimalist* model (section *top-down* (Left-Right) Minimalist Derivations), but not for the other models. According to Belletti and Rizzi (2013) pronouns are to a lesser extent interveners because of their lack of a lexical restriction. In fact, pronouns, given an appropriate context, can function as determiners (Postal, 1966, a.o.) and, unlike determiners, bear person features other than default 3rd person.

On the usage of pronouns as determiners¹⁴, we refer to Elbourne (2005) analysis and we consider them as (empty) definite determiners taking an index and an NP predicate as arguments. According to Elbourne (2005), this is the structure shared by all DPs referring to individuals, namely proper names, pronouns and definite descriptions. In this sense, the “lexical restriction” would be an NP predicate which is, semantically speaking, denoted by type $\langle e, t \rangle$, while the denotation of the pronominal determiner (“you”) is expressed as follows¹⁵:

$$(17) \llbracket you_i \rrbracket^{g,a} = \lambda f : f \in D_{\langle e, t \rangle} \ \& \ a \leq_i g(j) \ \& \ f(g(j)) = 1.g(j)$$

This means that “you,” when used as a determiner in the construction “you bankers,” takes the NP [“banker(s)”] with denotation f and returns, as the denotation of the full DP, some contextually salient plural individual j , such that the addressees a (deictic use of “you”) must be part of j and j must be f (i.e., a banker).

- (19) a. Sono [gli architetti]_i che [gli ingegneri] hanno consultato _i prima di iniziare i lavori.
are_{3P_PL} the architects that the engineers have_{3P_PL} consulted before beginning the works
 b. Sono [gli architetti]_i che [voi ingegneri] avete consultato _i prima di iniziare i lavori.
are_{3P_PL} the architects that you engineers have_{2P_PL} consulted before beginning the works
 c. Siete [voi architetti]_i che [gli ingegneri] hanno consultato _i prima di iniziare i lavori.
are_{2P_PL} you architects that the engineers have_{3P_PL} consulted before beginning the works
 d. Siete [voi architetti]_i che [voi ingegneri] avete consultato _i prima di iniziare i lavori.
are_{2P_PL} you architects that you engineers have_{2P_PL} consulted before beginning the works

¹⁴We do not have space here to discuss the restriction that seems to force pronouns used as determiners to be either 1st or 2nd (and marginally 3rd) person plural in most languages. The interested reader should refer to Postal (1966), but also check the footnote in (Elbourne, 2005):60 reporting usage of first person singular in Early Modern English.

¹⁵ g is a variable assignment, a is the addressee, \leq_i the individual part-of relation Link (1983).

The contextual salience of the relevant individuals is necessary for the sentence to be acceptable and it must be postulated in out of the blue sentences; this means that if a relevant context is not provided to the reader, s/he must infer by her/himself that a salient group of individuals is presupposed by the sentence even if s/he does not share this information with the speaker at the utterance time. On the one hand, we might expect this missing contextual information, related to unexpected saliency, to produce some slowdown in processing, forcing the reader to update his knowledge of the common ground in order to accept this specific utterance; on the other, this is a perfectly grammatical construction and it should be correctly interpreted even when an appropriate context is missing. As far as we can tell, the presence/absence of an appropriate context licensing this usage of second person pronouns as determiners has never been tested before.

Considering the cleft sentences under analysis, this could even happen twice:

- (18) It was [you/we bankers]_i that [you/we lawyers] avoided _i at the party

Both the focalized DP and the cleft subject require that both a group of bankers and a group of lawyers be salient in the context and that the speaker and/or the addressee be part of one specific group. If the context is provided and the two groups of bankers and lawyers are in the common ground, the sentence should sound perfectly acceptable, if not, the reader should postulate the presence of the two groups after she/he realizes that none of them was accessible in her/his contextual knowledge. To our knowledge, this again has never been tested before.

Assuming a given cost for definite descriptions, a pronominally restricted DP (“you bankers”) would pay either the same cost (default assumption), a minor cost as pronouns (coherently with their implicit referentiality in the *memory-load* model) or an extra cost (whenever they are non-salient in the context or they get re-introduced twice, as predicted in the *top-down* model).

Testing the Different Predictions

The paradigm expanded in (19) will be used to test the specific contribution of 2nd vs. 3rd person (default) in an overt subject verb agreement language, Italian, under the assumptions previously discussed:

Hereafter, we will refer to (19).a as the *Art₁-Art₂* matching condition, (19).b as *Art₁-Pro₂* mismatch condition, (19).c as *Pro₁-Art₂* mismatch condition and (19).d as *Pro₁-Pro₂* matching condition. DP₁ is the focalized object, DP₂ the OC subject.

TABLE 2 | Theory by theory overall (off-line) predictions on the paradigm (19).

Condition	Art ₁ -Art ₂	Pro ₁ -Pro ₂	Art ₁ -Pro ₂	Pro ₁ -Art ₂
Similarity-based prediction	Hard	Hard	Medium	Medium
Intervention-based prediction	Hard	Hard	Medium	Medium
Top-down prediction (FRC)—H1	Hard	Hard	Medium	Medium
Top-down prediction (FRC)—H2	Hard	Hardest	Medium	Hard
Memory-load prediction—A1	Hard	Hard	Hard	Hard
Memory-load prediction—A2	Harder	Hard	Hard	Harder
Memory-load prediction—A3	Hard	Harder	Harder	Hard
ACT-R-based prediction	Hard	Hard	Hard	Hard

TABLE 3 | On-line predictions on the paradigm (19); at the verb segment (encoding+retrieval).

	Condition	DP ₁	DP ₂	Verb	PP
Art ₁ -Art ₂	Memory-load prediction	2	2	4 (1+3)	4
	Top-down prediction	2	3	3.43 [2+log(27)]	5
Art ₁ -Pro ₂	Memory-load prediction	2	3	5 (1+4)	4
	Top-down prediction	2	4	3.38 [2+log(24)]	5
Pro ₁ -Art ₂	Memory-load prediction	3	2	4 (1+3)	4
	Top-down prediction	3	3	3.43 [2+log(27)]	5
Pro ₁ -Pro ₂	Memory-load prediction	3	3	5 (1+4)	4
	Top-down prediction	3	5	3.68 [2+log(48)]	5

Overall, in off-line terms, similarity-based (section Similarity-Based Accounts), intervention-based (section Intervention-Based Accounts) and *top-down Minimalist* (based on FRC, section *top-down* (Left-Right) Minimalist Derivations) models would all predict that matching conditions (*Art₁-Art₂* and *Pro₁-Pro₂*) should be more difficult than the mismatching conditions (*Pro₁-Art₂* and *Art₁-Pro₂*). The *top-down* model, in particular, would predict a facilitation under the mismatch condition due to the distinct cued features at the verbal predicate [higher *dF* coefficient in the FRC, as expressed in (13)]. By default (hypothesis H1 in **Table 2**), as discussed in section Pronouns as Determiners and Agreement, any feature mismatch could help, hence the facilitation should be similar for both *Pro₁-Art₂* and *Art₁-Pro₂* conditions. However, according to the 1st and 2nd person anchoring hypothesis (H2 in section Pronouns as Determiners and Agreement), a global facilitation for the *Art₁-Pro₂* condition compared with the *Pro₁-Art₂* is expected by the *top-down* model: retrieval of *Pro₂* due to the cued 2nd person should be favored (hypothesis H2 in **Table 2**). Similarly, a retrieval penalty due to 2nd person feature matching is expected under the *Pro₁-Pro₂* condition. Notice that H2 does not change in any relevant sense the predictions of any other model except for the *top-down* one.

As for *memory-load* hypotheses (section *memory-load* Accounts), *Pro* condition has never been discussed previously, so the model here needs further assumptions: on the one hand, we might expect *Pro* and *Art* conditions to be referentially similar; under this assumption (A1 in **Table 2**) no difference would be predicted whatsoever in the paradigm. If *Pro* and

Art are assumed to be referentially different, either *Art* turns out to be less accessible than *Pro* (A2 assumption in **Table 2**), then we should expect an extra encoding cost for *Art* and a related complexity signature for retrieving the focalized object in the *Art₂* condition, or *Pro* is less accessible than *Art* (A3 assumption in **Table 2**), then a greater effort should be paid in the *Pro₂* condition.

From a different perspective, ACT-R based approaches (section ACT-R-Based Predictions) would predict no difference across conditions since the distance between the focalized DP₂ and the OC predicate is always the same and no cue can help in retrieving/reactivating the focalized object.

In the end the retrieval/intervention cost predictions for all the models can be summarized in the table below [consider “hard” to be the baseline, based on the evidence discussed in section Processing Object Clefts (OCs)]:

In terms of on-line predictions, costs at the VERB regions should be proportional to the predictions expressed in **Table 2**. Moreover, *memory-load* and *top-down Minimalist* models also predict specific encoding costs that should have an effect in terms of on-line measures at the related DP₁, DP₂, predicate and PP final regions: according to Gibson (2000), the verb introduces an event referent (+1), while the integration cost must be calculated as crossing the verbal event (+1) and the cost of the intervening nominal referents (+1 in Gibson, 2000). According to the *top-down* approach, the predicate introduces two relevant features, a temporal index T and V predicate, in conformity with the DP encoding hypothesis: *Art* should have an encoding cost of 2, an index *D* and a *N* predicate (section Pronouns as Determiners and Agreement). We expect *Pro* to pay an extra encoding cost due to the out of the blue inclusion of a speech act participant into the N predicate (+1) as suggested in section Pronouns as Determiners and Agreement; as result, *Pro* condition would cost globally 3 units. Taking *Art* = 2 and *Pro* = 3 as encoding cost baseline, both for *memory-load* and *top-down* models, only the *top-down* model predicts an extra cost for any duplicated category (+1 for *N* in DP₂) and for a referential mismatch (+1 for II person matching at DP₂ in *Pro₁-Pro₂* condition).

Under these assumptions, we can compare on-line predictions at the relevant segments¹⁶ (**Table 3**).

¹⁶The integration cost is calculated as crossing the verbal event (+1) and 2 or 3 nominal referents (+2/+3 respectively for *Art* and *Pro*). The cost of +2 for

Crucially, the predicted cost at DP₂ is different since the *top-down* model, and not the *memory-load* one, predicts an extra encoding cost which is proportional to the number of matching features (and consequently to the necessity of updating an expectation, also in terms of speech act participants, otherwise salient, hence “free”). Moreover, at the verb segment, the predictions of the two models differ: the *memory-load* model predicts major efforts in the *Pro*₂ condition, while the *top-down* model expects milder differences and essentially a penalty to be paid for the *Pro*₁-*Pro*₂ condition. Under H2 (i.e., only 1st/2nd person features are cued by the verb), also a mild facilitation for the *Art*₁-*Pro*₂ would be expected by the *top-down* model as compared to the *Art*₁-*Art*₂ and *Pro*₁-*Art*₂ conditions. Under H1 (i.e., all person features count as distinct cues), also *Pro*₁-*Art*₂ would benefit by the cued feature mismatch.

MATERIALS AND METHODS

Ethics Statement

The experiment was approved by the Ethics Committee of the “Dipartimento di Scienze del Sistema Nervoso e del Comportamento” of the University of Pavia. A written informed consent was obtained from the participants of this study.

Participants

Fifty-three participants (age range 20–52, $M = 34.57$, $SD = 8.09$, 28 female; all speakers of center-north Italian variety) voluntarily signed up for the acceptability study (experiment 1).

A different sample of 33 Italian native speakers of the same Italian variety (age range = 19–35; 18 female) took part in the eye-tracking study (experiment 2). After the end of each experimental session, in experiment 2, we assessed the participants’ Verbal Working Memory Capacity (WM) using a test (in Lewandowsky et al., 2010) that is a variant of the Sentence Span test originally designed by Daneman and Carpenter (1980). Participants were asked to carry out a dual task: they were presented with series of 3 to 8 statements, each followed by a consonant. Participants had 4 sec to judge whether the statement was True or False then they were asked to remember the series of consonants that was presented after each sentence. At the end of the series a question mark appeared, signaling participants to type in all the consonants presented. A score ranging from 0 to 1 is obtained, indicating the individuals verbal working memory capacity. The scoring procedure takes into account the length of the series and the accuracy to the True/False judgment (see Lewandowsky et al., 2010, for more details).

Stimuli

We created 32 paradigms expressing the four possible conditions presented in (19), for a total of 128 items. DPs were introduced by articles and second person pronouns while keeping number

Art is considered for uniformity with respect the *top-down* model. These are the assumptions leading to the best possible predictions for the *memory-load* model. As for the *top-down* model, the FRC predicted is: $Art_1-Art_2 = 9.3 = 27$; $Art_1-Pro_2 = 8.3 = 24$; $Pro_1-Art_2 = 9.3 = 27$; $Pro_1-Pro_2 = 16.3 = 48$; See **Supplementary Materials**, section ACT-R-Based Model and section Top-Down Minimalist model for details.

(plural, in order to make all the oppositions sound) and gender (masculine) constant. The nouns within each sentence were balanced for (i) number of letters ($DP_1 = 8.86$, $SD = 1.46$; $DP_2 = 8.96$, $SD = 1.95$; $t < 1$), (ii) logarithmic frequency (based on Repubblica corpus, Baroni et al., 2004) of nominal items ($DP_1 = 7.66$, $SD = 1.68$; $DP_2 = 7.58$, $SD = 1.95$; $t < 1$), and (iii) concreteness and imageability (all were concrete nouns referring to professions). The plural masculine article in Italian is sensitive to the beginning of the following noun: when a vowel is present “gli” is used (“gli architetti,” the architects vs. “i giorni,” the days). We used “gli” to maximize the length of the determiner and match the length of the pronoun (“voi,” you) whenever possible, but in most cases (22/32), due to semantic congruity and to keep DP frequency comparable within sentences we used “i” determiner for nouns beginning with a consonant. Experimental lists included 32 critical items (eight per condition) and 112 fillers. As in previous experiments, we did not include any relevant context. Fillers included 64 declarative sentences, half of them perfectly grammatical and the other half presenting a subject-verb number agreement error; other 48 fillers were questions with various degrees of acceptability, ranging from perfect grammatical *wh*- long distance question (“what do you think that John saw?”) to violation of locality (“what do you wonder who see?”). Both experiments used these materials.

A Latin square design was used to counterbalance conditions across experimental lists, in a way that in each experiment participants were exposed to one only version of each paradigm, and each item within paradigms was presented to an almost equal number of subjects across lists.

Procedure

For experiment 1 (acceptability judgment), a web-based questionnaire was created using Osucre open source software (Van Acker, 2007). The experimental lists were presented one at a time on a single line and participants were asked to judge each for acceptability on a 7-points Likert-scale.

For experiment 2 (eye-tracking), participants were individually tested in a dimly lit room. They sat in front of a 17 inches computer screen and kept their head on a chin rest so that the distance between the display and their eyes was 56 cm. They were instructed to read the sentences carefully, as they would have to answer one question following each sentence. Each trial (presented in pseudo-randomized order) consisted of a fixation cross appearing at the center of the screen for 1,500 ms, and was followed by the sentence, displayed on one single line. Participants could press the space bar to signal the reading completion, or the sentence display timed out after 20 sec. After the sentence display a second fixation cross appeared for 1,500 ms, just before the presentation of the question, that remained on screen until the Yes/No response. Half of the questions concerned the subject (e.g., “did the architects consult someone?”) the other half the object of the cleft (e.g., “did someone consult the engineers?”); half of the questions included a relevant PP (e.g., “did the architects consult someone after the meeting?”); 50% of the questions required a positive answer, 50% a negative one.

Analysis

Data Acquisition and Pre-processing of Eye-Tracking Data

Eye-movements in experiment 2 were recorded with an Eyelink[®] 1000 system (SR-Research, Ottawa, CA), tracking the dominant eye and using the desktop mount configuration. Eye gaze was sampled at 1,000 Hz frequency. Consecutive fixations between 50 and 80 ms occurring at one character distance were grouped into one single fixation (1.35% of data). Fixations that were (a) shorter than 80 ms, (b) longer than 1,200 ms, (c) occurring within 20 ms from blink onset/offset, and (d) occurring outside sentence boundaries were excluded from the analyses (overall rejection rate 3.16%). Four participant were excluded on the basis of their performance on the comprehension questions (<60%), leaving 29 participants in the analyzed dataset.

Four canonical reading time measures (Rayner, 1998) were computed. *First Fixation* (FF) and *Gaze Duration* (GD) were defined as the time (ms) spent on each region when participants entered it (from the left side) for the first time: FF was the duration of the very first fixation only, whereas GD was defined as the time spent from the first time entering the region to the first time leaving it, to the right or to the left. Words in the sentences could be fixated after a regression (i.e., entering the region from the right) and the time spent in a region entering it from the right was defined as *Second Pass* (SP) reading time. Total duration Time (TT) was the total time spent on a given region.

We also analyzed regression data: a regression event occurred when participants looked back in the sentence. For each regression event, we determined the region from where the eye left (*R-from*), and the region where the gaze landed (*R-in*).

Sentences consisted of six different regions: BE, DP₁, C, DP₂, VERB, SPILLOVER. Regression analyses were carried out for each region, when the number of observations was sufficient for carrying out statistics. In particular, only Total Duration and Second Pass could be computed for BE and C regions.

Statistics

Results were analyzed using linear mixed-effects regression models (Baayen et al., 2008), using lme4 package (version 1.1.21) in R environment (R Core Team, 2018). Mixed models are widely used in eye-movements research (e.g., Staub, 2010; Kuperman and Van Dyke, 2011) as they conveniently handle imbalanced designs and missing values, typically occurring with eye-tracking data.

Linear mixed models were used for acceptability judgments of Experiment 1 and Reading Times in Experiment 2. The dependent variables for the linear mixed models were the single trials acceptability scores (ranging from 1 to 7) for Experiment 1, and the log transformed Reading Times for Experiment 2. Log transformation was needed because the distribution of times and the models' residuals were far from the normal distribution. For the analysis of regressions in Experiment 2 the dependent variable was binary (0,1) depending on whether a regression event was or was not recorded, and data were analyzed by fitting generalized mixed models using the logit response function (e.g., Jaeger, 2008)—as it was done

for the Accuracy data on the comprehension question of Experiment 2.

In Experiment 1 we used the “maximal” random structure allowing for by-subject and by-item intercept adjustments and by-subject slopes adjustments for DP₁ by DP₂ interaction. However, the maximal random structure of the models for Experiment 2 (where the number of participants was lower) often caused convergence issues. We therefore adopted a random structure chosen on grounds of feasibility (Matuschek et al., 2017). For reading measures (FF, GD, TT, and SP) the random structure was allowed by-subject and by-item intercept adjustments, and by-subject slope adjustments of DP₁ and DP₂. For regression measures the random structure was initially “minimal,” but, when possible, the model used to report the final estimates and the contrasts between conditions had the same structure used for the reading measures.

To evaluate the presence of significant main effects and interactions we used likelihood ratio tests (LRT) comparing the fit of (nested) models of increasing complexity (e.g., [factor A + B] compared to [factor A + B + AxB]). Tables in the Results section describe the outcome of the LRTs, reporting the value of chi square and the level of significance. In the analysis of Experiment 2 the models were further specified: the Null model consisted of the effect of Trial Order alone, and Trial Order was kept in all subsequent models. The contribution to the model fit was assessed for the factors DP₁ (Art, Pro), DP₂ (Art, Pro), and WM (continuous) and their interactions. To make the main effects interpretable as in standard ANOVAs we adopted contrast coding for categorical factors, while continuous predictors were z-centered around the mean (e.g., Levy et al., 2013). The effects of the factors of interest and their interactions are further described in the text and in the figures providing the size of the effect in the response measure (i.e., back transforming log(ms), in ms, and log odds into probability) using functions in the emmeans package (Lenth et al., 2019).

RESULTS

Off-Line Results (Acceptability Judgment and Accuracy in Comprehension Questions)

We compare here (Figure 1) the results of the offline data gathered from both experiments: the acceptability rate of experiment 1 and the accuracy in answering comprehension questions after eyetracking (experiment 2).

Summarizing the acceptability data collected in Experiment 1, we obtained the following pattern¹⁷:

$Art_1-Pro_2 (M = 5.31, SE = 0.19) \geq Art_1-Art_2 (M = 5.22, SE = 0.19) > Pro_1-Art_2 (M = 4.54, SE = 0.22) > Pro_1-Pro_2 (M = 2.50, SE = 0.18)$.

The analysis of Experiment 1 LRTs showed that judgments were influenced by the interaction between DP₁ and DP₂

¹⁷“≥” indicates a numerical, though non-significant difference; “>” indicates a significant difference.

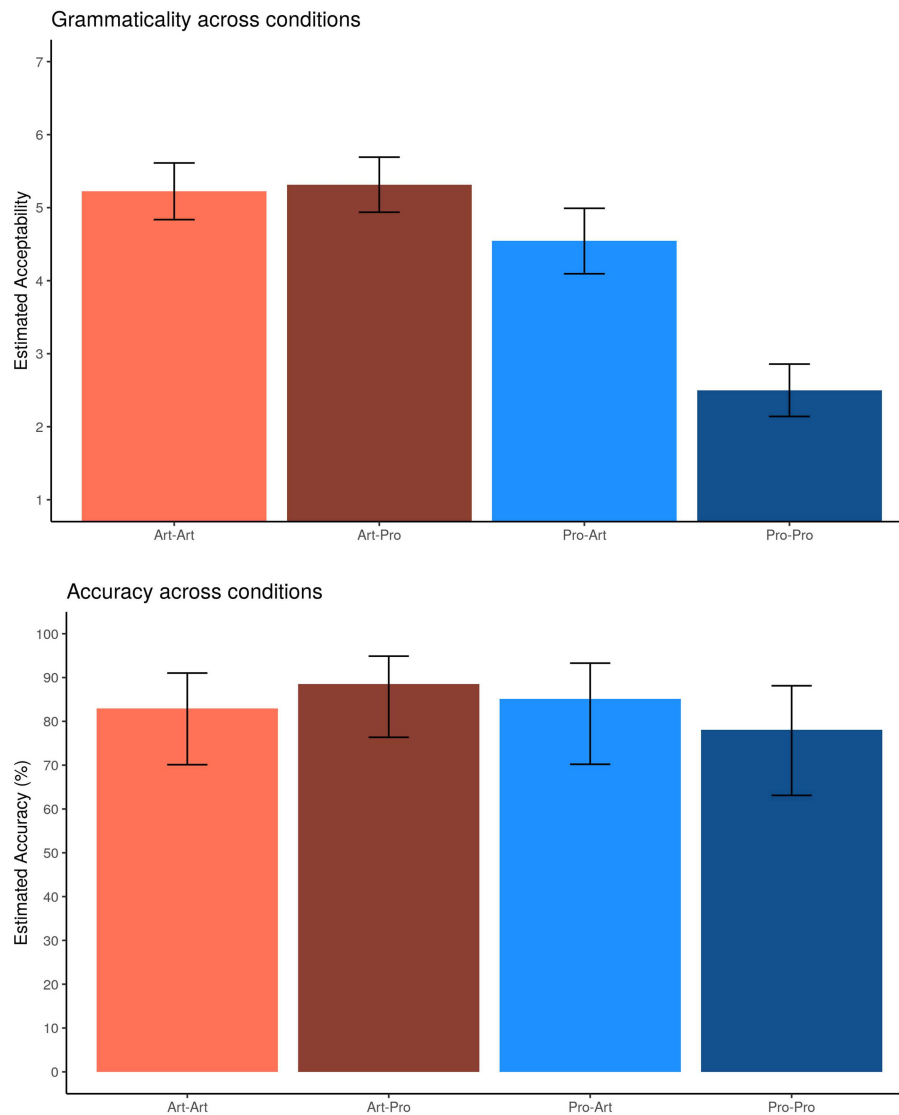


FIGURE 1 | Offline results: estimated acceptability (7-point Likert scale) judgments (experiment 1) and accuracy (%) in answering comprehension questions (experiment 2). Error bars represent 95% confidence intervals around the models' estimates.

TABLE 4 | Acceptability and Accuracy effects depending on DP₁-DP₂ types.

	χ^2	<i>p</i>
Acceptability		
DP ₁ type	22.09	<0.001
+ DP ₂ type	5.49	<0.01
+ DP ₁ :DP ₂	51.85	<0.001
Accuracy		
DP ₁ type	0.69	
+ DP ₂ type	0.00	
+ DP ₁ :DP ₂	3.76	<0.1

Bold values are the significant ones.

(**Table 4**) showing that when DP₁ was *Art* the subject type had no effect on acceptability ($Art_1-Art_2 = 5.22$; $Art_1-Pro_2 = 5.31$; $t = -1.04$), whereas when DP₁ was introduced by a pronoun, the effect of subject type was sensible being the matching condition much less acceptable ($Pro_1-Art_2 = 4.54$; $Pro_1-Pro_2 = 2.50$; $t = -9.28$, $p < 0.001$).

A similar numerical pattern indicating that Art_1-Pro_2 is better than Pro_1-Pro_2 is revealed in comprehension (while $Pro_1-Art_2 \geq Art_1-Art_2$ is non-significant):

Art_1-Pro_2 ($M = 88.6\%$) \geq Pro_1-Art_2 ($M = 85.1\%$) \geq Art_1-Art_2 (83%) \geq Pro_1-Pro_2 (78.1%).

The marginally significant interaction pointed to the effect of DP₁ when DP₂ is *Pro*, which describes the marginally significant

difference between Art_1-Pro_2 and Pro_1-Pro_2 (+10.5%, $z = 1.70$, $p < 0.1$).

On-Line Results (Eye Tracking Data, Experiment 2)

Reading Times

First fixation (FF)

Due to the small number of observations for FFs in BE ($n = 115$) and in C ($n = 307$), models did not converge and the results from these regions were omitted (Figure 2). Analyses (Table 5) revealed a main effect of DP₂ type, indicating longer FFs for Pro_2 , while reading DP₂ (+10 ms, $t = 2.06$, $p < 0.05$) and VERB (+10 ms, $t = 2.39$, $p < 0.05$) regions. Upon reading SPILL region, FFs were slightly longer with Pro_1 (+9 ms, $t = 1.98$, $p < 0.1$).

Considering the effect of WM, it was significant across the last three regions: FFs were generally shorter for Higher WM participants (β in DP₂: -0.038 , $t = -1.95$, $p < 0.1$; β in VERB: -0.070 , $t = -3.44$, $p < 0.01$; β in SPILL: -0.056 , $t = -2.51$, $p < 0.05$).

Gaze duration (GD)

Due to the small number of observations for GD in BE and in C, models did not converge and the results from these regions were omitted (Figure 3). In the analysis of GD (Table 6), a main effect of DP₁ type in DP₁, and a main effect of DP₂ type in DP₂, are significant. In DP₁ region, Pro_1 triggered longer GD compared to Art_1 [+64 ms, $t = 6.20$, $p < 0.001$]. In DP₂ region longer GD occurred with DP₂ [DP₂ region: +108 ms, $t = 7.28$, $p < 0.001$]. A marginal indication of a significant main effect of DP₂ emerged also in the VERB region, showing a slow down to Pro_2 [+26 ms, $t = 1.86$, $p < 0.1$].

The effect of WM, indicating shorter GD for participants with higher WM, was robust across regions: in DP₁ (-0.115 , $t = -3.45$, $p < 0.001$), DP₂ ($\beta = -0.097$, $t = -2.77$, $p < 0.01$), VERB ($\beta = -0.179$, $t = -4.37$, $p < 0.001$), and SPILL ($\beta = -0.181$, $t = -3.66$, $p < 0.001$).

Total time duration (TT)

As for TT (Figure 4), effects of the experimental factors were found on all Regions except for SPILL (Table 7).

In BE the significant interaction between DP types was due to the slow down (+43 ms, $t = 1.84$, $p < 0.1$) to Pro_1 compared to Art_1 that resulted marginally significant on the pairwise contrasts ($t = 1.84$, $p < 0.1$) in the Pro_2 condition and not in Art_2 (+2 ms, $t < 1$). In DP₁, reading Pro_1 took longer than Art_1 (+168 ms, $t = +6.71$, $p < 0.001$), and TT in this region is further modulated by the interaction between DP₂ type and WM: the facilitatory effect of WM was stronger ($\Delta\beta = -0.085$, $t = -1.90$, $p < 0.1$) when DP₂ was Art ($\beta = -0.07$) rather than Pro ($\beta = 0.01$). In DP₂ region, the effect of DP₂ (longer Pro_2 compared to Art_2 : +183 ms, $t = +5.41$, $p < 0.001$) was further modulated by DP₁ type: when DP₂ was Art , no effect of DP₁ emerged (-20 ms, $t < 1$), while when DP₂ was Pro , longer total times were observed for Pro_1-Pro_2 compared to Art_1-Pro_2 (+137 ms, $t = +3.31$, $p < 0.01$). In the VERB region, as in the previous region, the interaction between DP types revealed that the effect of DP₁ was present in the Pro_2 condition—longer TT for Pro_1-Pro_2 compared to

Art_1-Pro_2 (+83 ms, $t = +2.35$, $p < 0.05$)—and not in the Art_2 condition (-17 ms, $t < 1$).

Total times spent in BE and C were globally influenced by WM [in BE: $\beta = -0.095$, $t = -1.99$, $p < 0.1$; in C: $\beta = -0.064$, $t = -1.73$, $p < 0.1$].

Second pass duration (SP)

The presence of a pronominal restriction in DP₁ and DP₂ was sufficient to cause longer SP reading times in both DP₁ and DP₂ regions (Figure 5, Table 8): SP in DP₁ were longer for Pro_1 (+73 ms, $t = +2.89$, $p < 0.01$), and, similarly, in DP₂ region SP were longer for Pro_2 (+111 ms, $t = +3.21$, $p < 0.01$). In DP₂, however, participants show longer SP in the Pro_1 condition compared to Art_1 condition (+45 ms, $t = +1.75$, $p < 0.1$).

In the VERB region, a three-ways interaction emerged, showing that Pro generally causes a slowdown in SP (Pro_2 condition: +52 ms, $t = 1.78$, $p < 0.1$; Pro_1 condition: +57 ms, $t = 2.01$, $p < 0.05$), and that the effect of WM (faster reading times for higher WM) was much stronger in Pro_1-Art_2 ($\beta = -0.214$), compared to the other three conditions [Pro_1-Pro_2 $\beta = 0.025$, $\Delta\beta = -0.239$, $t = -2.75$, $p < 0.05$; Art_1-Art_2 $\beta = 0.036$, $\Delta\beta = -0.250$, $t = -3.04$, $p < 0.05$; Art_1-Pro_2 $\beta = 0.018$, $\Delta\beta = -0.232$, $t = -2.45$, $p < 0.1$].

In the end, SP in BE showed a significant DP₁ by DP₂ interaction. The time spent re-reading BE was not affected by DP₁ type when DP₂ was Art (18 ms, $t < 1$), while for Pro_2 the difference between Pro_1 and Art_1 was consistent (+49 ms, $t = 1.99$, $p = 0.05$).

Regressions

We first assessed the likelihood of performing a regression from each region, independently of the experimental factors, but distinguishing between regressions in first pass and overall regressions. The probability of performing a regression was largest from the rightmost region SPILL (80%), followed by DP₂ (46.8%, SPILL vs. DP₂: $z = +13.48^{***}$), VERB (29.9%, DP₂ vs. VERB: $z = +7.00^{***}$), DP₁ (23.3%, VERB vs. DP₁: $z = +3.15^*$), and C (19.8%, VERB vs. C: $z = +4.26^{***}$).

Notably, the probability of performing a regression during first pass was considerably reduced for the VERB region (9.42%).

VERB first pass regressions were less likely than those occurring in DP₁ (13.75%, $z = -2.79^*$) or DP₂ (19.30%, $z = -6.02^{***}$), suggesting that processing difficulties at the VERB may be different from those at DP regions. In particular, the integration efforts at DP₁ or DP₂ trigger immediate regressions, while encoding/retrieval costs do not immediately lead to a regressive saccade, but rather require more reading time.

All in all, the overall probability in regressions pattern suggests that the DP₂ region is particularly hard to process: upon reading this region, being it the first time or not, participants need to go back in the sentence to retrieve additional information.

Regressions from regions (R-from)

We first evaluated the probability of making a R-from each ROI on the total number of fixations on each region (Figure 6, Table 9). A main effect of DP₁ type in DP₁ region, showed that

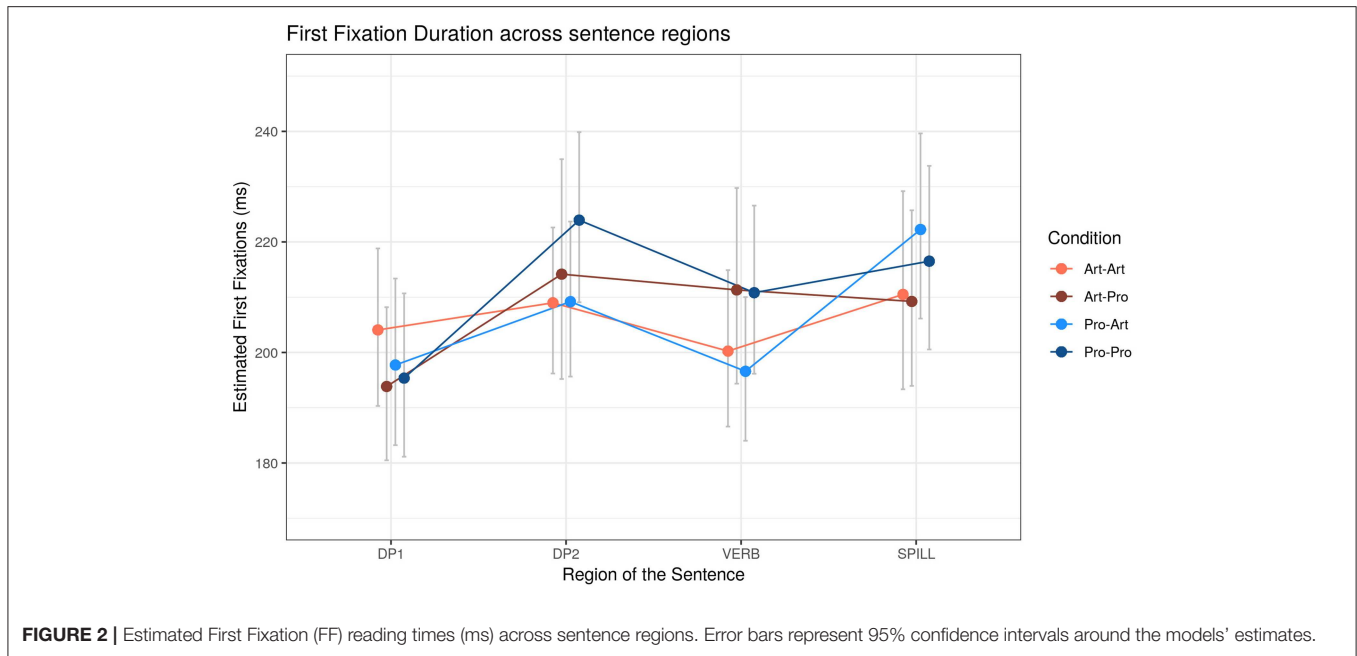


FIGURE 2 | Estimated First Fixation (FF) reading times (ms) across sentence regions. Error bars represent 95% confidence intervals around the models' estimates.

TABLE 5 | First Fixation (FF) effects depending on the DP₁-DP₂ types and Working Memory (WM).

First fixation times	Regions												
	BE		DP ₁		C		DP ₂		Verb		Spill		
	χ ²	p	χ ²	p	χ ²	p	χ ²	p	χ ²	p	χ ²	p	
Null (trial order)													
+ DP ₁ type							1.21					3.92	<0.05
+ DP ₂ type			2.14				4.98	<0.05	6.55	<0.05			
+ DP ₁ :DP ₂													
+ WM			1.29				4.37	<0.05	8.58	<0.01	6.41	<0.05	
+ DP ₁ :WM													
+ DP ₂ :WM									2.08				
+ DP ₁ :DP ₂ :WM									1.32				

χ² < 1 are omitted.
 Bold values are the significant ones.

regressions from DP₁ were more likely for DP₁ Pro (Pro₁ vs. Art₁: +0.11, z = 2.84, p < 0.01). In the pattern of Regressions from VERB and DP₂ regions significant interactions between DP₁ and DP₂ emerged. When DP₁ was Pro, more regressions out of VERB were associated to Pro₂ (Pro₂ vs. Art₂: +0.11, z = 2.06, p < 0.05), while when DP₁ was Art the pattern was numerically opposite (Pro₂ vs. Art₂: -0.06, z = -1.30, ns). Considering a post hoc mismatching vs. matching pairwise comparison, in addition to DP type comparison, we observe that mismatching conditions are associated to a much smaller regression probability from this region (mismatching vs. matching: -0.10, z = -2.754, p = 0.005). As for Regressions from DP₂ the interaction was explained by a similar pattern: here, though, when DP₁ was Pro, the higher proportion of regressions out for Pro₂ was not significant (Pro₂ vs. Art₂: +0.07, z = 1.45, ns), while when DP₁ was Art a higher

likelihood of making a regression was observed for DP₂ Art (Art₂ vs. Pro₂: +0.13, z = 2.25, p < 0.05).

The effect of WM, occurring in VERB (β = +0.403, t = 2.32, p < 0.05) and SPILL (β = +0.723, t = 2.58, p < 0.01) shows that participants with higher WM were more likely to make a regression out of these regions, whereas the marginally significant effect of WM in DP₁ had the opposite direction (β = -0.375, -1.82, p < 0.1), suggesting that low WM participants performed a regression much earlier on. In DP₂, the interaction between DP₂ type and WM was due to the more positive (Δβ = +0.31, z = +1.97, p < 0.05) slope of WM for Pro₂ (β = +.35) compared to Art₂ (β = +0.04).

Then we evaluated Regressions in first pass, considering the proportion of regressions from each region on the number of fixations made in each region during first pass only. Regressions

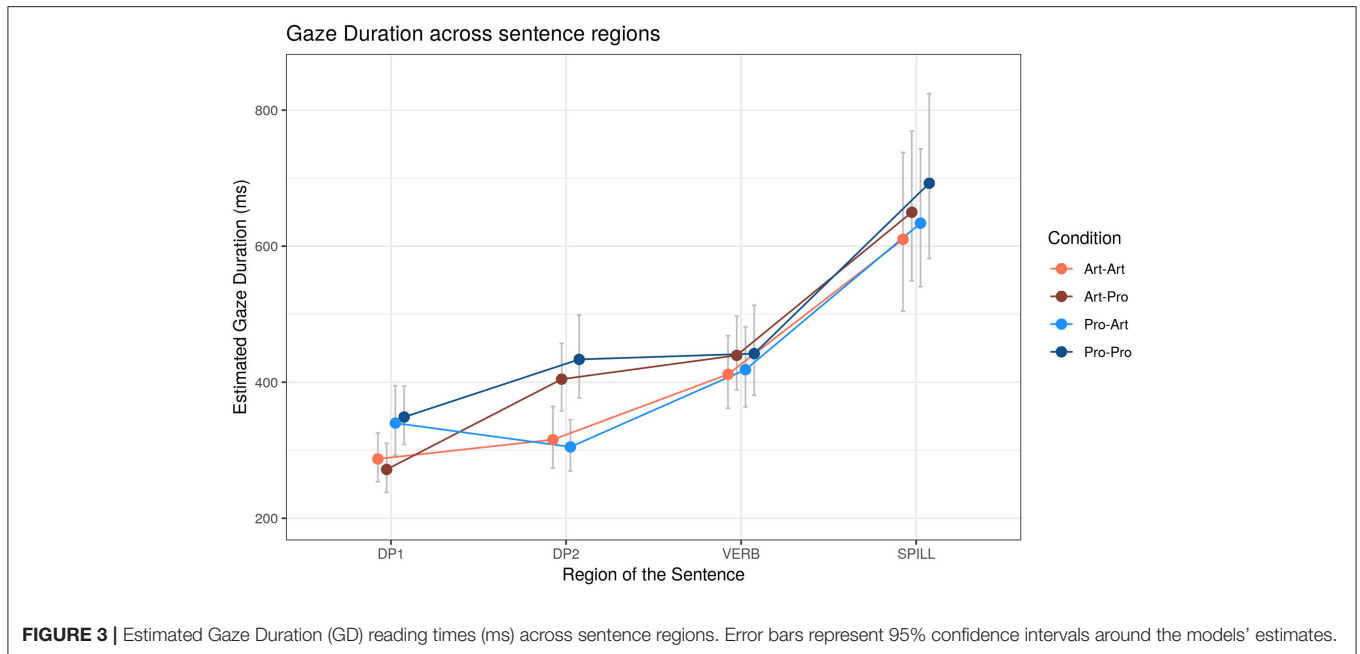


TABLE 6 | Gaze Duration (GD) effects depending on DP₁-DP₂ type and Working Memory (WM).

Gaze duration times	Regions											
	BE		DP ₁		C		DP ₂		Verb		Spill	
	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>
Null (trial order)												
+ DP ₁ type			30.1	<0.001								1.6
+ DP ₂ type							32.09	<0.001	3.42	<0.1	2.93	<0.1
+ DP ₁ :DP ₂			1.11				2.42					
+ WM			11.80	<0.001			8.12	<0.01	17.1	<0.001	12.59	<0.001
+ DP ₁ :WM												
+ DP ₂ :WM							1.93		1.01			
+ DP ₁ :DP ₂ :WM							1.13					

Bold values are the significant ones.

out of C were more likely for *Art*₁ (+0.08, $z = 1.93$, $p = 0.05$). First pass Regressions from DP₂ were more likely for *Art*₂ (+0.12, $z = 3.04$, $p < 0.05$). At the VERB, the DP₁ by DP₂ interaction was due to a larger number of regressions for *Art*₁ compared to *Pro*₁ (+0.07, $z = 2.08$, $p < 0.05$) when DP₂ was *Art*, which was absent when DP₂ was *Pro* (-0.02 , $z < 1$, ns).

Regressions into regions (R-in)

We first evaluated the likelihood of making a R-in for each ROI on the total number of Regressions events (Figure 6, Table 10). Regressions in C and DP₂ were affected by the main effect of DP₂ type: in C regressions were more likely for *Art*₂ (+0.05, $z = 2.84$, $p < 0.01$), while in DP₂ regressions were more likely for *Pro*₂ (+0.06, $z = 3.00$, $p < 0.01$). The effect of WM had a negative slope in BE ($\beta = -0.444$, $z = -4.01$, $p < 0.001$) and in C ($\beta = -0.257$,

$z = -2.32$, $p < 0.05$), suggesting that regressions in these regions were more likely for low WM participants, while the effect of WM had more positive slopes in DP₁ ($\beta = 0.135$, $z = 2.09$, $p < 0.05$), DP₂ ($\beta = 0.212$, $z = 2.55$, $p < 0.05$) and VERB ($\beta = +0.222$, $z = 2.46$, $p < 0.05$), showing that higher WM participants directed their regressions toward these sentence regions.

Summary

Acceptability judgments showed that matching conditions are significantly different (*Art*₁-*Art*₂ better than *Pro*₁-*Pro*₂). *Art*₁-*Art*₂ matching condition results slightly less grammatical than *Art*₁-*Pro*₂ mismatching condition. The other *Pro*₁-*Art*₂ mismatch condition ranking below them and above *Pro*₁-*Pro*₂ matching condition, which is unquestionably considered rather ungrammatical.

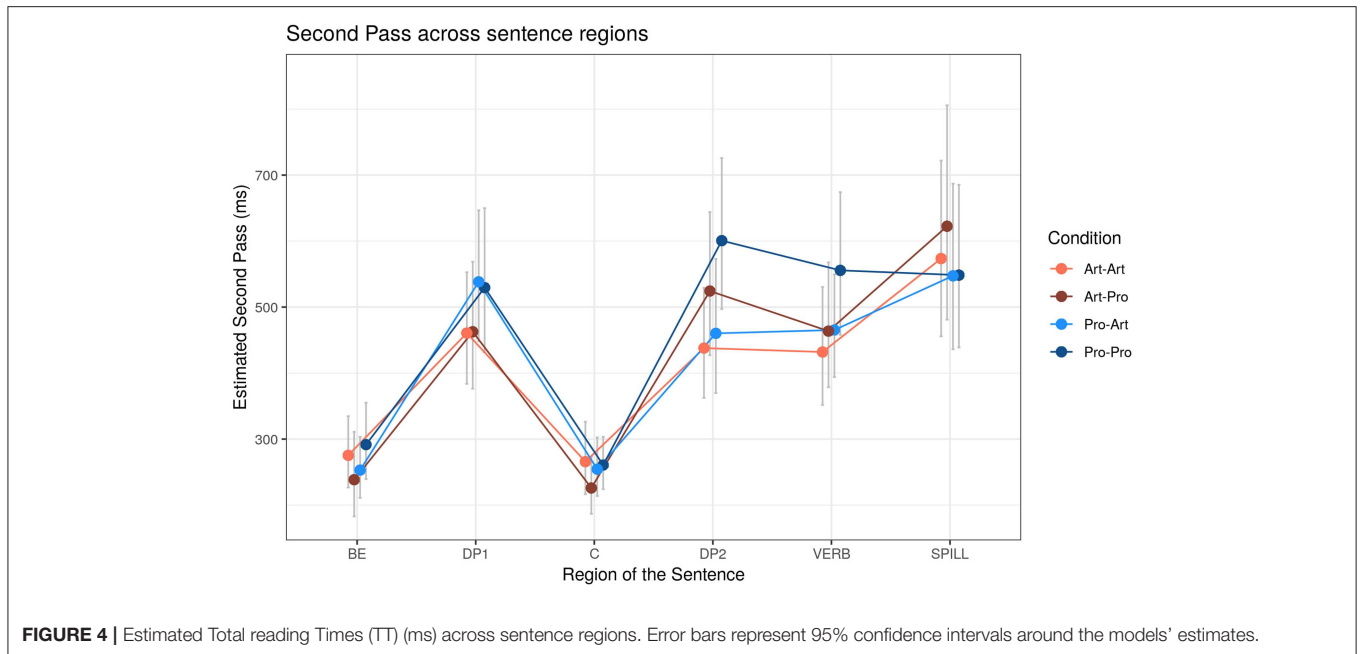


TABLE 7 | Total Time (TT) effects depending on the DP₁-DP₂ type and Working Memory (WM).

Total reading times	Regions											
	BE		DP ₁		C		DP ₂		Verb		Spill	
	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>
Null (trial order)												
+ DP ₁ type	1.52		28.22	<0.001	2.84	<0.1	1.04		2.34			
+ DP ₂ type							19.39	<0.001	2.77	<0.1		
+ DP ₁ :DP ₂	4.37	<0.05			1.58		8.95	<0.01	6.33	<0.05	1.32	
+ WM	3.53	<0.1			3.86	<0.05			1.21			
+ DP ₁ :WM												
+ DP ₂ :WM			3.60	<0.1			1.90					
+ DP ₁ :DP ₂ :WM												

Bold values are the significant ones.

Comprehension questions in the eye-tracking experiment, revealed only an interaction between DP types, showing a difference in accuracy between *Art*₁-*Pro*₂ and *Pro*₁-*Pro*₂, but overall, the experimental sample correctly (>85%) understood the sentences.

Concerning reading times, and looking at the global results we can summarize:

- Art*₁-*Art*₂ matching condition constitute the processing baseline in all measures;
- Art*₁-*Pro*₂ mismatch condition caused some slow-down, mainly at DP₂, where *Pro* was present (GD, TT, and SP), and marginally at VERB (FF only);
- Pro*₁-*Art*₂ mismatch condition caused some slow-down, but only on DP₁ region (TT and SP);
- Pro*₁-*Pro*₂ matching condition is numerically the most time consuming condition in all measures, and the numeric

differences emerge as statistically consistent with the DP₁ DP₂ interaction in TT for DP₂ and VERB regions.

About factors interaction:

- the interactions between DP₁ and DP₂ was found in DP₂ and VERB regions for TT; in BE and C for SP; in the regressions from DP₂ and VERB.
- the effect of DP₂ type is overwhelming; *Pro*₂ is problematic as revealed in FF, GD, TT, and SP both in DP₂ and VERB regions;
- mismatching conditions (*Art*₁-*Pro*₂ and *Pro*₁-*Art*₂), overall, are associated to a reduced probability to trigger regressions from DP₂ and VERB regions.
- the effect of WM is very strong, and, interestingly, the slope of WM is not always negative (higher WM associated with faster reading or fewer regressions). It has a negative slope in reading times measures FF, GD, TT, and SP. However,

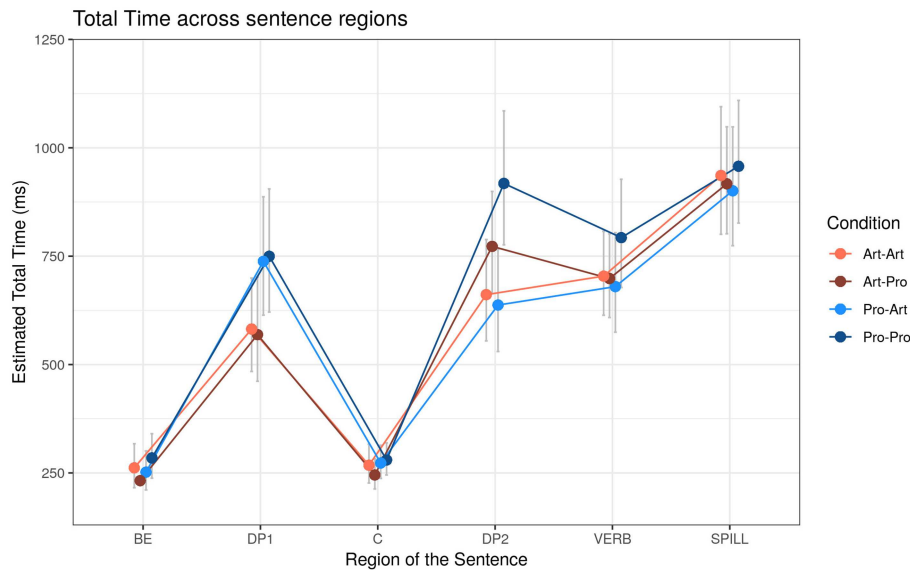


FIGURE 5 | Estimated Second Pass (SP) reading times (ms) across sentence regions. Error bars represent 95% confidence intervals around the models' estimates.

TABLE 8 | Second Pass (SP) reading times effects depending on the DP₁-DP₂ type and Working Memory (WM).

Second pass reading times	Regions											
	BE		DP ₁		C		DP ₂		Verb		Spill	
	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>
Null (trial order)												
+ DP ₁ type			6.68	<0.01			3.88	<0.05	6.33	<0.05	1.64	
+ DP ₂ type							9.70	<0.01	3.89	<0.05		
+ DP ₁ :DP ₂	5.06	<0.05			2.73	<0.1						
+ WM	1.89				2.21				2.66			
+ DP ₁ :WM			1.51				2.13		2.37			
+ DP ₂ :WM			2.08				1.27		2.47			
+ DP ₁ :DP ₂ :WM							2.13		3.82	=0.05		

$\chi^2 < 1$ are omitted.

Bold values are the significant ones.

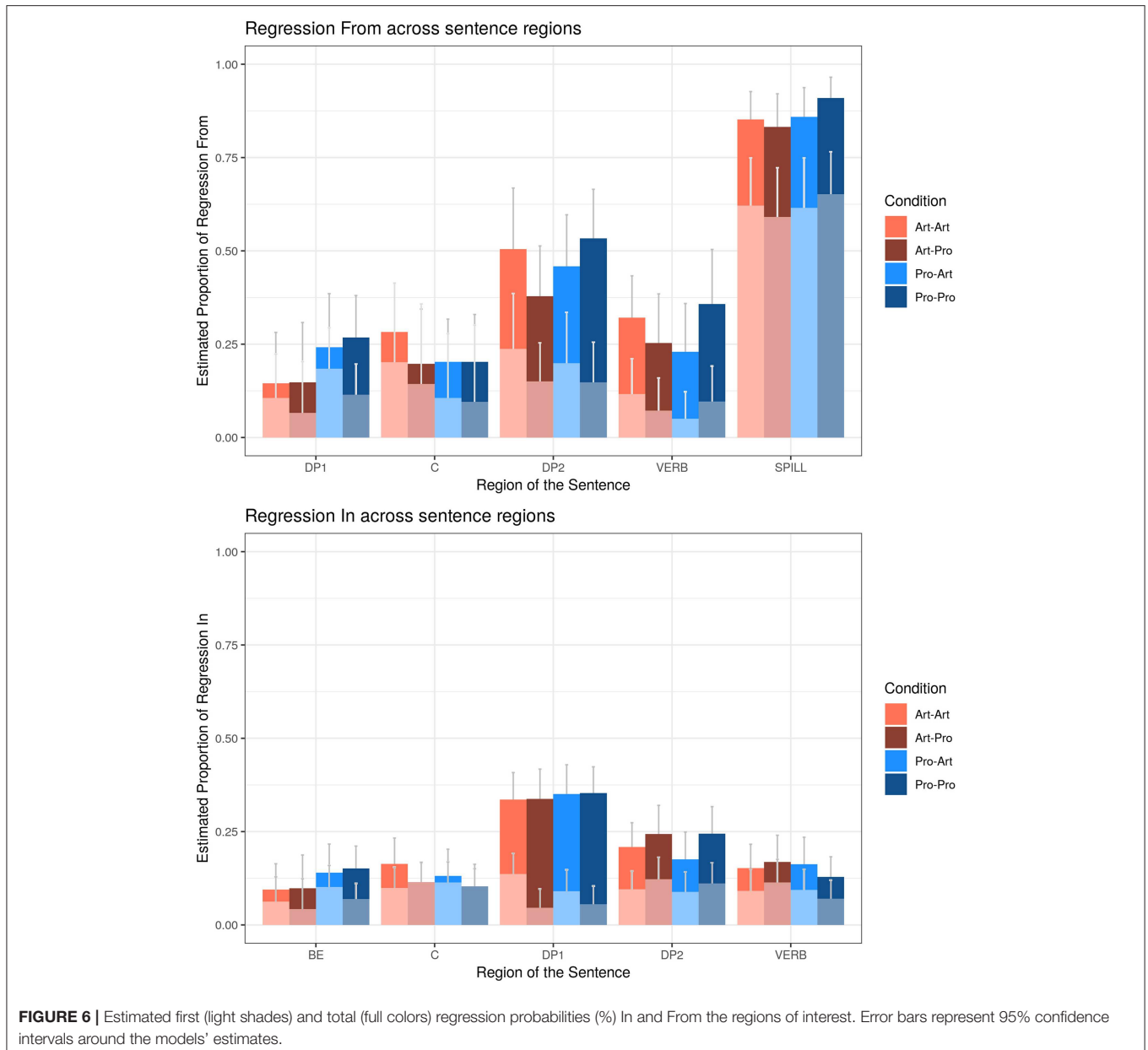
it shows different effects on regression probability. Higher WM is associated to a larger proportion of regressions out of VERB and SPILL regions, but to fewer regressions out (during first pass) of earlier sentence regions like DP1. No major interaction between WM * DP₁ * DP₂ type is revealed.

In **Table 11** a summary of the main results.

DISCUSSION

Starting with off-line considerations, the results of both experiments consistently show that a lexically restricted second person pronoun (“you linguists”) is at least as hard as a restricted definite article (“the linguist”) across all conditions, hence any advantage of the bare pronominal DPs (“you”),

revealed in previous experiments (Warren and Gibson, 2005, a.o.), is lost when a lexical restriction is present. This main result is consistent both with the intervention-based (Friedmann et al., 2009; Belletti and Rizzi, 2013) and with the similarity based prediction (Gordon et al., 2001). These models, however, fail to capture contrasts both in matching conditions (with *Art*₁-*Art*₂ condition “easier than” *Pro*₁-*Pro*₂ condition) and in mismatching ones (with *Pro*₁-*Art*₂ condition less acceptable than *Art*₁-*Pro*₂ condition). Both contrasts are predicted under the *memory-load* and *top-down* perspectives: *memory-load* approach (Gibson, 1998) can predict an extra cost in processing *Pro* conditions by relying on the absence of an appropriate context (assumption A3 in section Testing the Different Predictions), but it fails to predict the striking asymmetry found between *Art*₁-*Pro*₂ and *Pro*₁-*Pro*₂. This contrast is only correctly predicted



by the *top-down* model: the acceptability pattern revealed a significant difference in these conditions (Art_1-Pro_2 better than both Pro_1-Art_2 and, even more robustly, Pro_1-Pro_2) is predicted by this model under the H2 hypothesis: 2nd person on the subject, in a mismatching condition, induces a facilitation better than 3rd person (section Testing the Different Predictions). Similarly, under the same hypothesis, only the *top-down* model predicts a major effort in processing the Pro_1-Pro_2 with respect to the Art_1-Art_2 condition due to 2nd pronominal matching feature. This model also predicts a milder advantage of the Art_1-Pro_2 condition with respect to the Art_1-Art_2 condition which is numerically present in accuracy but not significant. Generally the pattern across conditions is similar both in acceptability and in accuracy in comprehension

questions after eye-tracking. However, the first pattern, but not the second (with the exception of Pro_1-Pro_2 vs. Art_1-Pro_2 , again coherently with H2), results in statistically significant contrasts. Notice that the Pro_1-Pro_2 condition is considered nearly ungrammatical by the subjects. This clearly differentiates this condition from the others. Also in comprehension questions this condition leads to the worst performance, but such performance is still surprisingly high (78.1%). This indicates that the subjects correctly answer to questions posed on sentences that they consider unacceptable, suggesting a milder discriminative power of the accuracy measure with respect to acceptability judgments.

Due to a constant set of factors (i.e., distance between the focalized object and the predicate, same context and same

TABLE 9 | Regression from (R-from) depending on the DP₁-DP₂ types and Working Memory (WM).

Regression from	DP1 (N = 198)		DP2 (N = 396)		Verb (N = 300)		Spill (N = 708)	
	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>
Null (trial order)								
+ DP ₁ type <i>fR-from</i>	6.16	<0.05	2.19		<1		1.64	
+ DP ₂ type <i>fR-from</i>	2.63		8.67	<0.01	1.29			
+ DP ₁ :DP ₂ <i>fR-from</i>	1.93		7.68	<0.05	7.80	<0.01		
					4.72	<0.05		
+ WM <i>fR-from</i>	3.31	<0.1			4.87	< 0.05	1.31	
	6.72	<0.01						
+ DP ₁ :WM <i>fR-from</i>								
+ DP ₂ :WM <i>fR-from</i>	3.48	<0.1	4.10	<0.05				
+ DP ₁ :DP ₂ :WM <i>fR-from</i>	1.81		2.38				2.89	<0.1

First Regression from are indicated (*fR-from*) under the total R-From estimate. $\chi^2 < 1$ are omitted.

Bold values are the significant ones.

Italics indicate the *fR-from* measures.

TABLE 10 | Regression In (R_{in}) depending on the DP₁-DP₂ types and Working Memory (WM).

Regression in	BE (N = 286)		DP1 (N = 534)		C (N = 243)		DP2 (N = 336)		C (N = 228)	
	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>
Null (trial order)										
+ DP ₁ type	3.00	<0.1								
+ DP ₂ type					7.01	<0.01	8.35	<0.01		
+ DP ₁ :DP ₂					2.43				1.09	
+ WM	13.42	<0.001	4.69	<0.05	5.90	<0.05	6.39	<0.05	5.60	<0.05
+ DP ₁ :WM			1.32		3.02	<0.1				
+ DP ₂ :WM	1.90				1.20					
+ DP ₁ :DP ₂ :WM					2.02					

$\chi^2 < 1$ are omitted.

Bold values are the significant ones.

cued-features for the focalized DP), ACT-R-based processing model (Lewis and Vasishth, 2005) flatly predicts no difference among any of the tested configurations suggesting that this model needs extra assumptions to account for the revealed asymmetries.

Considering the online predictions (Figure 7), assuming an encoding cost penalty for the *Pro* conditions (hypothesis A3 in section Testing the Different Predictions), *memory-load* model becomes competitive in predicting GD reading times ($r(14) = 0.65$, $p = 0.006$) and (less robustly) and TT ($r(14) = 0.53$, $p = 0.034$).

Feature Retrieval and Encoding Cost (FREC) (FRC+FEC) however, with the very same assumptions on *Pro* encoding penalty and under the hypothesis that only 2nd person features on the subject are cued by the verb (H2 in section Testing the Different Predictions), correlates much more precisely both with GD ($r(14) = 0.90$, $p < 0.001$) and (even better) with TT ($r(14) = 0.96$, $p < 0.001$). Considering the nature of retrieval, the higher correlation with respect to the most comprehensive, latest, measure (i.e., TT) is expected.

As predicted under Bianchi (2003, 2006), Sigurdhsson (2004) and Elbourne (2005) analyses, the difficulty associated to the processing of the *Pro* condition is mostly due to the out of the blue presentation of the second person feature restricted by a *N* predicate. Therefore, the processing cost revealed at the DP regions where *Pro* occurred is likely due to encoding (need of postulating the salience of the relevant referents and update the common ground accordingly): this is revealed by longer reading times in late measures (GD, TT, and SP) comparable at both DPs regions under the *Pro* condition. In DP₂ region, also an effect at the early FF measure was observed for *Pro*₁-*Pro*₂, suggesting an element of surprise as soon as the second restricted pronoun is encountered in the subject cleft position. A similar early effect, dependent on the presence of *Pro*₂, is revealed also at the verb segment possibly indicating a “spillover effect” of the previous region processing or a retrieval difficulty. The spill-over interpretation only is supported by the fact that the DP₂ effect disappears in later measures at the VERB segment (non-significant effect in GD and TT). On the other hand a retrieval problem

TABLE 11 | Main results summarized.

Region	Condition Measure	Art ₁ -Art ₂	Pro ₁ -Pro ₂	Art ₁ -Pro ₂	Pro ₁ -Art ₂
DP1 (focalized object)	Acceptability	Good	Bad	Good	Medium
	Comprehension	Good	Good	Good	Good
	First fixation	Baseline	Baseline	Baseline	Baseline
	Gaze	Baseline	Slower	Baseline	Slower
	Total	Baseline	Slower	Baseline	Slower
	Second pass	Baseline	Slower	Baseline	Slower
	Regressions from	Baseline	More	Baseline	More
	Regressions in	Baseline	Baseline	Baseline	Baseline
D2 (subject)	First fixation	Baseline	Slower	Slower	Baseline
	Gaze	Baseline	Slower	Slower	Baseline
	Total	Baseline	Slower	Mildly slower	Baseline
	Second pass	Baseline	Slower	Mildly slower	Baseline
	Regressions from	More	More	Baseline	Baseline
	Regressions in	Baseline	More	Baseline	Baseline
Verb	First fixation	Baseline	Slower	Slower	Baseline
	Gaze	Baseline	Slower	Baseline	Baseline
	Total	Baseline	Slower	Baseline	Baseline
	Second pass	Baseline	Slower	Baseline	Baseline
	Regressions from	Slightly more	More	Baseline	Baseline

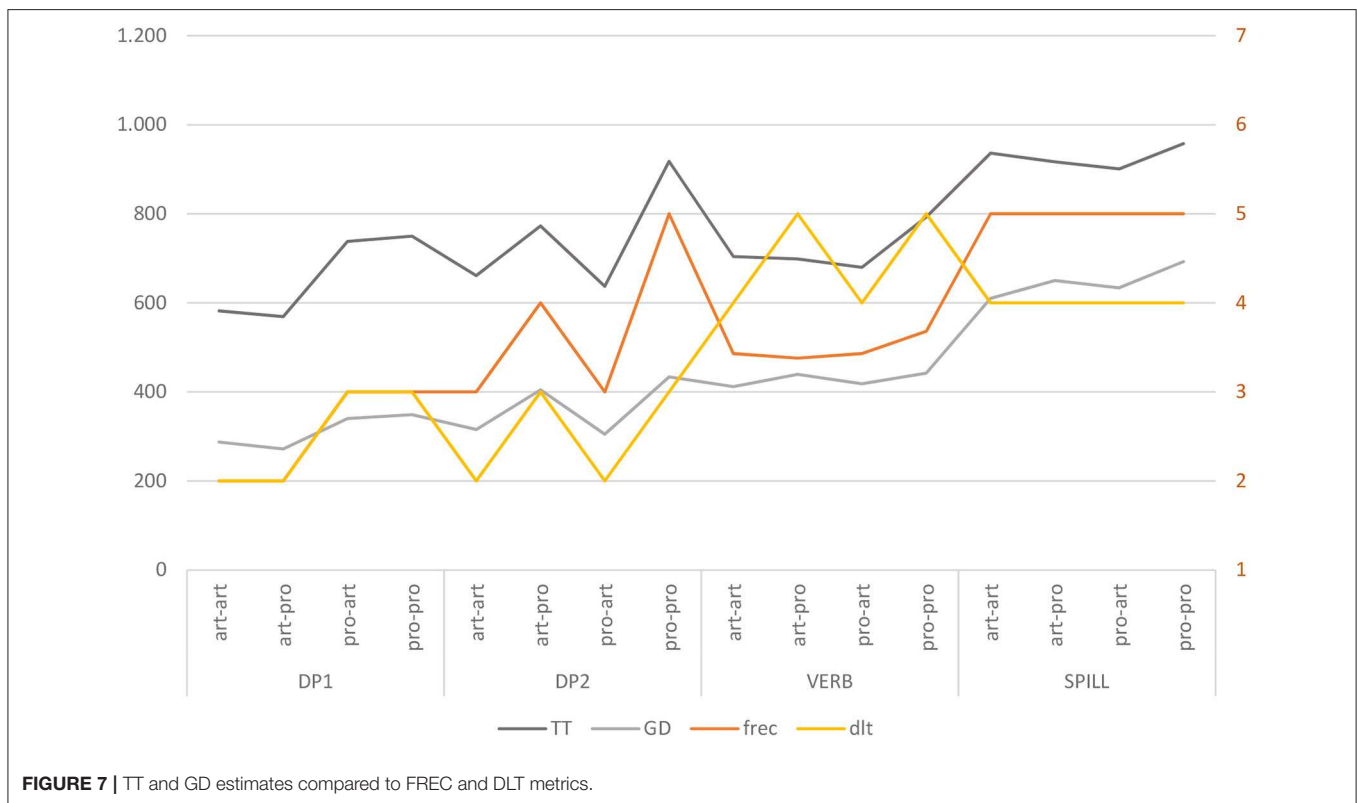


FIGURE 7 | TT and GD estimates compared to FREQ and DLT metrics.

in matching conditions is suggested by the first and total regression patterns leaving the VERB region: the significant *post hoc* pairwise comparison indicates less regressions from

the verb in the mismatching conditions, possibly revealing a difficulty in retrieving the correct argument in the matching cases (especially with second person feature matching). Considering

that the same effect is observed also at DP₂, we should conclude that the retrieval interpretation is not the sole possible analysis.

This difficulty in the matching conditions is in line with the *top-down*, *similarity*-based and *intervention*-based predictions, but not with the *memory-load* ones.

Under the H2 hypothesis (2nd person feature on the subject cued by verb agreement morphology should facilitate the integration of the subject, while 3rd person default agreement should represent just a minor facilitation) the top-down model would predict an asymmetry also in the mismatching condition. However, no significant on-line evidence of a facilitation in retrieval is encountered for the *Art*₁-*Pro*₂ vs. *Pro*₁-*Art*₂ neither in terms of reading times at VERB, or regression probability. The on-line results then support hypothesis H1 (section Testing the Different Predictions), namely that both 2nd and 3rd feature mismatch on the VERB might facilitate retrieval (quicker FF), but *Pro*₂ encoding penalty makes this effect hardly detectable. We agree with the reviewers suggesting that one way to tease apart the actual impact of the encoding penalty with respect to the retrieval effect would be to include in this experimental design the Subject Cleft condition where only encoding (and no retrieval penalty due to matching features to be re-merged) will be present. Another way to disentangle the two components would be to introduce a proper context, then removing the encoding penalty of *Pro* predicted under the assumption A3.

We can only suppose here that the encoding cost, responsible for the on-line slowdown revealed in the *Pro* conditions, could have been partially mitigated by the (mild) facilitation at retrieval in the mismatch *Art*₁-*Pro*₂ case, in the end producing an acceptability equivalence between the *Art*₁-*Art*₂ matching condition and the *Art*₁-*Pro*₂ mismatch condition. Everything being equal, removing the encoding cost, i.e., providing an appropriate context, we expect a difference *Art*₁-*Art*₂ vs. (worse than) *Art*₁-*Pro*₂ to appear, hence confirming the facilitation related to the usage of a deictic, mismatching, 2nd person feature in the intervening DP whose morphology is cued by the selecting verb (hence confirming the explanatory superiority of H2 over H1 also in on-line measures). Coherently with Staub (2010, p. 77-78), the complexity signatures at DP₂ revealed (also) by a generally higher probability of making a regression from this region, especially in the matching conditions case (and especially in the *Pro*₁-*Pro*₂ case) could be interpreted as an indication of “something is wrong” (as in E-Z reader 10 model, Reichle et al., 2009) or, more precisely, an integration failure due to time out (c.f. Staub, 2010, p. 83) because of a context-update request.

To conclude, a final crucial intent of this study was to provide some new evidence for disentangling the (complex) relation between off-line and on-line performance measures. Given the off-line results gathered, first, we observed that acceptability judgments are more discriminative than accuracy in comprehension questions (though both generally correlates on the numerical patterns), second, FRC metrics, based on the *top-down* model, is the one making the closest predictions with respect to the pattern revealed across conditions. This suggests that the retrieval effort, at least in this context, is the best predictor of the overall acceptability and that, despite heavy

encoding efforts (revealed by on-line measures), readers are fully rewarded by an adequate comprehension, revealed by accurate answers in all conditions.

As for the on-line data, again FRC+FEC (FREC) shows the best correlation with respect to the revealed “late” measures (GD and TT). Unexpected referents, introducing features that force a revision of the common ground assumptions are correctly predicted to affect performance by the FEC component at specific regions. These predictions crucially rely on a precise linguistic theory that takes into consideration the nature of the OC dependency and the relation between D, N types and person features. It is important to emphasize that no significant interaction between WM and our syntactic manipulation has been revealed: high WM participants simply show faster reading times and more regressive patterns compared to the low WM population across all conditions.

The actual usage of working memory during the processing of these specific constructions is still to be explored precisely. Nevertheless, we believe that the intuition that identical features that must be (re)merged within the active workspace are lighter to be processed than new ones or “similar” ones that must be kept distinct in memory (like an extra nominal restriction or an extra second person index) is worth further investigation: this idea is coherent with a “primed” active storage in which the “memory units” encoding a specific feature, being just activated would be more accessible than other units (on the line of ACT-R Lewis and Vasisht, 2005 intuition), while forcing a minimal diversification of a new pattern with respect to a pre-activated overlapping one has a considerably high cost.

ETHICS STATEMENT

The experiment was approved by the Ethics Committee of the Dipartimento di Scienze del Sistema Nervoso e del Comportamento of the University of Pavia.

AUTHOR CONTRIBUTIONS

CC designed and directed the project, developed the theoretical top-down framework, the FR(E)C complexity measure, and performed the acceptability judgment experiment. PC performed the eye-tracking experiment and analyzed the results of both experiments. Both authors discussed the results and contributed to the final manuscript.

FUNDING

This work was supported by IUSS as an internal research project (ProGraM-PC-A Processing-friendly Grammatical Model for Parsing and predicting on-line Complexity, PI: CC).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02105/full#supplementary-material>

REFERENCES

- Aho, A. V., and Ullman, J. D. (1972). *The theory of parsing, translation, and compiling*, Vol. 1, *Parsing*. Englewood Cliffs, NJ: Prentice Hall.
- Anderson, J. R. (2005). Human symbol manipulation within an integrated cognitive architecture. *Cogn. Sci.* 29, 313–341. doi: 10.1207/s15516709cog0000_22
- Anderson, J. R., and Matessa, M. P. (1997). A production system theory of serial memory. *Psychol. Rev.* 104, 728–748. doi: 10.1037/0033-295X.104.4.728
- Ariel, M. (1990). *Accessing Noun-Phrase Antecedents*. London: Routledge.
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005
- Baroni, M., Bernardini, S., Comastri, F., Piccioni, L., Volpi, A., Aston, G., et al. (2004). Introducing the La Repubblica corpus: a large, annotated, TEI (XML)-compliant corpus of newspaper Italian. *Issues* 2, 5–163. Available online at: <https://www.aclweb.org/anthology/L04-1128/>
- Belletti, A. (2008). The CP of clefts. *Rivista di Grammatica Generativa*. (2008) 33, 191–204. Available online at: <http://lear.unive.it/jspui/handle/11707/434>
- Belletti, A., and Chesi, C. (2014). A syntactic approach toward the interpretation of some distributional frequencies: comparing relative clauses in Italian corpora and in elicited production. *Rivista di Grammatica Generativa*. (2014) 36: 1–28. Available online at: <http://lear.unive.it/jspui/handle/11707/5239>
- Belletti, A., Friedmann, N., Brunato, D., and Rizzi, L. (2012). Does gender make a difference? Comparing the effect of gender on children's comprehension of relative clauses in Hebrew and Italian. *Lingua* 122, 1053–1069. doi: 10.1016/j.lingua.2012.02.007
- Belletti, A., and Rizzi, L. (2013). "Intervention in grammar and processing," in *From Grammar to Meaning: The Spontaneous Logicality of Language*, eds I. Caponigro and C. Cecchetto (Cambridge: Cambridge University Press), 293–311.
- Bever, T. G. (1974). "The ascent of the specious, or there's a lot we don't know about mirrors," in *Explaining Linguistic Phenomena* ed D. Cohen (Washington, DC: Hemisphere), 173–200.
- Bianchi, V. (2003). "On finiteness as logophoric anchoring," in *Temps et Point de Vue/Tense and Point of View*, eds J. Guéron, L. Tasmovski (Paris, Univ. Paris X Nanterre), 213–246.
- Bianchi, V. (2006). On the syntax of personal arguments. *Lingua* 116, 2023–2067. doi: 10.1016/j.lingua.2005.05.002
- Bianchi, V., and Chesi, C. (2014). Subject islands, reconstruction, and the flow of the computation. *Linguist Inq.* 45, 525–569. doi: 10.1162/LING_a_00166
- Chesi, C. (2015). On directionality of phrase structure building. *J. Psychol. Res.* 44, 65–89. doi: 10.1007/s10936-014-9330-6
- Chesi, C. (2017). Phase-based Minimalist Parsing and complexity in non-local dependencies. *Proceedings of CLIC-it 2017*. CEUR Workshop Proceedings, Roma:CEUR Workshop Proceedings.
- Chomsky, N. (1965). *Aspects of The Theory of Syntax*. Cambridge, MA: MIT Press.
- Crowder, R. G. (1976). *Principles of Learning and Memory*. Hillsdale, NJ: Erlbaum.
- Daneman, M., and Carpenter, P. A. (1980). Individual differences in working memory and reading. *J. Verb. Learn. Verb. Behav.* 19, 450–466. doi: 10.1016/S0022-5371(80)90312-6
- Elbourne, P. D. (2005). *Situations and Individuals*. Cambridge, MA: MIT Press.
- Friedmann, N., Belletti, A., and Rizzi, L. (2009). Relativized relatives: types of intervention in the acquisition of A-bar dependencies. *Lingua* 119, 67–88. doi: 10.1016/j.lingua.2008.09.002
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition* 68, 1–76. doi: 10.1016/S0010-0277(98)00034-1
- Gibson, E. (2000). "The dependency locality theory: a distance-based theory of linguistic complexity," in *Image, Language, Brain*, eds Y. Miyashita, A. P. Marantz, and W. O'Neil (Cambridge, MA: MIT Press), 95–126.
- Gordon, P. C., Hendrick, R., and Johnson, M. (2001). Memory interference during language processing. *J. Exp. Psychol.* 27:1411. doi: 10.1037//0278-7393.27.6.1411
- Gordon, P. C., Hendrick, R., and Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *J. Mem. Lang.* 51, 97–114. doi: 10.1016/j.jml.2004.02.003
- Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446. doi: 10.1016/j.jml.2007.11.007
- Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., and Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science* 274, 114–116. doi: 10.1126/science.274.5284.114
- Kidd, E., Brandt, S., Lieven, E., and Tomasello, M. (2007). Object relatives made easy: a cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Lang. Cogn. Proc.* 22, 860–897. doi: 10.1080/01690960601155284
- King, J., and Just, M. A. (1991). Individual differences in syntactic processing: the role of working memory. *J. Mem. Lang.* 30:580. doi: 10.1016/0749-596X(91)90027-H
- Kiss, K. E. (1998). Identificational focus versus information focus. *Language* 74, 2, 245–273. doi: 10.1353/lan.1998.0211
- Kuperman, V., and Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *J. Mem. Lang.* 65, 42–73. doi: 10.1016/j.jml.2011.03.002
- Lenth, R., Singmann H., Love J., Buerkner P., and Herve M. (2019). *Emmeans: Estimated Marginal Means, Aka Least-Squares Means*. R package version 1.4. Available online at: <https://github.com/rvnlenth/emmeans>
- Levy, R., Fedorenko, E., and Gibson, E. (2013). The syntactic complexity of Russian relative clauses. *J. Mem. Lang.* 69, 461–495. doi: 10.1016/j.jml.2012.10.005
- Lewandowsky, S., Oberauer, K., Yang, L. X., and Ecker, U. K. (2010). A working memory test battery for MATLAB. *Behav. Res. Methods* 42, 571–585. doi: 10.3758/BRM.42.2.571
- Lewis, R. L., and Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cogn. Sci.* 29, 375–419. doi: 10.1207/s15516709cog0000_25
- Link, G. (1983). "The logical analysis of plurals and mass terms: a lattice-theoretical approach," in *Formal Semantics: The Essential Readings*, eds P. Portner and B. Partee (Oxford: Blackwell Publishing), 127–146.
- Longobardi, G. (1994). Reference and proper names: a theory of N-movement in syntax and logical form. *Ling. Inq.* 609–665.
- Mancini, S., Molinaro, N., Rizzi, L., and Carreiras, M. (2011). A person is not a number: discourse involvement in subject-verb agreement computation. *Brain Res.* 1410, 64–76. doi: 10.1016/j.brainres.2011.06.055
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing Type I error and power in linear mixed models. *J. Mem. Lang.* 94, 305–315. doi: 10.1016/j.jml.2017.01.001
- Munaro, N. (1999). *Sintagmi Interrogativi nei Dialetti Italiani Settentrionali*. Padova: Unipress.
- Phillips, C. (1996). *Order and structure*, (Doctoral dissertation), Massachusetts Institute of Technology, Cambridge, MA, United States.
- Pollard, C., and Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*. Chicago, IL: University of Chicago Press.
- Postal, P. (1966). "On so-called pronouns in English," in *Modern Studies in English*, eds D. A. Reibel and S. A. Schane (Englewood Cliffs, NJ: Prentice-Hall), 177–206.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422.
- Reichle, E. D., Warren, T., and McConnell, K. (2009). Using EZ Reader to model the effects of higher level language processing on eye movements during reading. *Psychon. Bulletin*. Rev. 16, 1–21. doi: 10.3758/PBR.16.1.1
- Rizzi, L. (1990). *Relativized Minimality*. Cambridge, MA: MIT Press.
- Rizzi, L. (1997). "The fine structure of the left periphery," in *Elements of Grammar*, ed L. Haegeman (Dordrecht: Springer), 281–337.
- Sigurðsson, H. Á. (2004). "The syntax of person and speech features," in *The Syntax and Interpretation of Person Features*, eds V. Bianchi, K. Safir (Italian Journal of Linguistics), 16.1.
- Stabler, E. (1996). *Derivational minimalism*. In *International Conference on Logical Aspects of Computational Linguistics*. Berlin; Heidelberg: Springer. 68–95.
- Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition* 116, 71–86. doi: 10.1016/j.cognition.2010.04.002
- Stowell, T. (1983). Subjects across categories. *Ling. Rev.* 2, 285–312.

- Traxler, M. J., Morris, R. K., and Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *J. Mem. Lang.* 47, 69–90. doi: 10.1006/jmla.2001.2836
- Van Acker, F. (2007). *Online Survey Creator 7.4 for Windows. Ms.*
- Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *J. Exp. Psychol.* 33:407. doi: 10.1037/0278-7393.33.2.407
- Van Dyke, J. A., and McElree, B. (2006). Retrieval interference in sentence comprehension. *J. Mem. Lang.* 55, 157–166. doi: 10.1016/j.jml.2006.03.007
- Wanner, E., and Maratsos, M. (1978). “An ATN approach to comprehension,” in *Linguistic Theory and Psychological Reality*, eds M. Halle, J. Bresnan, and G. A. Miller (Cambridge, MA: MIT Press).
- Warren, T., and Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition* 85, 79–112. doi: 10.1016/S0010-0277(02)00087-2
- Warren, T., and Gibson, E. (2005). Effects of NP type in reading cleft sentences in English. *Lang. Cogn. Proc.* 20, 751–767. doi: 10.1080/01690960500051055
- Weckerly, J., and Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology* 36, 559–570. doi: 10.1111/1469-8986.3650559
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Copyright © 2019 Chesi and Canal. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*